May 2013- July 2013; Vol. 3, No. 3; 1847-1852. E- ISSN: 2249 –1929

Journal of Chemical, Biological and Physical Sciences

An International Peer Review E-3 Journal of Sciences

*Available online at*www.jcbsc.org

Section B: Biological Science

CODEN (USA): JCBPAT Research article

Molecular Mining of *M.paratuberculosis* Using Systems Biology Approach

M. Padmavathi

Department of Biotechnology,

DVR & Dr. HS MIC College of Technology, Kanchikacherla, A.P. India

Received: 29 April 2013; **Revised:** 24 May 2013; **Accepted:** 28 May 2013;

Abstract: Molecular mining studies of tuberculosis focused a number of molecular techniques in assessing the strains. This can be done by testing the genetic diversity of clinical strains of *Mycobacterium paratuberculosis.* These methods are used to control the tuberculosis. For example molecular techniques added accuracy, consistency, completeness and precision in explaining the dynamics transmission. In addition there is mounting evidence to suggest that specific strains such as *M. paratuberculosis* belonging to discrete phylogenetic clusters or lineages may differ in virulence, pathogenesis and epidemiologic characteristics all of which may significantly impact TB control. In the present study the current molecular tools and its approaches used to better understanding the epidemiology of tuberculosis.

Keywords: Molecular Mining, *M. paratuberculosis*, Insilico tools, Tuberculosis

INTRODUCTION

Molecules are represented by molecular graphs and is strongly related to graph mining and structured data mining. By using metrics the molecules discriminating the data instances chemically is a long tradition in the chemoinformatics field. In molecule topology the typical approach is to calculate the chemical similarities. There is an increasing interest in text mining and information extraction strategies applied to the biomedical and molecular biology literature due to the increasing number of electronically available publications stored in databases such as PubMed. The main developments in

this is related to the identification of biological entities such as protein and gene names in free text, the association of gene clusters obtained by microarray experiments with the biological context provided, automatic extraction of protein interactions and associations of proteins to functional concepts.

The problems include building efficient databases and indexes for sequence information, extracting the frequently occurring patterns, comparing sequences for similarity, and recovering missing sequence members. In general, sequence mining problems can be classified as *string mining* which is typically based on string processing algorithms and *itemset mining* which is typically based on association rule learning.

Systems biology is an emerging approach applied to biomedical and biological scientific research. Systems biology is a biology-based inter-disciplinary field of study that focuses on complex interactions within biological systems, using a more holistic perspective (holism instead of the more traditional reductionism) approach to biological and biomedical research. Particularly from year 2000 onwards, the concept has been used widely in the biosciences in a variety of contexts. One of the outreaching aims of systems biology is to model and discover emergent properties, properties of cells, tissues and organisms functioning as a system whose theoretical description is only possible using techniques which fall under the remit of systems biology. These typically involve metabolic networks or cell signaling networks[1].

