Advances in Computational Research, ISSN: 0975–3273, Vol. 1, Issue 1, 2009, pp. 5-8

Reaction Motifs- Assemblies and Significance

[1]Harshal Chaudhari and [1*]Somnath Tagore
[1]Department of Biotechnology & Bioinformatics, Dr D Y Patil University, Plot 50, Sec 15, CBD Belapur, Navi Mumbai 400614, India
*Corresponding Author Email: somnathtagore@yahoo.co.in

Abstract: The vast amounts of biological data present in standard repositories are the heart of bioinformatics today. This has been possible due to the sequence alignment, microarray, etc approaches over the years. The huge biochemical networks have certain assemblies of modules called the reaction motifs. There are different types of such motifs in a network and it is of foremost importance to identify such motifs and get insights of their functions and regularities. It is also required study the intensity of occurrences of these motifs to establish certain evolutionary and functional relationships in and amongst pathways. The inference that these motifs have been selected for function rests on the idea that their occurrences are significantly more frequent than random. Such motifs have not only been identified in a wide range of networks across many scientific disciplines and are suggested to be the basic building blocks of most complex networks.

Keywords: bioinformatics, sequence alignment, microarray, reaction motifs, complex networks.

1 INTRODUCTION

In biological organisms, networks of chemical reactions control the processing of information in a cell. A general approach to study the behavior of these networks is to analyze modules that are frequently observed in natural systems. Numerous network motifs that perform computational tasks have been discovered in biochemical reaction networks. Reaction networks are able to compute Boolean operations and implement simple binary computers **[1]** Cell signaling networks are known to exhibit parallelism, the integration and amplification of signals, bistable behavior and hysteresis through feedback and memory **[2]**.

The science of systems biology has the aim of understanding the functional constraints and design principles of biological networks. Alon and colleagues were the first to introduce the notion of "motifs" in biological networks **[3]**. Motifs are small patterns observed to recur throughout a network, with frequencies statistically higher than expected in random networks of similar connectivity parameters. Since the introduction of this concept, motifs have been reported in many biological networks: metabolic, signaling pathway, protein-protein interaction, and ecological networks amongst others **[4, 5]**. Moreover, the prevalence of motifs is often considered as evidence for evolutionary selection, for implementing a *specific* function **[6]**. Motifs are believed to be building blocks of the functional architecture of a biological network.

A major issue concerning the study of biochemical networks is the problem of their organization. Several attempts have been made to decompose the metabolic network into parts. These parts are called modules or “motifs”. Modules were first coined by Heartwell *et al* who outlined the general features of a module but did not give a specialized definition reguarding them. Motifs can be considered as the building blocks of the network. In terms of metabolic networks the definition of motifs could be based on the partition of the metabolic network into metabolic pathways as present in the databases. The advantage of such a technique is it helps the better understanding of metabolism. Graph based methods ranges from the study of graphs showing the simple connectivity amongst the metabolites to studying the maximization of the criteria expressing the modularities.

2 TYPES OF MOTIFS IN BIOCHEMICAL NETWORKS

A. Single-input modules (SIM): SIM’s contain a source node that is responsible for the transmission of inputs at the top and the intermediate and target nodes are present below the source nodes. The source node individually accesses all of these nodes. (Figure 1-a).

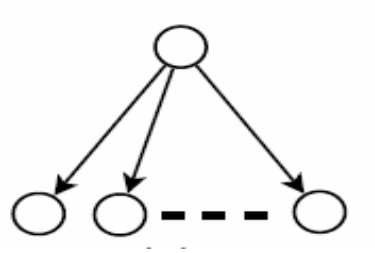

Figure 1-a: Single-input modules

B. Multiple-input modules (MIM): MIM’s contain many source nodes that can access same or different intermediate and target nodes at a specific given time **[7]**. All the nodes present are interconnected in a network. (Figure 1-b).

