Infection, Genetics and Evolution 11 (2011) 755–764

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Computational identification of signalling pathways in *Plasmodium falciparum*

Jelili Oyelade [a], Itunu Ewejobi [a], Benedikt Brors [b], Roland Eils [b,c], Ezekiel Adebiyi [a,*]

[a] Department of Computer and Information Sciences, Covenant University, Km 10 Idiroko Road, PMB 1023, Ota, Ogun State, Nigeria
[b] Division 'Theoretical Bioinformatics', German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
[c] Institute of Pharmacy and Molecular Biotechnology, and BIOQUANT Center, Im Neuenheimer Feld 267, Ruprecht-Karls-Universität Heidelberg, 69120 Heidelberg, Germany

ARTICLE INFO

Article history:
Received 10 May 2010
Received in revised form 17 November 2010
Accepted 18 November 2010
Available online 26 November 2010

Keywords:
Minimum paths
Signalling transduction pathways
Chloroquine resistance
Efflux process
Drug targets
Vaccine design
Malaria

ABSTRACT

Malaria is one of the world's most common and serious diseases causing death of about 3 million people each year. Its most severe occurrence is caused by the protozoan *Plasmodium falciparum*. Reports have shown that the resistance of the parasite to existing drugs is increasing. Therefore, there is a huge and urgent need to discover and validate new drug or vaccine targets to enable the development of new treatments for malaria. The ability to discover these drug or vaccine targets can only be enhanced from our deep understanding of the detailed biology of the parasite, for example how cells function and how proteins organize into modules such as metabolic, regulatory and signal transduction pathways. It has been noted that the knowledge of signalling transduction pathways in *Plasmodium* is fundamental to aid the design of new strategies against malaria. This work uses a linear-time algorithm for finding paths in a network under modified biologically motivated constraints. We predicted several important signalling transduction pathways in *Plasmodium falciparum*. We have predicted a viable signalling pathway characterized in terms of the genes responsible that may be the PfPKB pathway recently elucidated in *Plasmodium falciparum*. We obtained from the FIKK family, a signal transduction pathway that ends up on a chloroquine resistance marker protein, which indicates that interference with FIKK proteins might reverse *Plasmodium falciparum* from resistant to sensitive phenotype. We also proposed a hypothesis that showed the FIKK proteins in this pathway as enabling the resistance parasite to have a mechanism for releasing chloroquine (via an efflux process). Furthermore, we also predicted a signalling pathway that may have been responsible for signalling the start of the invasion process of Red Blood Cell (RBC) by the merozoites. It has been noted that the understanding of this pathway will give insight into the parasite virulence and will facilitate rational vaccine design against merozoites invasion. And we have a host of other predicted pathways, some of which have been used in this work to predict the functionality of some proteins.

1. Introduction

The most fatal and prevalent form of malaria is caused by the blood borne pathogen *Plasmodium falciparum* (henceforth *P. falciparum*). Annually, approximately three million people die of malaria. Also, hundreds of millions of people in a year become clinically ill. The negative influence of these results is huge and its socioeconomic impact is beyond measure. This influence is particularly prominent in the African continent, where an estimated US$12 billion is been lost yearly (Breman et al., 2004; Gallup and Sachs, 2001). Reports have shown that the rate of resistance of the parasite to existing drugs is increasing. Therefore, there is a huge and urgent need to discover and validate new drug or vaccine targets to enable the development of new treatments for malaria (Ben Mamoun et al., 2001). The ability to discover these drug or vaccine targets can only be enhanced from our deep understanding of the detailed biology of the parasite, for example how cells function and how proteins organize into modules such as metabolic, regulatory and signal transduction pathways. Biologically, a signal transduction pathway is the chain of processes by which a cell converts an extracellular signal into a response. In most unicellular organisms, the number of signal transduction pathways influences the number of ways the cell can react and respond to the environment. It has been noted that the knowledge of signalling transduction pathways in Plasmodium is fundamental to aid the design of new strategies against malaria (Doering, 1997; Koyama et al., 2009).

A major challenge of post-genomic biology is to understand the complex networks of interacting genes, proteins and small molecules that give rise to biological form and function. Knowledge on protein–protein interactions (for example, integrated with transcriptional data) is crucial to understand the

* Corresponding author.
E-mail address: e.adebiyi@daad-alumni.de (E. Adebiyi).

1567-1348/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2010.11.006

assembly of protein machinery and the formation of protein signalling cascades. Hence, the dissection of protein interaction networks has great potential to improve the understanding of cellular machinery and to assist in deciphering protein function.

