Efficiency Comparison Of Different Extreme Pathways By Genetic Algorithm

Gourav Mukherjee [1],
Aditya Dwarkani [2],
Mrs. Satarupa Bagchi [3]

Abstract— Extreme Pathway analysis method includes determination of extreme pathways consider all the necessary reaction steps of a network that must be used to complete the synthesis process.Extreme pathways can be characterized by their lengths . As an example of this application ,RBC metabolism and its metabolic physiology have been interpreted as an application of extreme pathway.Genetic Algorithm are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. It is an optimization method for searching optimums(global maximums or minimums).Through genetic algorithm it is possible to detect the global optimum of the extreme pathways which we have implemented.

Index Terms— Introduction on Metabolic Pathways and Methods, sources os data of Metabolic Pathways, Algorithm of EPA, Paths of EPA, Genetic Algorithm, Comparision of Different Extreme Pathways using Genetic Algorithm, Output of GAEPA

1 INTRODUCTION

System Biology[2] is a study of complex biological processes and it is a form of bioinformatics where it helps us theoretically to analyse cellular metabolism of various organisms. Cellular metabolism characterization is useful to understand the phenotypic capabilities of these organisms. It has been done quantitatively approaches for analyzing metabolic networks, including structural and stoichiometric analysis, Genetic Algorithm approach.

Structural and Stoichiometric Analysis is one of the Metabloic Pathway Analysis[1]. Metabolic pathways 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. Structural and Stoichiometric Analysis includes two types Network Based and Constraint Based Aprroaches. The metabolic pathway analysis that we have implemented is Extreme Pathway Analysis(EPA), it comes under Network Based Approach of Structural and Stoichiometric. 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.

Constraint Based Approach is one of the approaches under structural and stoichiometric modeling applies a set of constraints on a metabolic pathway to characterize its possible behaviours. There are many constraint based methods.

1. *Author: Gourav Mukherjee is currently pursuing bachelor degree program in information technology in Heritage Institute of Technology, India, PH-09830050527. E-mail: gouravmukherjee1993@gmail.com*
2. *Co-Author: Aditya Dwarkani is currently pursuing bachelor degree program in information technology in Heritage Institute of Technology, India, PH-08697721000. E-mail: aditya.dwarkani@gmail.com:*
3. *Co-Author: Mrs. Satarupa Bagchi Biswas is Assistant Professor in the department of Information Technology in Heritage Institute of Technolgy, Kolkata, India, PH-+919836699449. E-mail-satarupa.biswas@heritageit.edu*

The other approach under this structural and stoichiometric modeling is Network Based Approaches. In this modeling with the help of existing knowledge of cellular components and their connectivities systemic functions are described mathematically.

Under network based approaches there are few methods among which we have dealt with Extreme Pathway Analysis.

Extreme Pathway Analysis[2] are 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[3]. As a example of application, RBC metabolism and its metabolic physiology have been interpreted as an application of extreme pathway.

Genetic Algorithm are the adaptive heuristic search algorithm that are based on the evolutionary ideas of natural selection and genetics. It is an optimization method for searching optimums (global maximums or minimums).The next operation is the crossover.This is combining the genetic material of the randomly selected pair of subjects.Another operation is the mutation which is randomly changing the genetic material to individual subjects.This operation turns out to be very important ,because without it the genetic algorithm could get caught in local optimum. With the help of Genetic Algorithm it is possible to detect the global optimum of the extreme pathways which we implemented. They simulate the survival of the fittest among individuals over consecutive generation

for solving a problem. Each generation consists of a population of character strings that are analogous to the chromosome.

It itself contains all the basic evolutionary algorithm construction blocks. Therefore its study is also a basis for the study of other evolutionary algorithms.

2 Our Implementation

2.1 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. An extreme pathway matrix is considered 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.

