

# Analogizing And Investigating Some Applications of Metabolic Pathway Analysis Methods

Gourav Mukherjee<sup>1</sup>,  
Aditya Dwarkani<sup>2</sup>,  
Tanmoy Kumar Dawn<sup>2</sup>,  
Ms. Satarupa Bagchi Biswas<sup>4</sup>

**Abstract**— Comparing metabolic network analysis approaches which include the structural and stoichiometric modeling methods under which comes Network Based Approach and Constraint Based Approach. In Constraint Based Approach we have analyzed and implemented the Flux Balance Analysis Method and under Network Based Approach we have reviewed and analyzed the Extreme Pathway Analysis Approach.

**Index Terms**— Introduction on Metabolic Pathways and Methods, sources os data of Metabolic Pathways, Algorithms of FBA, Algorithm of EPA, Output of FBA, Paths of EPA

## 1 INTRODUCTION

System Biology[1] is a study of complex biological processes and the process of bringing together the component sub-systems into one system and ensuring that the subsystems function together as a system. The interactions involve a series of biochemical conversion and form different biochemical pathways. Major kinds of biochemical pathways are - signal transduction, gene regulatory and metabolic networks. The kind of the biochemical pathway we have dealt is metabolic networks.

Metabolic pathways[1] are a series of biochemical reactions, mostly catalyzed by enzymes. It involves step by step modification of an initial molecule (substrate) to form a product. Methods of metabolic pathways analysis that we have implemented are Flux Balance Analysis (FBA) and Extreme Pathway Analysis (EPA). FBA is a stoichiometric constraint basec analysis model that calculates the flow of metabolites through the metabolic network to predict the possible growth rate of an organism and the quantity of biotechnologically important metabolites to be produced during metabolism. EPA are a unique and minimal set of vectors that completely characterizes the steady state capabilities of a metabolic network and its implementation is provided to find out the length of extreme pathway and how individual reactions take part in it.

One of the approaches under stoichiometric modeling is constraint based that applies a set of constraints on a metabolic pathway[3] to characterize its possible behaviours. There are many constraint based methods among which we have

implemented Flux Balance Analysis(FBA) which is the classical starting point of constraint based modeling.

Flux Balance Analysis[2] is mathematical method of analyzing the metabolic capacity of the cell . The objective of FBA[3] is to find out the set of metabolic fluxes that maximizes the growth rate of the target metabolite, given some known available nutrients. The earliest work of FBA includes the way to construct flux balance equations using a metabolic map. Its success can be seen in the ability to accurately predict the growth of prokaryote.

The other approach under this stoichiometric modeling is Network Based Approaches. In this modeling with the help of existing knowledge of cellular components and their connectivities systemic functiond are described mathematically. Under network based approaches we have few methods from which we have dealt with Extreme Pathway Analysis.

Extreme Parhway Analysis[4] is a unique and minimal set of vectors that completely characterizes the steady state capabilities of genome scale metabolic network. The length of the extreme pathway is the number of reactions that comprises it. The mathematical framework is provided to find out length of extreme pathway and how individual reactions take part in it. This pathway has the following characteristics-

It generates unique and minimal systemic pathway. It describes all possible steady state flux distributions that the pathway network can achieve. It determines the time invariant topological properties of the network. Extreme pathway can be characterized by its length and reaction participation.

## 2 OUR IMPLEMENTATION

### 2.1 Flux Balance Analysis

The xml document location is taken as a parameterized input, using a DOM parser the document is parsed, where the attribute values are obtained from the xml file with which the node name is checked to find whether the substrate is distinct or product is distinct. With the rows considered as the total number of reactions. The columns considered as the substrate

