BTCH90009

Assignment 4: Analysing Systems Biology Datasets

Name: Jiayu Wang

Student ID: 1039580

Word Count: 1497 (excluding tables and figures)

Introduction

Malaria is a major life-threatening disease in tropical and sub-tropical regions [1]. According to the World Health Organization report, about 229 million malaria cases occurred worldwide in 2019 and caused 409,000 deaths in total, which shows a significant increase than in 2018 [2]. This disease not only harms human health but holds back the local economy and aggravates poverty. Therefore, improved control and treatment approaches for malaria are urgently needed.

As a typical infectious disease, malaria is caused by parasites in *Plasmodium* genus [1]. Within this genus, *Plasmodium falciparum* is the species that accounts for the most malaria cases [3]. Hence, *Plasmodium falciparum* should be the first target organism to eliminate when trying to conquer malaria. Recently, the drug resistance of *Plasmodium falciparum* has become more severe, and hence new antimalarial drugs need to be developed [4].

This research aims to find 3 most potential drug target genes expressed in the apicoplast, a vital organelle to the survival of *Plasmodium falciparum*. Druggability of genes can be measured in several aspects: (1) Essentiality. Essentiality means the inhibition of the protein encoded by the target gene is lethal and hence the inhibitors would be effective to kill *Plasmodium falciparum*. (2) Expression. Target genes should be expressed at human blood stages (late stages in human-infecting stages of *Plasmodium falciparum* as circled in **Fig.1**.), hence the inhibitors can cure late malaria. (3)Specificity. Target genes must lack human orthologs so that the drug would not affect human health. (4) Feasibility. Designing inhibitors based on genes with solved or modeled protein structures is more feasible for inhibitor screening and designing. Also, target protein that has desirable pocket structures is of high affinity for molecular inhibitors and hence more druggable.

All the four aspects of druggability mentioned above were assessed by combining two bioinformatics tools, TDRTargets and *PlasmoDB* [5, 6]. In this way, this research identified three promising targets that can be helpful to conquer malaria and wishes to get supported by Gates Foundation Grants.

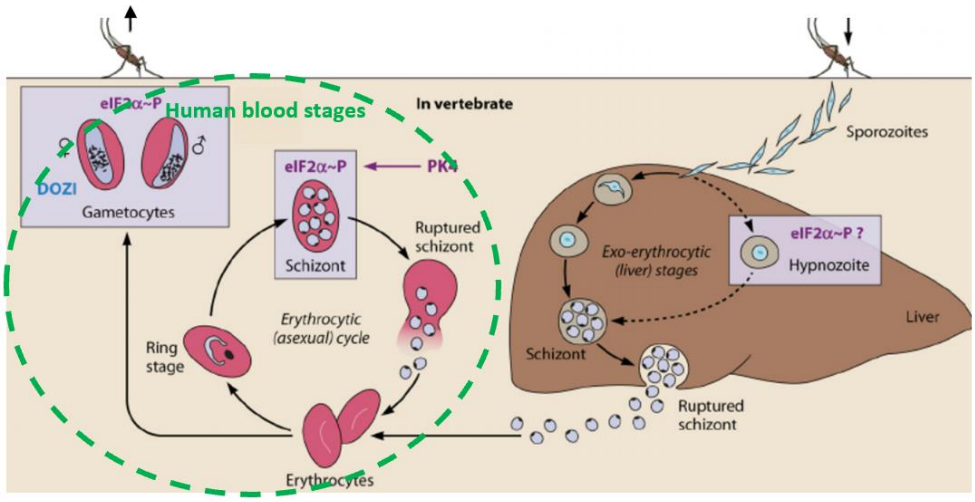

Fig.1. Schematic representation of human-infecting stages of *Plasmodium falciparum* [7]

Methods

TDRTargets

TDRTargets is a resource database that helps with the identification and prioritization of potential drug targets in different species [5]. In this research, I selected 'Targets' on TDRTargets, then added filter '*Plasmodium falciparum'* and 'druggability score >= 2' and got the 1,011 potential gene targets. In this way, all these 1,011 genes are potential targets with acceptable druggability scores in *Plasmodium falciparum.*

I also searched for other druggability evidence of candidate genes by browsing other content on TDRTarget including 'Orthologs', 'Essentiality Data', 'Phenotype & Validation', 'Druggability', 'Assayability'. Hence, I can learn further information about the candidate targets to make a comprehensive prioritization.