2.2 Genetic Algorithm

Considering the parameters of the extreme pathways we have performed genetic algorithm[4] which includes performing mutation, crossover and is continued till optimized global otpimum value is obtained. First the extreme pathway parameters are considered then intial random solutions are created. After the creation of initial random solutions **Z**(fitness value).We obtain the (Z) fitness value of the solution by multiplying the initial solution and the random variable and compute the summation of it.We perform crossover and create new individual and we find the fitness value of the new individual and we compare it with the previous value of the individual.If it is greater than that we again perform crossover to produce new individual and find their fitness value and comparing it.If it is not greater than the fitness value of the previous solution we perform mutation and we compare the value if the value is not greater than the previous value then we again peform mutation.This process continues until we find that after mutation the value is not changing or the value is nearby more or less the same we consider this solution optimum solution

2.3 Comparison of Extreme Pathways using Genetic Algorithm

For every optimum solution we obtain from the extreme pathways we take that and compare those values and we find the greatest optimum solution and we consider that solution as the globally optimum solution.

3 Sources of Data of Metabolic Pathways We Have Worked On

3.1 Kyoto Encyclopedia of Genes And Genomes[5]

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[6]

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 Extreme Pathway Analysis

4.1 Prequisite

Parse the XML file.

4.2 XML Processing

Step 1: Obtain the data of the substrate,product and reac tions.
Step 2: Check for distinct products and substrates.
Step 3: Store all these values.

4.3 Stoichiometric Matrix Formation

Step 1: Take a 2d array stoic size **m*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.

4.4 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.

4.5 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).

4.6 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 individu al nodes.
Run one more loop from 0 to total number of prod-ucts.
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.

4.7 Genetic Algorithm Extreme Pathway Analysis

GAEPA (Genetic Algorithm Extreme Pathway Analysis)

Step 1: Consider the parameters of genetic algorithm ob tained from the extreme pathway analysis
Step 2: Create random variables.
Step 3: Create the random initial solution.
Step 4: Obtain the fitness value of the solution.
Step 5: Perform Crossover to create new individual and calculate the fitness value.
Step 6: Check the fitness value with the initial solution whether it is greater than the initial fitness value if yes perform mutation.
Step 7: Compare the fitness value obtained from the muta tion with the fitness value of the previous solution if it is not greater then goto step 5.
Step 8: Continue Step 5 to 8 until we find a solution of fit ness which is greatest or recurring for a certain number of generation and we consider the solution as the optimum solution.
Step 9: For each extreme pathway an optimum solution is produced and is stored in an array.
Step 10: Find the max element of the array which denotes the global optimum solution for the pathway.

5 Result

5.1 Output From Genetic Algorithm Extreme Pathway Analysis (GAEPA)

i. For Pathway hsa00520

Path Id 1: 93-2-92-4-94-6-95-8-96-10-97
Path Id 2: 93-63-96-10-97
EP0: optimum value 24.56472048286715
EP1: optimum value 23.08146824857662

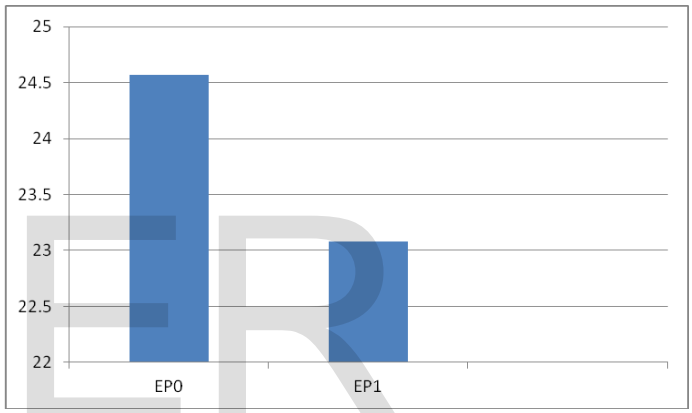


Fig 2: Graphically comparing the optimum values of the extreme pathway for pathway hsa00520

