Advances in Computational Research, ISSN: 0975–3273, Vol. 1, Issue 1, 2009, pp. 1-4

Functional module analysis in metabolomics: *Chokes*

[1]Shaily Mehta 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: Since recent years the work on biological and metabolic network has been increasing due to the new biological discoveries and essential metabolites. Metabolomics being a burgeoning field, which produces voluminous data that, like other 'omics' data, should be seen as a resource that contributes specifically to the former half of an iterative cycle of hypothesis-generating and hypothesis- testing phases. It is becoming increasingly apparent that our ability to generate large quantities of metabolomics or metabolic profiling data will help to open up many previously inaccessible areas of biology various high-throughput techniques are leading to an explosive growth in the size of biological databases and creating the opportunity to revolutionize our understanding of life and disease. Interpretation of these data remains, however, a major scientific challenge. With the study of enzymes and metabolites new pathways can be discovered, which can help in the analysis of the various process taking place in the organism. In order to identify potential drug targets the concept of choke points was used to find enzymes which uniquely consume or produce a particular metabolite. Hence the study of these choke are taken into consideration.

Keywords: Biological, metabolic network, metabolites, enzymes, choke points

1 INTRODUCTION

The present scenario tells us that network analysis is essential for the analysis of genetic, proteomics and metabolomics data **[1, 2]**. The present discovery of enzymes and metabolites has made the study of metabolomics very much in need. Since past couple of decades we have understood the basic idea of the formation of metabolites. It all starts with the process of ingestion were the organism takes the material such as food into their bodies. These material become compounds and energy necessary for sustaining the activity of the organism by various chemical reactions. The whole of such chemical reaction taking place in the body is called metabolism. Here the substrate is converted to a product i.e. one compound is converted to another and chain of reactions is generated which forms a large scale network. This network is known as Metabolic Pathway.

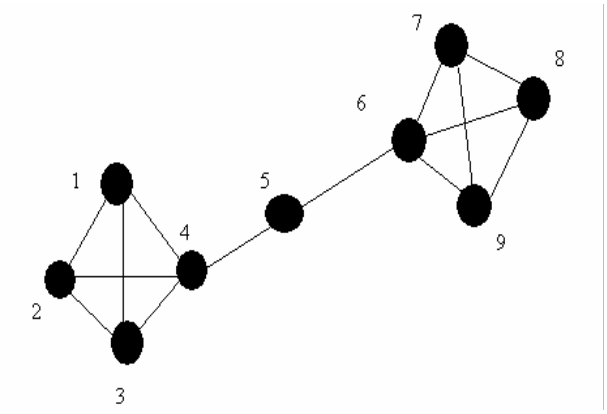
With the increase in research work large no of metabolic pathways are continuously being discovered and there activity is studied. Due to which large no of enzymatic databases were build to store in these data. Also the activity of each of these enzymes taking part in the metabolic pathways is studied carefully. Various graph theories, mathematical, computational and programming aspects are taken into consideration for the verification of these metabolites and chemical reactions and to demonstrate the intrinsic hierarchical modularity of metabolic networks **[3]** and their robustness based on the shortest path analysis of the metabolic networks **[4-6]**.

A typical metabolic network consists of reactions, metabolites and enzymes, which can be modeled using graph theory **[7-11]**. These representations lead from a simple graph consisting of edges (reactions) and nodes (metabolites) or vice versa to a complex bipartite graph where two nodes (metabolites) share a common node (reaction/enzymes) **[12]**. Enzyme-centric networks can be created by joining enzymes that share a common metabolite in a path. The enzyme-centric view **[13]** simplifies the representation of the metabolic network by removing loose ends in the network (metabolites at the periphery of the network) and forming clusters of interacting enzymes. The gene-centric view has been successfully used in determining co-regulated genes in the metabolic and regulatory networks **[14-18]**.

