Development of A visualization tool for *Plasmodium falciparum* Metabolic Networks

Angela Makolo[1] **and Segun Fatumo**[2]

[1]Department of Computer Science, University of Ibadan.

[2]H3Africa Bioinformatics Network (H3ABioNet) Node, National Biotechnology Development Agency (NABDA), Federal Ministry of Science and Technology (FMST), Abuja, Nigeria.

Abstract

Plasmodium falciparum, the deadly malaria parasite, causes the most severe form of malaria, resulting in about 1.5-2.7 million deaths annually mostly in Africa. Metabolism is one of the most complex cellular processes. Connections between biochemical reactions via substrate and product metabolites create complex metabolic networks that may be analyzed using network theory which plays a major role discovery of new drug targets. This is especially important as the malaria parasite has become resistant to existing malaria drugs. Discoveries in biomedical research increases the possibility of treating the disease by effectively and specifically targeting essential enzymes of this parasite. In this work, we developed a simple but efficient tool for visualisation of metabolic pathways of *Plasmodium falciparum.* This web based software enables translation of text format descriptions of biological networks into standardized diagrams and readable to users. It allows users to view their biological network in a pathway or sub-pathway and how reactions of interest are linked. Our tool was implemented in Java and tested on Windows and Unix systems.

Keywords: Metabolic Pathway, visualization, bipartite graph, graphml.

1.0 Introduction

The complete genomes of many organisms are now available; this has paved the way for several types of analysis and experiments. The new approaches and questions that the genomes make possible are usually referred to as functional genomics. The task is to define the function of a gene or its protein in the life processes of the organism, where function refers to the role it plays in a larger context. The most obvious of life processes is the metabolism of an organism, the basic chemical system that generates essential components such as amino acids, sugars and lipids, and the energy required to synthesize them and to use them in creating proteins and cellular structures. This system of connected chemical reactions is a metabolic network. The work complements our previous works (Fatumo [4], Fatumo et al 2011) which investigated the metabolic network of Plasmodium falciparum and thus predicted potential novel drug targets. In this work a web based software was implemented using graph visualization tool implemented in Java programming language with Netbeans ID as the development environment.

Key processes in life depends on the capacity of individual cells to respond effectively to insights about their changing internal and external environments. Decision making about Cellular activities and responses are coordinated by complex molecular networks consisting of entities such as genes, proteins or RNAs connected by interactions such as activation or synthesis. Primary databases contain extensive information and also experimental literature relevant to these networks is rapidly growing that it is increasingly difficult to integrate or provide a platform for its visualization. A useful aid to theoretical and experimental research, aimed at presenting the inferences contained in the experimental literature and databases would be an enhanced visualization of biological pathways. The modelling of biochemical pathway as a bipartite graph makes it possible to see the reactions involved as a network structure with the reactants and the compound connected, thus providing information on some pertinent questions that arise during metabolic/biochemical pathway analysis. This includes knowledge about the most essential reaction to a particular organism, the alternative pathways between two compounds and the shortest path between them. Biochemical pathways such as metabolic, regulatory or signal transduction pathways can be viewed as interconnected process forming an

intricate network of functional and physical interactions between molecular species in the cell.

A metabolic network is a set of interconnected metabolic pathways that shows the pathway of reactions occurring in a particular organism. Metabolism is the chemical engine that drives the living process of organisms. With the aid of a vast repertoire of enzymatic reactions and transport processes, organisms can process and convert thousands of organic compounds into the various biomolecules necessary to support their existence. Consequently, the organism direct the distribution and processing of metabolites throughout its extensive map of pathways. Metabolism corresponds to the set of molecular transformations or energy transfers that occurs in the cell of living organism. [8]

