IACSIT International Journal of Engineering and Technology, Vol.2, No.2, April 2010
ISSN: 1793-8236

In-silico evaluation of malaria drug targets

Chinwe Ekenna, Segun Fatumo and Ezekiel Adebiyi

***Abstract*—The most severe *form* of malaria, a disease that affects over 300 million people annually, is caused by the single-celled parasite *Plasmodium falciparum*. It is most prominent in Africa and has led to the death of millions of people. Studies have also shown that the low level of resistance to this disease in children has made them susceptible to malaria. Anti-malarial drugs have been developed to target specific sites in the pathway of *Plasmodium falciparum* but due to the level of resistance that the organism has developed, essential drug target sites have to be identified of which *Plasmodium falciparum* would have low or no resistance. In a recent publication, 22 potential drug targets based on an automated metabolic pathway database called PlasmoCyc were predicted. However, in a more recent publication, a critical evaluation of the comparison of a manual reconstruction database (Malaria Parasite Metabolic Pathways) against pathways generated automatically like PlasmoCyc, MetaSHARK and KEGG (Kyoto Encyclopedia for Genes and Genomes) was done. The study shows that the automatically generated pathways/databases need an expert manual verification. We employed extraction programming techniques to create an enhanced *PlasmoCyc* database and a comparison technique to identify and evaluate these drug targets in their pathways and then employed the homology modeling technique to model their structures.**

***Index Terms*—Drug targets, Malaria, Plasmodium falciparum, Protein, PlasmoCyc.**

I. Introduction

Malaria is one of the world's most common and severe tropical diseases. It is caused by a protozoan belonging to the genus *Plasmodium* and transmitted by the anopheles mosquito. Malaria continues to be a major cause of mortality and morbidity in tropical countries and affecting around 100 countries of the world. Malaria kills about three million people each year. More than 1 million of these are children under the age of 5. In addition to the millions who die, up to a half billion suffer the effects of malaria. Because mothers are more likely to suffer malarial relapses during pregnancy, malaria is an important cause of low-weight births and stillbirths. More than half of the miscarriages in endemic areas are caused by malaria. Ninety per cent of malaria cases occur in Africa, south of the Sahara. (Netmark, 2005).There are over 150 species of *Plasmodium* of which *Plasmodium falciparum* is considered the most dangerous of the species. It has the highest rates of complications and mortality, it accounts for about 80% of all human malarial infections and approximately 90% of the deaths (Keiser family foundation, 2007). It is more prevalent in sub-Saharan Africa than in other regions of the world. Forty percent of the world's populations live in endemic areas. Epidemics have devastated large populations and malaria poses a serious barrier to economic progress in many developing countries. Overall, malaria accounts for 10% of Africa's disease burden, and it is estimated that malaria costs the continent more than $12 billion annually (Keiser family foundation, 2007). Although Africa is hit hardest, it is estimated that more than one-third of clinical malaria cases occur in Asia and 3% occur in the Americas. The estimated cost to effectively control malaria in the 82 countries with the highest burden is about $3.2 billion annually (Keiser family foundation, 2007). The prevalence of resistance to known anti-malarial drugs has resulted in the expansion of anti-malarial drug discovery efforts. In a bid to further advance the work currently being done on malaria we evaluated the drug targets previously predicted (Fatumo *et al*., 2008). *et al*. We also present an evaluation of the PlasmoCyc repository and the creation of an enhanced result on the data stored and finally we modeled the structures for this verified drug targets using MODELLER(Andrej *et al*.,1993). In section 2, we introduced the methodology of this work, flowchart representing the extraction program, design and methods for creating the enhanced database, the criteria used to determine viable and existing drug targets among the already identified targets,. We discussed our result and findings in section 3. In the last section we discussed the persistent challenges, and other areas of similar research and advancement to the work.

Computer and Information Sciences Department, Covenant University, Ota, Nigeria.

II. Methodology

A. Biological Databases

There are a lot of databases to gain information about metabolic pathways, reactions, and their enzymes about organisms. One of such databases is BioCyc which is the collection of Pathway/Genome Databases that provides electronic reference sources on the pathways and genomes of different organisms including plasmodium. We used the *Plasmodium* collection of the Biocyc database called PlasmoCyc to retrieve information for *Plasmodium falciparum* and the Humancyc database to retrieve information for Homo sapiens. The relevance of retrieving information for Homo sapiens is because it serves as a host for the parasite and we need to glean similar information that exist both for the host (Homo sapiens) and the parasite (*Plasmodium falciparum*).

