# Computational Modeling of Metabolic Networks

S. N. Kalyankar, N. V. Kalyankar, and Mohseena Thaseen

***Abstract*—Metabolism is the set of biochemical reactions occurring in living organisms. Metabolites are usually small molecules like glucose, amino acids etc. These biochemical inter-conversions are generally catalyzed by enzymes. The sequencing of genomes and development of functional genomics make it now possible to reconstruct and understand the structure and function of metabolic networks at large scale. New computational tools and biological concepts are being developed to understand these metabolic networks more precisely. Here attempts were made for reconstruction, visualization, and graph representation of metabolic networks for structural analysis i.e. connectivity and centrality analyses, modularity and decomposition of the networks to fundamental level. Reconstruction, visualization, and graphical representation of glycolysis for structural analysis and decomposition of the network to fundamental is done as an example of basic metabolic network in cells. The methods and concepts presented deals with static properties and functions of glycolysis and more complex networks can be represented following similar methods.**

***Index Terms*— Metabolic networks, modeling, visualization.**

## I. Introduction

Metabolism is the set of biochemical reactions occurring in living organisms. Metabolites are usually small molecules like glucose, amino acids etc. These biochemical inter-conversions are generally catalyzed by proteins which have specialized functions known as enzymes. All biochemical reactions are catalysed by enzymes with a few exceptions that are spontaneous and may be non enzymatic. Metabolic pathway is fundamental concept in biochemistry comprising of arrays of successive or closely associated biochemical reactions for a specific metabolic function. Metabolic network is the network composed of metabolites and their interconversions in a living organism. Glycolysis, lysine degradation, and penicillin biosynthesis etc. are some of the example of simple metabolic pathways. A metabolic pathway can be considered as a small local area of a metabolic network, whereas a metabolic network gives a better and more complete view of the cellular metabolism.

Thus, in the general method of visualization and analysis, it can be represented as a graph with two different kinds of vertices - metabolite and enzyme. Thus in the most simplest method of analysis, the metabolic networks can be represented as a simple direct graph with vertices representing metabolites and edges corresponding to reactions converting one metabolite into the other [1] [2]. As a complement, a reaction graph can also be used [3]. In the reaction graph, the vertices represent reactions and there will be an edge between two vertices if the product of one reaction is the substrate of the other. This simplification is helpful for structural analysis as many graph algorithms do not consider different types of vertices as in a bipartite graph. The drawback of this representation is that if the utilized metabolites are not removed, biologically meaningless shortcuts may be introduced in the network.

Manuscript received June 27, 2012; revised September 3, 2012.

S. N. Kalyankar is with the Department of Chemistry, Yeshwant College, Nanded 431602(M.S.) India (e-mail: drkalyankarsn@ yahoo.com).

N. V. Kalyankar is with the Department of Physics, Yeshwant College, Nanded 431602(M.S.) India (e-mail: drkalyankarnv@rediffmail.com)

M. Thaseen is with Computer Science and Information Technology Department, Yeshwant College, Nanded 431602(M.S.) India (e-mail:mohseena@gmail.com)

## II. Reconstruction of Metabolic Network

A simple algorithm for metabolic network reconstruction can be introduced. Fig. 1 shows the workflow to reconstruct the metabolic network from genome annotation. The EC numbers (Enzyme Commission numbers, a number representing the biochemical reaction catalyzed by the corresponding enzyme) can be extracted or enzyme name can also be used where EC numbers were not assigned[4].

After representation of the reaction list, these reactions are further converted to connection matrix: the substrate–product pairs are recoded in a format that can be easily recognized and processed by the computer. The connection matrix file can be directly interpreted by many network visualization tools for visualization and analysis. The connection matrix is further treated by removing connections via conversion metabolites such as $H_2O$, $CO_2$, ATP, etc [2]. Conversion metabolites are mainly used as carriers for transferring electrons and certain functional groups such as hydrogen, phosphate, amino group, one carbon unit, methyl group, etc. in the given reaction Glucose + ATP = G6P + ADP. ADP and ATP are conversion metabolites for transferring phosphate to glucose.