The study of metabolic pathways is important in understanding the operations of life processes making it possible to manipulate and exploit the pathways the cell has at its disposal. This result in the ability to develop strategies to effectively eliminate or enhance certain pathways and introduce entirely novel routes for the production of various biochemicals of interest as indicated in the work of Fatumo [4]. An understanding of the structural design and capabilities of the cellular metabolic network clearly places scientist in an advantageous position to manipulate the cell for various purposes. In addition, metabolic pathway is necessary in understanding the biological mechanisms of organism and in discovering viable drug targets aimed at eradicating the survival of harmful parasites such as P.*falciparum*in humans. This was shown in the work of Makolo [6], which in addition to comparing the similarity checks of two popular motif discovery tools also identified genes in the metabolic pathways. Specifically, the work focused on identifying motifs in the glycolytic pathway of the P.*falciparum*, which is the core of the parasite's adaptability and consequently, its high resistance to the existing anti-malaria drugs.

BioCyc [5] http://www.biocyc.org, Version 10.5) is a collection of over 200 pathway/genome databases, containing whole databases dedicated to certain organisms. For example, EcoCyc which falls under the giant umbrella of BioCyc, is a highly detailed bioinformatics database on the genome and metabolic reconstruction of Escherichia Coli, including thorough descriptions of various signaling pathways. The EcoCyc database can serve as a paradigm and model for any reconstruction. Additionally, MetaCyc, an encyclopedia of metabolic pathways, contains a wealth of information on metabolic reactions derived from over 600 different organisms including *Plasmodium falciparum*. There are some other databases, for example PATIKA, WIT, BRITE, CSNDB and SPAD which describe the reactions in metabolic pathway for plasmodium, our choice database for the construction and visualisation of metabolic network is PlasmoCyc (BioCyc) for the fact that we found more metabolites, and reactions for *Plasmodium* than in elsewhere. However, our tool is able to fetch metabolic information from other databases and present its visualization tool. Metabolic pathways and other biochemical pathways such as signal transduction pathways, Protein-Protein interaction and regulatory pathway can be viewed as interconnected process, forming an intricate network of functional and physical interactions between molecular species in a cell. The amount of information available on such pathways for different organisms is increasing rapidly and this offers the possibility of performing various analyses on the structure of the full network of pathways for one organism as well as across different organisms.

Some of the data on metabolic networks includes resources for information on protein function such as SWISS-PROT and more sophisticated databases such as KEGG, EMP/WIT, EcoCYC, MetaCYC, and the enzyme resource BRENDA.

Evangelos [3] carried out an analysis on metabolic networks using a pathway distance metric through linear programming. In their work, they studied a linear programming based algorithm that calculates the minimum pathway distances in metabolic networks. Minimal pathways distances are identified as the smallest number of molecular steps separating two enzymes in metabolic pathways. They illustrated the applicability of the algorithm calculating the minimal pathway distances for *Escherica coli* small molecule metabolism enzymes, and then considering their correlations with genome distance (genome distance is the distance separating two genes on a chromosome) and enzyme function. The showed the effectiveness of the linear programming model with the result of their work.

Borqui [1] used constraint planar graph drawing algorithm for metabolic network visualization. The graph algorithm developed was applied to an entire metabolic network rather than the individual pathway that make up the network. They collaborated with biologists in their work and they introduce drawing constraints which takes into account the decomposition of the network into metabolic pathways as well as biochemical textbook drawing conventions. The drawing constraints introduced

raised numerous graph drawing problems which they solved by first recursively decomposing the network then applying suitable graph drawing algorithms. Their algorithm is useful for visualizing groups of reactions which span several metabolic pathways. Decomposition of the network into metabolic pathways was used as the backbone of their representation and this notion is central in biological studies on metabolic networks.

Rahman[8] in the paper titled "Observing local and global properties of metabolic pathways: load points and choke points in the metabolic networks" showed how Local and global aspects of metabolic networks analyses allows the identification of enzymes or reactions that are crucial for the survival of the organisms. This also gives an insight into the potential drug targets that can be discovered. In order to identify potential drug targets (based on the biochemical lethality of metabolic networks), the concept of choke points and load points was used to find enzymes (edges) which uniquely consume or produce a particular metabolite (nodes).

