Original Research Article

IN SILICO ANTIMALARIAL TARGET SELECTION CONSERVED IN FOUR PLASMODIUM SPECIES

ABSTRACT

The need for new antimalarials drugs and drug targets is pertinent due to the emergence of drug resistant strains of the parasites. Improper target selection has resulted in therapeutic failure. The genomic/post genomic era has made possible the deciphering of the 3D crystal structures of proteins and DNA which are drug targets and are deposited in the protein data bank. Novel antimalarial targets obtained from evolutionary conserved short sequence motifs are utilised and are essential in transcription processes in the parasite. The motifs TGCATGCA, GTGCAC and GTGCGTGC were curated from experimental work, validated and analysed via phylogenomics genomics and comparative genomics. PlasmoDB blastn was applied to determine their similarity in Plasmodium vivax, knowlesi, Ovale and yoeli. The complete genome of Plasmodium falciparum vivax, knowlesi, Ovale and yoeli was downloaded from the plasmoDB and their positions determined. The targets are essential, conserved in rodent and mammalian species via phylogenomics with percentage identity and similarity greater than 80%, have no similar genes in the same genome and also found to be selective in the parasites vis-à-vis the Homo sapiens via comparative genomics with 0% identity and similarity in the human genome. The targets reveal at the molecular and biochemical level, the vulnerable regions in the parasite while safe in human hence their choices in subsequent rationale drug discovery and design protocols.

Key

INTRODUCTION

Human malaria with estimated 515 million cases of is generally caused by four species of Plasmodium namely P. falciparum, P.vivax, P. ovale and P.malariae and are transmitted by the bite of the female anopheles mosquito (Snow, Guerra et al. 2005). Other mammalian Plasmodium species implicated include: P. knowlesi and P. yoelii. P. falciparum has the greatest toll on human health most especially in children under five years of age. The severity of malaria apart from the characteristic fever and anaemia include neurological involvement which may result in coma and subsequently to death. In the case of cerebral malaria, there could be damage to the hippocampus evident by impaired learning ability and cognitive function. The lifelong potentials of the child are therefore reduced(Boivin, Bangirana et al. 2007). Novel drugs have been launched into the drug discovery pipelines by the Medicines for Malaria Venture MMV and currently MMV pathogen box (Van Voorhis, Adams et al. 2016). Notwithstanding the propensity of the malaria parasites to become drug resistant still poses a major problem which may threaten the efficacy of promising new antimalarial chemotypes such as the synthetic peroxide OZ439 (Charman, Arbe-Barnes et al. 2011) and the spiroindolone NITD 609(Rottmann, McNamara et al. 2010) over time and this scenario will persist until the human-pathogenic Plasmodium species are eventually eradicated (Müller and Hyde 2010).

Despite recent advances in the “omics” field Lack of sufficient understanding of biologic networks, incomplete understanding of the biology and dynamics of drug-target interaction, lack of efficient and accurate ways to determine clinical relevance and the need to integrate massive amounts of sequence and phenotype data from the public and private sector. Focus is placed on the generation of an inventory of potential targets for therapeutics based on human and pathogen genome sequence studies to replenish the depleted pipeline. Target identification and selection is a lengthy and complex process that is now expedited and improved by bioinformatic (Shaikh, Ahmed et al. 2004).

A persistent bottleneck for target-based approaches is the identification of a suitable drug target in the first place(Hubbard 2011). This enzyme should be essential for survival of the parasite and sufficiently different from its closest counterpart in the human host to be inhibited selectively. Experimental tools to validate candidate drug targets are limited for the malaria parasites(Greenwood, Fidock et al. 2008). Several bioinformatics approaches have previously been employed to help identify or prioritise drug targets for Plasmodium parasites. These include methods based on automated identification of important steps in metabolic pathways(Plata, Hsiao et al. 2010), methods that combine chemical starting points and protein-based queries(Joubert, Harrison et al. 2009) , as well as the use of the TDRtargets web-resource (Magariños, Carmona et al. 2011) (http://www.tdrtargets.org) to prioritise drug targets through the combination of multiple data types relevant to drug development (Guiguemde, Shelat et al. 2010). A drug target must satisfy the following conditions to be selected (i) It must have conserved orthologues in all of the mammalian-pathogenic Plasmodium. (ii) It must have no other match in the Plasmodium falciparum. (iii) It must not occur or have a good match in all the human genome (chromosomes 1-23 and X,Y). (iv)Functions in the transcription process resulting in formation of gametocytes. The basis behind this principle is that conserved genes usually perform essential functions and in this case the motifs are essential in the regulation of gene(Doyle, Gasser et al. 2010). These criteria were applied in the selection of candidate drug targets that satisfy the conditions stated above.

