Microbial Cell Factories

Review

Metabolic networks of microbial systems

Sumana Bhattacharya[1], Subhra Chakrabarti[1], Amiya Nayak[1] and Sanjoy K Bhattacharya*[2]

Address: [1]Environmental Biotechnology Division, ABRD Company LLC, 1555 Wood Road, Cleveland, Ohio 44121, USA and [2]Dept. of Ophthalmic Research, Cleveland Clinic Foundation, Area I 31, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA

Email: Sumana Bhattacharya - sbc-abrd@usa.net; Subhra Chakrabarti - sbc-abrd@usa.net; Amiya Nayak - nayakamiya@hotmail.com; Sanjoy K Bhattacharya* - bhattas@ccf.org

* Corresponding author

Abstract

In contrast to bioreactors the metabolites within the microbial cells are converted in an impure atmosphere, yet the productivity seems to be well regulated and not affected by changes in operation variables. These features are attributed to integral metabolic network within the microorganism. With the advent of neo-integrative proteomic approaches the understanding of integration of metabolic and protein-protein interaction networks have began. In this article we review the methods employed to determine the protein-protein interaction and their integration to define metabolite networks. We further present a review of current understanding of network properties, and benefit of studying the networks. The predictions using network structure, for example, *in silico* experiments help illustrate the importance of studying the network properties. The cells are regarded as complex system but their elements unlike complex systems interact selectively and nonlinearly to produce coherent rather than complex behaviors.

Review

The microbial proteome and proteomic approaches

The analyses of proteome, the ensemble of proteins associated with a given cell, organelle or other sub-cellular component at a given physiological state of cells have been made possible by the sequencing of genome of a number of organisms. The proteins bring about the functions of a cell, the study of their ensemble would help better understand the function of cell and also enable integration of the function at molecular level. Understanding the proteome of microorganisms provides two broad advantages, one is the utilization of microorganisms or biotechnological exploitation, that is, for the production of metabolites [1] and the other is to deterring the pathogenic and parasitic microbes by understanding host-pathogen interactions or interaction of the microorganisms with multi-cellular systems [2]. The integration of

message transfer at different levels within the cells is not completely understood and analyses of proteome may help improve that understanding [3], which is expected to improve metabolite production and also better targeting of molecular interactions as therapeutic strategy. Biotechnology is the exploitation of the living organisms either directly or using derivatives of them for the benefit of humankind. The derivatives are often exploited is different ways compared to their counter natural processes within the cells.

The chemical or biochemical engineering processes operate with purified compounds (substrates or reactants and catalysts or enzymes). The processes with purified catalysts usually turn the substrates into product(s). The engineering processes are difficult to maintain at a steady state of production, require very careful maintenance of the op-

eration variables. The presence of impurities and variation in any parameters results in dramatic decrease in productivity despite employing the same cellular catalyst in engineered processes. A cell lack perfect compartmentalization, although a definite degree of separation or compartmentalization exists within the cells and produces a large number of chemicals at any given time in a mixed state. In contrast to engineered systems which has a great deal of emphasis on purity of reagents, lack robustness and not rapid in response to perturbation in operation variables, all cells demonstrates a great degree of tolerance to impurity, are impervious to disturbances and mostly have a rapid response to an external stimuli. The perturbations in operation variables are often detrimental for chemical and bioengineered systems. Adapting control features in engineered processes from cells would enable better process control and may enable designing methods for new chemicals employing cell derived catalysts [3].

The cell can be viewed as ensemble of network of interactions such as protein-protein, protein-nucleic acids and protein-small molecule. Several protein and small molecules constitute anabolic, catabolic and amphibolic pathways which are often interconnected and referred to as metabolic pathways [4,5]. Inside the cells, specific metabolic networks are driven by demand for robustness to uncertain environments and use often imprecise components, yet in response to an external stimuli it produces the precise yield of chemicals necessary, with a great deal of molecular understanding these paradoxical features remains to be understood even today [6,7]. It has been put forth that these paradoxical features are neither accidental nor artificial but derive from a deep and necessary interplay between complexity and robustness, modularity, feedback and fragility [6–8]. In anabolic pathways a substrate enters into the pathway and acted upon by a number of proteins in series for a particular conversion, which is necessary for a catalytic reaction involving another substrate received from external cellular environment. All of this occurs in presence of entities unrelated to pathway. The importance of metabolic networks within the cell has been realized and they have been found interesting and complex enough to be the subject of intense mathematical and computational analysis. Thousands of components of a living cell are interconnected and are in the dynamic state [4,5]. Thus the functional properties of the cells are encoded by a set of complex intracellular interactions at molecular level [9,10]. The architecture of metabolic networks inside the cell has been found to have a highly heterogeneous scale-free topology [9]. This structure of network also allows simultaneous tolerance to random errors [4,5,8] and determines the key aspects of functionality and regulation [11].