Global Optimum Value 24.56472048286715

ii. 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-

EP0: optimum value 9.73019849801107
EP1: optimum value 8.32697956217664
EP2: optimum value 9.45545754395765
Global Optimum Value : 9.73019849801107

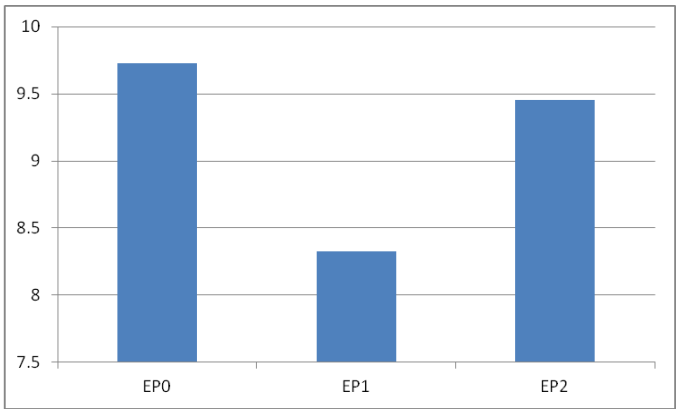
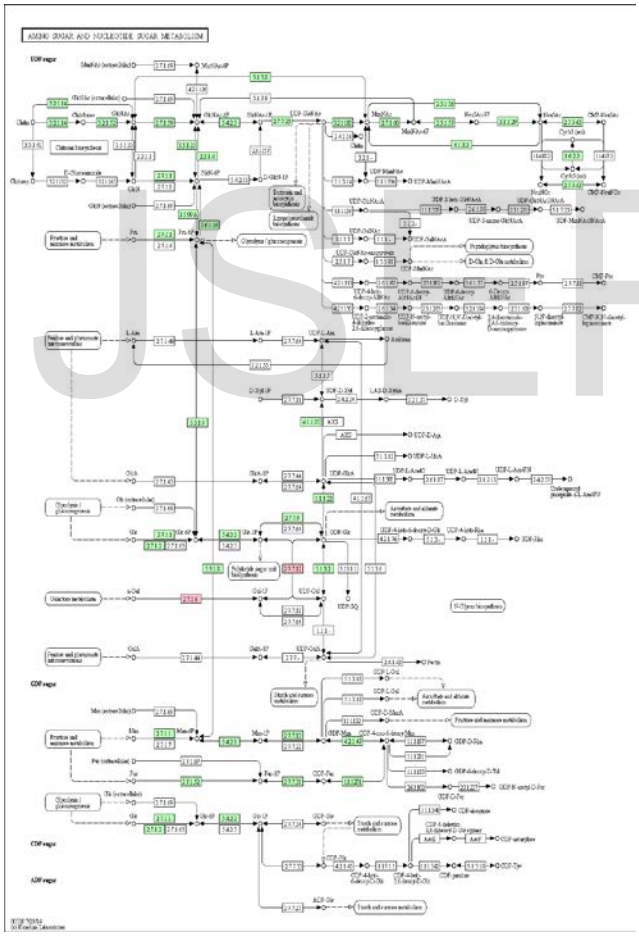
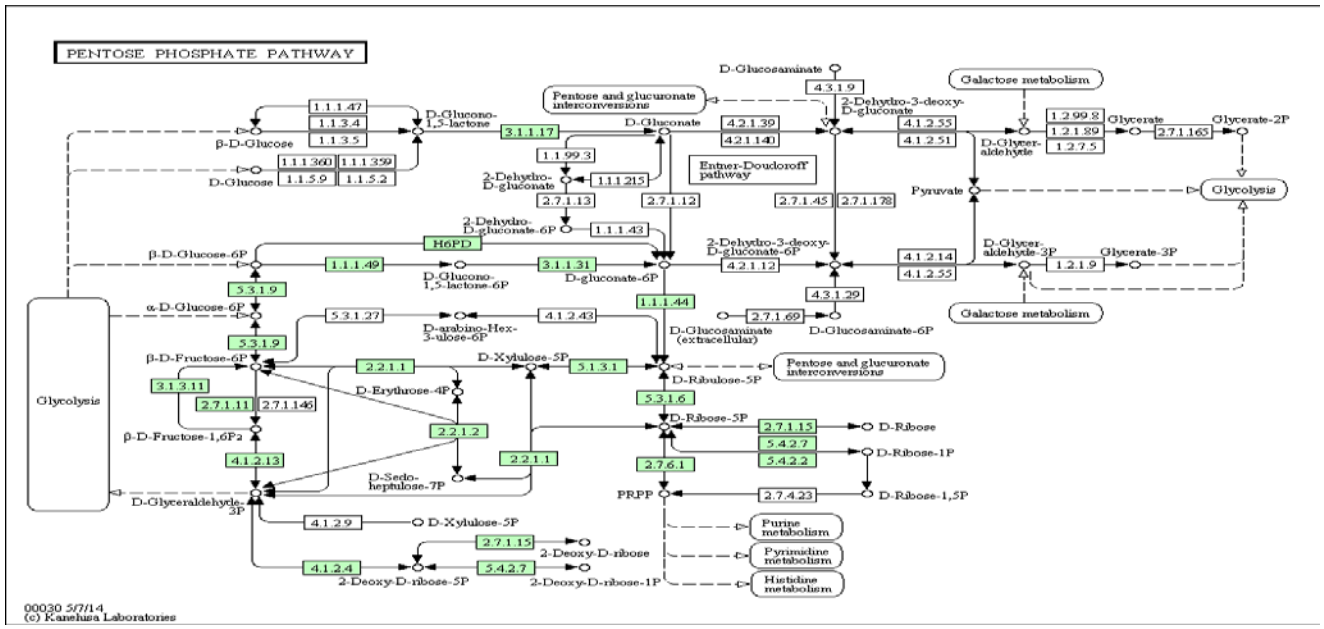


