Several decades of research in biochemistry and molecular biology have been devoted for studies on isolated enzymes and proteins. Recent high throughput technologies in genomics and proteomics have resulted in avalanche of information about several genes, proteins and enzymes in variety of living systems. Though these efforts have greatly contributed to the detailed understanding of a large number of individual genes and proteins, this explosion of information has simultaneously brought out the limitations of reductionism in understanding complex biological processes. The genes or gene products do not function in isolation in vivo. A delicate and dynamic molecular architecture is required for precision of the chemical reactions associated with “life”. In future, a paradigm shift is, therefore, envisaged, in biology leading to exploration of molecular organizations in physical and genomic context, a subtle transition from conventional molecular biology to modular biology. A module can be defined as an organization of macromolecules performing a synchronous function in a given metabolic pathway. In modular biology, the biological processes of interest are explored as complex systems of functionally interacting macromolecules. The present article describes the perceptions of the concept of modularity, in terms of associations among genes and proteins, presenting a link between reductionist approach and system biology.

Keywords: Genome, Modular biology, Molecular machines, Metabolon, Proteome, STRINGS, Systems biology

Introduction

Discovery of isolatable enzymes in early 20th century led to the consideration of living cells as ‘bags of soluble enzymes’. The precise chemical reactions associated with all life processes were thought to occur by chance encounters of the metabolites and intermediates with randomly distributed intracellular enzymes. Biochemical research on isolated purified enzymes and proteins strengthened the reductionist approach in biology, putting forth the philosophy that understanding the individual enzymes was sufficient to understand the metabolic functions of the organism. The structural associations and the cellular integrity were thought as not relevant for it. Not much attention was paid to the differences between metabolic rates in vitro and in vivo in the cells or to the fact that intracellular milieu is water-limited, resulting in macromolecular crowding. Ease of measurement of biochemical reactions in vitro overpowered the idea of special properties such as metabolite channeling, conformational diversity and fine regulation of enzyme activities by reversible association-dissociation being conferred by interacting systems, which might be significant in vivo.

There have been several observations which indicated that soluble proteins in the cytoplasm exist in the organized state. The existence of an intricate network of proteins called as “macromolecular lattice” was reported by high-voltage electron microscopy. This lattice was thought to function in localizing proteins involved in the metabolic processes. The intracellular water molecules were also shown to be structurally organized resulting in low diffusion coefficients of the proteins in living cells as compared to aqueous solutions. Observations on supramolecular organization of sequential enzymes along with the kinetic and physicochemical analysis of multi-enzyme complexes supported the idea that in vivo super complexes of intracellular proteins, which could also be associated with large cellular structures like membranes, might exist.

Interestingly, in a review on complexes of sequential enzymes, Sere represented schematically about 500 metabolic reactions, revealing that majority of the intermediates in these pathways have just one fate or single use in the cell. It would be in vain, if these intermediates have to diffuse randomly all over the cell to reach the destined enzymes. It is rather
obvious that the sequential enzymes would associate to facilitate the channeling and micro-compartmentation of metabolic intermediates. This perception about organization of metabolic pathways has been further advanced with more focused investigations for different metabolic pathways.

**Metabolons and Molecular Machines**

The concept of ‘metabolon’ was first introduced by Srere\(^{15}\) to describe supramolecular complexes of sequential enzymes. Parallel studies revealed a tendency of multimolecular organizations and metabolite channeling among sequential enzymes of several metabolic pathways such as Krebs cycle\(^{13}\), glycolysis\(^{14}\), DNA repair and replication\(^{15}\), Calvin cycle\(^7\) and other metabolic pathways in plants\(^8\). Conventional biochemical techniques, such as co-purification, co-sedimentation, fluorescence techniques, and kinetic analysis were used to demonstrate the organization at macromolecular levels.

