# Extreme Pathway Analysis of Mycobacterium Tuberculosis

Jelili Oyelade[1], Itunuoluwa Isewon[1], Femi Aromolaran[1]
[1]Department of Computer and information Sciences
Covenant University
Ota, Nigeria
ola.oyelade@covenantuniversity.edu.ng

Moses Achas[2]
[2]Department of Computer Science
Bells University of Technology,
Ota, Nigeria

***Abstract*—Tuberculosis is a multisystem disorder characterized by the formation of swelling that is filled with blood which is caused from breakage in the wall of a blood vessel. This breakage can occur in different organs of the body and is caused by a bacterium known as *Mycobacterium Tuberculosis (mtb)* which can be represented as a metabolic network with genes, proteins and enzymes that are interconnected. This interconnection defines the uniqueness of any bacteria. The analysis of a metabolic network system is achievable through different computational techniques depending on what information is available. The flux balance analysis is mostly used for analyzing this type of network because of the little amount of information required. The application of flux balance analysis to *mtb* involves the conversion of the metabolic network into a stoichiometric matrix where the rows represents the metabolites and the columns represent the reactions. In this study, the stoichiometric matrix is an 828 by 1027 matrix. The analysis generated a linear problem having more unknowns than the number of equations. This type of problem is normally solved using an extreme pathways algorithm to extract independent paths and simplex method for optimization of biomass. Here, the extreme pathways analysis was used in the categorization of the metabolite of *mtb* while biomass was employed as the objective function using default constraints. The output represents three categories of metabolite: 31 metabolites that form part of the biomass component that are inactive, 14 metabolites that remain active and 32 metabolites that are activated after the optimization process.**

***Keywords*—** *Mycobacterium Tuberculosis; Metabolic network system; Flux balance; Extreme pathways algorithm; Stoichiometric matrix; Biomass*

## I. INTRODUCTION

One of the diseases that has caused a lot of infection and have also led to the death of a lot of people is Tuberculosis. This disease is caused by Mycobacterium Tuberculosis (*mtb*). Tuberculosis is a multisystem disorder that involves the formations of hematomas in multiple organs of the body [1]. In order to achieve a successful evaluation of *mtb*, the extreme pathways analysis which is focused on extracting the sets of metabolites pathways that are independent must be implemented. The output is used for the optimization of the biological system using an objective function. Presently, the early discovery of this disease and a cure still remain a threat in medicine because there is no single symptom that can be used to uniquely identify the disease [1].

Most of the cases of *mtb* aside from the fact that in claims the lives of most of its victims, is the difficulty that exist in the early diagnosis of this disease. The disease is very difficult to diagnose at early stages because the tuberculosis disease has the potential to remain in a dormant state without detection. Adding to the peculiarity of this disease is the fact that it is not hereditary so there is no means of genetically tracing the disease.

Metabolic network analysis carried out in most literatures use a part of a metabolic network as a basis to evaluate the overall behaviour of the network without putting into consideration what other metabolites are affected by that section of the network. So in any metabolic network analysis, it is important to examine what component of the metabolic network is affected by a target before using the target to analyze the overall behaviour of the network when such target is manipulated. It is this ignored aspect that is the focus of this research with application to *Mycobacterium Tuberculosis.*

The study therefore applied flux balance analysis to the categorization of the metabolites of *mtb* using biomass. This was achieved by the construction of metabolic network for *mtb,* then, the extraction of Stoichiometric matrix for the organism and next was the development of a flux balance analysis system. Finally, the flux balance analysis was applied to an objective function.

## II. RELATED WORKS

Extreme pathways analysis is an approach that has been applied to different organism for prediction purposes. Such as to discover if the organism's predicted computational growth rate is similar to the experimental growth rate in the laboratory as conducted by [2] where they used the extreme pathways approach to measure and predict the secretion rate of Escherichia coli using aerobic and anaerobic cells.