Hence it becomes a necessity to understand the choke points which are enzymes which play a crucial role in the metabolic pathway. Understanding these enzymes will widen the scope of analyzing the pathway considerably. Choke points are those enzymes which uniquely consume and/or produce a certain metabolite. They are ranked by the number of k-shortest paths (in/out) passing through it and the load point (in/out) on it. Since it is a reasonable assumption that a large number of the biochemical reactions follow the shortest path, we assume that the shortest path count can good indicator of biochemical activities. Inactivation of choke points may lead to an organism's failure to produce or consume particular metabolites which could cause serious problems for fitness or survival of the organism hence they are considered crirtical points in the metabolic networks.

Chokepoint analysis has several advantages. First, it allows us to test the consistency between experimental data and assumptions about the organization and regulation of the biochemical pathway and of its interdependencies with other processes. Second, it can be used to predict the consequences of various mutations or inhibitors.The concept of choke points was used in our study to find potential drug targets in the metabolic network of Bacillus anthracis Sterne. The metabolites and enzymes are further ranked on the basis of their loads in the given network. A comparative study was performed between the human metabolic network and pathogen choke points to discriminate human choke points from the pathogenic bacterial choke points. A homology search was performed against human genome to find non-homologus potential drug targets from the pathogen choke points.
A new method to analyse choke points by screening the entire metabolic network of pathogens and report the probable choke points in the network was discovered by group of scientists **[19]**. This extended graph theory model ranks the choke points according to the k-shortest path passing through it and the load (in/out) on it. This ranking has a major advantage as this measure may help determine the biochemical essentiality of a metabolite/enzyme. For example, in Plasmodium falciparum—a parasite causing malaria in humans—a host cell enzyme 4.2.1.24 (d-aminolevulinate dehydratase; ALAD) involved in heme biosynthesis was suggested as an antimalarial target. This enzyme is also a choke point enzyme and identifying such potential targets in the pathogens can accelerate the drug discovery. Also all three clinically validated drug targets for malaria are chokepoint enzymes. A total of 87.5% of proposed drug targets with biological evidence in the literature are chokepoint reactions.

Fig 1:- Metabolic-centric view of a graph model. Node (5) is a choke point (metabolite) and thinner edges adjacent to this node (enzymes) are also choke points.



The above diagram clearly specifies that if node no 5 is an important metabolite without which the reactions then the reactions are not possible and we can't get the desired output.

2 CURRENT SCENARIO

In order to confer biological meaning to the graph-based approach of finding choke points, the present studies deal with the following steps:-

For building the biochemical network we used the LIGAND database from KEGG as this data model is the backbone for the Pathway Hunter Tool in addition to BRENDA. For the predicted choke points in the pathogen we performed a homology search against the human genome using BLAST.

Calculations of the top choke points which are ranked by number incoming shortest paths along with important load points are reported in the metabolic network of the organism. A network based comparative study of the important choke points between a model organism and Homo sapiens is performed using Pathway Hunter Tool (PHT). '+' implies that a particular enzyme acts as a choke point in the human biochemical network as well as in the pathogen whereas '–' indicates that this enzyme is only a choke point in the pathogen and not in the human biochemical network.

A homology search is performed between the human and model organism. choke point enzymes using BLAST and chokepoints with a closest homologue with e-values <1.0e–02 are removed.

The above method was one of the techniques proposed by the scientist working in this field. There may be many more methods which have potential in identification of these choke points which are yet to be discovered or they are in the process to reach there out.

3 IMPLEMENTATION

Chokepoint analysis can be implemented in various pathway analysis.

First, it allows us to test the consistency between experimental data and assumptions about the organization and regulation of the biochemical pathway and of its interdependencies with other processes.

Second, it can be used to predict the consequences of various mutations or inhibitors.

The targeting of metabolic pathways has several advantages on its own. Each step in the pathway is well validated as an essential function for pathogen growth. The target enzymes from the pathogen which are discarded and which share a similarity with the host proteins ensures that the targets have nothing in common with the host proteins, thereby, eliminating undesired host protein–drug interactions. Metabolic pathway analysis is becoming increasingly important for assessing inherent network properties in reconstructed biochemical reaction networks.