The methods for determination of ensemble of interacting partners
A number for high-throughput methods for detecting protein interactions especially for microbial systems are now available. Protein-protein interactions play crucial roles in the execution of various biological functions. Interactions with other proteins are of extreme importance and can serve as highly informative hints for functional prediction: physical association between a novel protein and a well characterized one readily indicates that the former has a function related to that of the latter. Comprehensive analyses of protein-protein interactions thus constitute an integral part of functional genomics. The natures of proteins are far more complex than nucleic acids. It is apparent that a variety of methods would be needed to determine genome wide protein-protein interactions for any organism or even for subcellular organelles. These approaches can be divided into two categories: the top-down proteomic approach and the bottom-up genomic ones. The proteomic approaches are represented by the mass spectrometric and protein chip analyses of native protein complexes. The genomic approaches are where protein encoded in the genome of interest is expressed to evaluate interactions. These include the techniques such as yeast two-hybrid system and phage display.

Mass spectrometry of purified complexes
The purifications of protein complexes are achieved by immunoprecipitation using column of beads containing coupled antibodies to one of the components of complex. Often a variation of method is utilized where either a bait protein or tagged components are used. The constituent protein components of the complex is then separated by gel electrophoresis and subjected to analyses by mass spectrometry. Usually the excised gel bands (1-D gels) or spots (2-D gels) of proteins are subjected to in-gel proteolytic digestion (usually trypsin) and subjected to mass spectrometric identification. Two techniques are commonly used the first, Tandem affinity purification (TAP) allows rapid purification under native conditions of complexes, even when expressed at their natural level [12,13]. The TAP method requires fusion of the TAP tag, either N- or C-terminally, to the target protein of interest. The TAP method was initially developed in yeast but can be successfully adapted to various organisms. Its simplicity, high yield, and wide applicability make the TAP method a very useful procedure for protein purification and proteome exploration [12–14]. The second technique is, high-throughput mass spectrometric protein complex identification (HMS-PCI) protocol where a number of predicted interacting partner proteins are used as baits and associated proteins are identified. This method was also initially developed for yeasts [15]. Numerous protein complexes have been identified using this method including several new interactions in a number of signaling pathways and

in the DNA damage response [15]. The TAP approach has the advantage of working with the protein in its native state within the cells and the tagged proteins provides avenue to determine internal consistency. However, it has the disadvantage of missing the weak binding partners or weak association complexes and may not detect complex partners that associate with protein that undergoes transient modification subject to irreversibility on changing the conditions. The HMS-PCI approach may result in formation of some unnatural complexes and render detection of some partners impossible due to blockage of sites on protein as a result of association with the bait present in excess amount than in native state.

Protein chip
Protein chips coated with antibodies or chemicals help capture the protein or biomarkers. The captured proteins or biomarkers can be identified by using surface enhanced laser desorption/ionization (SELDI) mass spectrometry. The technique of SELDI is also referred as Protein Chip Technology [16]. Currently protein biochips can be utilized to profile and compare protein expression in normal and diseased states particularly in the area of cancer, infectious disease and toxicology [17]. The technique has the limitation at level of capture and the ability of ionization of low abundance proteins.

Yeast two-hybrid assay
Pairs of proteins to be tested for interaction are expressed as fusion proteins (or hybrids) in yeast with one protein being fused to a DNA-binding domain (bait) the other to a transcriptional activator domain (prey). The plasmids carrying the DNA sequences coding for the bait and prey proteins are introduced into the same yeast cell using a mating strategy [18–21]. The system is designed so that the bait and prey proteins interactions leads to transcriptional activation resulting in expression of a reporter gene that allows detection of the interaction. This technique enables detection of transient and unstable interactions and also *in vivo* interactions. However, this technique allows only pair-wise testing proteins and interaction of multiple proteins at a given time cannot be investigated. The investigation of proteins resident of sub-cellular organelles by this technique deviates from their native physiological setting and often provides results that are not true.

Correlated mRNA expression (synexpression)
The genes showing similar transcriptional response under different cellular conditions are grouped by systematic measurement of the mRNA levels. Usually such gene groups consist of a large number of genes that codes for physically interacting proteins [22]. In eukaryotes, highly coordinated expression of genes functioning in common processes is a widespread phenomenon. These sets of co-regulated genes ('synexpression groups') may be the key determinant facilitating evolutionary changes that lead to species diversity. The development of DNA micro-arrays for the simultaneous monitoring of thousands of transcripts has allowed global insights into gene expression [23]. The technique of synexpression has helped determine a number of interacting proteins for yeast cell-cycle analysis [22] and potential drug targets in rat brain by EST analyses [24]. The results indicate the usefulness of synexpression for exploring the functional genomics for information forthcoming from various genome projects. This is an indirect but *in vivo* technique and may cover a wide range of cellular conditions than other methods, but is a relatively inaccurate predictor of direct physical interaction.