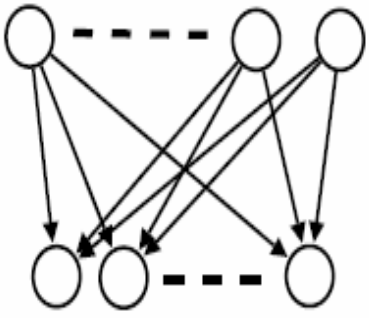

Figure 1-b: Multiple-input modules

C. Feed-forward loops (FFL): FFL’s contain the target node at the top, the intermediate node at the bottom left and the target node at the bottom right. (Figure 1-c). The top node can access both the nodes **[8]**.

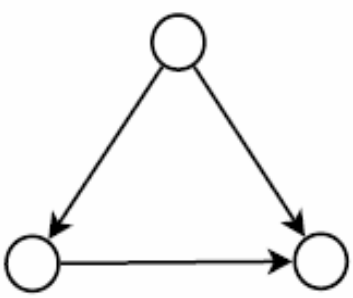

Figure 1-c: Feed forward loops

D. 3-Cycle: It is a three node directed cyclic graph. (Figure 1-d).

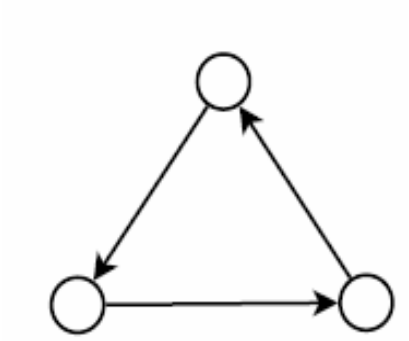

Figure 1-d: 3-cycle

E. Bifan: SIM, MIM, and Bifan are two-layered graphs with edges from nodes in top- to bottom-layer **[9]**. A Bifan is a MIM with exactly 2 parent and 2 child nodes. (Figure 1-e).

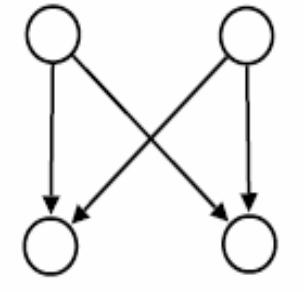

Figure 1-e: Bifan motif

3 ASSEMBLIES OF MOTIFS

Kashtan *et al.* observed that regulatory networks contain multi-output FFL generalizations (see Figure 2(a)) in frequencies much higher than multi-input (Figure 2(d)) and multi-intermediate (Figure 2(f)) generalizations **[10]**.

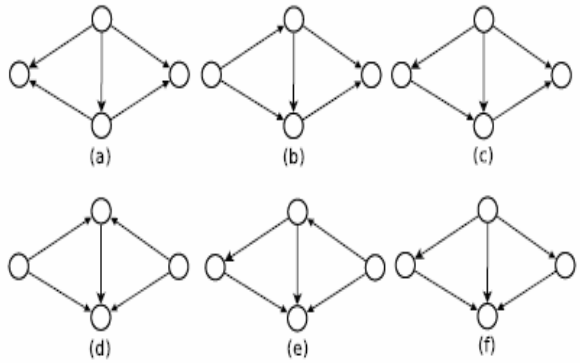

Figure 2: Self-assemblies of two FFL's

Here the varied frequencies of motifs found due to the occurrence of FFL's around the hubs. i.e. the nodes share a common edge. SIMs and MIMs are variable sized sub graphs. Alon *et al* defined the dense overlapping regulon (DOR) as a two-layered sub graph with not necessarily complete connections between them. MIMs are special DORs, a concept that arose as a generalization of the Bifan (Figure 1(e)) sub graph **[11]**. These Bifans were observed to be present in large numbers in biological networks. However, some investigators fail to impose the criterion of maximality while counting MIMs. This can lead to significant inflation of counts . A maximal MIM with m parents and n children contains $[2m - (m + 1)] \times [2n - (n + 1)] - 1$ easily enumerable nonmaximal "subMIMs". Our definition of a Bifan ensures that we are only counting (maximal) MIMs that contain 2 parents and 2 children. Counting subMIMs as Bifans will combinatorially increase their counts, as each maximal MIM will contribute to $mC2 \times nC2$ Bifans. It is also studied that biparate cliques appear sufficiently dense in biparate graphs **[12]**.