Experimental Motif detection

Evolutionary conserved short sequence motifs in a set of promoters were curated from experimental work. The regulatory motifs were validated from the database. The motifs as follows:
TGCATGCA
GTGCAC
GTGCGTG

Datasets

The entire genome sequence of the Plasmodium falciparum was downloaded via the file transfer protocol ftp from the database of the National Centre for Biotechnology Information (NCBI) (ftp) site (ftp//.ncbi/nlm.nihgov/blast/executables/)..The download was done in the fasta format. The PlasmoDB http:// www.plasmodb.org/common/download(Aurrecoechea et al .,2009) and Kyoto ecyclopedia of genes of genes and genomes http:// www.genome.jp /kegg/ (Aurrecoechea). The frequency of occurrence of the raw sequence was determined in the mammalian and rodent Plasmodium species. Their upstream and downstream sequences were determined after the sequences of the Plasmodium falciprum hence the final sequences were not less than 16base pairs long (Roche, Thomson et al. 1994). Blastn was applied to determine the percentage similarity and identity of the selected sequences in the other plasmodium species selected.

The entire Human genome was downloaded via the file transfer protocol ftp from the database of the National Centre for Biotechnology Information (NCBI) ftp site. Homo_Sapiens.NCBI34.pep.fa.downloaded from ftpsite (ftp.// ftp.ensembl.org/pub/current_human/data/fasta/pep/). The download was done in the fasta format. Comparative genomics was applied to determine the percentage identity and similarity of the final 16 base pair sequences in the human genome. The sequences of interest above were compared with all known sequences in a database to identify homologous sequences. The genomic sequences for Plasmodium falciparum and their functional annotations were obtained from plasma DB (version 4.1) software sequence similarity between P. falciparum sequence and human sequence was run on a standalone computer. The default parameters were selected, and the p-value was calculated to validate sequence similarity to a human chromosome across the 23 chromosome including the X and Y

chromosome. The statistical significance of the similarity search in the human genome was obtained from the BLASTn results.

Results

Table 1.0 Summary of the various experimetally derived DNA Sequence Motiffs (Rs) to Target Sequence oligonucleotides (Ts1) via computational analysis

Raw Sequence in sPlasmodium genome	Oligomers obtained from Motifs TS	Organism in which motif is CONSERVED	Base Pair no (TS)	Biological function
TGCATGCA (RS1)	CTGA**TGCATGCA**TTTA (TS1)	All	16	Sequence Found in Sporozoites
GTGCAC(RS2)	ACCT**GTGCATGCA**GGGAA	P. vivax	18	Sequence formation of Merozoite
	(TS3)	All	17	Sequence found in merozoites
	ACCAGG**GTGCAC**AAGCA	All	18	Dephosphorylation
GTGCGTG(RS4)	(TS2b)	*P.falciparum*	16	Not a good target
	CATTTC**GTGCAC**CTACAT (TS2)			
	TGTGT**GTGCGTG**CGGA (TS4)			

Table 2.0 Experimetally derived DNA Sequence Motiff (Rs) TGCATGCA to Target Sequence 16base pair oligonucleotide (Ts1) via computational analysis

Query sequence	Number of base pairs	Frequency
TGCATGCA	8	26
ATGCATGCAT	10	8
GATGCATGCATT	12	1
TGATGCATGCATTT	14	1
CTGATGCATGCATTTA	16	1

Table 3.0 Comparative blastn results for 16bp oligonucleotide CTGATGCATGCATTTA to establish identity and homology of sequence in human plasmodium species(*P.falciparum, vivax, malariae,berghei, knowlesi and yoeli*)