Genetic interactions (synthetic lethality)
Two mutations are synthetically lethal if cells with either of the single mutations are viable but cells with both mutations are nonviable. Usually these genes are functionally associated and their protein products undergo physical interaction. Such genetic interactions are being studied in an all-versus-all approach in yeast [25]. As with the suppressor analysis, synthetic lethal mutations often indicate that the two mutations affect a single function or pathway. Synthetic lethal phenotypes are diagnostic of an interaction between the products of the two mutant genes in the cell [26]. Combinations of synthetic lethal mutants can also give information about what products are needed to complete a cellular process. This is also an indirect *in vivo* technique, amenable to unbiased genome-wide screening for protein-protein interaction.

Phage display
The phage display approach is a powerful tool, which was initially developed to detect antibodies [27]. Phage carrying V genes were displayed on the surface of fd bacteriophage which binds specifically to antigen and can be isolated by affinity chromatography. The use of the technique has expanded and phage libraries displaying random peptides can be used to select for ligands able to bind specific target molecules, to mimic non-proteinaceous ligands and also as a tool to map epitopes recognized by the antibodies. The display of the proteins or their functional domains provides a system for the analysis of structure-function relationships, novel peptide vaccines and others [28,29]. The technique suffers from limitation where interaction depends on conformation or needs correct folding for binding, which is not conferred by expressed truncated peptides.

In silico *predictions through genome analysis*
Three types of evidence may be found in a genome wide scan indicating interacting partners or potential interaction among the proteins in question, the first type of evi-

dence being, the proteins that are interaction partners are often encoded by conserved operons in prokaryotes [30–32]. The second type of evidence being, since interacting proteins interact together to provide a function or absence of a partner may impair the others and therefore in the genome wide scan the interacting proteins are likely to be either present or absent together and unlikely to be found alone in absence of other key partners [33]. The interacting proteins are likely to have a similar 'phylogenetic profile'. The third type of evidence being, the presence of fusion of seemingly unrelated proteins into a single polypeptide chain suggests a physical interaction [34,35]. The *in silico* techniques are rapid, relatively inexpensive and with the rapid sequencing of genomes their utility is ever expanding. However, *in silico* experiments requires assigning orthology relationship between proteins, and fails in prediction where such relationship cannot be established and this has led their application remaining confined mainly on prokaryotes.

A large number of data pertaining to interacting proteins generated by proteomic methods described above are available in public databases such as DIP http://dip.doe-mbi.ucla.edu/, BIND http://bind.ca, MIPS http://www.mips.biochem.mpg.de, WIT http://wit.mcs.anl.gov/WIT2 and IBIOINFORMATICS http://www.ibioinformatics.org/research.htm. A comprehensive listing of databases can be found at Molecular Information Agent http://mia.sdsc.edu/mia/html/bioDBs.html and also at human proteome organization http://www.hupo.org.

Metabolic and protein–protein interaction network
The cellular processes in a microorganism are seamlessly integrated through a complex network of constituent components and reactions. Systematic comparative mathematical analysis of the metabolic networks of several microorganisms have shown that, despite significant variation in their individual constituents and pathways, these metabolic networks have the same topological scaling properties and show striking similarities to the inherent organization of complex non-biological systems. This indicates that organization of metabolic networks is identical for all living organisms, and also has similar design principles such as being robust, error-tolerant and being scale-free. The metabolic network entails large-scale organization of interactions [36] among all cellular constituents and probably has a common blueprint among all organisms.

The robustness of various cellular processes is rooted in the dynamic interactions among its many constituents, such as proteins, DNA, RNA and small molecules. Scientific developments have improved our ability to identify the design principles that integrate these interactions into complex systems. Discovery of genomes now allow devel-

opment of organism-specific connectivity maps of metabolic and, to a lesser extent, other cellular networks, however, the large number and diversity of the constituents and reactions that form such networks, these maps are extremely complex, offering only limited insight into the organizational principles. The study of network properties has endowed ability to address in quantitative terms the structure of these cellular networks and has advanced our understanding of the generic properties of complex networks.

Properties of network
Complex networks occurring in nature share common global statistical features including "small-world" property of short-paths between two nodes and highly clustered connections. In many such networks there are few nodes with many more connections than the average node connections and such networks have been termed as "scale-free". In scale-free networks the fraction of nodes having k edges, p(k), decays as power law $p(k) \sim k^{\sim y}$ (y is between 2 and 3). Understanding of basic structural elements particular to each class of network is necessary to define the networks properly. Patterns of interconnections between the nodes have been termed as "network motifs". Understanding of the network motifs is expected to help uncover the structural design principle of the networks. Motifs are expected to define the universal classes of networks and unraveling them may help uncover the basic building blocks of most networks [37].