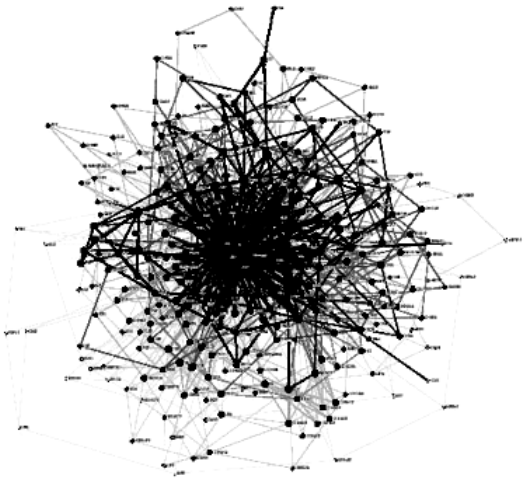


Fig. 1. The metabolic network of *Escherichia coli.* The nodes correspond to chemicals (metabolites) linked by a metabolic reaction

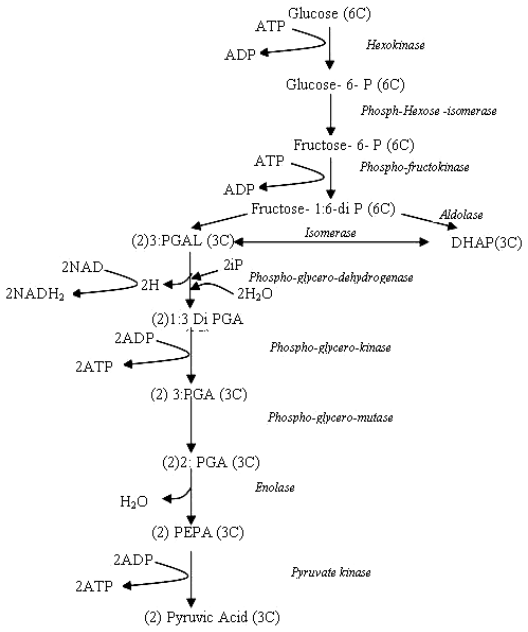



Fig. 2. Conventional representation of Glycolysis pathway

## III. Visualization

The number of reaction steps in the glycolysis pathway describes the path length, i.e. the conversion of glucose to pyruvate should be nine steps in terms of biochemistry (Fig. 2). Whereas considering ATP and ADP as vertices in the network there would be only two steps from glucose to pyruvate, in the first reaction, the conversion of glucose and produces ADP, while the last reaction consumes ADP and produces pyruvate. Such short annotation in biological reference will be meaningless. A number of approaches are proposed to address this problem. A simple method is to exclude the terminal metabolites based on their connection degree i.e. the number of edges connected with a metabolite. The problem is that certain primary metabolites such as pyruvate may also have high connection degrees. Hence the conversion metabolites cannot be defined *per se* by compounds but defined according to the reaction.

## IV. Participating Component

There are various ways of visualizing the metabolic network according to the focus selected. The complexity of the network is such that there is no single approach that can include all its aspects. This is not evident immediately when the fundamental unit of the network is considered, but when a reaction catalyzed by an enzyme in where one of the chemical reactants in metabolism is metabolic intermediate or *substrate* and is transformed into another intermediate or *product*. For instance, if water is overlooked as a reactant being it presences everywhere in the cells, three quarters of the biochemical reactions actually involve two or more reactant / substrates interacting to form two or more products. Considering the example of two substrates – A and Y reacting to form two products – B and Z. This can be presented in very simple methods. Considering one pair of metabolites A and B can be shown as the basic link while the involvement of Y and Z is concealed or presented as a byproduct of the main conversion. This is due to the facts that metabolism involves carbon, which form the bulk and the largest component of living systems. The linkage between A and B is that the most of the carbon in A are also present in B. If the maps represent two substrates with equal importance it means that the both the reactants contribute carbon atoms to the product leading to formation of two products, since the substrate has split into two fragments. Thus, there is basic understanding of chemical and biochemical knowledge involved considering the statement with the reaction 'A + Y → B + Z ' occurs in the network and converting that to a network link 'A → B' on the basis being the principal conversion involved in the reaction.

