Journal of Microbiological Methods 109 (2015) 1–8

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

In silico comparative genomics analysis of *Plasmodium falciparum* for the identification of putative essential genes and therapeutic candidates

Subhashree Rout [a], David Charles Warhurst [b], Mrutyunjay Suar [a], Rajani Kanta Mahapatra [a,*]

[a] School of Biotechnology, KIIT University, Bhubaneswar 751024, Orissa, India
[b] Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, London, United Kingdom

ARTICLE INFO

Article history:
Received 3 November 2014
Received in revised form 27 November 2014
Accepted 27 November 2014
Available online 5 December 2014

Keywords:
P. falciparum
Comparative genomics
Drug targets
Molecular docking

ABSTRACT

A sequence of computational methods was used for predicting novel drug targets against drug resistant malaria parasite *Plasmodium falciparum*. Comparative genomics, orthologous protein analysis among same and other malaria parasites and protein–protein interaction study provide us new insights into determining the essential genes and novel therapeutic candidates. Among the predicted list of 21 essential proteins from unique pathways, 11 proteins were prioritized as anti-malarial drug targets. As a case study, we built homology models of two uncharacterized proteins using MODELLER v9.13 software from possible templates. Functional annotation of these proteins was done by the InterPro databases and from ProBiS server by comparison of predicted binding site residues. The model has been subjected to *in silico* docking study with screened potent lead compounds from the ZINC database by Dock Blaster software using AutoDock 4. Results from this study facilitate the selection of proteins and putative inhibitors for entry into drug design production pipelines.

1. Introduction

Malaria, the widespread tropical parasitic disease, needs new antimalarial drugs and vaccines urgently, particularly to prevent its deadly effects seen mostly in children and during pregnancy. In 2013, 97 countries had ongoing malaria transmission with an estimated 3.4 billion people currently at risk of malaria (World Health Organization, 2013). Five *Plasmodium* species, namely, *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and *Plasmodium knowlesi* cause malaria in humans (Arama and Blomberg, 2014) among which *P. falciparum* is responsible for most morbidity and mortality (Miller et al., 2013). Current malaria chemotherapies are subject to resistance, and now even the artemisinins are seen as possibly a fading hope (Andrews et al., 2014). There is a continuous need to search for additional drug targets for better protection and long term effectiveness.

This study employs computational approaches for finding suitable antimalarial drug targets through comparative metabolic pathway analysis of pathogen and host. Essentiality of proteins of interest which are non-homologous to the human host can be predicted if the protein is found in falciparum and other malaria parasite proteomes (Ludin et al., 2012) and has a high functional association with other proteins there through protein–protein interaction network(s) (Kushwaha and Shakya, 2010). Despite the global importance of *P. falciparum*, most components of the pathogen proteome have not yet been characterized experimentally. In the present study we have made an attempt to determine the structure and functions of some uncharacterized proteins computationally, as suitable drug targets.

2. Methods

A systematic workflow was defined that involved several bioinformatics tools, databases and drug target prioritization parameters (Fig. 1), with the goal of obtaining information about the drug targets in the *P. falciparum* genome but absent in its host, therefore avoiding any potential side effects.

2.1. Identification of metabolic pathways of pathogen and host

Metabolic pathway information of *P. falciparum* 3D7 and *Homo sapiens* was taken from the PlasmoDB (Yeh et al., 2004) and KEGG pathway databases (Kanehisa et al., 2010) respectively. Manual comparisons were made between pathogen and host pathways. Pathways which appeared in both host and pathogen were considered as common and those which did not were considered as unique in nature.

2.2. Identification of non-homologous proteins

The corresponding protein sequences from unique pathways of pathogen were taken from the Uniprot database (Boeckmann et al., 2003) with reference to Uniprot accession number from PlasmoDB. They were subjected to a BLASTP search (Altschul et al., 1997) against

* Corresponding author. Tel.: +91 9337627961.
E-mail address: rmahapatra@kiitbiotech.ac.in (R.K. Mahapatra).