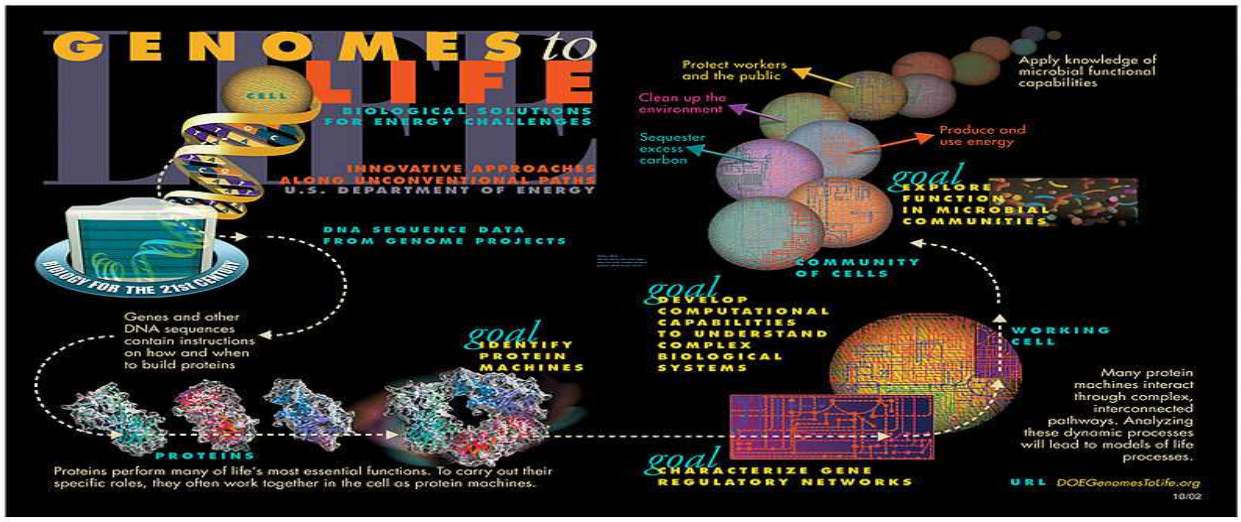

Fig. 1: Genomics GTL Pictorial Programmed.

The large increase in data from the omics (e.g. genomics and proteomics) and the accompanying advances in high-throughput experiments and bioinformatics. Since then, various research institutes dedicated to systems biology have been developed. For example, the NIGMS of NIH established a project grant that is currently supporting over ten systems biology centers in the United States[2]. As of summer 2006, due to a shortage of people in systems biology[3], several doctoral training programs in systems biology have been established in many parts of the world. In that same year, the National Science Foundation (NSF) put forward a grand challenge for systems biology in the 21st century to build a mathematical model of the whole cell[4].

Systems biology approach is one of the main impotant application can be distinguished from the tumorigenesis and treatment of cancer. It works with the specific a data such as sample of the patients, high-throughput data with characterizing cancer genome in patients and tools like xenograft models, next generation sequencing methods, siNA based knocking down screenings, genome instability[5]. The

main objective of systems biology of cancer is ability to better diagnosis cancer, classification and prediction for treatment which is a basis of personalized cancer medicine and virtual cancer patient. Significant efforts in Computational systems Biology of Cancer have been made in creating realistic multi-scale in silico models of various tumours[6].