In our work, we used the data of the metabolic reaction database PlasmoCyc and developed a tool that provides a visualization, modelling the reactions as a bipartite graph of the metabolic network for P. falciparum

2.0 Materials and Methods

2.1 Constructing the metabolic network

In this work, we constructed the metabolic network of *P.falciparum* using metabolic reactions from the PlasmoCyc Database Version 10.5 [5], available at http://www.biocyc.org. By defining neighbours of reactions, we established a connected graph such that two reactions were neighbours if a metabolite existed that was the product of one reaction and the substrate for the other. This yielded a bipartite graph of alternating reaction and metabolic compound nodes.

2.2 The Tool

The Java visualization library used in this work is the Prefuse Java Visualization Library which takes a file in graphml format as input. The input file containing the reactions were thus processed first and converted to a graphml format. Graphml is an XML representation of a graph which contains tags for graph nodes, edges and it uses attributes to specify other graph attributes such as: edge weights, node color and edge direction. The graphml output obtained from processing the reaction file in turn serves as input to the algorithm that uses the Prefuse Java Visualization Library to create a graphical visualization of the data.

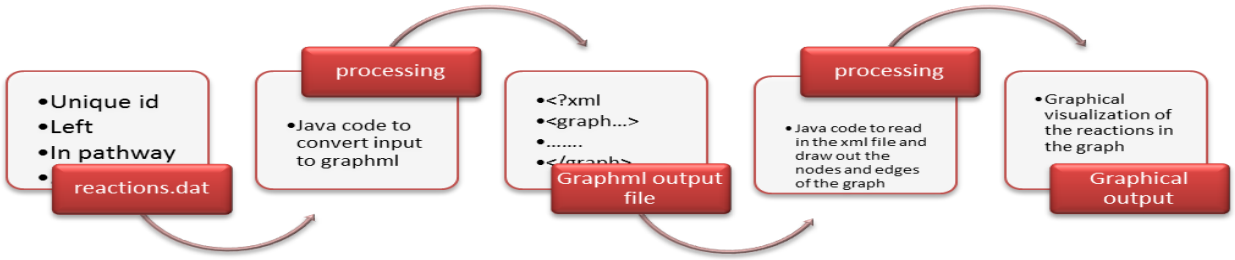

Fig 1. The framework of the Tool

2.3 System design

A formal model of the proposed system was developed using Unified Modeling Language (UML). The system consists of a form and a database of the reaction files. The form enables an administrator to select the file to visualize.

2.3.1 Use Case Diagram of the System

A Use Case diagram graphically depicts the interactions between the system, the external system and the user, defining who will use the system and the ways users are expected to interact with the system

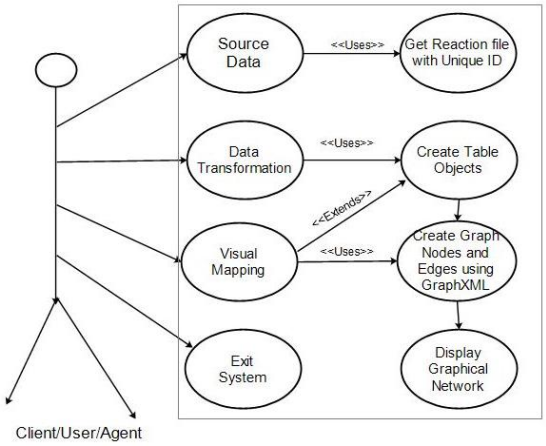

Fig 2: Use Case Diagram for the System

In this work we introduce four main Use Cases which extend, include and uses other Use Cases. They are the source data, data transformation, visual mapping and Exit System. The actor is the user that is using the application. The Source Data represents the interface where the users can feed data into the system by uploading a reactions file to be visualized. This file is transformed into a database and thereafter the visual mapping is achieved using Java GraphML library. With the exit system, the user of the system can decide when to leave the application.