Another database used in this work is the Protein Data Bank (Berman et al.,2000) database where all the template structures used to model the structure of the protein was retrieved. Homology modeling works with already existing

structures to develop a new structure for the sequence based on the level of similarity that exist between the protein of known structure and protein of unknown structure. The protein databank serves as a repository for experimentally determined structures done (x-ray or NMR methods)

We also retrieved information about the Protein sequences from Genebank. This sequence serve as input into the BLAST software (Altschul *et al*.,1990). The BLAST result shows the level of similarity that exist between our protein that have unidentified structures and those that exist within the Protein data bank.

B. *Metabolic Network reconstruction*

The reconstruction of a metabolic network allows for an in-depth insight into the molecular mechanisms of a particular organism, especially when correlating the genome with molecular physiology (Francke *et al*., 2005). A reconstruction breaks down the metabolism into pathways and their respective reactions and enzymes, and enables analyzing them within the perspective of the entire network. Examples of various metabolic pathways include glycolysis, Krebs cycle, pentose phosphate pathway, etc. In simplified terms, a reconstruction involves collecting all of the relevant metabolic information of an organism and then compiling it in a way that makes sense for various types of analyses to be performed. The correlation between the genome and metabolism is made by searching gene databases, such as KEGG (Kanehisa *et al*., 2004). GeneDB (Benson *et al*., 2003). For particular genes by imputing enzyme or protein names. For example, a search can be conducted based on the protein name or the EC number (a standardized term that represents the catalytic function of the enzyme of interest) in order to find the associated gene (Francke *et al*., 2005).

For our reconstruction of the metabolic network for *Plasmodium falciparum*, we explored PlasmoCyc which provides adequate metabolic information for *Plasmodium falciparum*.

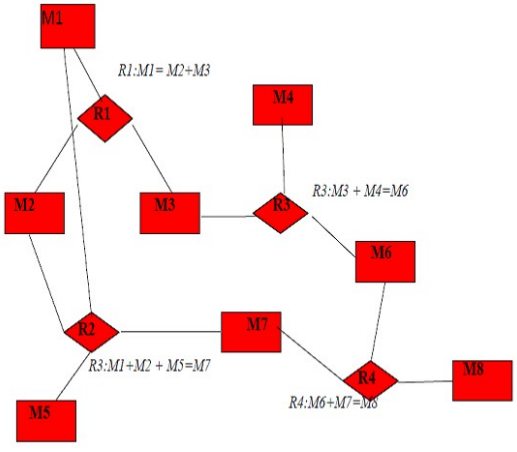

Figure 1.1 Bipartite gragh of the metabolic network

We found 811 reactions for *Plasmodium falciparum* in PlasmoCyc version 11.6. Unspecific compounds were discarded. Such compounds were for example ATP, NADH, and WATER. Each compound can be substrate and product of a certain reaction. The metabolites and reactions/enzymes are considered as alternating nodes. We selected the largest connected graph out of these. Finally we yielded network (*Figure 1.1*) with 691 reactions/enzymes. Each reaction is considered to be reversible.

C. *Verification of PlasmoCyc*

In a recent work,an analysis of the PlasmoCyc database was done and the result showed that there are reactions which have no gene in some of the automatically generated pathways/databases. Such reactions do not make meaning biologically. In an attempt to build an enhanced PlasmoCyc which will make better meaning biologically, we identified 184 pathways in PlasmoCyc which consist of over 800 hundreds of reactions, we extracted all reactions and classified then according to their corresponding pathways, reactions that have got no genes were discarded. We also discarded pathways that have got less than 3 reactions within it.

This analysis is done because a pathway should be a called a pathway if it has at least three reactions within its' network.Our enhanced PlasmoCyc produced result Table1 and finally, all reactions that do not fall into any pathway were also discarded. We present the flowchart representation for the verification and enhancement of the PlasmoCyc database.