Fig1: Graphically comparing the optimum values of the extreme pathways for pathway hsa00030



Pathway hsa00520

.



Pathway of hsa00030

6 Conclusion

Therefore we can conclude that with the help of the implementation of these methods of different metabolic pathway analysis we can mathematically and programmatically analyze the flow of metabolites through a metabolic network.We also understood the method of genetic algorithm where we can obtain the global optimum value and the usefulness as well as the accuracy of this method can be used to maximize the objective function .We even analyzed the extreme pathways and how it can be used to obtain the pathways which can denote the proper desired rate of the metabolites in the particular reactions.

The optimized Z value can be used to compare the extreme pathways and which pathway produces the optimized result of objective function .Analysis of biochemical pathways is one of the key topics in the post genomic era.In order to understand the cellular mechanisms, to automatically retrieve metabolic information from the predicted metabolic pathways , we have to develop and implement useful methodologies.

There should be enough development of the techniques in the future which can be search and extract the information and biochemical reactions that can be incorporated into standard systems.

7 Acknowledgment

We are thankful to Mrs. Satarupa Bagchi 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] Gourav Mukherjee, Aditya Dwarkani, Tanmay Kumar Dawn, Assistant Professor Satarupa Bagchi Analogizing and Investigating Some Applications of Metabolic Pathway, International Journal of Scientific & Engineering Research, Volume 6, Issue 2, February-2015 273

[2] Namrata Tomar And Rajat K De Comparing Methods of metabolic Network Analysis And An Application To Metabolic Engineering

[3]Jason A.Papin ,Nathan D.Price ,Bernard Palsson Extreme Pathway Lengths and Reaction Participation in Genome Scale Metabolic Networks.

[4]Kiran Raosaheb Patil,Isabel Rocha,Jochen Forster and Jens Nielson Evolutionary programming as a platform for in silico metabolic engineering

[5]http://jjj.mib.ac.uk/models/

[6]http://www.genome.jp/kegg/