2.3.2 Activity diagram of the system

The activity diagram of the proposed system is shown in fig 3. The diagram depicts the actions or sub activities as well as the transactions that are triggered by the completion of these actions or sub activities. With this we describe the system logical design using the workflow of the sub activities in the system.

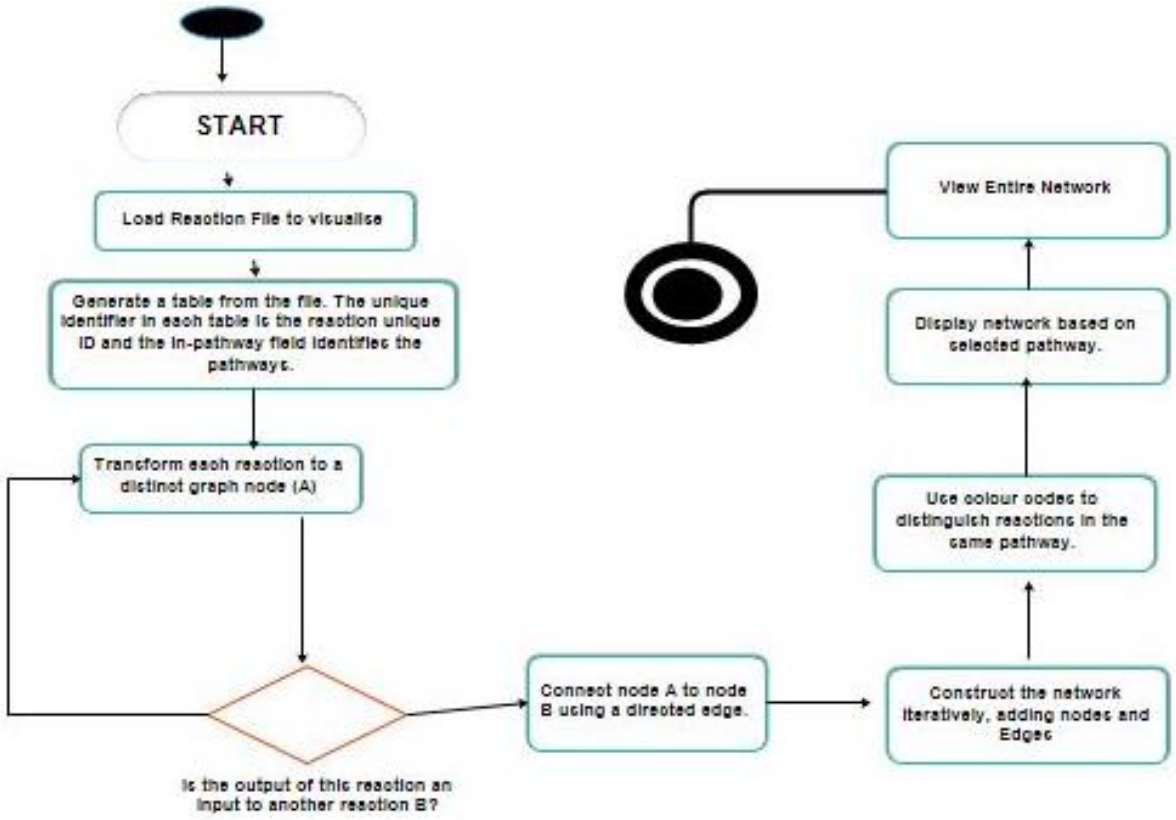

Fig 3: Activity Diagram for the System

3.0 Result

The result of our implementation is a visualized bipartite graph that shows the connection between the reactants and the products of the reaction.