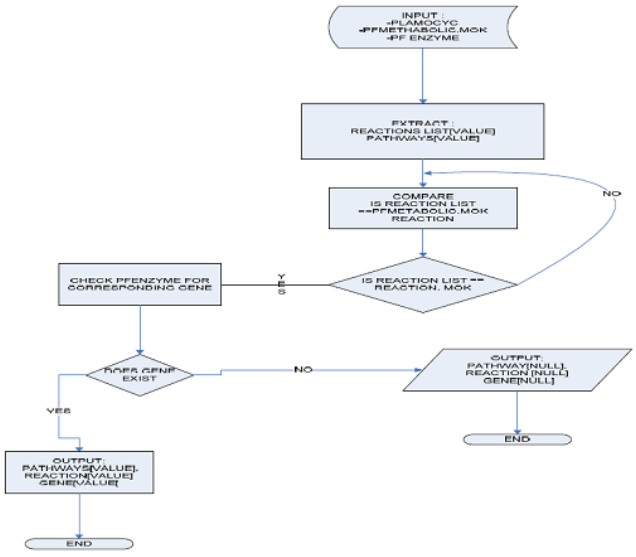

Figure 1.2 Flowchart representation for the Enhanced PlasmoCyc creation

A design representation of the method we used to achieve the enhanced PlasmoCyc database is as shown. The flowchart shows the implementation of the algorithm with the PlasmoCyc flatfile as input with special emphasis to the pathway.dat file and reactions.dat file. that if a protein or enzyme has no gene that encodes for it then according to the data produced by PlasmoCyc it cannot be said to exist and subsequently be a part of a reaction hence it would be discarded .this program runs through all the pathways within the PlasmoCyc and subsequently produces another flatfile that has pathways with proteins that have genes that encode for them. .

D. Protein Structure Prediction

We used Modeller for our protein structure predictions in this case we modeled the structure of the drug target that are also are enzymes that exist with the pathway of *Plasmodium flaciparum*. Modeller implements a technique inspired by nuclear magnetic resonance known as satisfaction of spatial restraints, by which a set of geometrical criteria are used to create a probability density function for the location of each atom in the protein. The method relies on an input sequence alignment between the target amino acid sequence to be modeled and a template protein whose structure has been solved. The program also incorporates limited functionality for *ab initio structure prediction* of loop regions of proteins, which are often highly variable even among homologous proteins and therefore difficult to predict by homology modeling

III. Results

A. Analysis of the extraction result

In the course of the extraction we noticed that some enzymatic reactions do not have genes that encode them such was retrieved and tabularized as shown in table 1.Also we found out that some enzymatic reactions do not have identifiable pathways, this was also path of our result.

TABLE 1. Tabular Representation of Analysis

Specie	Enzymatic Reaction with gene	Enzymatic Reaction without gene	Enzymatic Reaction without pathway	Total
Humans	440(33.5%)	256(19.50%)	617(46.99%)	1313
Plasmodium falciparum	305(37.62%)	386(47.59%)	120(14.17%)	811

B. Insilico verified Drug targets

Out of the 22 drug potential drug target identified by Fatumo *et al*., 2008. 16 of those drug targets were validated and their corresponding Protein structures modeled and created. Table 2 shows the result of this 16 drug targets, their encoding genes, their pathways and the compounds that have been identified as potential drugs active within that pathway where this drug targets reside.

TABLE 2. Tabular Representation of the 16 Drug Targets and Their Encoding Genes

ENZYME NAME	GENE ID	ENZYMATIC REACTION
Nicotinate phosphoribosyltransferase	MAL6P1.137	NICOTINATEPRIBOSYLTRANS-RXN
Hexokinase	MAL6P1.189	R81-RXN
Pyridoxal kinase	MAL6P1.266	PYRIDOXKIN-RXN
Hydroxyethylthiazole kinase	PFL1920C	THIAZOLSYN3-RXN
Phosphomethylpyrimidine kinase	PFE1030C	PYRIMSYN3-RXN
Thymidylate kinase	PFL2465C	DTMPKI-RXN
FMN adenylyltransferase	PF10_0147	FADSYN-RXN
Cardiolipin synthetase	MAL6P1.97	CARDIOLIPSYN-RXN
CDP-diacylglycerol–inositol 3-phosphatidyltransferase	MAL13P1.82	2.7.8.11-RXN
Hydroxyacylglutathione hydrolase	PFL0285W	GLYOXII-RXN
Nicotinamidase	PFC0910W	NICOTINAMID-RXN
Deoxyribose-phosphate aldolase	PF10_0210	DEOXYRIBOSE-P-ALD-RXN
Enoyl-CoA hydratase	PF10_0167	TIGLYLCOA-HYDROXY-RXN
3-Hydroxydecanoyl-[acyl-carrier protein] dehydratase	PF13_0128	3-HYDROXYDECANOYL-ACP-DEHYDR-RXN
Arginine-tRNA ligase	PFL0900C	ARGININE--TRNA-LIGASE-RXN
Long-chain-fatty-acid-	PF14_0761	R223-RXN

C. Homology Modeling Result

We Successfully built the structures for the 16 drug targets using Modeller and identified their Inhibitors.The pictorial representation is an example of structure modeled for the hexokinase enzyme.