PlasmoDB

PlasmoDB is a *Plasmodium* genome database that contains the latest annotation on *Plasmodium falciparum* genome [6]. In this research, I used PlasmoDB to gradually eliminate the less druggable target candidates by adding several filtering steps on the 1,011 genes obtained from TDRTargets as mentioned above (the actual number is 1,002 because 9 genes are not included in PlasmoDB).

These filters include: (1) 'Protein targeting and localization - P.f. Subcellular localization - apicoplast'. This filter limits the target organelle to be apicoplast. (2) 'Transcriptomics –> Microarray Evidence -> Erythrocytic expression time series (3D7,DD2, &HB3)-'S'/'P'. This filter only kept the most expressed genes and the genes most similar to known targets during human-blood stages, hence drugs based on the targets can be effective for late-stage malaria. (3) 'Orthology and Synteny - Orthology Phylogenetic profile – exclude homos sapiens. This filter excludes genes having human orthologs so drugs based on the targets would not hurt humans. (4) 'Structure analysis - PDB 3D structures'. This filter eliminated the genes without solved protein structure, so the left gene targets are easier to be assayed. (5) 'Phenotype – Phenotype Evidence – piggyBec insertion mutagenesis – select mutagenesis index score '0 to 0.6'. This filter filtered out the less conserved genes, so inhibition of left genes is more likely to be lethal and the drugs are more likely to be effective.

Results

(i)Identifying genes with positive druggability scores

On the website of TDRTargets, 1,011 potential *P.falciparum* targets with druggability scores equal to or greater than 2 are identified.

(ii) Eliminating less druggable genes

On the website of PlasmoDB, several features are combined as filters to eliminate the less desirable drug targets as shown in **Fig.2.** 9 genes are left after filtration as shown in **Table.1**.

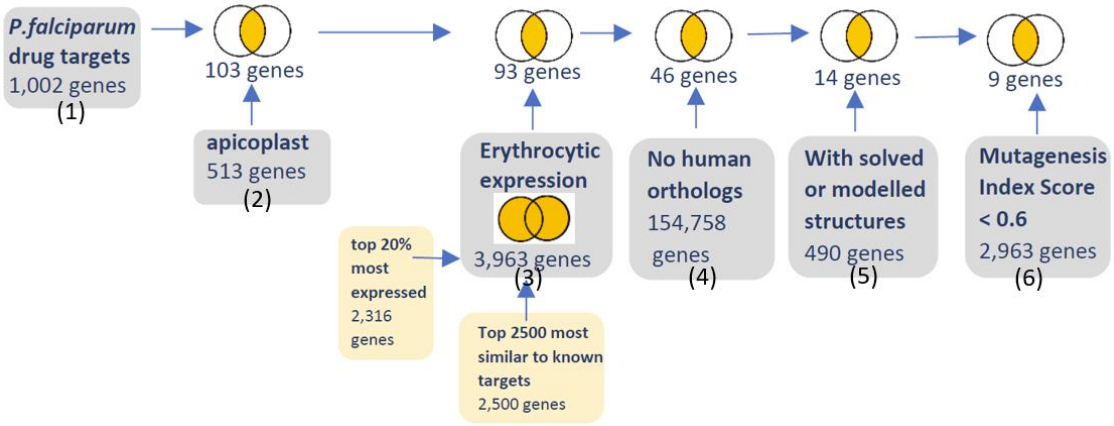

Fig.2. Schematic representation of combining intersection or union queries in PlasmoDB to eliminate the less desirable targets

Table.1. **ID and the protein product of the 9 sifted genes after filtration**

Gene ID	Protein product
PF3D7_0207500	serine repeat antigen 6
PF3D7_0209300	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
PF3D7_0615100	enoyl-acyl carrier reductase
PF3D7_0623200	ferredoxin--NADP reductase
PF3D7_0804400	methionine aminopeptidase 1c, putative
PF3D7_1323500	plasmepsin V
PF3D7_1344800	aspartate carbamoyltransferase
PF3D7_1436800	ATP-dependent Clp protease proteolytic subunit
PF3D7_1443900	heat shock protein 90, putative

These 9 genes have all the 6 desirable features including: (1) Existing in *P.falciparum.* (2) Within apicoplast. (3)Expressed during human-blood stages. (4) Without human orthologs. (5) With solved or modeled protein structures. (6) With mutagenesis index less than 0.6. Specifically, feature(3) is obtained from the union of the top 20% most expressed genes and top 2500 genes most similar to known targets, in human-blood stages.