Here, the implications of including or deleting *co–substrate* / *co–product* pairs such as Y and Z as they have immediate implications for the connectivity of the network can be considered. Since, these pairs generally take part in a large number of reactions where the identities reactants A and B differ. Some of these reactions will convert Z back to Y, so that the total cyclic flux converting Y to Z and back again is much more higher than the rates of conversion of Y and Z other different compounds.

For instance, the pair ATP and ADP are standard abbreviations for the most connected metabolites in the metabolic network. They have a major role in the energetics of the cell, but chemically ATP differs from ADP having three phosphate groups which is one more than ADP. So these reactions involve transfer of phosphate group without change in carbon atom among them. But in most of these biochemical reaction, which involves change in state of carbon atom of the metabolite are stressed much rather than the reaction like ADP and ATP or NADP/NAD / $NADPH_2$/ $NADH_2$ are taken into consideration. This also implies that generally, the total rate of conversion of ATP to ADP is equal to the rate of conversion of ADP back to ATP or rate of conversion of NADP/NAD to $NADPH_2$/ $NADH_2$ are always equal. Therefore, the addition to the links in the network with flows and there are additional linkages across the network pointing toward the possibility of interpreting an invalid pathway that are controlled by mass conservation of carbon [5] (Fig. 3)

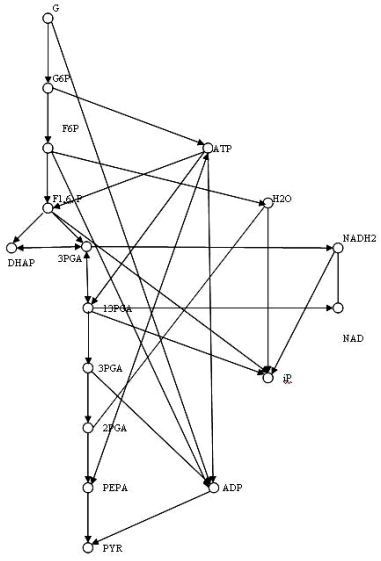
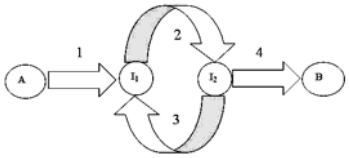



Fig. 3. The graphical representation of Glycolysis with metabolites shown as nodes and vertices and reactions as connections

## V. Mathematical Representation

Mathematical representations of all such methods are necessitated for representation of the metabolic network. The starting point for this is provided by the stoichiometry matrix. To illustrate this, consider the following small network




This stoichiometry matrix can also represented in two different following methods

$$\begin{array}{c} I_1 \\ I_2 \end{array} \begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{pmatrix}$$

Or

R1: $A \rightarrow I_1$

R2: $I_1 \rightarrow I_2$

R3: $I_2 \rightarrow I_1$

R4: $I_2 \rightarrow B$

The upper one is the stoichiometry matrix where the numbers represent the moles of the substrate in row taking part in the reaction specified by the column. The negative numbers indicating substrates of the reaction. The lower form is a symbolic list of reactions, starting with a reaction number (R1, 2, 3, 4) representing the balanced equation of the reaction in text form. All these three forms can be inter-converted where the point of initiation of these reactions is to be ascertained in the set of reactions taking part in the network. Here, the stoichiometry matrix contains rows of the *intermediates* of the network. The system needs to define to such an extent that many of the modes of analysis of the steady states of the system are in balance. This implies that the synthesis and utilization of each *intermediate* is at equilibrium. The *products or out put of the reaction* are not at equilibrium since there should be a mass flow through the metabolic network which maintains its shift from chemical equilibrium. If the synthesis and utilization of these *products or out put of the reaction* is to be computed, then a full stoichiometry matrix is required, with additional rows describing the involvement of other reactants/products in the reactions. Here it is becomes necessary to decide whether the intermediates of each reaction are treated as products or reactants.