Organism	Similarity
Pf3D7 14v3	100%
PfIT 14 v 3	100%
P.yoeli	100%

P.vivax	100%
P.berghei	100%
P.malariae	100%

Table 4.0 Experimetally derived DNA Sequence Motiff (Rs2) GTGCAC to Target Sequence 16base pair oligonucleotide ATTTCGTGCACCTACA (Ts2) via computational analysis

DNA SEQUENCE (5-3), (3-5)	No Of Base Pair	Frequency in PG
GTGCAC	6	89
CGTGCACC	8	1
TCGTGCACCT	10	1
TTCGTGCACCTA	12	1
TTTCGTGCACCTAC	14	1
ATTTCGTGCACCTACA	16	1

Table 4.1 Comparative blastn results for 16bp oligonucleotide ATTTCGTGCACCTACA to establish identity and homology of sequence in human plasmodium species(*P.falciparum, vivax, malariae,berghei, knowlesi and yoeli*)

Organism	Similarity
Pf3D7 14v3	100%
PfIT 14 v 3	100%
P.yoeli	94%
P.vivax	100%
P.berghei	100%
P.malariae	100%

Table 5.0 Experimetally derived DNA Sequence Motiff (Rs4) GTGCGTGC to Target Sequence 16base pair oligonucleotide GTGTGTGCGTGCGGAT (Ts4) via computational analysis

DNA sequence (5-3)(3-5)	No of Base Pair	Frequency in pG
GTGCGTGC	8	2
TGTGCGTGCG	10	1
TGTGTGCGTGCGG	12	1
TGTGTGCGTGCGGA	14	1
TGTGT**GTGCGTG**CGGA	16	1

Table 5.1 Comparative blastn results for 16bp oligonucleotide GTGTGTGCGTGCGGAT to establish identity and homology of sequence in human plasmodium species(*P.falciparum, vivax, malariae,berghei, knowlesi and yoeli*)

Organism	Similarity
Pf3D7 14v3	0%
PfIT 14 v 3	0%
P.yoeli	0%
P.vivax	0%
P.berghei	0%
P.malariae	0%

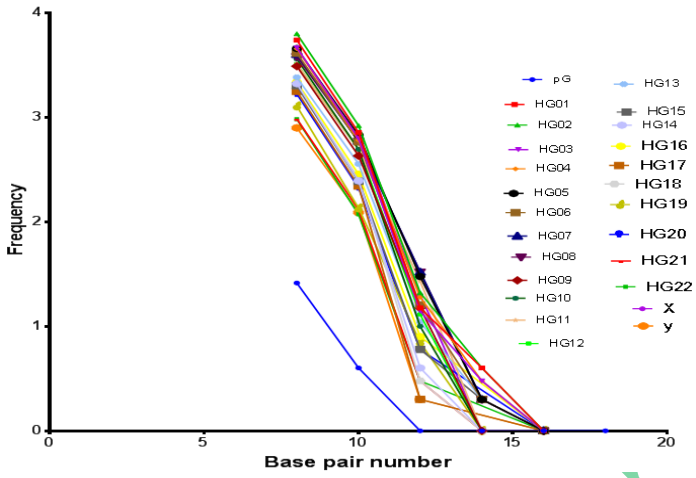

Figure 1: Graph of the frequency of Occurrence of CTGA**TGCATGCA**TTTA (TS1)against base pair number in the *Plasmodium falciparum genome* (pG) and Human Genome HG1-23.