The classical random network theory was introduced by Erdös-Rényi, based on the assumption that each pair of nodes (that is, constituents) in the network is connected randomly with probability *p*, leading to a statistically homogeneous network in which, despite the fundamental randomness of the model, most nodes have the same number of links, *<k>*. In particular, the connectivity follows a Poisson distribution that peaks strongly at *<k>* implying that the probability of finding a highly connected node decays exponentially ($P(k) \approx e^{-k}$ for $k >> <k>$). On the other hand, empirical studies on the structure of the World-Wide Web, Internet and social networks have shown serious deviations from this random structure, demonstrating that these systems are described by scale-free networks for which $P(k)$ follows a power-law, $P(k) \approx k^{-\gamma}$ [5,9]. Unlike exponential networks, scale-free networks are extremely heterogeneous their topology being dominated by a few highly connected nodes (hubs), which link the rest of the less connected nodes to the system. The distinction between scale-free and exponential networks emerges as a result of simple dynamical principles. Understanding the large-scale structure of cellular networks can provide not only valuable universal structural information, but could also lead to a better understanding of the dynamical processes that generated these networks. Any

two nodes in the system can be connected by relatively short paths along existing links also known as "small-world character" is a general feature of many complex networks. In metabolic networks these paths correspond to the biochemical pathway connecting two substrates. The degree of interconnectivity of a metabolic network can be characterized by the network diameter, defined as the shortest biochemical pathway averaged over all pairs of substrates.

Large-scale, scale-free topology, governed by power law

An important consequence of the power-law connectivity distribution is that a few hubs dominate the overall connectivity of the network. The sequential removal of the most connected nodes leads to sharp rise in the diameter of the network eventually disintegrating the network into isolated clusters that are no longer functional. Scale-free networks also demonstrate unexpected robustness against random errors. A unique feature of metabolic networks, as opposed to non-biological scale-free networks, is the apparent conservation of the network diameter in all-living organisms. Within the special characteristics of living systems this attribute may represent an additional survival and growth advantage, as a larger diameter would attenuate the organism's ability to respond efficiently to external changes or internal errors. For example, if the concentration of a substrate were to suddenly diminish owing to a mutation in its main catalyzing enzyme, offsetting the changes would involve the activation of longer alternative biochemical pathways, and consequently the synthesis of more new enzymes, than within a metabolic network with a smaller diameter. An important known consequence of the inhomogeneous structure is the network's simultaneous tolerance to random errors, coupled with fragility against the removal of the most connected nodes.

The simultaneous emergence of an inhomogeneous structure in both metabolic [9,38] and protein interaction networks suggests that there has been evolutionary selection of a common large-scale structure of biological networks and indicates that future systematic protein-protein interaction studies in other organisms will uncover an essentially identical protein-network topology. The correlation between the connectivity and indispensability of a given protein confirms that, despite the importance of individual biochemical function and genetic redundancy, the robustness against mutations in organisms is also derived from the organization of interactions and the topological positions of individual proteins

Clustering and modularity and hierarchical organization

An important feature of scale-free networks is the existence of a few highly connected nodes that participate in a very large number of metabolic reactions. With a large number of links, these hubs integrate all substrates into a single, integrated web and therefore the existence of fully separated modules is prohibited in highly integrated module-free metabolic network organization. The analyses of the functional capabilities of metabolic networks by a number of approaches have indicated the existence of separable functional modules. The metabolic network of many organisms (for example,*E. coli*) is known to possess a highclustering coefficient that suggests modular organization.

The apparent dichotomy of presence of modular organization in contrast to scale-free nature of metabolic networks was reconciled by hierarchical organization. Average clustering coefficient defined as $C_i = 2n/k_i(k_i - 1)$, was calculated for 43 organisms, where n denotes the number of direct links (distinct substrates present in the metabolic network) connecting the k_i nearest neighbors of node i, is equal to 1 for a node at the center of a fully interlinked cluster, and it is 0 for a metabolite that is part of a loosely connected group. Therefore, C_i averaged over all nodes i of a metabolic network is a measure of the network's potential modularity. We found that, for all, The average clustering coefficient was determined for 43 organisms and was found to be an order of magnitude larger than that expected for a scale-free network of similar size suggesting that metabolic networks in all organisms are characterized by a high intrinsic potential modularity. The clustering coefficient of metabolic networks is independent of their size. To resolve the apparent contradiction of existence of modularity and scale-free network a simple heuristic model of metabolic organization was proposed and referred to as a "hierarchical" network. The architecture of such a network integrates a scale-free topology with an inherent modular structure. It has a power law degree distribution with degree exponent $\gamma = 1 + (\ln 4)/(\ln 3) = 2.26$, in agreement with $\gamma = 2.2$ and has a clustering coefficient C is about 0.6 (independent of the size of the network) is also comparable with that observed in metabolic networks.