4 CURRENT STRATEGIES FOR MOTIF IDENTIFICATION

4.1. **Using probability**:

4.1.1. Network motifs also have uncertainties.

4.1.2. A stochastic network can be thought of as coming into being by embedding a family of mutually similar interconnection patterns (subgraphs) in a background random ensemble with a probability alpha.

4.1.3. The set of patterns defines a foreground stochastic network motif and is described by a probability matrix $\Theta_1 = (\theta_{ij})_{n\times n}$, $0 \le \theta_{ij} \le 1$. is the probability that node i and node j are connected. The background ensemble is characterized by the degree distributions of the given stochastic network (Θ_0).With such a mixture model, the network motif can be recovered by fitting the stochastic network with a foreground motif and a suitable background ensemble **[13]**.

4.2. **Based on Hardness results**:

4.2.1. Every motif is given color. Hence each motif while identification be recognized on the basis of color.

4.2.2. It may assume the graph is connected and all vertices have colors that appear in the motif.

4.2.3. Otherwise, we preprocess the graph throwing away all the vertices having no color appearing in the motif and solve the problem in each component of the resulting graph.

4.2.4. Given a motif and a labeled graph, checking for the motif and deciding whether the motif occurs in the graph or not.

4.2.5. If 'yes' then the motif is validated and the type is identified, if 'no' then the search is continued until motifs are found **[14]**.

4.3. **Topological approach**:

4.3.1. This method was suggested by Schuster *et al.* and Holme *et al.* It works by removing the metabolites that participate in many reactions (nodes with a high degree).
4.3.2. Create modules by removing entities.
4.3.3. Remove nodes that exhibit large betweeness centrality. The most central node has to be identified and removed.
4.3.4. Number of shortest paths between pairs of nodes that pass through the considered node.
4.3.5. The highest degree or the central node is checked for interactions within other reactions in the same pathway or other pathways.
4.3.6. Found results are validated and fixed **[15]**.

.

5 APPLICATION IN PATHWAY ENGINEERING

Many networks have been shown to share global statistical features, such as the ''small world'' property of short paths between any two nodes and highly clustered connections **[16]**. It has also been shown that many networks are ''scale-free'' networks, in which the node degrees follow a power-law distribution. Recent studies have shown that many networks contain a small set of ''network motifs,'' that is, patterns of interconnections occurring in networks at numbers that are significantly higher than those in randomized networks that are uniformly drawn from the networks with the same degree distributions as the original networks . These network motifs may define universal classes of networks in that similar motifs have been found in a wide variety of networks, ranging from the World Wide Web to the electronic circuits, from the transcriptional regulatory networks of *Escherichia coli* to the neural network of *Caenorhabditis elegans*. The research on network motifs is therefore promising in uncovering the basic building blocks of most complex networks **[17]**.

Two seminal studies recently have shown that topological networks indeed contain statistically significant patterns indicative of biological functions. These *motifs* are patterns that occur more frequently in the observed network than expected in a suitable null ensemble. The motifs found so far have been identified because they occur *identically* at different positions in a network. If network evolution is a stochastic process, however, functionally related motifs do not need to be topologically identical. Hence, the notion of a motif has to be generalized to a stochastic one as well. Variations arise because of uncertainties in the network data, or, more importantly, because some of the interactions can change without affecting the functionality of the motif. This ''noise'' is an important characteristic of biological systems, familiar from sequence analysis, where one searches for local sequence similarities blurred by mutations, insertions and deletions, rather than for identical subsequences. It leads us to the notion of a *probabilistic motif* in which each link occurs with certain likelihood. Probabilistic motifs arise as consensus from finding a family of ''sufficiently'' similar sub-graphs in a network **[18]**.