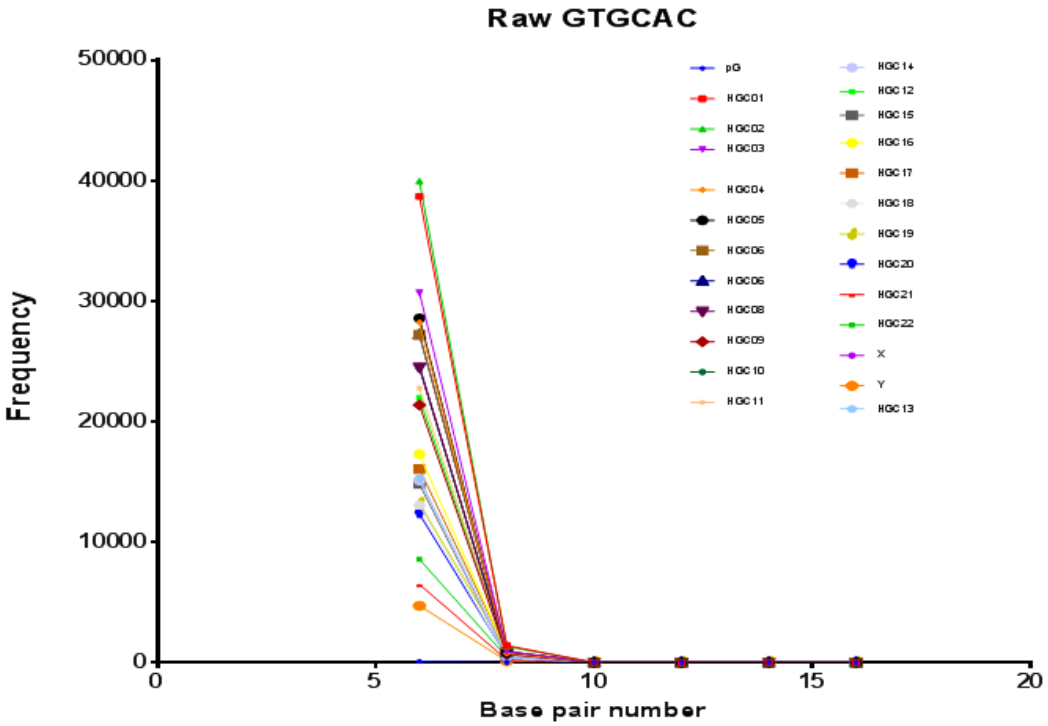

Figure 2: Graph of the frequency of Occurrence of ATTTC**GTGCAC**CTACA (TS2) vs base pair number in the Plasmodium *falciparum genome* (pG) and Human Genome