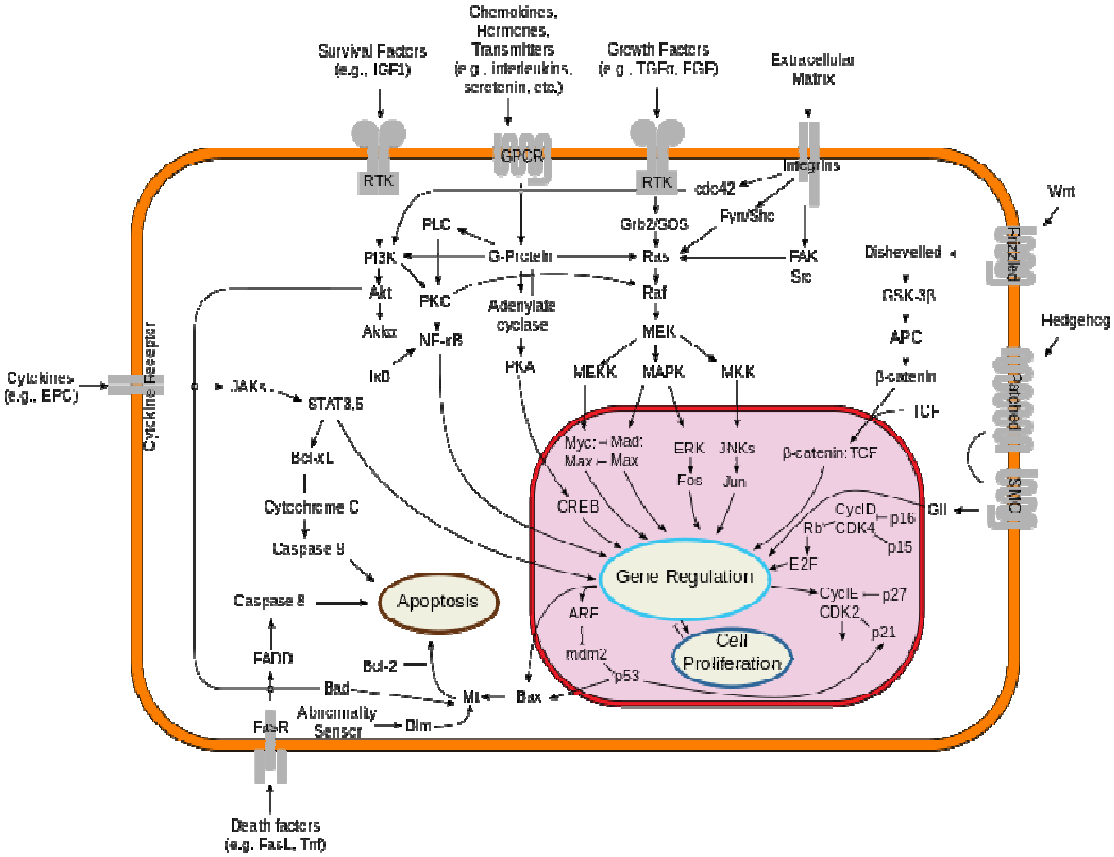

Fig. 2: Signal Transduction Pathways

Mycobacterium avium subspecies *paratuberculosis* (MAP) is an obligate pathogenic bacterium in the genus *Mycobacterium*. It is often abbreviated *M. paratuberculosis* or *M. avium* ssp. *Paratuberculosis*. It is the causative agent of Johne's disease, which affects ruminants such as cattle, and also perhaps the human disease Crohn's disease. MAP causes Johne's disease in cattle and other ruminants, and it has long been suspected as a causative agent in Crohn's disease in humans[5] this connection is controversial[7].

Recent studies have shown that MAP present in milk can survive pasteurization, which has raised human health concerns due to the widespread nature of MAP in modern dairy herds. MAP survival during pasteurization is dependent on the D_{72C}-value of the strains present and their concentration in milk. It is heat resistant and is capable of sequestering itself inside white blood cells, which may contribute to its persistence in milk. It has also been reported to survive chlorination in municipal water supplies.

MAP is a slow growing organism and is difficult to culture. Bacterial cultures were regarded as Gold standards for detection of MAP. Detection is very limited in fresh tissues, food, and water.

It is not susceptible to antituberculosis drugs (which can generally kill *Mycobacterium tuberculosis*). MAP is susceptible to antibiotics used to treat Mycobacterium avium disease, such asrifabutin and clarithromycin. MAP is recognized as a multi-host mycobacterial pathogen with a proven specific ability to initiate and maintain systemic infection and chronic inflammation of the intestine of a range of histopathological types in many animal species, including primates[8].

On the assumption that MAP is a causative agent in Crohn's disease, the Australian biotechnology company Giaconda is seeking to commercialize a combination of rifabutin, clarithromycin, and clofazimine as a potential drug therapy, called Myoconda[9], for Crohn's[10]. MAP has been found in larger numbers within the intestines of Crohn's disease patients[11] than those with ulcerative colitis and healthy controls[12].

MATERIALS AND METHODS

The process of identifying the location of genes and the coding regions and identifying the function of these genes. The genes were identified by gene finder. Basic level of annotation was done by using BLAST for finding the similar sequences.

CGAS: **Comparative Genome Annotation System**: This is to determine the sequence of whole genome in an effective manner by genome sequencing technology. Comparing genome requires a number of computational tools and procedures for a large amount of output. To alleviate the requirement for the computational tools and databases, comparative genome annotation was used.

RAST: (Rapid Annotations Using Subsystems Technology): This technology gives accuracy, consistency and completeness on the use of a growing library of subsystems that are curated. The RAST distinguishes these two classes of annotatation and uses the relatively reliable subsystem-based assertions as the basis for a metabolic reconstruction makes the RAST annotations an exceptionally good starting for a more comprehensive annotation effort.