1. Author: Gourav Mukherjee is currently pursuing bachelor degree prgram in information technology in Heritage Institute of Technology, Kolkata, India, PH-+919830050527. E-mail: [gouravmukherjee1993@gmail.com](mailto:gouravmukherjee1993@gmail.com)
2. Co-Author: Aditya Dwarkani is currently pursuing bachelor degree program in information technology in Heritage Institute of Technology, Kolkata, India, PH- +918697721000. E-mail: [aditya.dwarkani@gmail.com](mailto:aditya.dwarkani@gmail.com)
3. Co-Author: Tanmoy Kumar Dawn is currently pursuing bachelor degree program in Information Technology in Heritage Institute of Technology, Kolkata, India, PH-+918293449763. E-mail-[tanmoydawn@gmail.com](mailto:tanmoydawn@gmail.com)
4. Co-Author: Ms. Satarupa Bagchi Biswas is Assistant Professor in the department of Information Technology in Heritage Institute of Technology, Kolkata, India, PH-+919836699449. E-mail-[satarupa.biswas@heritageit.edu](mailto:satarupa.biswas@heritageit.edu)

and products. If it is a product then 1 is inserted into the stoichiometric matrix, if it is substrate then -1 is inserted into the matrix. In this way the stoichiometric matrix is generated. The mass balance constraint equations are thus formed with the help of it. The system of mass balance equations at steady state is made equal to 0. The reaction is set as objective with the objective function  $Z=c_1V_1+c_2V_2+.....+c_nV_n$ . The reaction set as objective will have a corresponding value 1 and others will have value as 0.  $c^T$  is  $1*n$  matrix. We have maximized the objective function and generate the output. We have also found out the constraints for our convenience.

## 2.2 Extreme Pathway Analysis

The xml document location is taken as parameterized input using DOM parser the document is parsed, where the attribute values are obtained from the xml file with which the node name is checked whether it is a substrate or a product. With the rows considered as the total number of reactions. The columns are considered as the substrate and product. If it is a substrate or product 1 is inserted into the corresponding position in the stoichiometric matrix, in this way the stoichiometric matrix formed. We have also taken the source node and destination node as the input and we try to find all the possible paths from the source node to destination node. Here the nodes are the substrates and products. In this way the extreme pathway matrix in which extreme pathways are considered as column and the substrates and products together constitutes row. The substrate or products constituting the paths when considered will be checked if present then 1 is inserted to corresponding position in the extreme pathway matrix and if not present then 0 is inserted. We call this matrix as pathway matrix. After getting this matrix we transpose it and then find the product of pathway matrix and its transposed matrix. The resultant matrix which we obtain after this multiplication is called Pathway Length Matrix. The diagonal elements in this Pathway Length Matrix represent the length of corresponding pathway. The transpose of pathway length matrix is calculated, then pathway length matrix and its transpose are multiplied, the result obtained is the reaction participation matrix. The off diagonal elements refer to the number of extreme pathways that contain both corresponding reactions.

## 3 SOURCES OF DATA OF METABOLIC PATHWAYS WE HAVE WORKED ON

### 3.1 Kyoto Encyclopedia of Genes And Genomes

It is a database containing information of genes, proteins, reactions, pathways and useful for building association among enzymes, reactions, genes. Information about human diseases, drugs and other health related substances. KEGG is queried through a language based on XML, called KEGG Markup Language (KGML).

### 3.2 JWS Online Cellular Systems Modeling And Microbiology

Online website aims in providing a user friendly internet based repository of such pathway models and also an application for running kinetic models of biological systems.

## 4 ALGORITHM OF FLUX BALANCE ANALYSIS

### Prerequisite

Parse the XML file.

### XML Processing

**Step 1:** Obtain the data of the substrate, product and reactions.

**Step 2:** Check for distinct products and substrates.

**Step 3:** Store all these values.

### Stoichiometric Matrix Formation

**Step 1:** Form a 2d array stoic of size  $m*n$ .

**Step 2:**  $m$  is the number of reactions.

**Step 3:**  $n$  is the number of substrate and products.

**Step 4:** Run a loop from 0 to  $m$

Run one more loop from 0 to  $n$ .