For the malaria parasites, the most commonly used computational method for analysing microarray gene expression data is clustering. This has been used by Le Roch et al. (2003) and Bozdech et al. (2003). The results obtained have been used to classify and support gene classification into functional modules, namely metabolic pathways. The results obtained have left us with many genes of putative functional assignments. The Malaria Parasite Metabolic Pathways (http://sites.huji.ac.il/malaria/), also accessible from www.plasmodb.org, provides limited information about this. Recent studies like Gangman et al. (2007) and Zhou et al. (2005) introduce the use of Gene Ontology, but the results are also still very limited in their application to *P. falciparum* (Oyelade et al., 2008). This is because only a minority of *P. falciparum* proteins are annotated by GO terms. An extensive analysis of the available protein–protein interaction information (LaCount et al., 2005) for *P. falciparum* is not available. Apart from the Ca^{2+}/Calmodulin-PfPKB signalling pathway of Vaid and Sharma (2006), no other chain of signal transduction pathways in *P. falciparum* is presently known.

The available knowledge about protein interactions and gene co-regulation in a single specie can be represented as a weighted graph of protein interactions, whose vertices represent proteins and whose edges represent interactions; each edge is assigned a weight from available experimental data (using the transcriptomic and protein interaction data), indicating the strength of evidence for the existence of the corresponding interaction. Protein signalling cascades (or signal transduction pathways) can be described as chains of interacting proteins, in which protein interactions enable each protein in the path to modify its successor so as to transmit biological information. Such structures correspond to simple paths in the protein interaction weighted graph (Scott et al., 2006). Identifying biologically meaningful simple paths corresponding to signalling pathways is very straightforward since in most of the signalling cascades the proteins would transmit the signal from the plasma membrane, where the signal is initiated, towards the nucleus by activation of transcription factors, which in turn lead to transcription of the final effectors.

For the first time, to mine out chains of signal transduction pathways in *P. falciparum*, we applied the techniques developed and deployed to the yeast protein network by Scott et al. (2006). We consider a new, modified and biologically motivated extension of the basic path-finding problem. This is essential for application to organisms where not many experimentally validated protein interactions are known, such as *P. falciparum*. Recent work by Bebek and Yang (2007) presented alternative techniques to solving the path-finding problem, but all of these methods suffer from the problem of data sparsity. In *P. falciparum*, 60% of its proteins lack resemblance to any existing annotated organism (Gardner et al., 2002), which illustrates the dimension of this problem.

The snapshot of our results is presented in Tables 1a–1e with the overall results given in Tables 2–8 of the supplementary materials. From these results, we have been able to predict several important signalling transduction pathways in *P. falciparum*. We have predicted a viable signalling pathway characterized in term of the genes responsible that may be the PfPKB pathway recently biologically elucidated by Vaid and Sharma (2006) and Vaid et al. (2008). We obtained from the FIKK family, a signal transduction pathway that ends up on a chloroquine resistance marker protein, which indicates that interference with FIKK proteins might reverse *P. falciparum* from resistant to sensitive phenotype. We also proposed a hypothesis that showed the FIKK proteins in this pathway as enabling the resistance parasite to have a mechanism for releasing chloroquine (via an efflux process) (Krogstad et al., 1987, 1988). Furthermore, we also predicted a signalling pathway that may have been responsible for signalling the start of the invasion process of Red Blood Cell (RBC) by the merozoites. It has been noted that the understanding of this pathway will give insight into the parasite virulence and will facilitate rational vaccine design against merozoites invasion (Miller et al., 2002). And we have a host of other predicted pathways, some of which have been used in this work to predict the functionality of some proteins in the malaria parasite.

The paper is structured as follows. In the next section, we describe briefly the techniques employed in this work. Section 3 shows our predicted chains of signal transduction pathways in *P. falciparum* encapsulated in tables. Section 4 discusses the results obtained, and we conclude the paper in Section 5.

2. Materials and methods

2.1. Constructing a protein–protein interaction network

Protein–protein interaction data was obtained from the work of LaCount et al. (2005). Their results comprise 2846 interactions between 1308 proteins of *P. falciparum* in its intra-erythrocytic cycle. In addition to the protein–protein interaction data, the transcriptional data from Le Roch et al. (2003) and Bozdech et al. (2003) were integrated to contributing to weighting the interaction reliabilities, depicted by the edges of the interaction graph.

2.2. The algorithm

Essentially, we applied the techniques developed and deployed to the yeast protein network by Scott et al. (2006). We consider a new, modified and biologically motivated extension of the basic path-finding problem to be able to deal with organisms of sparsely populated experimentally verified protein interactions such as the malaria parasite.