As of now, identifying choke point reactions, identification of enzymes has been done that are essential to the parasite's survival. There is an enrichment of drug targets in chokepoints as compared with non-chokepoints.This leads to the conclusion that the classification of an enzyme as a chokepoint has some bearing on whether or not it would make a good drug target.

Another approach could be combining Choke Point analysis with chemo-genomic profiling (micro-array data), Providing a complete and better annotation in vivo (thereby reducing the identification of false Choke Point enzymes and providing previously unreported Choke Point enzymes in the metabolic network) is one of the first steps in this direction.

4 CHOKE POINT ANALYSIS

Choke point analysis was successfully performed on large no of organisms to discover potential drug targets. e. g. Plasmodium falciparum, B.antracis, Corynebacterium glutamicum, E.histolytica.

If an enzyme catalyzes at least one chokepoint reaction, it is classified as a potential drug target. chokepoints and non chokepoints against proposed drug targets from the literature is compared to assess the usefulness of identifying chokepoint enzymes for proposing drug targets.

A complete literature search for proposed amoebiasis drug targets is attempted that were metabolic enzymes and met the criteria discussed above.
Chokepoints may not be essential. One reason could be that they create unique intermediates to an essential product which are not essential themselves and finally, there could be chokepoint reactions that are not essential due to other pathways that achieve the same metabolic goal within the organism. One example could be blocking the reaction that has no deleterious effects on the parasite. Due to the high percentage of enzymes identified as choke points, one additional criteria observed in addition to being a choke point enzyme for identifying potential metabolic drug targets is that an enzyme not having isozymes would make it more likely to be a good drug target.

An analysis of the top 10 choke points in B.anthracis, a pathogen, is presented. In a number of possible drug targets against infection of B.anthracis are identified. It was found that the enzymes

tryptophan synthase (EC: 4.2.1.20) and anthranilate phosphoribosyltransferase (EC: 2.4.2.18) could be effective potential drug targets. Neither of these enzymes are choke points in the human metabolic network nor do they share a significant homology with the human genome. This means that blocking these enzymes might affect the pathogen but not the human as there exists an alternate pathway

To identify potential drug targets, a chokepoint analysis of the metabolic network of E.histolytica is performed. A "chokepoint reaction" is a reaction that either uniquely consumes a specific substrate or uniquely produces a specific product in the Entamoeba metabolic network. It is expected that the inhibition of an enzyme that consumes a unique substrate result in the accumulation of the unique substrate which is potentially toxic to the cell and the inhibition of an enzyme that produces a unique product to result in the starvation of the unique product which potentially cripple essential cell functions **[20]**. Thus, it is believed that chokepoint enzymes may be essential to the parasite and are therefore potential drug targets.

5 CONTRIBUTION OF CHOKE POINTS IN THE HUMAN METABOLIC NETWORK

While treating disease like diabetes, obesity, cancer, HIV etc it is very important that the drug enables target specific action. This includes the fact that the drug wold act directly on the metabolic pathway in whole or he enzyme which is responsible for spreading the disease.

Many important drug target specific metabolic reactions has been discovered. Drug target identification based on "omics" is a very promising approach that has only recently become possible. The concept of choke points in a given network contributes effectively in the identification of the lethality/bottleneck (here potential drug targets) in a network. Since a high load on a certain enzyme means that a large number of shortest paths go through it, therefore indicating a position in the central metabolism, ranking choke points on the basis of load will move enzymes with a higher probability of biochemical lethality to the top of the candidate list.

A comparative study of choke points with the human metabolic network is essential to identify possible interference of the drugs with the human metabolism which might lead to side effects. It has to be kept in mind though, that presently a large number of genes have unidentified functions which could lead to erroneous prediction of choke points. For example, often drug targets are identified by a unique pathogen-specific metabolic activity, as in the case of reverse transcriptase in the case of HIV.

Hence, the study of theses choke points is very much in demand and had potential to act on large no of targets thereby giving favorable results.