This concept of ‘metabolon’ was widened further, when the term “molecular machines” was coined to describe the complexes of interacting proteins\(^{16}\). The cell was described as a factory containing an elaborate network of interlocking assembly lines, comprising the large molecular machines. The machines involved in many vital processes such as DNA replisome\(^7\), chromatin remodeling\(^8\), nuclear pore complex machinery\(^9\), ribosomes\(^{10}\), chaperone\(^{11}\), proteasome\(^{12}\), respirasome\(^{13}\), and photosynthesome\(^{14,15}\) have been explored. Advanced methods in proteomics and electron microscopy using cryo technology and digital imaging have been recommended to understand the molecular architecture and dynamics of some of these machines\(^{16,17}\).

**Concept of Modularity in Biological Systems**

With realization of the need for supramolecular organization in the biological systems, a unique concept of modularity is emerging in biology to describe different levels of organization and relate them to intrinsic properties of biological systems. Hartwell et al.\(^{28}\) advocated the need for progressing from the “molecular to modular biology” and hypothesized the concept of a module. A module is defined as a discrete structural and functional entity, an ensemble of physically interacting proteins that deals with a specific metabolic process and can be co-isolated. The composition, location and function of a module may vary depending on the cellular requirements. Thus, modules symbolize a dynamic entity of a group of molecules that occur and function together. It represents a minimum critical level of organized macromolecular interactome and is anticipated to involve a small fraction of the cell components, accomplishing a relatively autonomous function\(^{29}\). Specific spatio-temporal composition, organization and interactions of the underlying components form characteristics of a module. The functional properties of a module arise from the properties of the underlying components and their interactions.

A striking advantage of considering modularity in metabolic networks lies in its structural and functional flexibility. Modularity offers freedom to assume that the same molecular species of protein/enzyme may belong to different modules at different time and in different intracellular locations and may differ in function depending on the environment. Such functional segregation of proteins in different modules also explains the term ‘moonlighting’ of a protein, a phenomenon where a single protein exhibits different functions\(^{29}\). The molecules of the same enzyme present in different modules may or may not be structurally identical. Small structural differences in the enzyme molecules may be significant for their involvement in different modules and also in isolating different intersecting metabolic pathways functioning simultaneously in the cell.

The function of a module is expected to be quantitatively or qualitatively regulated by various physical and physiological factors. Modules may have different lifetimes and can be insulated or interconnected with each other. Rearrangements or conformational changes of the components within a module may have significant effect on its function. The flexibility of modular structure can be significantly employed for sensitive regulation of biochemical activities. Modules, representing the restricted sub-networks of the complex biological systems help to deduce the principles underlying the cellular physiology such as signal transduction\(^{28}\).

Modularity can decide scope for “Selection for a function”, thereby playing important role in evolution\(^{30}\). Mutations, resulting in improvement of a protein, but impairing its interaction with other proteins would not be favored, if there is no desired advantage to the whole system. Such mutations could be counter-balanced and hence tolerated by subsequent changes in the respective modules. Thus, modularity might play an important role in...
evolvability of living systems. Though modularity has been strongly supported at genomic level, direct experimental evidences for the existence of modules in physiological context are inadequate. Limitation for identification and isolation of structurally and functionally intact modules is the main hurdle that needs to be overcome.

**Modules in the Genomic Context**

Advances in genomics, proteomics and bio-informatics also advocate the concept of modules or networks. A protein has now been defined as an element in the network of interactions. Several software tools have been developed to predict associations and networking of proteins. Functional links in networks have been identified, based on ordering of related genes on genome sequences, following the notion that gene proximity might result from the need to associate genes that are co-regulated and hence the potential interacting partners. Genetic associations can be predicted using various parameters like gene fusion, phylogenetic profile, co-expression, conserved neighborhood or the study of interologues. The in silico search tools, based on the individual or integrative approaches may predict a network of interactions among proteins.

An example of such predicted genomic associations among Calvin cycle enzymes using the search tool STRING (Search Tool for Recurring Instances of Neighboring Genes) is shown in Fig. 1. These interconnections among the Calvin cycle enzymes and thylakoid membrane proteins in genomic context support the concept of photosynthetic module, for which evidence has also been obtained by co-isolation of various components using conventional biochemical procedures. In this study, a thylakoid fraction having all the components of electron transport chain and ATP synthase, as well as other Calvin cycle enzymes, is shown in the network diagram.