For completeness, we briefly discuss here the Scott et al. (2006) techniques. Note that a recent work by Bebek and Yang (2007) presented an alternative method to solve the path-finding problem. In organisms such as *P. falciparum*, where 60% of its proteins lack resemblance to any existing annotated organism (Gardner et al., 2002) and a number of its signalling pathways are unknown, these methods cannot operate reliably without adaptation to this situation.

Consider a weighted interaction graph in which each vertex is a protein and each edge (u, v) represents an experimentally observed interaction between proteins u and v, and is assigned a numerical value $p(u, v)$ representing the probability that u and v interact. This probability is calculated under the logistic distribution given three variables, namely (i) the number of times an interaction between two proteins was observed (Deng et al., 2003; Sharan et al., 2005), (ii) the Pearson correlation of expression measurements for the corresponding genes, and (iii) the proteins' small world clustering coefficient (Sharan et al., 2005). Note that the product of the values assigned to the edges of a simple path can be use to score it. Among paths of a given length, those with the highest scores are plausible candidates for linear signal transduction pathways. Given a set I of possible start vertices, we would like to find the highest-scoring paths from I to each vertex of the graph. In the case of signalling pathways, I might be the set of all receptor proteins or a single protein of particular interest.

The problem can be framed mathematically as follows. In order to work with an additive weight rather than a multiplicative one, we assign each edge (u, v) a weight $w(u, v) = -\log p(u, v)$. We define the weight of a path as the sum of the weights of its edges, and the length k of a path as the number of vertices it contains.

Table 1a
Extracted potential important signalling transduction pathways from calcium modulated and signalling proteins. Column one indicates the name of the signalling pathway, the second column shows minimum paths extracted, while optimizing the identified number of proteins in the pathway under consideration. The third column shows the weight *p*-value and column four detailed the products (from plasmodb) of the proteins in the identified potential signalling pathways.

Name/ Figure tag	Minimum path	*p*-Value	Gene IDs	Products
Calcium modulated				
	PFB0540w → PFB0815w → PFF0220w → PFF0590c → PF14_0632	0.044	PFB0540w PFB0815w PFF0220w PFF0590c PF14_0632	Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 t conserved Plasmodium protein, unknown function homologue of human HSPC025 26S proteasome subunit, putative
	PFA0110w → PFB0540w → FB0815w → FD0090c FF0220w → PFF0590c	0.049	PFA0110w PFB0540w PFB0815w PFD0090c PFF0220w PFF0590c	DNAJ protein, putative Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 Plasmodium exported protein (PHISTa), unknown function conserved Plasmodium protein, unknown function homologue of human HSPC025
	PFB0540w → PFB0815w → PFE0070w → PFF0675c → PF11_0111	0.044	PFB0540w PFB0815w PFE0070w PFF0675c PF11_0111	Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 Interspersed repeat antigen, putative Myosin E asparagine-rich antigen
	PFB0540w → PFB0815w → PFD0985w → PFF0590c → PFF0785w	0.044	PFB0540w PFB0815w PFD0985w PFF0590c PFF0785w	Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 Transcription factor with AP2 domain(s), putative Homologue of human HSPC025 Ndc80 homologue, putative
Calcium Signalling				
	PF10_0143 → PF11_0142 → PF11_0239 → MAL13P1.206	0.033	PF10_0143 PF11_0142 PF11_0239 MAL13P1.206	Transcriptional activator ADA2, putative ubiquitin domain containing protein calcium-dependent protein kinase, putative Na^+-dependent Pi transporter, sodium -dependent phosphate transporter
	PF11_0142 → PF11_0239 → PF13_0197 → MAL13P1.206	0.038	PF11_0142 PF11_0239 PF13_0197 MAL13P1.206	ubiquitin domain containing protein calcium-dependent protein kinase, putative Merozoite Surface Protein 7 precursor, MSP7 Na^+ -dependent Pi transporter, sodium-dependent phosphate transporter
	PF11_0142 → PF11_0239 → MAL13P1.206	0.022	PF11_0142 PF11_0239 MAL13P1.206	ubiquitin domain containing protein Calcium-dependent protein kinase, putative Na^+-dependent Pi transporter, sodium-dependent phosphate transporter
	PF11_0142 → PF11_0239 → MAL13P1.206 → PF14_0059	0.046	PF11_0142 PF11_0239 MAL13P1.206 PF14_0059	Ubiquitin domain containing protein Calcium-dependent protein kinase, putative Na^+-dependent Pi transporter, sodium-dependent phosphate transporter Conserved plasmodium protein, unknown function
	PF11_0142 → PF11_0239 → MAL13P1.206 → PF14_0678	0.033	PF11_0142 PF11_0239 MAL13P1.206 PF14_0678	Ubiquitin domain containing protein Calcium-dependent protein kinase, putative Na^+-dependent Pi transporter, sodium-dependent phosphate transporter Exported protein 2
	PFD0090c → PFF0670w → PF08_0048 → PF11_0239	0.041	PFD0090c PFF0670w PF08_0048 PF11_0239	Plasmodium exported protein (PHISTa), unknown function Transcription factor with AP2 domain(s), putative ATP-dependent helicase, putative Calcium-dependent protein kinase, putative