An extreme pathway is also used for predicting drug targets for different bacterium. These drugs could serve as a cure for those diseases or limit the damage the bacterium causes. Study by [3] applied the extreme pathways to predicting possible drugs for the treatment of malaria. Apart from drug prediction,

978-1-5386-2652-8/17 $31.00 © 2017 IEEE
DOI 10.1109/CSCI.2017.223

the distinct characteristics of an organism can also be derived through the extreme pathways [4].

Other area of use of extreme pathway analysis is in the area of reaction rate [5] and also to investigate the relationship between an organism and its environment [6].

## III. MATERIALS AND METHODS

### A. Metabolic Network Construction

A metabolic network reconstruction collects all of the relevant metabolic information of an organism and compiles it in a mathematical model. "Validation and analysis of reconstructions can allow identification of key features of metabolism such as growth yield, resource distribution, network robustness, and gene essentiality" [7]. "The Advancement of sequencing technologies and high throughput "omics" methods has greatly aided the reconstrution of metabolic networks at a genomic scale" [8]. Latest development in genome annotations and the progress of omics have made it easier to extract metabolism and possible information at genomic scale [8]. This has produced great efforts in network reconstruction and predictive model building. Re-constructions for organisms network has been possible with inadequate biochemical data, most of whose networks are reconstructed depending on genome annotations [8].The model includes the listed information below:

- The list of all possible enzymes existing in the organism of interest
- The substrates and products of enzymatic reactions
- The stoichiometric coefficients of each of these compounds
- Information about thermodynamic reversibility or irreversibility of each reaction
- Conditions for the activity or inactivity of each enzyme
- Sub cellular localizations of each enzyme

### B. Stoichiometric Matrix

Stoichiometric matrix is a component of flux balance analysis that is used to represent the attributes that makes up a metabolic network usually denoted by S. It is an *m x n* matrix in which m (row) represents the individual metabolites involved in every reaction that exists in the metabolic network why n (column) represent the number reactions involved in the metabolic system [9]. Building a stoichiometric matrix from a metabolic network follow some standard rules such as;

- In every reaction that takes place within a metabolic network, a metabolite is either produced or consumed.
- In a particular reaction,
    - If a metabolite is produced (gain), the coefficient of that metabolite is used to represent that metabolite under that reaction.
    - If a metabolite is consumed (loss), the coefficient of that metabolite is also used to represent that metabolite but with a negative sign.
- If a metabolite does not take part in a particular reaction, it is represented with a zero value.
- If a reaction is bi-directional, it is represented as two different reactions in which one is positive while the other is negative.

The stoichiometric matrix of any metabolic network represents the metabolites involved in a reaction, the status of such metabolites in that reaction (loss or gain?) and the quantity of the metabolite involved in that reaction. The point of interception of the rows and column in the stoichiometric matrix indicates the status of every metabolite. It shows if that metabolite partakes of a particular reaction, and if it does, the sign (+/-) of the value at the point of interception will indicate if the metabolite is consumed or produced. The value at the point of interception also indicates the quantity of that metabolite involved in the reaction [10].

### C. Flux Balance Analysis Formulation

The Impose constraints setting stage is the stage where the flux balance analysis is formed. This involves three other stage which are:
Dynamic mass balancing
Steady state assumption
Linear problem formulation
At the dynamic mass balancing stage, the formula:

$S.v = \frac{dc}{dt}$ is applied to the constructed stoichiometric matrix.

*c = concentration*
*t = time*
*s = stoichiometric matrix*
*v = flux vector*

That is, the differentiation of compound concentration over time equals the multiplication of stoichiometric matrix and flux vector.

After the mass balancing, the next stage which is the steady state assumption is implemented. The steady state assumption states that the differentiation of compound concentration over time is zero. That is;

$$\frac{dc}{dt} = 0$$

Hence the mass balancing in the previous step results to:

$$S.v = 0$$

That is, the multiplication of the stoichiometric matrix and the flux vector results to zero. It is this steady state assumption that gives the flux balance analysis method an advantage over the other metabolic network analysis techniques.

The resolving of the steady state equation (S.v = 0) leads to the linear problem formulation. It's a linear problem because the numbers of unknowns are more than the number of equations. In order to evaluate this steady state equation, an

algorithm known as the Extreme pathways algorithm used originally by [11] is applied.