Motifs undergo repetitions throughout the network and are thought to be conserved through evolution. It has also been hypothesized that similar repeated elements may also be functionally related. So these studies imply that if we find such elements in the network we can establish evolutionary and functional relationships among pathways even if they may possess different topologies and may not seem to be related in the beginning.

6 CONCLUSION

It has been invariably studied that motifs are building blocks of a network and hence have to be studied extensively. Until now no specific definition of a motif has not been formulated hence studies in this field is very important. Though studies have been incited but unfortunately they are not sufficient. It is evident that motifs are biologically significant from structural and functional point of studies and efforts further studies would in turn add up to metabolic studies and engineering.

REFERENCES

[1] Arkin, A. and Ross, J. (1994). 'Computational Functions in Biochemical Reaction Networks.' *Biophysical Journal,* **67(2)**, 560–78.
[2] M. Velasco, J. I. Leguina, and A, (2002). 'Lazcano. Molecular evolution of the lysine biosynthetic pathways.' *J. Mol. Evol.*, **55**, 445–459.
[3] Bhalla, U. and Iyengar, R. (1999) 'Emergent Properties of Networks of Biological Signaling Pathways' *Science,* **283**, 381– 387.
[4] Bray, D. (1995).'Protein Molecules as Computational Elements in Living Cells.' *Nature*, **376(6538)**, 307–12.
[5] D. Wolf and A. Arkin, (2003) 'Motifs, modules and games in bacteria.' *Curr. Opin. Microbiol.*, **6(2)**, 125–134.
[6] Kashtan N, Itzkovitz S, Milo R, Alon U,(2004). 'Topological generalizations of network motifs.' *Phys Rev E Stat Nonlin Soft Matter*, **70(3 Pt 1),** 031909.
[7] K. Voss, M. Heiner, and I. Koch, (2003) 'Steady state analysis of metabolic pathways using Petri nets.' *In Silico Biol.*, **3(3)**, 367–387.
[8] Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA,(2002). 'Transcriptional regulatory networks in Saccharomyces cerevisiae.' *Science*, **298(5594)**, 799-804.
[9] Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M,(2004) 'Genomic analysis of regulatory

network dynamics reveals large topological changes.' *Nature*, **431(7006)**, 308-312.

[10] Máayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, Kershenbaum A, Stolovitzky GA, Blitzer RD, Iyengar R,(2005).'Formation of regulatory patterns during signal propagation in a mammalian cellular network.' *Science*, **309**, 1078-1083.

[11] Mangan S, Alon U,(2003) 'Structure and function of the feed-forward loop network motif.' *Proc Natl Acad Sci USA*, **100**, 11980-11985.

[12] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U,(2002) 'Network motifs: simple building blocks of complex networks.' *Science*, **298(5594)**, 824-827.

[13] Sauro, H. and Kholodenko, B. (2004) 'Quantitative Analysis of Signaling Networks.' *Progress in Biophysics and Molecular Biology*, **86**, 5–43.

[14] Steijaert, M., Liekens, A., ten Eikelder, H.,and Hilbers, P. (2007). 'Multiple Functionalities of Biochemical Reaction Networks.' *International Conference on Systems Biology (ICSB 2007).*

[15] Steijaert, M.,ten Eikelder, H., Liekens, A., Bosnacki, D., and Hilbers, P. (2008). 'Stochastic Switching Behavior of a Bistable Auto-phosphorylation Network.' *12th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2008).*

[16] Shen-Orr SS, Milo R, Mangan S, Alon U,(2002). 'Network motifs in the transcriptional regulation network of Escherichia coli.' *Nature Genetics*, **31**, 64-68.

[17] Tyson, J., Chen, K., and Novak, B. (2003). 'Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathways in the Cell.' *Current Opinion in Cell Biology*, **15**, 221–231.

[18] Ullmann JR, (1976) 'An Algorithm for Subgraph Isomorphism.' *J. ACM*, **23**, 31-42.