---

**Fig. 1**—Network diagram for predicted genomic associations of Calvin cycle enzymes [The diagram shows the predicted genomic associations of five sequential enzymes of Calvin cycle viz., phosphoribokinase, phosphoglyceratekinase and glyceraldehyde-3-phosphate dehydrogenase (shown bold in boxes). The search tool for recurring instances of neighboring genes (STRING) available on-line (http://www.bork.embl-heidelberg.de/STRING), which uses gene neighborhood, gene fusion, co-occurrence, co-expression, text mining for prediction of association is used. Dashed lines show association with proteins of non-photosynthetic function and continuous lines show association of proteins involved in photosynthesis. The length of lines has no bearing on the strength of association. The associations involving components of photosynthetic apparatus have functional significance and have been isolated as photosynthesosome using conventional biochemical techniques. Photosynthesosome represents a photosynthetic module. The network diagram shows the interacting partners in photosynthesosome and predicts some more putative associations. However, the validity and functional significance of such predicted associations needs to be tested biochemically]
the five sequential enzymes of Calvin cycle viz., phosphoribosiltransferase, phosphoribulokinase, RuBP carboxylase, 3-phosphoglycerate kinase and glyceraldehyde-3-phosphate dehydrogenase, along with the soluble proteins such as phycobilisomes, ferredoxin-NADP reductase and CP12 have been isolated. In vitro studies on this fraction have shown photophosphorylation-dependent CO₂ fixation. This fraction, therefore, can be viewed as a functional photosynthetic module or photosynthesome. This super-molecular organization facilitates the coupling between photosynthetic electron transport and CO₂ fixation.

When analyzed in the genomic context with STRING, the proteins like RuBP carboxylase, 3-phosphoglycerate kinase and glyceraldehyde-3-phosphate dehydrogenase have shown multiple associations with proteins of photosynthetic as well as non-photosynthetic functions such as sulfite reductase, respiratory enzymes, groEL, trigger factor and superoxide dismutase (Fig. 1). This indicates a possibility of interactions or cross talks among different modules, which needs to be further established. STRING tool, however, has limitations as it could not detect some of the experimentally verified associations such as CF1 ATP synthase-Rubisco, phosphoribulokinase-glyceraldehyde-3-phosphate dehydrogenase, and RuBP carboxylase-phosphoribulokinase.

Transcription profiling using microarrays, DNA chips or serial analysis of gene expression allows simultaneous monitoring of expression of thousands of genes under various experimental conditions. Snel et al. have identified a level of organization of proteins that lies between pair-wise interactions and the complete network by studying the properties of large number of interactions in the networks obtained by the in silico prediction-based analysis of genomic associations. The topology of such networks indicates the presence of small sub-clusters of homogenous function, thus representing the functional modules. Genes that are regulated together can be viewed as associated genes, representing a network of activities or a module. However, there could be many different cascades involving different gene families, which make the in silico prediction of modules difficult. Since the predictions are often global in nature, functional significance of involvement of an isoform/isozyme in the specific associations is difficult to envisage.

Isolation and characterization of modules, therefore, would cast light on the finer differences in molecular architecture of the modules, with reference to space and time. Studies on mapping of protein-protein interactions have been successfully carried out using several techniques such as glutathione-s-transferase (GST) fusion proteins, tandem affinity purification (TAP) or yeast two-hybrid systems. Microsequencing of gel purified proteins, mass spectrometry coupled with MALDI and use of protein chips are crucial to detect the interacting partners of the proteins. The high throughput methods dealing with the large proteomes have revealed organized biochemical networks. However, these large-scale screening procedures often face the problem of false positives and negatives. Efforts have been made to overcome these drawbacks by experimental validation and integrated 'omic' approaches.