(iii) Exploring further druggability evidence of the sifted 9 genes

The sifted 9 genes were furtherly explored in TDRtargets website as shown in **Table.2**. All the characters (shown as the columns in **Table.2**) are evidence of druggability, and hence genes with

more these characters are more druggable. We can see only plasmepsin V, heat shock protein 90, enoyl-acyl carrier reductase have more than 2 druggable characters, other genes are shown less significant regarding these druggable characters.

Table.2. Further druggability evidence of the sifted 9 genes

Gene ID	Protein product	Druggability				Essentiality		
		Drugg-ability index	Known inhibitors for this target	orthologs that are druggable targets	homologs that are druggable targets	Essential orthologs	Annotated phenotypes	Interact with other proteins
PF3D7_0207500	serine repeat antigen 6	Not exist	0	0	0	3	growth; lethal	Not exist
PF3D7_0209300	2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	Not exist	0	3	0	2	catalytic activity; decreased	Not exist
PF3D7_0615100	enoyl-acyl carrier reductase	0.8	100	10	0	3	growth; lethal	Not exist
PF3D7_0623200	ferredoxin--NADP reductase	0.3	0	0	0	2	Not exist	Not exist
PF3D7_0804400	methionine aminopeptidase 1c, putative	0.3	0	0	0	1	Not exist	Not exist
PF3D7_1323500	plasmepsin V	0.1	20 chemical compounds	1	2, with the highest aa sequence idendity 22.6%	2	Not exist	Not exist
PF3D7_1344800	aspartate carbamoyltransferase	0.3	0	0	0	8	Not exist	Not exist
PF3D7_1436800	ATP-dependent Clp protease proteolytic subunit	Not exist	0	0	2, with the highest aa sequence idendity 36.1%	0	Not exist	Not exist
PF3D7_1443900	heat shock protein 90, putative	0.3	874 chemical compounds	0	2, with the highest aa sequence idendity 37.6%	2	Not exist	Not exist

Discussion

By comprehensively considering different druggability characteristics of these genes in the 'Results' section, gene enoyl-acyl-carrier-protein reductase (*PfENR*), plasmepsin V (*PMV*), heat shock protein 90 (*Hsp90*) are the three most promising target genes.

Gene *PfENR* has the highest TDRtargets druggability index, the most known inhibitors, the most druggable homologs. Inhibition of enoyl-acyl carrier reductase is lethal. The limitation of it is lacking druggable orthologs so it might lack existing chemical drugs to use for reference. Gene *PMV* has 20 known inhibitors, druggable orthologs and homologs, and essential orthologs, but it lacks validated phenotype indicating the inhibition is lethal. Gene *Hsp90* has 874 known inhibitors, druggable homologs and essential orthologs, though it also lacks known druggable orthologs for reference. A

common limitation of these 3 genes is lacking evidence of interaction with other proteins in PlasmoDB. Besides, these 3 genes have published literature supporting them as targets, which is strong supporting evidence.

I furtherly explored the major metabolic pathways involved and the biological functions of these 3 genes. Protein PfENR is an enzyme within Type II fatty acid biosynthesis system (FAS II) [8]. Fatty acid biosynthesis is essential for the growth and survival of *P.falciparum* since fatty acids are important components of cell membranes. This suggests the essentiality of PfENR. Protein PMV is an enzyme recognizing and processing the Plasmodium-Export-Element (PEXEL)-containing proteins [9]. In this way, PMV enables these proteins can be exported out of *P.falciparum* to remodel the host erythrocytes and hence are critical for the intraerythrocytic growth of *P.falciparum.* Protein Hsp90 is an abundant chaperone [10]. It is responsible for the folding and functioning of many proteins, especially those involved in signal transduction and cell-cycle regulation. This suggests the essentiality of Hsp90.