When the user clicks on the button to view the entire graph, the result is displayed as shown in fig 5. A cluster can be seen in the middle and some other reactions floating on the side. The cluster in the middle represents the reactions that are essential to the organism. They represent the major reactions that are the core processes in the reactions data file. The other reactions floating on the left are those reactions that are not very important and can be knocked-off or discarded. They represent the reactions that can be traded off in other to make the entire network

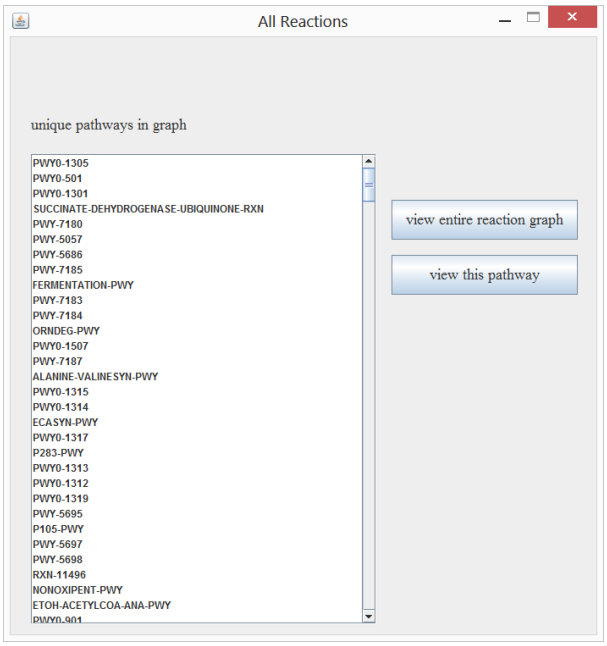

Fig 4 : The user interface for reaction selection

simpler. A color code is used to separate the reactants from the actual reaction and the product. For easy comprehension, the reactants which are the right of any reaction are represented with a pink outline, the reactions are represented with a green outline while the product which are to the left of a reaction are represented with a blue outline. The edges are directed edges that connect the reactants to the reactions and the reactions to the product. The diagram in fig 5 shows reactant connected to two different reactions at a time producing two different product.

In addition to viewing the entire network as shown in fig 6, users can pick a particular pathway of interest from the list and click on *the view this pathway button*, this allows individual pathway in the network to be visualized.