At steady state, there are some groups of reactions (pathways) that remain active in the existence of the organism. These are referred to as extreme pathways. It is believed that pathways, particularly the extreme pathways are what cells aim to regulate. Hence, rather than guessing, it is more beneficial to understand the regulatory mechanism of these pathways and tries to relate it to theoretical pathway structure. These pathways are the ultimate objective of cellular regulation.

The unique extreme pathways can be used to make deductions on the functionality in terms of capability and also how flux are distributed [11]. Extreme pathways are used to analyze an organism metabolic network at steady state. At steady state, the metabolic network of an organism which is represented by a stoichiometric matrix forms a convex shape which results from some pathways that forms the existence of that organism. The apex of this convex are referred to as extreme pathways.

The space within these extreme pathways (convex) forms the basis of the existence of such organism at steady state. It also forms the solution space from where optimization can be evaluated [12]. The extreme pathway algorithm involves three major categories: the initialization stage, the main calculation loop and the balance External Metabolites.

Initialization Stage: This begins with the appending of an $nxn$ matrix to the transpose of the stoichiometric matrix $S^T$ of the given metabolic network to form an initial table T. with the new matrix T, the external fluxes that are without constraints are transferred to a temporary matrix $T^{E.}$ After enforced this separation, the metabolites {M} without an external flux connected with it must also be identified.

Main Calculation loop stage: Here, the extreme pathway algorithm tries to balance the fluxes on each of the metabolites. This is done through iterations such that the $i^{th}$ iteration copies all the rows from $T^{i-1}$ that contains a zero in the column of the transposed stoichiometric $S^T$. This column becomes the pivoting column. The rows of the column in the pivoting row with values of opposite signs are added together to form a new matrix $T^i$ [6].

Balance External Metabolite: This begins by appending the $T^E$ to the resulting matrix from the previous phase. From the $S^T$ matrix, the first non-zero column is added to the corresponding row in $T^E$. This is done so as to achieve a zero in the non-zero position in the column from $S^T$. When zeros have been achieved in the metabolites, the $T^E$ can then be removed. The resulting matrix T forms the extreme pathways of the metabolic network [6].

### *D. Simplex Method*

The simplex method is an approach to solving a linear equation problem such that there is an objective function that needs to be optimized using some given constraints. The stages involved are:

Step 1

All the equations must be standardized by introducing a slack variable. For the objective function, this is done by making the variable to be optimized the subject of formula.

Stage 2

The second stage of the operation of the simplex method is a test for negative elements in the bottom row of the matrix. The aim of the simplex linear solving problem is to optimize by eliminating negative values from the last row of a given matrix. If negative value(s) exist, an optimization procedure is carried out otherwise; there is already an optimized value for the given objective function.

Stage 3 and 4

This stage is implemented when there exists a negative element in the last row of the given matrix haven appended the objective function. At this phase, a pivot column is selected by picking the most negative element from the last row of the matrix. The column of such element becomes the pivot column.

Stage 5

The goal of getting a pivot row is to produce a 1 for the coefficient above the most negative element that gives the smallest quotient after division. This is achieved through the use of an elementary row operation. That is dividing the pivot row by the number in the position where a 1 in needed. After obtaining a 1, there is need to obtain a zero in the other positions of the pivot column. This is achieved by subtracting multiples of the pivot rows from the other rows. The process returns to stage 2 to test for the most negative element in the last row then repeat the process till there is no negative element in the last row.

## IV. RESULTS AND DISCUSSION

The optimization approach in this research is the metabolite optimization approach where a set of target metabolite (biomass) is used as the optimized metabolite to extract metabolites that becomes active and inactive at the given biomass metabolite. This optimization is confined to the solution space formed by the extreme pathways. The biomass component of *mtb* is comprised of 45 metabolites.

With reference to the biomass metabolites, the set of metabolites that are active is extracted using the equation:

$$v' = \sum_{i=1}^{k} w_i E P_i, \quad w_i \geq 0$$

Where;

$V$ = biomass metabolite
$W$ = activated metabolite (desired output)
$EP$ = Extreme pathways
$i - k = 1$
$w \geq 0$

Using the above equation, the output 'w' represents a list of metabolites that becomes active when '$v$' is activated.