Check if a reaction contains substrate then

set the corresponding position in the stoic as -1.

Check if a reaction contains product then

set the corresponding position in the stoic a +1.

Otherwise set it as 0.

**Step 5:** End

### Mass Balance Constraint And Maximize Objective Function

**Step 1:** Run a loop from 0 to total number of products and total number of substrates

Run one more loop from 0 to total number of reactions.

Check each and every value of the stoichiometric matrix if it is +1 mark it with +V

Otherwise if it is -1 mark it with -V

**Step 2:** Set the objective function

Run a loop from 0 to total number of products

Run one more loop from 0 to total number of reactions, now we add both constraints +V and -V.

**Step 3:** Using Simplex method of linear programming maximize the objective function given the constraints

## ALGORITHM OF EXTREME PATHWAY ANALYSIS

### Prerequisite

Parse the XML file.

### XML Processing

**Step 1:** Obtain the data of the substrate, product and reactions.

**Step 2:** Check for distinct products and substrates.

**Step 3:** Store all these values.

### Stoichiometric Matrix Formation

- Step 1:** Take a 2d array stoic size  $m \times n$ .
- Step 2:**  $m$  is the number of substrate and number of products.
- Step 3:**  $n$  is the number of reactions.
- Step 4:** Run a loop from 0 to  $m$   
Run one more loop from 0 to  $n$ .  
Check the reaction if contains substrate  
Then set the values in stoic at corresponding position as  $-1$ .  
If it contains products then set the values in stoic at corresponding position as  $+1$ .  
Otherwise set it as 0.
- Step 5:** End.

### Graph

- Step 1: Considering two string nodes as parameters add them as adjacent node and form an edge in hash.
- Step 2: If the two vertex are two way then add them as Edges in a hash.
- Step 3: Store this hash in a linked list.

### Search

- Search(Graph graph,LinkedList visited)
- Step 1: Set the START and END nodes.
- Step 2: If visited contains the node then continue
- Step 3: Check if the node contains end node or if the list contains the visited node then add the node.
- Step 4: Recursively call Search(graph,visited).

### Extreme Pathway Matrix Formation

- Extremepathway (XMLProcessor proc)
- Step 1: Run a loop from 0 to pathway string length.  
Run one more loop from 0 to individual node length.
- Step 2: Check whether the id of the substrate matches with the element obtained from processing if it matches then set the corresponding position value in EPA matrix as 1.
- Step 3: Run a loop from 0 to the total number of individual nodes.  
Run one more loop from 0 to total number of products.  
Now check whether product id matches with the element obtained from processing.  
If it matches then set the corresponding value in the EPA matrix as 1.

### Pathway Length Matrix And Reaction Participation Matrix

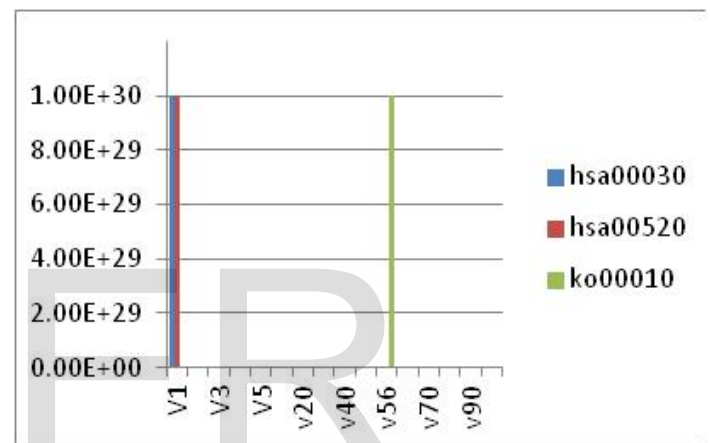
- Step 1: Take a 2d array EPATr which is the transpose of EPA Matrix.
- Step 2: Now, multiply both EPATr and EPA. Store this result in a new array named PL which is Pathway Length matrix.
- Step 3: Take the diagonal elements which represents the corresponding pathway.
- Step 4: Now we find the product of EPA and EPATr.