**Integrating Modularity to Systems Biology**

Study of individual modules provides only a fragmented picture. Thus for a more complete picture of any biochemical process, it is necessary to understand the ways, in which modules can be integrated into larger networks. In "systems biology", a network of complexes comprising different metabolic paths is considered as a system to understand function and regulation. Systems biology allows integration of datasets obtained by the individual genomic and proteomic approaches. Modularity of biological activities can be extended to systems level to obtain information on directionality of interactions in the modules and depict the functional outcomes of the modules. It would help in implementation of systems approach, as the modularity deals with proteins that are functionally inter-dependent. With this in focus, it is easy to screen and select from a large proteome, the targets for functional disruption and to monitor the effect on the whole system.

Using a high-quality network obtained by integrating five different datasets, Han et al. have identified “date and party hubs”, a small, but significant proportion of proteins that interact with many partners. The “date hubs” represent connections between the modules, while the “party hubs” function inside the modules at lower level of organization of proteome. Biological networks have been found to be extremely sensitive to the targeted removal of hubs. Thus, understanding the functions at modular and systems level will provide an improved vision for research in biology.
Prospects and Problems in Modular Biology

Understanding the regulation of biochemical networks as viewed by specific spatio-temporal associations of participating components can answer several fundamental questions in biology. For example, it explains why several variants of the same protein are required in the cell. Small structural differences in the enzyme molecules will be significant for their involvement in different modules ascertaining the need for presence of isozymes and isoforms produced by multi-gene families, alternate splicing of genes, single nucleotide polymorphisms (SNPs) or variety of post-translational modifications. The necessity for interaction with other proteins may drive the need for such minor variations without affecting the functionality.

The consortium of biochemical reactions at modular level supports the requirement of conformational diversity and also intrinsically unstructured proteins\(^1\), which in turn confers structural and functional flexibility to proteins. Since modularity allows dynamism of biochemical networks, it would be pivotal in offering robustness, a fundamental feature of the complex, as well as evolvable biological systems\(^5\).

Modules result in caging and isolation of intermediates to avoid unwanted reactions in living cells. Such microenvironment explains how several competing and contrasting metabolic reactions occur simultaneously in crowded atmosphere in vivo. Metabolic regulation also needs to be viewed from a different perspective, in the backdrop of modularity. Differential degree of recruitment of modules would explain the phenomenon of up and down regulation of metabolic processes as also the metabolic compensation.

While thinking of the prospects one should not overlook the limitations for undertaking studies on the modules experimentally. Since the interactions among components of modules are weak, it is often difficult to isolate them as a single functional unit. The dynamic nature of these interactions poses additional problem. Understanding modules in genomic context using high throughput technologies may or may not suggest exact physical associations.

Purification techniques such as TAP, in conjunction with use of cross linkers and mass spectrometry may be useful in mapping the physical interactions between nearest neighbors in the modules. However, it is imperative to demarcate between the native-specific associations of proteins against the non-physiological, non-specific aggregates.

Identification of exact contact points and interacting surfaces of macromolecules present in the modules and understanding their functional significance is the next big challenge technologically. The tools in cryoelectron microscopy like single particle 3-D reconstruction, 2-D crystallography and electron tomography will facilitate visualization and localization of the modules in the biological systems. Study of the modules in genomic context would overcome the limitations of physical isolation of modules due to weak associations of the components, bringing in light many putative interactions and cross talks among the modules, belonging to different intersecting pathways. Modular biology will thereby amalgamate the conventional biochemical studies with most recent technology to understand the hierarchy of networks and molecular architectures of biological nanomachines.

Thus, in summary, modularity aims to decipher the complex biochemical networks by resolving them in smaller sub-networks. The slow, but subtle transition from conventional molecular biology towards modular biology will move the focus of research in the biology from studies of individual genes and proteins to more complex sets of molecules that interact to form functional modules. Elucidating the network of the modules is a challenge to upcoming technologies, whereby different structural, biochemical, biophysical and molecular approaches could be combined to get a descriptive and complete picture of functioning of a cell. Modular biology thus offers a new, intellectually challenging perception of metabolism, an avenue for research that would stretch the limits of yesteryears and modern technologies in biology.

Acknowledgement

DD was a recipient of Department of Atomic Energy graduate fellowship.

References
yeast proteome by systematic analysis of protein complexes. Nature 415, 141-145