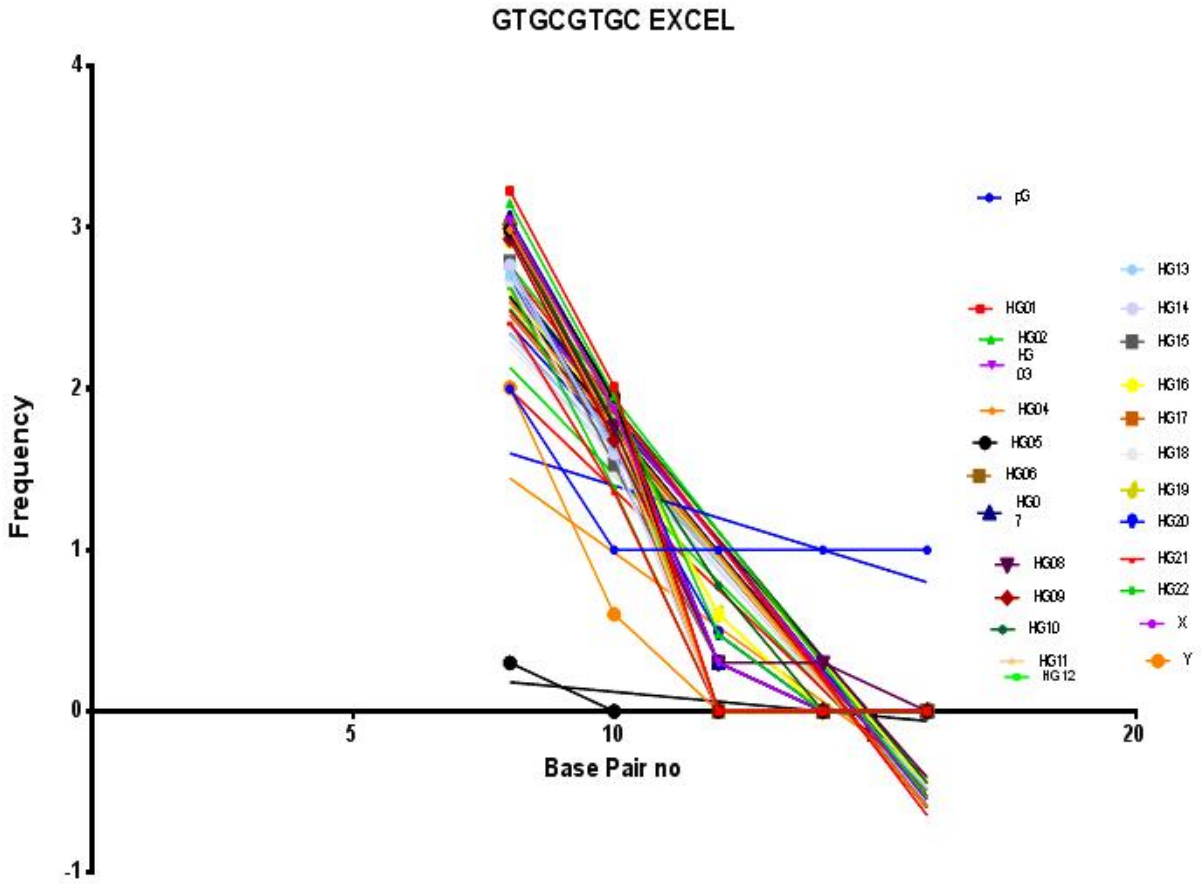

Figure 3: Graph of the frequency of Occurrence of TGTGT**GTGCGTG**CGGA (TS4) against base pair number in the Plasmodium *falciparum genome* (pG) and Human Genome HG1-23.

Discussion

The process of searching and obtaining clearly defined targets was done via extensive literature search in which the motifs were finally selected, based on their various roles in the biochemical and metabolic processes within the Plasmodium parasites (Carlton, Adams et al. 2008). The motifs are sequences covered by transcription factors and are conserved between evolutionary related species(Toucan ref). The 16 base pairs was selected as this length will represent one and a half turn of DNA so as to achieve specificity and selectivity in the Plasmodium parasite (Roche, Thomson et al. 1994). The evolutionary conserved short sequence motifs SSM TGCATGCA, GTGCAC and GTGCGTGC were experimentally determined and validated. Computational analyses carried out on these SSM via location in the downloaded Plasmodium falciparum genome gave the 16 base pair oligonucleotide sequences as follows CTGATGCATGCATTTA, ATTTCGTGCACCTACA and GTGTGTGCGTGCGGAT (Roche, Thomson et al. 1994) . The 16 base pair oligonucleotides were selected based on the criteria that a conserved single copy gene that lacks close matches in the same genome is most likely to be indispensable(Doyle, Gasser et al. 2010) . They were selected as targets based on their specific roles in the regulation of transcription process involved in the process of sporozoite formation as in the TGCATGCA motif. The second motif GTGCACACAC was essential and regulatory for the process of merozoite invasion into the host cell while GCACGCGTGC is involved in the regulation of dephosphorylation process. These motifs were found to be conserved in the four species of Plasmodium namely Plasmodium falciparum, P.vivax, P.knowlesi and P.yoelii yoelii. ACCTGTGCATGCAGGGAA (TS3) derived from the GTGCAC motif and occurs in P.vivax and identified in literature (Carlton, Adams