Sequence Analysis of Proteins: Proteins are an important class of biological macromolecules present in all biological organisms. To be able to perform their biological function, proteins fold into one or more, specific spatial conformations, driven by a number of no covalent interactions such as hydrogen bonding, ionic interactions, Vanderwaal's forces and hydrophobic packing. X-ay crystallography or NMR spectroscopy to determine the structure of proteins. Te levels of protein structures, organization of polypeptides, the forces between the them and the spatial arrangement of subunits and nature of their interactions can be determined by the sequence analysis.

Expasy: (Expert Protein Analysis System): In this Expasy Proteome Tools server helps to determine identification and characterization of proteins, translate DNA sequence to protein sequence, Similarity searches, protein and profile searches, Prediction of post-translational modification, Prediction of topology, Structural prediction, Alignment of the sequence and biological text analysis.

The parameters under the Swissprot were molecular weight, Theoretical pl, Composition of amino acids, extinction co-efficient, Half-life estimation, Instability index, Aliphatic index, GRAVY (Grand Average of Hydropathy)

RESULTS

Using RAST the prediction of genome annotation provides a total of 4500 coding regions. In this some are protein coding regions and some are RNA coding regions. By NCBI structural RNA's annotation was done. The comparison the information of newly predicted proteins as well as number of modified proteins. High throughput screening of drug targets in this species shows total number of genes are 4500 680 were found to be found in pathogen. These are drug targets and the remaining

3820 are non drug targets.A number of proteins fall under various categories of subsystems are amino acid derivatives, carbohydrates, cofactors, Nitrogen, Phosphoorus, Potassium, Protein metabolism, Regulation and signaling, Isoprenoids, Virulence. 680 proteins that are only present in pathogen are to be treated as the drug targets for pathogen *M. paratuberculosis.* These are categorized into various subsystems. A subsystem is a set of functional roles that an annotator has decided should be thought of elated one. These subsystems represent the collection of functional roles that make up a metabolic pathway, a complex, a class of proteins.

Table- 1: Representing the statistics of gene annotation

Total number of proteins obtained by performing RAST	4500
Number of newly function predicted proteins	1530
Total number of modified proteins	450
Hypothetical	1500