I also explored the structural characters of the 3 selected targets. As shown in **Fig.3**, active sites of the 3 targets are pointed by arrows. The active site of protein PfENR is a narrow and deep cavity (**Fig.3a**). N-terminal ATP-binding site of Hsp90 is one of its most druggable active sites, shown as a beautiful deep pocket desirable for small-molecule inhibitors. The active site of PMV is also a long narrow cavity. These structural characters are all desirable for the affinity of inhibitors and hence are druggable.

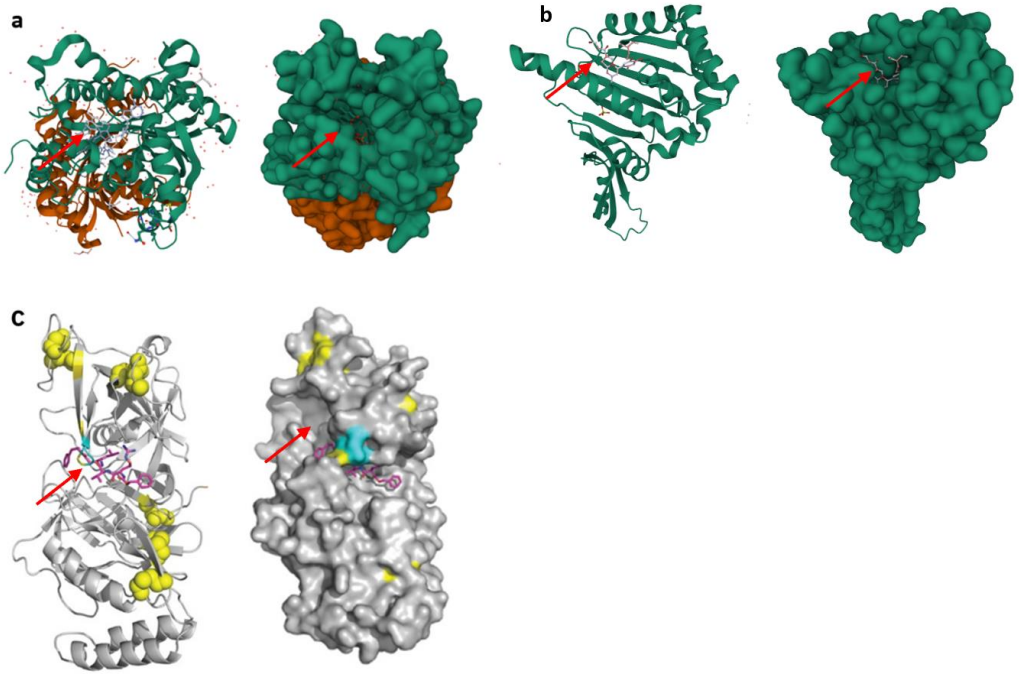

Fig.3. Schematic graphs of solved 3D-structure of the 3 selected targets and their active sites from PDB database [11-13]. Arrows point at the active sites of the proteins. **a**. Structure of protein *P.falciparum* PfENR. **b**. Structure of protein *P.falciparum* Hsp90. **c**. Structure of protein *P.falciparum* PMV.

By reading relevant literature, I found PMV and Hsp90 have validated inhibitors proved effective both *in vitro* and *in* vivo (mouse model), for example, WEHI-916 (targeting PMV), geldanamycin (targeting Hsp90)[11, 14]. PfENR also has published inhibitors that have been proved effective *in*

silico and *in vitro,* for example, compounds from the rhodanine class [15]. However, there is no *in vivo* evidence supporting these PfENR inhibitors.

This research combines TDRTargets and PlasmoDB to enable a rapid and simple-operated identification and prioritization of the drug targets, by adding multiple filters to eliminate the target candidates lacking essential druggable characters. This saved time to a great extent, but might have filtered out some potential desirable targets due to the incomprehensive or late updates of the two databases. I might have also referred more to the UniProt and PDB databases to explore more for structural characters and protein interaction.

As for future drug identification based on these 3 targets, I plan to first search for existing inhibitors for reference. Then after studying the structure of both the target proteins and the reference inhibitors, I will identify or design more potential inhibitors with high structural similarity to the reference inhibitors, with consideration regarding covalent interaction and lock-and-key model [16]. Before *in vitro* and *in vivo* validation, *in silico* virtual drug screening can be performed using molecular docking tools like MolDock [17].