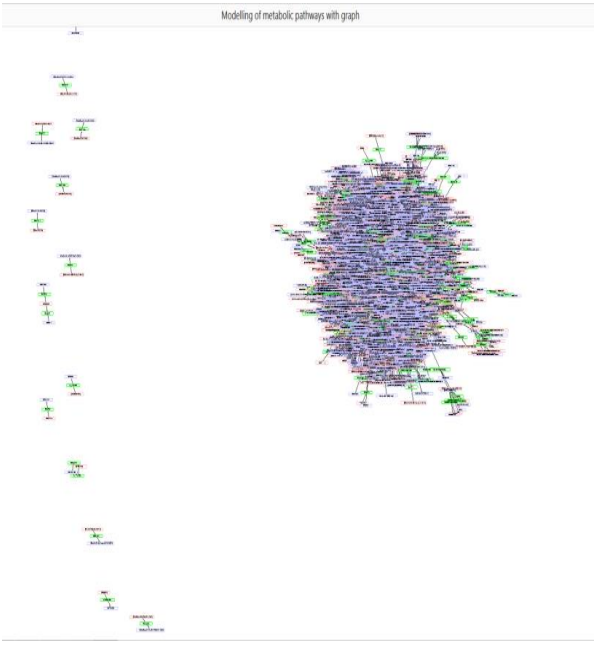

Fig 5: Clustered reactions with floating reactions

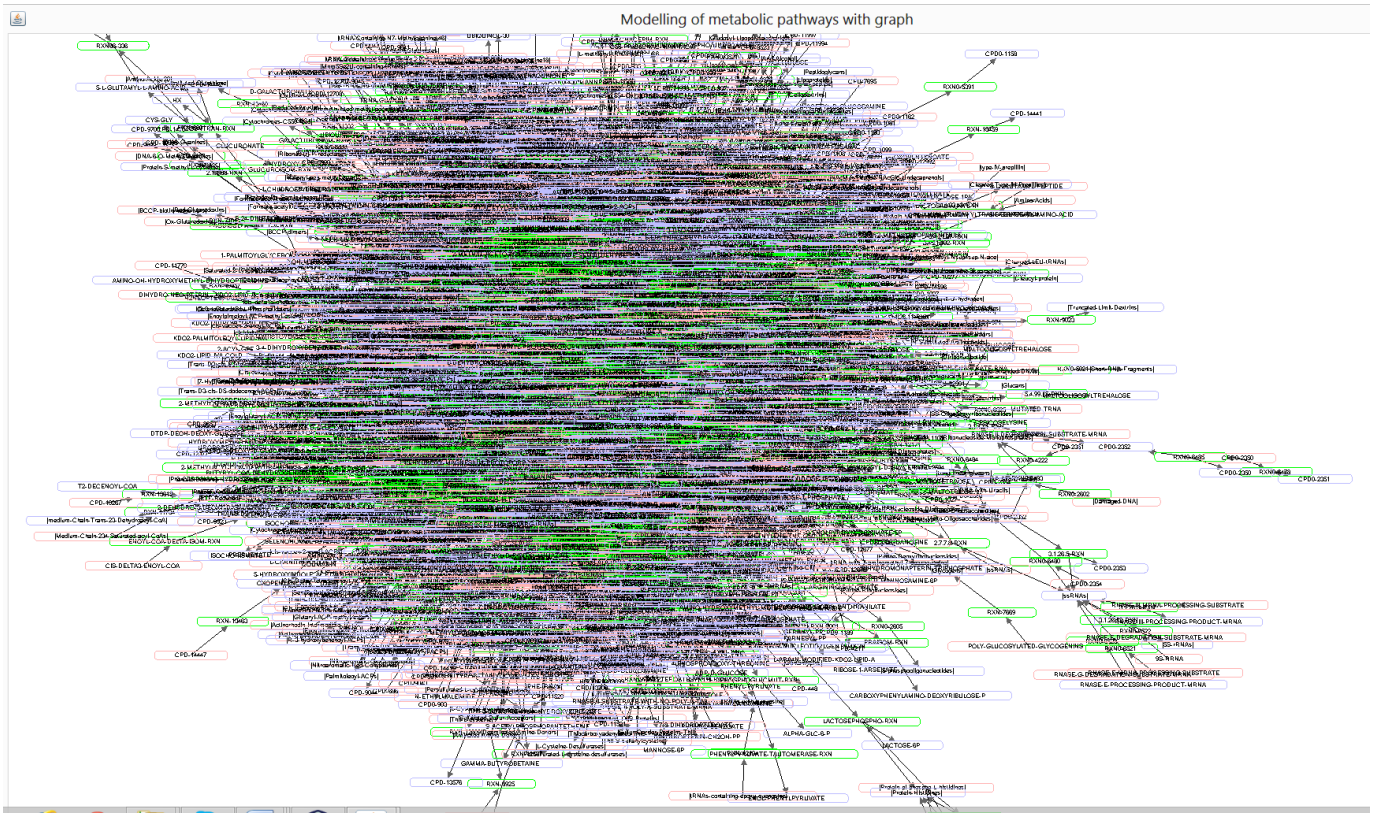

Fig 6 : The clustered reactions of the entire network

4.0 Discussion

Information about biological network structure can be used to derive far reaching conclusions about performance and robustness of metabolic pathways. The analysis of metabolic pathways provides the fundamental knowledge of how the cell meets its metabolic objectives either by destroying, creating, or enhancing the production capabilities of the organism for some bio-molecular substances.

In this paper, we present a tool developed to visualize and analyze biochemical pathways of any organism using the Prefuse Java Visualization Library. The entire pathway of the organism or just some selected pathways can be viewed. Using this algorithm, pertinent questions about an organism such as the number of reactions in a pathway, the core reactions etc can be answered. The result produced is a bipartite graph that shows the connection between the reactants and the compounds in the reactions.