6 CONCLUSION

Choke points are important points in a reaction; they are reaction which consumes/produces certain metabolites which play important role in a given reaction. In absence of these choke points an organism cannot survive. The current analysis includes only the completely annotated enzymes in each organism. Including all the available enzymes for the organisms, such as putative enzymes, may complete the analysis of the metabolic network. The extended graph-based choke point concept can facilitate drug discovery and ranking choke points based on their load values may be a likely pointer to the lethality level of such potential drug targets in the network. Further study and comparative analysis of various metabolic networks based on our network model can be beneficial for in vivo and in vitro studies.

There are further aspects on which the list of potential drug targets can be narrowed down. The drug should adversely affect the parasite but not the human host which means that if the drug target has a homologous enzyme in human, it should not be essential or have differential inhibition in human. In other way, it can be said that potential drug targets should be expressed in the human stages of the parasite. The rapid emergence of multi-drug resistant strains of these potentially lethal pathogens calls for the identification of new targets. The discovery of new targets with help of choke point analysis may lead to a drug formulation that would be able to counteract the resurgence of these diseases.
Also with further studies we can prove choke points to be helpful in the discovery of important regions in the pathway along with a better approach towards understanding the system well.

Its can be a key in the research of large metabolites and can provide with extremely important information which were hidden and needed to be discovered. The two most promising concepts for pathway analysis focused here are closely related. Assessing metabolic systems by the set of extreme pathways can, in general, give misleading results owing to the exclusion of possibly important routes. A full assessment of the proposed listed steps will require intense further effort. It is to be expected that some experiments may be significant which will stimulate the next phase of amendments and refinements. As it stands, it is hoped to serve the scientific community as a starting point for further data collection and experimentation in concert with, and based on, pathway analysis. One of the most important step is to reduce the Probability of identifying false choke points which can be done by undergoing annotation. Studies say that, Choke Point analysis with chemo-genomic profiling (micro-array data), can Provide a complete and better annotation in vivo.

REFERENCES

[1] Grigorov, M.G. (2005) Global properties of biological networks. *Drug Discov. Today*, 10, 365–372.

[2] Hartwell, L.H., et al. (1999) From molecular to modular cell biology. *Nature*, **402**, C47–52.

[3] Ravasz, E., et al. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–1555.

[4] Arita, M. (2004) The metabolic world of *Escherichia coli* is not small. *Proc. Natl Acad. Sci. USA*, **101**, 1543–1547.

[5] Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet*, . **5**, 101–113.

[6] Papin, J.A., et al. (2004) Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem Sci*, . **29**, 641–647.

[7] Girvan, M. and Newman, M.E. (2002) Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA*, **99**, 7821–7826.

[8] Jeong, H., et al. (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.

[9] Ma, H. and Zeng, A.P. (2003) Reconstruction of metabolic networks from genome data and analysis of

their global structure for various organisms. *Bioinformatics*, **19**, 270–277.

[10] Oltvai, Z.N. and Barabasi, A.L. (2002) Systems biology. Life's complexity pyramid. *Science*, **298**, 763–764.

[11] Schuster, S., et al. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol*, . **18**, 326–332.

[12] Rahman, S.A., et al. (2005) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics*, **21**, 1189–1193.

[13] Horne, A.B., et al. (2004) Constructing an enzyme-centric view of metabolism. *Bioinformatics*, **20**, 2050–2055.

[14] Barrett, C.L., et al. (2005) The global transcriptional regulatory network for metabolism in *Escherichia coli* exhibits few dominant functional states. *Proc. Natl Acad. Sci. USA*, **102**, 19103–19108.

[15] Covert, M.W., et al. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, **429**, 92–96.

[16] Levchenko, A. (2003) Dynamical and integrative cell signaling: challenges for the new biology. *Biotechnol Bioeng*, **84**, 773–782.

[17] Luscombe, N.M., et al. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, **431**, 308–312.

[18] Ozbudak, E.M., et al. (2004) Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, **427**, 737–740.

[19] Yeh, I., et al. (2004) Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome Res*, . **14**, 917–924.

[20] I. Yeh, *et al*. (2004) Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.,* **14**: 917.