Conclusion

Enoyl-acyl-carrier-protein reductase (PfENR), plasmepsin V (PMV), heat shock protein 90 (Hsp90) are the three most promising targets within the apicoplast of *P.falciparum*. According to this discovery, further inhibitor design and identification will be performed if this research can be supported by Gates Foundation grants.

Reference

1. Alegana VA, Okiro EA, Snow RW: **Routine data for malaria morbidity estimation in Africa: challenges and prospects.** *BMC medicine* 2020, **18:**1-13.
2. Fornace KM, Diaz AV, Lines J, Drakeley CJ: **Achieving global malaria eradication in changing landscapes.** *Malaria journal* 2021, **20:**1-14.
3. Ipa M, Widawati M, Laksono AD, Kusrini I, Dhewantara PW: **Variation of preventive practices and its association with malaria infection in eastern Indonesia: Findings from community-based survey.** *PloS one* 2020, **15:**e0232909.
4. Ghosh M, Olaniyi S, Obabiyi OS: **Mathematical analysis of reinfection and relapse in malaria dynamics.** *Applied Mathematics and Computation* 2020, **373:**125044.
5. Magariños MP, Carmona SJ, Crowther GJ, Ralph SA, Roos DS, Shanmugam D, Van Voorhis WC, Agüero F: **TDR Targets: a chemogenomics resource for neglected diseases.** *Nucleic Acids Res* 2012, **40:**D1118-1127.
6. Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, Ginsburg H, Gupta D, Kissinger JC, Labo P, et al: **PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data.** *Nucleic Acids Res* 2003, **31:**212-215.
7. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL: **A proteomic view of the Plasmodium falciparum life cycle.** *Nature* 2002, **419:**520-526.
8. de Medeiros PSdM, Ducati RG, Basso LA, Santos DS, da Silva LHP: **Enzyme Mechanism and Slow-Onset Inhibition of Plasmodium falciparum Enoyl-Acyl Carrier Protein Reductase by an Inorganic Complex.** *Enzyme research* 2011, **2011:**642758-642758.
9. Russo I, Goldberg DE: **Chapter 18 - Plasmepsin V.** In *Handbook of Proteolytic Enzymes (Third Edition).* Edited by Rawlings ND, Salvesen G: Academic Press; 2013: 103-105
10. Kumar R, Musiyenko A, Barik S: **The heat shock protein 90 of Plasmodium falciparum and antimalarial activity of its inhibitor, geldanamycin.** *Malaria Journal* 2003, **2:**30.
11. Sleebs BE, Lopaticki S, Marapana DS, O'Neill MT, Rajasekaran P, Gazdik M, Günther S, Whitehead LW, Lowes KN, Barfod L, et al: **Inhibition of Plasmepsin V activity demonstrates its essential role in protein export, PfEMP1 display, and survival of malaria parasites.** *PLoS biology* 2014, **12:**e1001897-e1001897.
12. Corbett KD, Berger JM: **Structure of the ATP-binding domain of Plasmodium falciparum Hsp90.** *Proteins: Structure, Function, and Bioinformatics* 2010, **78:**2738-2744.
13. Perozzo R, Kuo M, Sidhu AbS, Valiyaveettil JT, Bittman R, Jacobs WR, Jr., Fidock DA, Sacchettini JC: **Structural Elucidation of the Specificity of the Antibacterial Agent Triclosan for Malarial Enoyl Acyl Carrier Protein Reductase &#x2217.** *Journal of Biological Chemistry* 2002, **277:**13106-13114.
14. Roe SM, Prodromou C, O'Brien R, Ladbury JE, Piper PW, Pearl LH: **Structural Basis for Inhibition of the Hsp90 Molecular Chaperone by the Antitumor Antibiotics Radicicol and Geldanamycin.** *Journal of Medicinal Chemistry* 1999, **42:**260-266.
15. Kumar G, Banerjee T, Kapoor N, Surolia N, Surolia A: **SAR and pharmacophore models for the rhodanine inhibitors of Plasmodium falciparum enoyl-acyl carrier protein reductase.** *IUBMB Life* 2010, **62:**204-213.
16. Jorgensen WL: **Rusting of the lock and key model for protein-ligand binding.** *Science* 1991, **254:**954-956.
17. Thomsen R, Christensen MH: **MolDock: a new technique for high-accuracy molecular docking.** *Journal of medicinal chemistry* 2006, **49:**3315-3321.