The optimization result is made up of different metabolites that are active at the given objective targets. The most important aspect of extreme pathway optimization are the extractions of metabolites that become activated at the

optimization phase. Biologically, they represent a further reduction in the solution space created by the extreme pathway analysis with respect to the objective function. In this research, these represent a set of metabolite that is linked with biomass in *mtb*. That is, production of biomass also links to the production of these set of metabolites. Hence to reduce metabolite production which is to reduce the growth of the bacterium, this group of metabolites should be considered. The metabolites produced in this study are listed in Table I with the abbreviation and their full name.

TABLE I. UNIQUELY ACTIVE METABOLITES

| **Abbreviation** | **Full Name** |
|---|---|
| arachACP(c)- | eicosanoyl-ACP |
| cmcbtt(c)- | carboxymycobactin T |
| dhpt(c) | Dihydropteroate |
| hdcoa(c)- | Hexadecenoyl-CoA |
| hexccoa(c)- | Hexacosanoyl-CoA |
| lac-L(c) | L-Lactate |
| meroacidACP(c)- | alpha meroacid ACP |
| meroacidcyc1ACP(c)- | cyclopropanated alpha meroacid ACP (1 cyclopropane ring |
| mi1p-D(c)- | 1D-myo-Inositol 1-phosphate |
| mppp9(c)- | Magnesium protoporphyrin |
| mql6(c)- | Menaquinol 6 |
| mqn6(c)- | Menaquinone 6 |
| mshfald(c)- | mycothiol conjugated methanol |
| odecoa(c)- | Octadecenoyl-CoA |
| ohpb(c)- | 2-Oxo-3-hydroxy-4-phosphobutanoate |
| omdtria(c)- | octa-methyl dotriacontanoic acid (C32:0 with 7 methyl branches) |
| palmACP(c)- | Palmitoyl-ACP (n-C16:0ACP) |
| pan4p(c)- | Pantetheine 4'-phosphate |
| pant-R(c)- | (R)-Pantoate |
| pep(c)- | Phosphoenolpyruvate |
| peptido_TB1(c)- | peptidoglycan subunit |
| pgp190(c)- | Phosphatidylglycerophosphate (dituberculostearoyl, C19:0) |
| phdca(c)- | phenol palmitic acid |
| phdcaACP(c)- | phenol palmitic acid ACP |
| ppp9(c)- | Protoporphyrin |
| tmhexc(c)- | tri-methyl hexacosanoate |
| tmlgnc(c)- | tri-methyl lignoceric acid |
| trdox(c)- | Oxidized thioredoxin |
| trdrd(c)- | Reduced thioredoxin |
| uaAgla(c)- | Undecaprenyl-diphospho-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-D- |
| alanyl-D- | alanine |
| pyr(e)- | Pyruvate |

Table II is a collection of metabolites that are part of the biomass component and also remain active as part of the component of the extreme pathways. The list of metabolites in Table II includes abbreviation and their full name.

There are also a group of metabolites that are used as part of the objective flux in optimizing the extreme pathways that are not present in the output of the optimization result. These comprised of some set of metabolites that were active as a part of the objective flux but become inactive and as a result are not parts of the optimization result. There are 31 of such metabolite that became in-active. The metabolites in this category are listed in Table III below with their abbreviation and the full name.

TABLE II. COMMON ACTIVE METABOLITES BETWEEN BIOMASS AND ACTIVE FLUX

| **Abbreviation** | **Full name** |
|---|---|
| Ac1PIM1- | acyl phosphatidylinositol mannoside |
| Ac1PIM2- | acyl phosphatidylinositol mannoside di-mannose |
| Ac1PIM3- | acyl phosphatidylinositol mannoside tri-mannose |
| Ac1PIM4- | acyl phosphatidylinositol mannoside tetra-mannose |
| Ac2PIM2- | di-acyl phosphatidylinositol mannoside di-mannose |
| PIM1- | phosphatidylinositol mannoside |
| PIM2- | phosphatidylinositol mannoside di-mannose |
| PIM3- | phosphatidylinositol mannoside tri-mannose |
| PIM4- | phosphatidylinositol mannoside tetra-mannose |
| PIM5- | phosphatidylinositol mannosidepenta-mannose |
| PIM6- | phosphatidylinositol mannosidehexa-mannose |
| Arach- | Arachidic acid |
| Dhpmp- | Dihydroneopterin monophosphate |
| Hdcea- | Hexadecenoate |