In deterministic scale-free networks, the clustering coefficient of a node with k links follows the scaling law $C(k) \sim k^{-1}$. This scaling law quantifies the coexistence of a hierarchy of nodes with different degrees of modularity. The higher a node's connectivity, the smaller its clustering coefficient that asymptotically follows the $1/k$ law [5,10].The $C(k)$ function for the metabolic networks of 43 organisms was calculated to determine whether hierarchical organization is present in cellular metabolism. For each organism, $C(k)$ was represented by $C(k) \sim k^{-1}$, in contrast to the k-independent $C(k)$ predicted by both the scale-free and modular networks providing evidence for an inherently hierarchical organization. Hierarchical modularity reconciles all observed properties of metabolic networks [10]: their scale-free topology; system size-in-

dependent clustering coefficient; and the power law scaling of $C(k)$.

The organization of metabolic networks is likely to combine a capacity for rapid flux reorganization with a dynamic integration with all other cellular function. The system-level structure of cellular metabolism is approximated by a hierarchical network organization with seamlessly embedded modularity. In contrast to commonly assumed existence of a set of modules with a non-uniform size potentially separated from other modules, the metabolic networks has many highly integrated small modules that group into a few larger modules, that in turn is integrated into even larger modules. The scale-free topology has been found at many different organizational levels, genetic protein interaction and protein domain networks within the biological networks. For biological systems, hierarchical modularity is consistent with the notion that evolution may act at many organizational levels simultaneously. The accumulation of many local changes, which affect the small, highly integrated modules, could slowly impact the properties of the larger, less integrated modules. The emergence of the hierarchical topology through copying and reusing existing modules and motifs reminds the results of gene duplication and offers a key role to the modules that appeared first.

Utilities and future of network-structure knowledge
The prediction ability network structure
The network structures allow predictions using efflux measurements alone that could be experimentally verified for simple organisms. All necessary mechanistic details including kinetic parameters are unlikely to be available for all constituent members of large-scale metabolic networks, the ones that are generally present in the cells. The lack of such mechanistic details renders the dynamic modeling of networks extremely difficult if not impossible. The structure-oriented analyses requiring only network topology have been recently attempted. The method involves breaking down a network into elementary flux modes based on simple accounting of metabolic inputs and outputs and enables predicting possible properties of network [11,39]. Elementary mode analysis establishes a link between structural analysis and metabolic flux analysis. Elementary flux modes are defined as smallest subnetworks enabling the metabolic system to operate in steady state. The annotated genome sequences now provide the ability to reconstruct the metabolic network for the whole cell, which in turn provide the ability to modeling and computation of complex biological functions [31,40,41]. Whole cell constraints-based (and constraint-independent) *in silico* models have been constructed enabling prediction of optimal growth rates on common carbon substrates as well as adaptive evolution that are consistent with experimental data [42,43]. The knowledge

of metabolic pathways have also led to development of number of computational tools for different purposes, network construction as well as prediction, some of these utilities are PathFinder http://www.zsolutions.com/pathfind.htm, BioMiner http://www.microdiscovery.de/biominer_e.html, MetaTool http://www.bioinf.mdc-berlin.de/projects/metabolic/metatool/, FluxAnalyzer http://www.mpi-magdeburg.mpg.de/en/research/projects/1010/1014/1020/mfaeng/fluxanaly.html and FluxVision http://www.metabolic-explorer.com/flux.html.

Prediction about potential behavior of mutants
As pointed out that network structure provide the ability to predict the outcomes with minimum available kinetic details pertaining to constituent catalytic proteins. With the knowledge of network structure the mutants becomes a special case of such prediction ability. Mutants also enables test of the ability of predictions [11]. Predicting whether a mutant would be able to grow involves identifying elementary flux modes that do not use the missing reaction. [39].

Deriving understanding of the non-homologues metabolic pathway in organisms
In many microorganisms the classical metabolic pathway may be different. Understanding of the network structure and comparison of different modules may help predict the intermediate connections in the metabolic pathway. The proteomic methods may help in identifying the interacting partners and define an interactome or module, which would fit with the rest of the pathway components in the organism. For example, in a number of thermophiles, the classical Calvin-Benson pathway members for fixation of CO_2 are absent yet Rubisco is found in a number of members [44]. The proteomic approach may help identify the proteins interacting with Rubisco and thus help determine as to how 3-phosphoglycerate is utilized or how D-ribulose-1,5-bisphosphate is generated in these thermophiles in absence of other members of the classical cycle.

Production of biochemicals using microbes or microbe derived catalysts
Currently in chemical as well as in biochemical engineering the approach to produce chemicals are using pure components (reagents or substrates and catalysts). However, with the advent of knowledge of protein complexes, interactomes, modules and network it is increasing becoming clear that future bioengineering would depend on use of interactomes or truncated networks. It is evident that future bioengineering system would involve combination of interactomes or truncated modules of network based on systems design principles derived from the knowledge of system level understanding in biology. Metabolic engineering embraces the art of joining the differ-

ent modules or truncated interactomes using the underlying concept of applied system biology, which has been used in pathway engineering to produce new compounds and cell engineering for their use as catalysts in multi-step biotransformations [45,46]. Although such approaches have been utilized in metabolic engineering in past [45–47], however, with the emerging knowledge of global network properties, utilization of the embedded protocols connecting the interactomes will provide new dimensions to metabolic engineering.