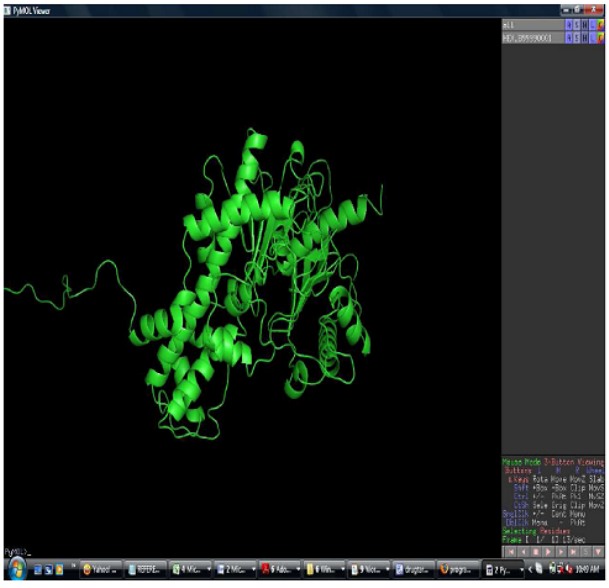

Figure 1.3 Modeled structure for the hexokinase drug target

The identified compounds in Figure1.3 are the chemical substances that have effectively been identified as inhibitors or receptors to that drug target. Showing also are the pathways where these drug targets reside and carry out their function.

IV. Discussion

Identification of drug targets from their individual pathways, modeling the Structures and identifying the enzymatic reactions where these drug targets (which are special enzymes) actually function is a task that is intensive and rigorous if done simply by biological experiments .We built an extraction program using the C programming language and using the *PlasmoCyc and Humancyc* flat file as input successfully extracted the enzymatic reactions, their genes, their enzymes and the pathways where these enzymes function. We further analyzed this preliminary result and retrieved the enzymes that had genes that encode them and also have pathways where they belong. We compared the result for Humans and *Plasmodium falciparum* and did a comparison where we validated our drug targets based on the enzymes occurred both in humans and *Plasmodium falciparum.* The structure of this drug target was modeled eventually using the homology modeling technique built on the modeler platform.

During the research on this work ideas for future work came and they include:

- Creating an interface where all the various modules used for this work can be merged under one platform and the user can easily enter a potential enzyme and search for all the necessary information needed.
- In terms of structure modeling, an efficient algorithm that uses the minimal energy and spatial restraints which during my research have come to find out are the two major yardsticks used to identify protein modeling points can be incorporated.Finding an effective drug to curb malaria coming from a computational platform involved a high level of accuracy and consistency has lead to various validation and reevaluation of result gotten. We had to identify the drug target through an extraction process in the human and *Plasmodium falciparum* database and then set out to characterize these drug targets into pathways, we also identified the genes that encode for each of this drug target and then identified their corresponding inhibitors.

V. Conclusion

Drug target prediction and analysis using computational approaches has helped tremendously in drug design. Also modelling of protein structures has brought us closer into understanding the inner structure of Protein and its' function. The level of accuracy of the results gotten using computational approach still has a long way to becoming hundred percent but gives an insight into results that would take biologist months to get result and in some cases years.

Acknowledgment

The authors thank colleagues of Covenant University whose input was most helpful.and also the management of Covenant University for their financial assistance.

Reference

[1] Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215:403-410.
[2] Andrej Sali and Tom Blundell J.Mol Biol (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints 234, 779-815.
[3] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2003),:GenBank. Nucleic Acids Res 31:23-27.
[4] Fatumo, S, Kitiporn Plaimas , Jan-Philipp Mallm , Gunnar Schramm ,Ezekiel Adebiyi, Marcus Oswald, Roland Eils , Rainer Ko¨nig., (2005). Estimating novel potential drug targets of Plasmodium falciparum by analysing the metabolic network of knock-out strains in silico, Infect. Genet. Evol. (2008), doi:10.1016/j.meegid.2008.01.007.
[5] Francke,C,siezen, R.J and Teusink,BReconstructing the metabolic network of the bacterium from it's genome. Trends in Microbiology 13(11):550-558.
[6] Karp, P.D., Ouzounis, C.A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P.,Ahren, D., Tsoka, S., Darzentas, N., Kunin, V., Lopez-Bigas, N., (2005).Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res. 33, 6083–.
[7] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M., (2004) The KEGG resource for deciphering the genome, Nucleic Acids Res. Jan 1;32(Database issue):D277-80.
[8] Keiser family foundation (2007) http://www.globalhealthreporting.org.
[9] Netmark"keyissues-malaria" http://www.rph.wa.gov.au/labs/haem/malaria/history.html.
[10] Plasmodium falciparum Wikipedia.Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B. (2000).The Protein Data Bank. Nucleic Acids Research, 28 pp. 235-242
[11] Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M., (2004) The KEGG resource for deciphering the genome, Nucleic Acids Res. Jan 1;32(Database issue):D277-80.