## VI. Connectivity and Centrality in Metabolic Network

Metabolic networks derived from the information embedded in the genome have imparted a remarkable insight into the fundamental aspects of cellular metabolism. The key finding is that metabolic networks exhibit typical characteristics of small-world networks like many non -biological complex systems viz. a power law connection degree distribution [6], high cluster coefficients and a short network diameter [7] [8] [9]. This small world structure is regarded as one of the design principles of many robust and error-tolerant networks such as the computer network, neural network, and certain social and economic networks [6] [10]. Metabolic networks were scale free networks in terms of the connections through currency metabolites [9]. The structure of metabolic networks still has the characteristics of a scale-free network even after deleting the linkages through currency metabolites [2]. One important feature of scale-free network is the power law distribution of connection degree among the vertices [6]. A random network has a Poisson distribution of the connection degree as opposite to the scale-free network. The connection degree is defined as the number of connections linked with each metabolite (vertex). Considering the direction, the number of connections starting from the metabolite is called output degree, and the number of connections ending at the metabolite is called input degree. The input degree distributions for these organisms also have similar power law relations. The power law degree distribution exists in networks of all the organisms studied in the work of Ma and Zeng [2]. This indicates that the metabolic network without connections through currency metabolites is still a scale- free network.

## VII. Modularity and Decomposition of Metabolic Networks

In biochemistry, modules consisting of several interacting biological reactions or metabolic pathways build discrete functional units of metabolism. Thus such smaller modules are linked to form a complex metabolic network. However, the structural analysis of metabolic networks indicates a small world structure [1][2] where all the vertices in the whole network are linked through a short path. The modular organization seems missing in this small-world structure. To resolve the apparent contradiction between the small-world structure and modularity organization Ravasz et al. [11] proposed a hierarchical modularity model for metabolic networks. According to this model, metabolic networks of organisms are organized as many small, but highly connected modules that combine in a hierarchical manner to larger, less cohesive units. Several studies using concepts such as the reaction connectivity centrality distribution and the dependency of metabolites have further verified that metabolic networks are organized in a hierarchical way [12] [13]. These results indicate that hierarchical modularity is also an important feature of metabolic networks. Modularity has been shown to be common in the organization of robust and sustainable complex systems [14]. Therefore, identifying the modular organization of metabolic network by certain network decomposition methods can help us in better understanding the organization principle of complex systems.

## VIII. Conclusion and Perspective

The recent interest in the structural properties of networks has highlighted the limitations of viewing metabolism as the sum of a small number of underlying pathways. Metabolism

is the set of biochemical reactions occurring in living organisms. Metabolites are usually small molecules like glucose, amino acids etc. These biochemical inter-conversions are generally catalyzed by proteins which have specialized functions known as enzymes. The sequencing of genomes and development of functional genomics make it now possible to reconstruct and understand the structure and function of metabolic networks at large scale. Elucidating the substructure that is helpful to biologists in understanding the different metabolic phenotypes of different organisms still remains a challenge. New computational tools and biological concepts are being developed to understand these metabolic networks more precisely.

## ACKNOWLEDGMENT

The authors are thankful to Dr. M. M. V. Baig, Dept. of Biotechnology, Yeshwant College, Nanded for his critical analysis and help.

## REFERENCES

[1] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. "The large-scale organization of metabolic networks." *Nature*, vol. 407, no. 6804, pp. 651–654, Oct. 2000.

[2] H. Ma and A. P. Zeng. "Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms." *Bioinformatics*, vol.19, no. 2, pp. 270–277, Jan. 2003.

[3] H. W. Ma, X. M. Zhao, Y. J. Yuan, and A. P. Zeng. "Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph." *Bioinformatics*, vol. 20, no. 12, pp. 1870–1876, Mar. 2004.

[4] M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. "From genomics to chemical genomics: New developments in KEGG." *Nucleic Acids Research*, vol. 34(Database issue), pp. 354–357, Jan. 2006.