Chanelling metabolite to prepare new compounds
In contrast to chemical and biochemical systems, cells have several properties. A metabolite is often at a branch point and in different organisms these branch points are different. Different proteins act upon the same metabolite in different organisms. It is likely that truncated interactomes from different organisms could be used to generate a novel metabolite from a given metabolite [45]. This will lead to a new path by stabilization of different but truncated interactomes from different organisms channeling the intermediates in a new direction.

Preparation of non-living miniature multi-step catalytic bags from microbial cells
The knowledge of metabolic network or modules and how to keep a portion of that intact, cocooned within the cell may render preparation of non-live cells that may not only allow efficient conversion but may enable extended functioning. Why go to bother of cost intensive purification and immobilization of enzyme, if the same cohort of enzyme trapped within non-live cells can provide for the same level of conversion. More importantly it is possible that some salvage pathway is retained within the non-live cells enabling efficient conversion using redundant path even if the catalytic ability of one pathway component in the main pathway is impaired [45–47].

Knowledge of networks as a tool for effective disruption of pathogenic microorganisms
Although an overwhelming number of microbes are useful for the mankind but a number of them are very harmful as well. They are the pathogens for human, livestock and plants. The fight against pathogenic microbes uses the arsenal of compounds that impairs their metabolic system. However, the tiny microorganisms are extremely versatile, something their multi-cellular counterparts are always lacking. The microbes undergo a selection process against the chemicals used for their destruction. They become resistant to chemicals used against them. The increased resistance to microbicidal agents had been a problem of great concern in medical and pharmaceutical community. The knowledge of network also confers the ability to attack the system in a much better and concerted manner. For harmful microbes disruption of their metab-

olism, based on their differences from hosts would be easier and effect with the advent of knowledge of their metabolic network structure and their constituent components. Disruption of metabolic network has been proven to be a successful therapeutic strategy for Mycoplasma [48], Candida and Leishmania [49] and against plant pathogen Ustilago [50].

Conclusions
A number of recent advances have been made in quantitative experimental approaches, includes new advances in mass spectrometry, electrophoresis with different fluroesent labels. These proteomic approaches, however, does allow only a static glimpse of abundant proteins. Whereas these powerful approaches has difficulty in identification of proteins at a given point enabling in static glimpse, inside the cell the networks not only identifies the right protein at right time but also enables performing functions so as to enable steady flow of metabolites. How such recognition and functionality is achieved amidst myriads of ensemble of proteins and other components? The current powerful methods are much away from that understanding. What is needed is equivalent of an ultrasonogram in embryology/neonatology at molecular level. What is needed is a dynamic picture with molecular details where each molecule could be identified separately but at the same time sufficient molecular details of interactions could also be sufficiently decipherable. Quantitative flux analyses [51,52] approaches to understand metabolome [53,54] or fluxome [47,55] have begun, which provides dynamic data and has the potential to provide dynamic information of interactome and modules and eventually that of the network. In future with superior organized databases are expected to give ability to incorporate information derived from the flux analyses and the kinetic details of the protein components in elemental flux analyses models [11,39]. Also, based on the static information for the same system analyzed under several different conditions may help construction of dynamic maps of the interaction networks. Each proteome analyzed under a given condition will be equivalent to a static picture but their combination will enable simulation of dynamic situations.

It is now gaining acceptance that understanding secrets of cell lies in understanding complex networks which a cell employs. The future designs should be putting artificial metabolic pathways and not linear addition of catalysts in bioreactors with all advantage of stability and channeling that is available inside the cells. Such approaches may deliver productivity and rapidity akin to cells and remain impervious to disturbances in operation variables. Stabilization of such truncated modules (or combining different interactomes) and their being in a state of continuous functionality are important issues. The knowledge gained

from studies on protein interaction and metabolic networks [19,25,56,57] is expected to result in system-level [6,7,58] understanding in biology, (a) system structure which comprises of gene interaction and biochemical pathways and the mechanisms by which such interactions modulate the physical properties of intra- and inter-cellular structures, (b) system dynamics that is how the system responds to external stimuli or operation variables, (c) the control methods, the mechanisms that systematically control the state of cell and last and most important (d) the design principles or methods [7]. The design methods are the underlying foundations based on which biological systems have derived their desired properties. The knowledge of cellular design methods would enable bioengineers to simulate rather than blind trial-and-error [7]. [A number of websites help in familiarizing with networks and systems biology and a few of them are: http://www.AfCS.org, http://www.sbml.org, http://www.cellml.org, http://www.stke.org, http://www.nd.edu/~networks/gallery.htm]

The protocols and modules are complementary, interwined and as systems become more complex protocols facilitate the layering of additional protocols involving feedback and signaling. Pathway engineering in bioreactors using underlying design principles would need deeper understanding of system level biology and its implicit layered protocols. As pointed out earlier, system level understanding of network not only makes better ability to produce new entities or devising new routes for production but also enables novel disruption strategies for treatment of pathogen perhaps using the hitherto known chemical entities. There is a realization that biological systems has large number of functionally diverse multifunctional interacting set of elements that interact together in a nonlinear fashion to produce coherent behavior [58]. This has led to approach understanding biological systems as systems, targeting their identification of their structure and unraveling their dynamics. As opposed to controlled complex systems where systems designer often provides target values for the system, in biological systems such targets are created by the system and renewed continuously by the system. Thus there is self-determined evolution in biological systems [58].