The visualization provides useful analysis .This work is of immense value to biologists as it helps to visualize metabolic or a sub-pathway data of different organisms. The structural model of the visualization data provide answers to some metabolic pathway analysis questions such as: The possible paths from a compound to another, How many paths and how many within each path leads from compound A to compound b, how many unique pathways are there in the entire reaction, which reaction(s) is more important to the organism, shortest path between two compounds, load point and choke points in the network, the effect of knocking out a particular compound and the alternate route to generate a particular product.

The tool was evaluated by testing it with the metabolic pathway data of Plasmodium falciparum. The tool takes in the reactions data of a metabolic pathway file and convert the text in the file to nodes and edges of a graph. In addition to been able to view the entire reaction complex network, provision was made for the visualization of the unique pathways in the network.

References

[1] R. Bourqui, LaBRI UBI, D.Auber, V. Lacroix, F.Jourdan. Metabolic network visualization using constraint planar graph drawing algorithm.Information Visualisation (IV), 2006 10th International Conference page(s): 48 – 53.

[2] C.Didier, C.Fabian, J. Shoshana, V.H Jacques, "Metabolic PathFinding: inferring relevant pathways in biochemical networks" .Nucleic Acids Res. 2005, W326 W330.

[3] S. Evangelos , C.G Stuart, Janet MT, David L.B, Lazaros , "Analysis of metabolic networks using a pathway distance metric through linear programming", Metabolic Engineering, 2003 5(3):211-9.

[4] S. Fatumo, K. Plaisma, J.P Mallm, G. Schramm E. Adebiyi, M.Oswald, R. Eils, R.Koenig. "Estimating potential drug targets of plasmodium falciparum by analyzing the metabolic network of knock-out strains in silico".ScienceDirect Infection, Genetics and Evolution 2009, 9(3), 351-358.

[5] P.D Karp, C.A Ouzounis, C.A., Moore-Kochlacs, C. Goldovsky, L. Kaipa, P. Ahren, D., Tsoka, S. Darzentas, N. Kunin, V. Lopez-Bigas, "Expansion of the BioCyc collection of pathway/genome databases to 160 genomes" Nucleic Acids Res. 2005, 33, 6083–6089.

[6] A.U Makolo, E.F Adebiyi and A.O Osofisan. "Comparative Analysis of Similarity Check Mechanism for Motif Extraction" IEEE Afr J Comp & ICT. 2012, 5(1):34-41

[7] S.A Rahman, D. Schomburg. "Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks", Bioinformatics 2006, 15;22(14):1767-74.

[8] C.H Schilling, S.Schuster, B.O Palsson. "Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-genomic Era". Biotechnol.Bioeng. 1999, 52, 396-403.

[9] G.A Viswanathan, J. Seto, S. Patil, G. Nudelman, S.C Sealfon "Getting Started in Biological Pathway Construction and Analysis" PLoSComputBiol 2008, 4(2): e16.

[10] I.Yeh, T.Hanekamp, S.Tsoka, P.D Karp, R.B Altman, "Computational analysis of Plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery", Genome Res. 2010, 14, 917–924.

[11] S Fatumo, K.Plaimas, E. Adebiyi, R König Comparing metabolic network models based on genomic and automatically inferred enzyme information from Plasmodium and its human host to define drug targets *in silico*. Infect. Genet. Evol. 2011, *11*, 201-208

Angela Makolo is with Computer Science Department of the University of Ibadan and the University of Ibadan Bioinformatics Research Group (ui.bioinformatics.edu.ng). She has a PhD in Computer Science with Bioinformatics Option from the University of Ibadan.

Segun Fatumo is with H3Africa Bioinformatics Network (H3ABioNet) Node, National Biotechnology Development Agency (NABDA), Federal Ministry of Science and Technology (FMST), Abuja, Nigeria