http://dx.doi.org/10.1016/j.mimet.2014.11.016
0167-7012/

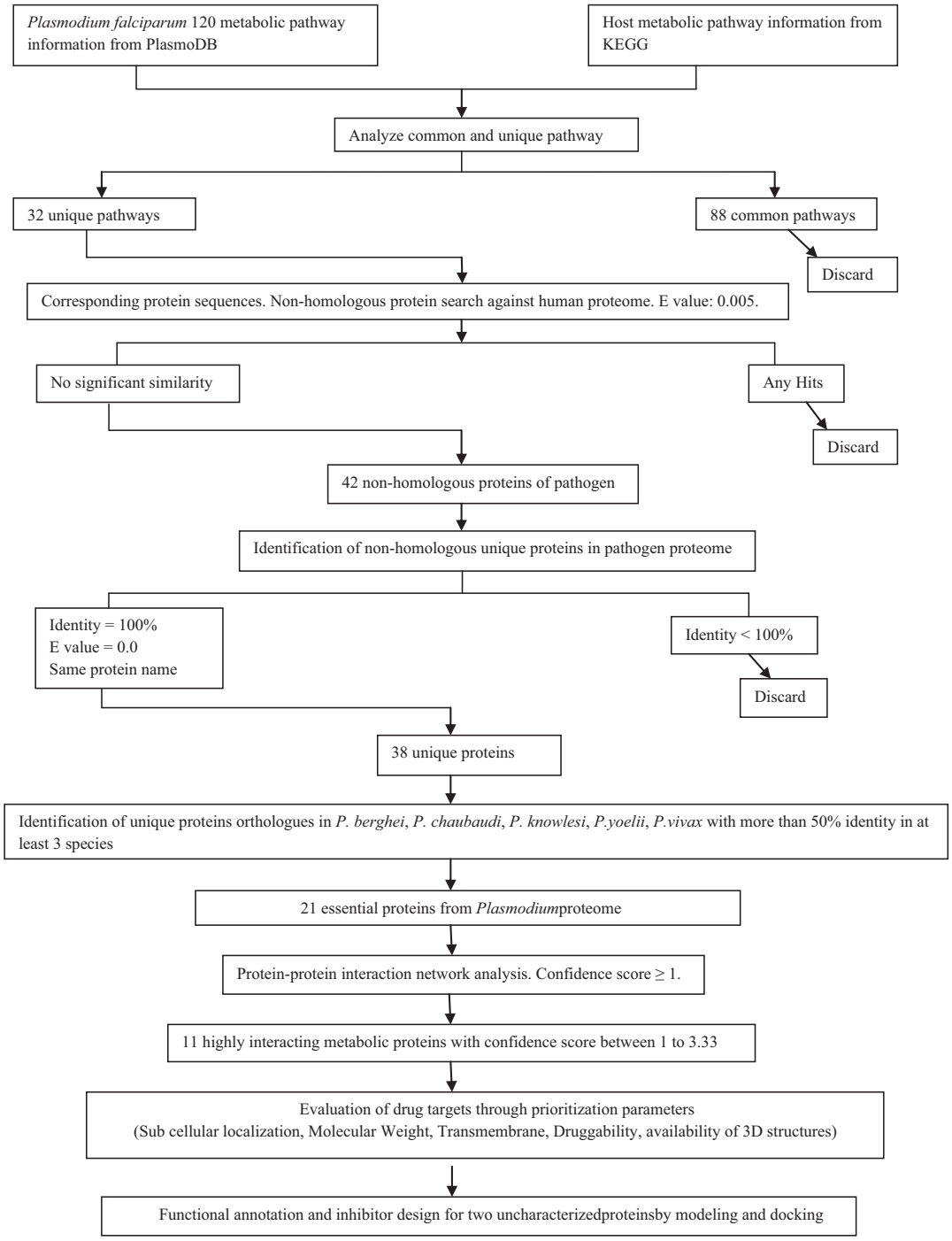

Fig. 1. Schematic representation of steps involved in drug target identification of *P. falciparum* through computational methods.

the human proteome with the e-value threshold set to 0.005 (Anishetty et al., 2005; Ghosh et al., 2014; Chawley et al., 2014). We adopted a rigorous way of determining no significant similarity for non-homologous proteins after applying filters.

2.3. Unique and conserved (essential) protein prediction in P. falciparum

Essentiality can be further predicted by identifying unique proteins in *P. falciparum* and other *Plasmodium* proteomes (Ludin et al., 2012). Non-homologous protein sequences from unique pathways were subjected to a manual BLASTP search for no other match in *P. falciparum* proteome with a default e-value threshold. These proteins are regarded as unique proteins in the *P. falciparum* genome. We used an OrthoMCL search (Chen et al., 2006) to identify unique proteins that are having the same orthologous group in other *Plasmodium* species. We considered the default e-value for putative orthologous based study. Additionally, a BLASTP search was employed for identification of conserved proteins against other *Plasmodium* proteomes namely *Plasmodium berghei*, *Plasmodium chabaudi*, *P. knowlesi*, *Plasmodium yoelii*, and *P. vivax* with an e-value threshold of $1e^{-5}$. We used a cut-off score of 50% homology in at least three *Plasmodium* species. This is a validation study searched by the OrthoMCL database for identification