- Step 5: Store this result in a new array named Reac. which is Reaction Participation Matrix
- Step 6: Take the off diagonal elements which represents the number of extreme pathways that contains both the corresponding reactions.

## 5 RESULT

### 5.1 Output From Flux Balance Analysis

From the graphical representation we can see the maximized output of the pathway  $V_1, V_2, \dots$  these are constraints and maximized value is obtained from the linear programming is able to denote the value of  $V$  which is displayed in the graph.



### 5.2 Output From Extreme Pathway Analysis

#### i. For Pathway ko00010

- Path Id 1: 93-76-94-129-125-140-85-54-92-52-98-46-99-43-101-37-40-
- Path Id 2: 93-76-94-129-125-140-85-54-92-52-98-46-99-43-101-39-40
- Path Id 3: 93-76-94-129-125-140-85-54-92-52-98-46-99-43-101-138-40-
- Path Id 4: 93-76-94-129-125-140-85-54-92-52-98-47-99-43-101-37-40-
- Path Id 5: 93-76-94-129-125-140-85-54-92-52-98-47-99-43-101-39-40-
- Path Id 6: 93-76-94-129-125-140-85-54-92-52-98-47-99-43-101-138-40-

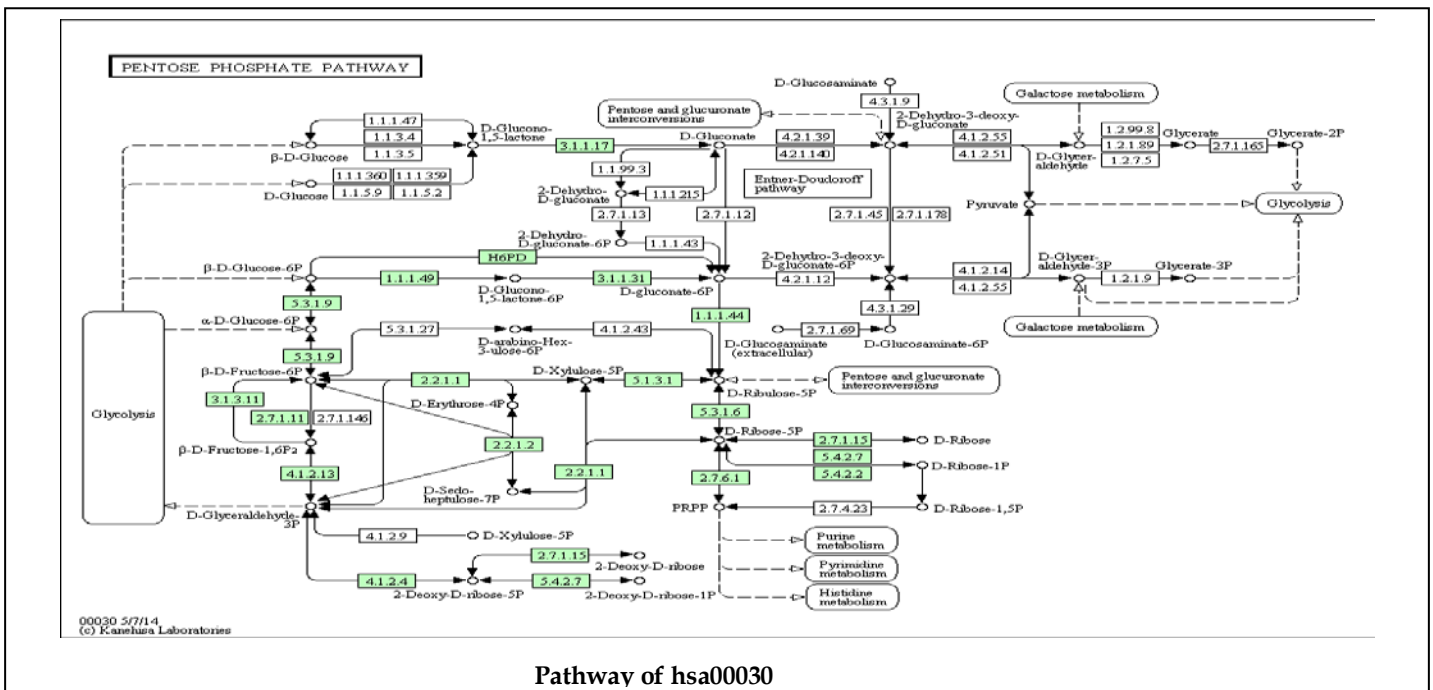
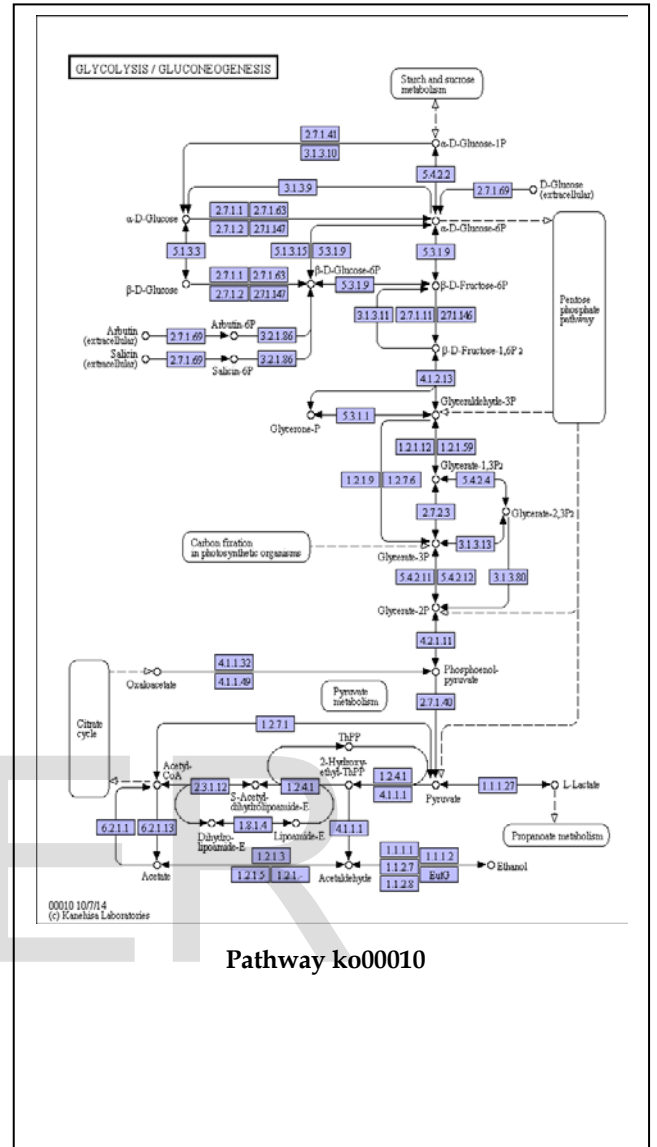
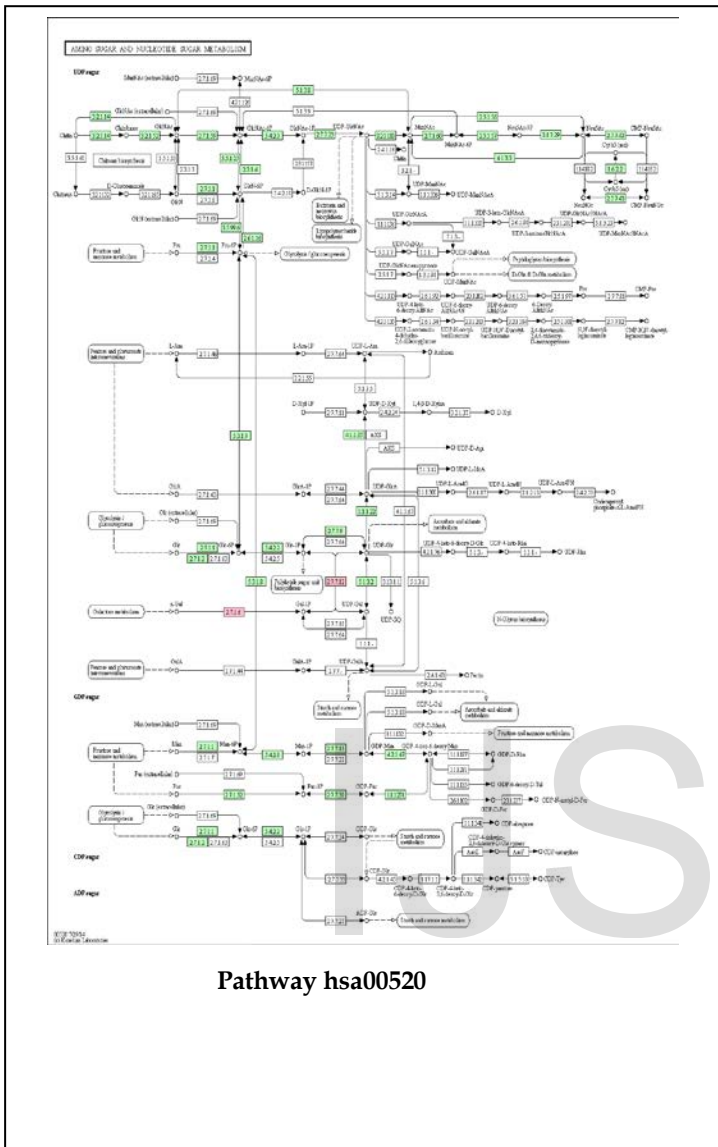
#### ii. For Pathway hsa00520