References

1. Villaverde A **Old bugs for new tasks; the microbial offer in the proteomics era** *Microb Cell Fact* 2002, **1**:1-4
2. Moxon R and Rappuoli R **Bacterial pathogen genomics and vaccines** *Br Med Bull* 2002, **62**:45-58
3. Chakrabarti S, Bhattacharya S and Bhattacharya SK **Biochemical Engineering: cues from cells** *Trends in Biotechnol* 2003,
4. Albert R, Jeong H and Barabasi AL **Error and attack tolerance of complex networks** *Nature* 2000, **406**:378-382
5. Jeong H, Mason SP, Barabasi AL and Oltvai ZN **Lethality and centrality in protein networks** *Nature* 2001, **411**:41-42
6. Csete ME and Doyle JC **Reverse engineering of biological complexity** *Science* 2002, **295**:1664-1669
7. Kitano H **Systems biology: a brief overview** *Science* 2002, **295**:1662-1664
8. Barkai N and Leibler S **Robustness in simple biochemical networks** *Nature* 1997, **387**:913-917
9. Jeong H, Tombor B, Albert R, Oltvai ZN and Barabasi AL **The large-scale organization of metabolic networks** *Nature* 2000, **407**:651-654
10. Ravasz E, Somera AL, Mongru DA, Oltvai ZN and Barabasi AL **Hierarchical organization of modularity in metabolic networks** *Science* 2002, **297**:1551-1555
11. Stelling J, Klamt S, Bettenbrock K, Schuster K and Gilles ED **Metabolic network structure determines key aspects of functionality and regulation** *Nature* 2002, **420**:190-193
12. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M and Seraphin B **A generic protein purification method for protein complex characterization and proteome exploration** *Nat Biotechnol* 1999, **17**:1030-1032
13. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM and Cruciat CM **Functional organization of the yeast proteome by systematic analysis of protein complexes** *Nature* 2002, **415**:141-147
14. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M and Seraphin B **The tandem affinity purification (TAP) method: a general procedure of protein complex purification** *Methods* 2001, **24**:218-229
15. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K and Boutilier K **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry** *Nature* 2002, **415**:180-183
16. Weinburger SR, Morris TS and Pawlak M **Recent trends in protein biochip technology** *Pharmacogenomics* 2000, **1**:395-416
17. Fung ET, Thulasiraman V, Weinburger SR and Dalmasso EA **Protein biochips for differential profiling** *Curr Opin Biotechnol* 2001, **12**:65-69
18. Fields S and Song O **A novel genetic system to detect protein-protein interactions** *Nature* 1989, **340**:245-246
19. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M and Sakaki Y **A comprehensive two-hybrid analysis to explore the yeast protein interactome** *Proc Natl Acad Sci U S A* 2001, **98**:4569-4574
20. Cagney G, Uetz P and Fields S **High-throughput screening for protein-protein interactions using two-hybrid assay** *Methods Enzymol* 2000, **328**:3-14
21. Kolonin MG, Zhong J and Finley RL **Interaction mating methods in two-hybrid systems** *Methods Enzymol* 2000, **328**:26-46
22. Qian J, Dolled-Filhart M, Lin J, Yu H and Gerstein M **Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions** *J Mol Biol* 2001, **314**:1053-1066
23. Niehrs C and Pollet N **Synexpression groups in eukaryotes** *Nature* 1999, **402**:483-487
24. Marvanova M, Toronen P, Storvik M, Lakso M, Castren E and Wong G **Synexpression analysis of ESTs in the rat brain reveals distinct patterns and potential drug targets** *Brain Res Mol Brain Res* 2002, **104**:176-183
25. Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B and Paoluzi S **A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules** *Science* 2002, **295**:321-324
26. Flower A, Osborne R and Silhavy T **The allele-specific synthetic lethality of prlA-prlG double mutants predicts interactive domains of SecY and SecE** *EMBO J* 1995, **14**:884-893
27. McCafferty J, Griffiths AD, Winter G and Chiswell DJ **Phage antibodies: filamentous phage displaying antibody variable domains** *Nature* 1990, **348**:552-554
28. Hill HR and Stockley PG **Phage presentation** *Mol Microbiol* 1996, **20**:685-692
29. Cabilly S **The basic structure of filamentous phage and its use in the display of combinatorial peptide libraries** *Mol Biotechnol* 1999, **12**:143-148
30. Dandekar T, Snel B, Huynen M and Bork P **Conservation of gene order: a fingerprint of proteins that physically interact** *Trends Biochem Sci* 1998, **23**:324-328
31. Overbeek R, Larsen N, Pusch GD, D'Souza M, Selkov N, Kyrpides N, Fonstein M, Maltsev N and Selkov E **WIT: integrated system for**

high-throughput genome sequence analysis and metabolic reconstruction *Nucleic Acids Res* 2000, **28**:123-125