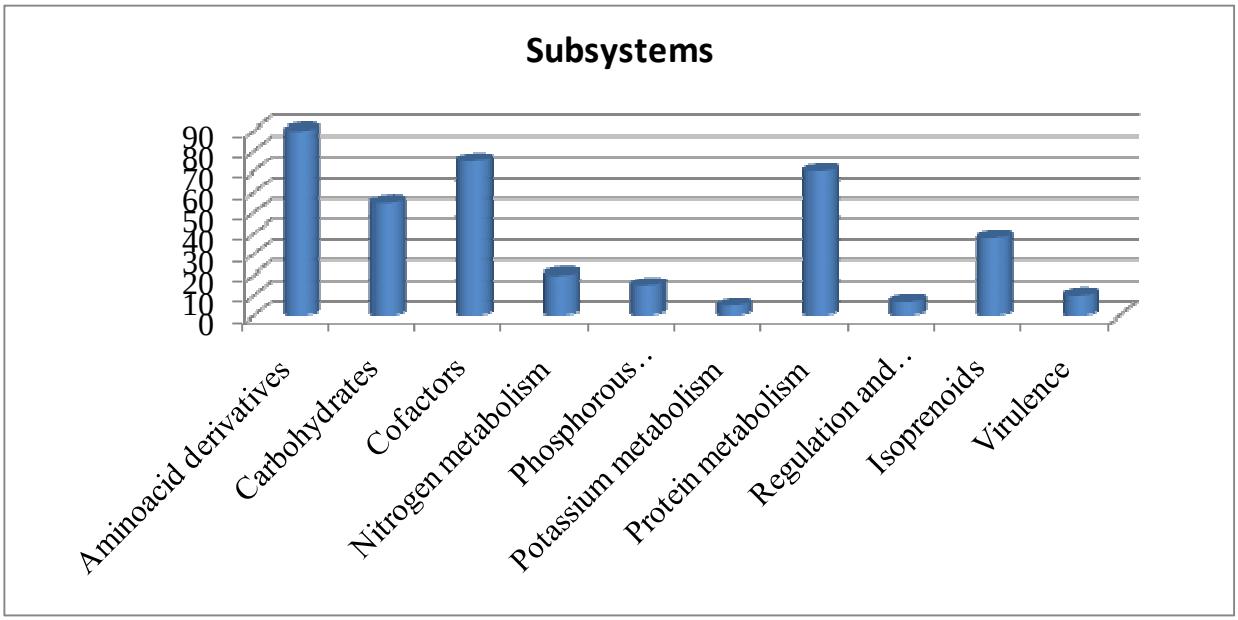

Fig. 3: Various categories of subsystems

CONCLUSION

M. paratuberculosis is a deadly pathogen causing tuberculosis. A number of drugs which are available in the market shows side effects and the strains are more resistant to these drugs. So to generate an alternate dug for the treatment of the pathogen an efficient technology was high through put screening. This technology was used to mine the genome of pathogen for new drug targets. Genome was annotated, mined by comparing the metabolic pathways of pathogen and host classified into potential drug targets and sequence analysis was performed. 680 proteins which are present in pathogen can be treated as drug targets. By using these insilico tools we can determine the potential drug targets can be used to screen leads against the diseases. These drugs can be used in clinical trials.

REFERENCES

1. H. Kashima, K. Tsuda, A. Inokuchi, Marginalized Kernels Between Labeled Graphs, The 20th International Conference on Machine Learning, 2003, pp 235-242
2. M. Deshpande, M. Kuramochi, N. Wale, G. Karypis, *Frequent Substructure-Based Approaches for Classifying Chemical Compounds*, IEEE Transactions on Knowledge and Data Engineering, 2005, 17(8), 1036-1050.
3. C. Helma, T. Cramer, S. Kramer, L. de Raedt, *Data Mining and Machine Learning Techniques for the Identification of Mutagenicity Inducing Substructures and Structure Activity Relationships of Noncongeneric Compounds*, J. Chem. Inf. Comput. Sci., 2004, 44, 1402-1411.
4. T. Meinl, C. Borgelt, M. R. Berthold, *Discriminative Closed Fragment Mining and Perfect Extensions in MoFa*, Proceedings of the Second Starting AI Researchers Symposium 2004, pp210-218
5. Barillot, Emmanuel; Calzone, Laurence; Hupe, Philippe; Vert, Jean-Philippe; Zinovyev, Andrei *Computational Systems Biology of Cancer*. Chapman & Hall/CRC Mathematical & Computational Biology. (2012). p. 461
6. Byrne, Helen M. "Dissecting cancer through mathematics: from the cell to the animal model". *Nature Reviews Cancer* (2010). 10 (3): 221–230.
7. H. Fröhlich, J. K. Wegner, A. Zell, *Kernel Functions for Attributed Molecular Graphs - A New Similarity Based Approach To ADME Prediction in Classification and Regression*, QSAR Comb. Sci., 2006, 25, 317-326.
8. H. Fröhlich, J. K. Wegner, A. Zell, *Assignment Kernels For Chemical Compounds*, International Joint Conference on Neural Networks 2005 (IJCNN'05), 2005, 913-918.
9. P. Mahe, L. Ralaivola, V. Stoven, J. Vert, *The pharmacophore kernel for virtual screening with support vector machines*, J Chem Inf Model, 2006, 46, 2003-2014.
10. J. K. Wegner, H. Fröhlich, H. Mielenz, A. Zell, Data and Graph Mining in Chemical Space for ADME and Activity Data Sets, QSAR Comb. Sci., 2006, 25, 205-220.
11. Xiaohong Wang, Jun Huan , Aaron Smalter, Gerald Lushington, *Application of Kernel Functions for Accurate Similarity Search in Large Chemical Databases* , in BMC Bioinformatics Vol. 11, *2010*, p878-888
12. A.Goulon, T. Picot, A. Duprat, G. Dreyfus "Predicting activities without computing descriptors: Graph machines for QSAR". *SAR and QSAR in Environmental Research* (2007). 18 (1-2): 141–153.

Corresponding author: M. Padmavathi, Department of Biotechnology, DVR & Dr. HS MIC College of Technology, Kanchikacherla, A.P. India