- Path Id 1: 93-2-92-19-100-

#### iii. For Pathway hsa00030

- Path Id 1: 99-37-121-58-122-57-106-56-131-
- Path Id 2: 99-37-121-97-106-56-131-
- Path Id 3: 99-53-124-51-120-50-19-46-110-47-

IJSER



## 6 CONCLUSION

Therefore we would like to conclude that Flux Balance Analysis is an approach for analyzing the flow of metabolites through a metabolic network and it is able to calculate the possible growth rate of an organism. It doesn't need kinetic parameters and can be computed quickly even for large network. The other approach which is Extreme Pathway Analysis describes the conversion of substrate into product and it accounts for all reaction steps. Extreme Pathway specifies theoretical upper and lower bounds of conversion of any substrate into product.

## 7 ACKNOWLEDGMENT

We are thankful to Mrs. Satarupa Bagchi Biswas for providing us facilities and kind support throughout the research work and also for constantly guiding us for giving us invaluable insights. Her guidance and words of encouragement motivated us to achieve our objective and impetus to excel and also playing a vital and decisive role during the development of our research work.

## 8 REFERENCES

- [1] Tomar N, De R.K.: Comparing methods for metabolic network analysis and an application to metabolic engineering - Gene (Elsevier), Volume 521, Issue 1, 25 May 2013, Pages 1-14
- [2] Jeffrey D Orth, Ines Thiele & Bernhard Ø Palsson: What is flux balance analysis? - Nature Biotechnology 28, (2010), Page-245-248
- [3] Karthik Raman and Nagasuma Chandra: Flux Balance analysis of biological systems: applications and challenges- BRIEFINGS IN BIOINFORMATICS. VOL 10. NO 4., March 15, 2009, Page No: 435- 449
- [4] Price ND, Reed JL, Papin JA, Famili I, Palsson BO.: Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices - Biophys Journal. Vol: 84(2) Feb 2003, Page No: 794-804