32. Huynen M, Dandekar T and Bork P **Differential genome analysis applied to the species-specific features of *Helicobacter pylori*** *FEBS Lett* 1998, **426**:1-5

33. Huynen M and Bork P **Measuring genome evolution** *Proc Natl Acad Sci U S A* 1998, **95**:5849-5856

34. Enright AJ, Iliopoulos I, Kyrpides NC and Ouzounis CA **Protein interaction maps for complete genomes based on gene fusion events** *Nature* 1999, **402**:86-90

35. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO and Eisenberg D **Detecting protein function and protein-protein interactions from genome sequences** *Science* 1999, **285**:751-753

36. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S and Bork P **Comparative assessment of large-scale data sets of protein-protein interactions** *Nature* 2002, **417**:399-403

37. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U **Network motifs: simple building blocks of complex networks** *Science* 2002, **298**:824-827

38. Fell DA and Wagner A **Animating the Cellular Map** *Stellenbosch Univ Press* 2000,

39. Cornish-Bowden A and Luz Cardenas M **Metabolic balance sheets** *Nature* 2002, **420**:129-130

40. Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin E, Selkov E and Palsson BO **Metabolic modeling of microbial strains *in silico*** *Trends Biochem Sci* 2001, **26**:179-186

41. Karp PD, Riley M, Saier M, Paulsen IT, Collado-Vides J, Paley SM, Pellegrini-Toole A, Bonavides C and Gama-Castro S **The EcoCyc Database** *Nucleic Acids Res* 2002, **30**:56-58

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

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

44. Imanaka T and Atomi H **Thermostable archeal Rubisco with novel structure** *In Recent Research Developments in Biotechnol. Bioengg (Edited by: Bhattacharya SK, Chakrabarti S, Mal TK) Kerala, Research Signpost* 2003, 89-108

45. Du L, Sanchez C and Shen B **Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules** *Metab Eng* 2001, **3**:78-95

46. Strohl WR **Biochemical engineering of natural product biosynthesis pathways** *Metab Eng* 2001, **3**:4-14

47. Stephanopoulos G **Metabolic fluxes and metabolic engineering** *Metab Eng* 1999, **1**:1-11

48. Dhandayuthapani S, Blaylock MW, Bebear CM, Rasmussen WG and Baseman JB **Peptide methionine sulfoxide reductase (MsrA) is a virulence determinant in *Mycoplasma genitalium*** *J Bacteriol* 2001, **183**:5645-5650

49. Garami A and Ilg T **Disruption of mannose activation in *Leishmania mexicana*: GDP-mannose pyrophosphorylase is required for virulence, but not for viability** *EMBO J* 2001, **20**:3657-3666

50. Kim SH, Virmani D, Wake K, MacDonald K, Kronstad JW and Ellis BE **Cloning and disruption of a phenylalanine ammonia-lyase gene from *Ustilago maydis*** *Curr Genet* 2001, **40**:40-48

51. Mankoff DA, Shields AF, Graham MM, Link JM, Eary JF and Krohn KA **Kinetic analysis of 2-[carbon-11] thymidine PET imaging studies: compartmental model and mathematical analysis** *J Nucl Med* 1998, **39**:1043-1055

52. Wendisch VF, de Graaf AA, Sahm H and Eikmanns BJ **Quantitative determination of metabolic fluxes during coutilization of two carbon sources: comparative analyses with *Corynebacterium glutamicum* during growth on acetate and/or glucose** *J Bacteriol* 2000, **182**:3088-3096

53. Jiao Z, Baba T, Mori H and Shimizu K **Analysis of metabolic and physiological responses to gnd knockout in *Escherichia coli* by using C-13 tracer experiment and enzyme activity measurement** *FEMS Microbiol Lett* 2003, **220**:295-301

54. Prasad Maharjan R and Ferenci T **Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*** *Anal Biochem* 2003, **313**:145-54

55. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, Wuthrich K and Bailey JE **Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism** *J Bacteriol* 1999, **181**:6679-6688

56. Resing KA **Analysis of signaling pathways using functional proteomics** *Ann N Y Acad Sci* 2002, **971**:608-614

57. Sanford K **Genomics to fluomics and physiomics-pathway engineering** *Curr Opin Microbiol* 2002, **5**:318-322

58. Kitano H **Computational systems biology** *Nature* 2002, **420**:206-210