[5] Croes, D., Couche, F., Wodak, S.J. and Van Helden, J. "Metabolic PathFinding: Inferring relevant pathways in biochemical networks." *Nucleic Acids Research*. vol. 33, no. 2, pp. W326-330. 2006

[6] S. H. Strogatz. "Exploring complex networks." *Nature*, vol. 410, no. 6825, pp. 268–276, Mar. 2001.

[7] A. Wagner and D. A. Fell. "The small world inside large metabolic networks." *Proceeding of the Biological Sciences*, vol. 268, no. 1478, pp. 1803–1810, Sep. 2001.

[8] D. A. Fell and A. Wagner. "The small world of metabolism." *Nature Biotechnology*, vol. 18, no. 11, pp. 1121–1122, Nov. 2000.

[9] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. "The large-scale organization of metabolic networks." *Nature*, vol. 407, no. 6804, pp. 651–654, Oct. 2000.

[10] R. Albert and A. L. Barabasi. "Statistical mechanics of complex networks." *Reviews of Modern Physics*, vol. 74, no. 1, pp. 47–97, Jan. 2002.

[11] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. "Hierarchical organization of modularity in metabolic networks." *Science*, vol. 297, no. 5586, pp. 1551–1555, Aug. 2002.

[12] J. Gagneur,D.B. Jackson, and G. Casari. "Hierarchical analysis of dependency in metabolic networks." *Bioinformatics*, vol. 19, no. 8, pp. 1027–1034, May. 2003.

[13] P. Holme, M. Huss, and H. Jeong. "Subnetwork hierarchies of biochemical pathways." *Bioinformatics*, vol. 19, no. 4, pp. 532–538, Mar. 2003.

[14] C. Francke, J.S. Siezen, and B. Teusink. "Reconstructing the metabolic network of a bacterium from its genome." *Trends in Microbiology*. vol. 13, no. 11, pp. 550–558. Nov. 2005

**Sarla N. Kalyankar** is an Asst. Professor in the Department of Chemistry, Yeshwant College, Nanded (M.S.) India. She received her Master degree (1983) and Ph.D. degree (1999) in Chemistry from Dr. B. A. M. University, Aurangabad, India.

Her areas of specialization and interests are Cheminformatics, Computer Modeling, Drug designing. She is currently working on project related to drug designing using nanotechnology.

Dr. Kalyankar is actively engaged in research supervising more than 10 students for their MSc and PhD degree. She has published more than 20 publications in peer reviewed journals. She is member of various professional societies including ISCA, IACSIT.

**Namdeo V. Kalyankar,** FLS**,** is the Principal of Yeshwant College, Nanded (M.S.) India. He received his Ph. D. in Physics from Dr. B. A. M. University, Aurangabad, India in 1995. His research interests are Computer Graphics, Pattern Recognition, Computer Vision, Image Processing, Networking and Soft Computing. He is currently working on various projects related to academia and industry.

He is actively engaged in research supervising more than 15 students for their PhD theses. He has published more than 70 publications in the form of various Books, Book Chapters, journal papers and conference papers.

Dr. Kalyankar is member of various professional societies including ISCA, IAE, ILA, IACSIT. He is a Chair, member of the International Advisory Committees and Organizing Committees of various international Conferences, Symposiums and Workshops. He is the reviewer, for many international Journals, Conferences, meetings, and workshops around the world. He has achieved various awards in education, research, and administrative services like Best Teacher Award, Best Principal Award etc.

**Mohseena Thaseen** obtained her B.Sc. degree in 2000 and Master in Computer Applications (MCA) degree from Osmania University, Hyderabad, India in 2004.

Currently, she is pursuing research leading to PhD in Computer Science from Swami Ramanand Teerth Marathwada University, Nanded, India. Her field of study is machine – human interface, mathematical and computational modeling of biological systems, artificial intelligent system, Artificial Neural Networks and biological information processing. She worked as an Assistant Professor in ICBM, Hyderabad, India and ITM, Nanded India.