TABLE III. LIST OF DEACTIVATED METABOLITES IN OBJECTIVE FUNCTION

| Abbreviation | Full name |
|---|---|
| Hexc- | hexacosanoate |
| Kmycolate- | ketomycolate (2 cyclopropanated rings) |
| Mbhn- | methyl behenic acid |
| Mcbts- | mycobactin S |
| Mfrrppdima- | phenolic glycolipid (Mtb) |
| Mmmycolate- | methoxymycolate (1 cyclopropanated ring) |
| Mmycolate- | methoxymycolate (2 cyclopropanated rings) |
| Mocdca- | 10-methylstearic acid |
| mql8- | Menaquinol |
| ocdca- | octadecanoate |
| ocdcea- | octadecenoate |
| pa160-1,2- | dihexadecanoyl-sn-glycerol 3-phosphate |
| pa160190- | 1,2-sn-glycerol 3-phosphate |
| pa190190- | 1,2-sn-glycerol 3-phosphate |
| pdima- | phthioceroldimycocerosate |
| pe160- | phosphatidylethanolamine |
| pg160- | Phosphatidylglycerol |
| pg160190- | Phosphatidylglycerol (hexadecanoyl, methylstearoyl) |
| pg190- | Phosphatidylglycerol |
| ppdima- | phenol phthioceroldimycocerosate |
| tmha1- | tetramycolylhexaarabinoside |
| tmha2- | tetramycolylhexaarabinoside |
| tmha3- | tetramycolylhexaarabinoside |
| tmha4- | tetramycolylhexaarabinoside |
| ttdca- | tetradecanoate |
| phdca- | phenol palmitic acid |
| fcmcbtt- | iron(III) chelated carboxymycobactin T |
| hdca- | Hexadecanoate |
| clpn160190- | cardiolipin (dihexadecanoyl, dimethylstearoyl) |
| dhor | Dihydroorotate |
| arabinanagalfragund- | arabinofuranose-galactofuranosyl(30)-rhamanosyl-N- |
| arabinan- | acetylglucosamyl-undecaprenyl diphosphate |

The summary of the observation from this research is summarized in the graph below in Figure 1.