et al. 2008) is also a DNA target that is explored in this study. The raw sequence motifs were found through computational analysis to be selective. Their frequency of occurrence in the Plasmodium falciparum genome was computed and then simultaneously the target sequences were obtained through the location in the various chromosomes.The frequency of TGCATGCA is 26 (Table 2.0) while GTGCAC occurs 89 times in the Plasmodium falciparum (Table 4.0) and GTGCGTGC occurs only two times (Table 5.0) in the Plasmodium falciparum. These frequencies of occurrence are as opposed to the 16 base pair oligonucleotides derived from these motifs which are CTGATGCATGCATTTA, ATTTCGTGCACCTACA and GTGTGTGCGTGCGGAT and occur only with a frequency of one in the Plasmodium falciparum (Table 2.0, Table 4.0 and Table 5.0) hence fulfilling the criteria of essentiality in the Plasmodium falciparum and the fact that it is a single copy gene. Further bioinformatics analysis also allowed the frequency of occurrence in the Human genome to be calculated in chromosomes number 1-22 and the X and Y chromosomes. The graphical representations are as presented in Figures 1, 2 and 3 respectively for the 16 base pair oligonucleotides CTGATGCATGCATTTA (TS1), ATTTCGTGCACCTACA(TS2), and GTGTGTGCGTGCGGAT (TS4). The plot of the Frequency of occurrence of the raw motifs against the base pair numbers revealed in the case of TS1 and TS2 a clear distinction in the occurrence of the Plasmodium falciparum pG as distinct from the human chromosomes HG 1-22 and the X and Y chromosomes. These oligonucleotides are therefore selective as drug targets in the Plasmodium species in which they are conserved and drugs developed against them will not be toxic in the human.The formation of sporozoites via the sequence TS1 CTGATGCATGCATTTA. Formation of merozoites, however, involves ACCTGTGCATGCAGGCAA (TS2) and ACCAGGGTGCACAAGCA (TS3). CATTTCGTGCACCTACAT (TS4) regulates the dephosphosrylation process in the plasmodium parasite. CATTTCGTGCACCTACAT (TS4) The sequences were located in their respective chromosomes in the genome and the 16 base pair were determined via a program to obtain sequences in the 5’-3’position from the motif as well as the 3’-5’ position from the motif (Tables 1-3). The blastn revealed a similarity and identity of not less than 80% in the other species CTGATGCATGCATTTA has 100% similarity in P.vivax, malariae, ovale, knowlesi and 94% in p.yoeli.

References

Boivin, M. J., et al. (2007). "Cognitive impairment after cerebral malaria in children: a prospective study." Pediatrics 119(2): e360-e366.

Carlton, J. M., et al. (2008). "Comparative genomics of the neglected human malaria parasite Plasmodium vivax." Nature 455(7214): 757-763.

Charman, S. A., et al. (2011). "Synthetic ozonide drug candidate OZ439 offers new hope for a single-dose cure of uncomplicated malaria." Proceedings of the National Academy of Sciences 108(11): 4400-4405.

Doyle, M. A., et al. (2010). "Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes." BMC Genomics 11(1): 222.

Greenwood, B. M., et al. (2008). "Malaria: progress, perils, and prospects for eradication." The Journal of clinical investigation 118(4): 1266.

Guiguemde, W. A., et al. (2010). "Chemical genetics of Plasmodium falciparum." Nature 465(7296): 311-315.

Hubbard, R. E. (2011). "Structure-based drug discovery and protein targets in the CNS." Neuropharmacology 60(1): 7-23.

Joubert, F., et al. (2009). "Discovery: an interactive resource for the rational selection and comparison of putative drug target proteins in malaria." Malaria Journal 8(1): 178.

Magariños, M. P., et al. (2011). "TDR Targets: a chemogenomics resource for neglected diseases." Nucleic Acids Research 40(D1): D1118-D1127.

Müller, I. B. and J. E. Hyde (2010). "Antimalarial drugs: modes of action and mechanisms of parasite resistance." Future Microbiology 5(12): 1857-1873.

Plata, G., et al. (2010). "Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network." Molecular Systems Biology 6(1): 408.

Roche, C. J., et al. (1994). "Site selectivity of daunomycin." Biochemistry 33(4): 926-935.

Rottmann, M., et al. (2010). "Spiroindolones, a potent compound class for the treatment of malaria." Science 329(5996): 1175-1180.

Shaikh, S. A., et al. (2004). "A molecular thermodynamic view of DNA–drug interactions: a case study of 25 minor-groove binders." Archives of Biochemistry and Biophysics 429(1): 81-99.

Snow, R. W., et al. (2005). "The global distribution of clinical episodes of Plasmodium falciparum malaria." Nature 434(7030): 214-217.

Van Voorhis, W. C., et al. (2016). "Open source drug discovery with the malaria box compound collection for neglected diseases and beyond." PLoS Pathogens 12(7): e1005763.