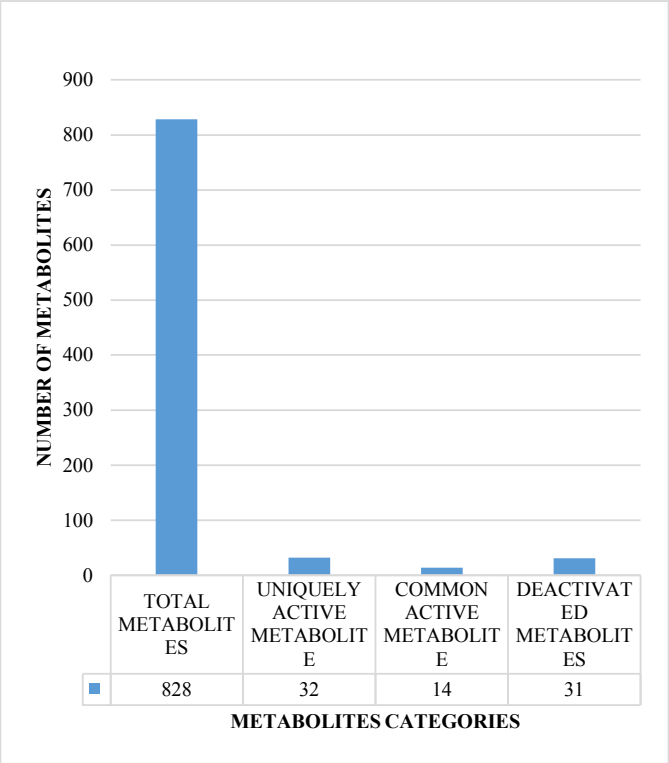


Figure. 1: Graphical representation of result summary

## V. CONCLUSION

This study has been carried out to investigate the capabilities of computational methods and to predict the metabolic activities of the *mtb*. In line with this objective, a metabolic network was extracted from various biological databases. The extracted metabolic network was represented in a stoichiometric matrix. A familiar method to getting an understanding into the metabolism of an organism is to analyze the network using a computational approach; Extreme pathway analysis was used in this work to investigate the biomass component of *mtb*. This was achieved by extracting the extreme pathways and then using these pathways to evaluate the behaviour of the metabolites in the network at certain conditions. The result provided us with information listing all the extreme pathways that are active at the given biomass component objectives.

Tuberculosis is a major infection that has claimed a lot of victims. Its peculiarity is in the ability to remain dormant for a

long time without being detected. A lot of research is ongoing with aims to reduce the impact of this disease on humanity. A biological understanding of the nature of the bacteria that causes this disease (on which this study is based) will provide additional information to help develop cure for the disease. In this work, the nature of the cause of tuberculosis in terms of pathway analysis has been identified. The results therefore help in the analysis of this bacterium in terms of pathway redundancy, correlated reaction set and extreme pathway reaction participation.

## ACKNOWLEDGMENT

We acknowledge the support of Covenant University Center for Research, Innovation and Discovery, Covenant University, Ota, Nigeria.

## REFERENCES

[1] B.A. Stanley, E.A. Vail and E.A. Thiele, "Tuberous Sclerosis Complex: Diagnostic challenges, presenting symptoms and commonly missed signs. Paediatrics publications," Rediatrics, vol. 127, pp. 117-125, 2011.

[2] X. Chen, A.P.Alonso AP, Allen, J.L. Reed, Y. Shachar-Hill, "Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*," Metab Eng, vol. 13 pp. 38–48, 2011.

[3] C. Huthmacher, A. Hoppe, S. Bulik and H. Holzhutter, "Antimalarial drug targets in Plasmodium falciparum predicted by stage-specific metabolic network analysis," BMC System Biology, vol.4, pp. 1-27, 2010.

[4] J.S. Edwards, R.U. Ibarra and B.O. Palsson "Insilico predictions of Escherichia coli Metabolic capabilities are consistent with experimental data," Nature Biotechnology, vol. 19, pp. 125-130, 2011.

[5] R. Mahadevan, J.S. Edwards and F.J. Doyle, "Dynamic flux balance analysis of Diauxic Growth in Escherichia Coli," Biophysical Journal, vol. 83, pp. 1331-1340, 2011.

[6] L. Lee, J. Varner and K. ko, "Parallel Extreme Pathway Computation for Metabolic Networks," IEEE Computational System Bioinformatics, pp. 636-639, 2004.

[7] C. Franke, R.J. Siezen and B. Teusink, "Reconstructing the metabolic network of a bacterium from its genome," Trends in Microbiology, vol. 13, pp. 550-558, 2005.

[8] M.F Adam J.H. Markus, T. Ines, L.R. Jennie and P. Bernhard, "Reconstruction biochemical networks in microorganisms," Nature Reviews Microbiology, vol. 7, 129 – 143, 2005.

[9] L. Nakhleh, "Bioinformatics: Network Analysis - Analysing stoichiometric matrices. Molecular," System Biology, 2013.

[10] F. Llaneras F and J. Pico, "Stoichiometric modelling of cell metabolism," Journal of Bioscience and Bioengineering, vol. 105, pp. 1-11, 2008.

[11] C.H. Schilling and B.Q. Palsson, "Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis,". J. Theor.Biol., vol. 203, pp. 249–283, 2000.

[12] J.M. Lee E. P.Gianchandani and A.P. Jason, "Flux balance analysis in the era of metabolomics," Briefings in Bioinformatics, vol. 7, pp. 140-150, 2006.