Table 1b
Extracted potential important signalling transduction pathways from cell cycle, cyclic nucleotide and phosphatidylinositol cycle proteins. Column one indicates the name of the signalling pathway, the second column shows minimum paths extracted, while optimizing the identified number of proteins in the pathway under consideration. The third column shows the weight *p*-value and column four detailed the products (from plasmodb) of the proteins in the identified potential signalling pathways.

Name/ Figure tag	Minimum paths	*p*-Value	Gene ID	Products
Cell Cycle				
	PFE1370w → PF10_0143 → PF10_0272	0.021	PFE1370w	hsp70 interacting protein, putative
			PF10_0143	Transcriptional activator ADA2, putative
			PF10_0272	60S ribosomal protein L3, putative
Cyclic nucleotide				
	PFB0190c → PFC0435w → PFE0660c → PF10_0254 → MAL13P1.202	0.033	PFB0190c	Conserved Plasmodium protein, unknown function
			PFC0435w	Conserved Plasmodium protein, unknown function
			PFE0660c	Purine nucleotide phosphorylase, putative
			PF10_0254	Conserved Plasmodium protein, unknown function
			MAL13P1.202	Conserved Plasmodium protein, unknown function
	PFC0435w → PFE0660c → PF08_0129 → PF11_0111 → MAL13P1.202	0.028	PFC0435w	Conserved Plasmodium protein, unknown function
			PFE0660c	purine nucleotide phosphorylase, putative
			PF08_0129	Serine/threonine protein phosphatase, putative
			PF11_0111	asparagine-rich antigen
			MAL13P1.202	Conserved Plasmodium protein, unknown function
	PFC0435w → PFE0660c → PFL2520w → MAL13P1.202	0.016	PFC0435w	Conserved Plasmodium protein, unknown function
			PFE0660c	Purine nucleotide phosphorylase, putative
			PFL2520w	Reticulocyte-binding protein 3 homologue
			MAL13P1.202	Conserved Plasmodium protein, unknown function
Phosphatidylinositol Cycle				
	PFE0750c → PFL1930w → MAL13P1.256	0.005	PFE0750c PFL1930w	RNA recognition motif, putative
			MAL13P1.256	Conserved Plasmodium protein, unknown function
				Phosphatidylinositol transfer protein, putative
	PFA0110w → PFE0750c → MAL13P1.256 → PF14_0257	0.046	PFA0110w	DNAJ protein, putative
			PFE0750c MAL13P1.256	RNA recognition motif, putative
			PF14_0257	phosphatidylinositol transfer protein, putative
				conserved protein, unknown function
	PFA0110w → PFD0090c → PFE0750c → MAL13P1.256	0.029	PFA0110w PFD0090c	DNAJ protein, putative
			PFE0750c	Plasmodium exported protein (PHISTa),
			MAL13P1.256	unknown function
				RNA recognition motif, putative
				Phosphatidylinositol transfer protein, putative
	PFE0750c → PFF1050w → PF10_0115 → MAL13P1.256	0.043	PFE0750c	RNA recognition motif, putative
			PFF1050w PF10_0115	Nascent polypeptide associated complex
			MAL13P1.256	alpha chain, putative
				QF122 antigen
				Phosphatidylinositol transfer protein, putative

Given an undirected weighted graph $G = (V, E, w)$ with n vertices, m edges and a set I of start vertices, we wish to find, for each vertex v, a minimum-weight simple path of length k that starts within I and ends at v. If no such simple path exists, this should be reported. In general, this problem is NP-hard, as the travelling-salesman problem is reducible to it. A standard dynamic programming algorithm exists for this problem, which runs in $O(kn^k)$ and requires also $O(kn^k)$ memory (Scott et al., 2006).

In an attempt to reduce the time and space complexity demanded by the technique above, the colour coding idea was introduced. The idea of colour coding is to assign each vertex a random colour between 1 and k and, instead of searching for paths with distinct vertices, search for paths with distinct colours. The introduction of this greatly reduced the complexity of the dynamic programming algorithm, and the paths extracted are necessarily simple. However, a path fails to be discovered if any two of its vertices receive the same colour, so many random colourings need to be tried to ensure that the desired paths are not missed. The running time of the colour coding algorithm is exponential in k and linear in m, and the storage requirement is exponential in k and linear in n. This method is much more cost-effective when n is much larger than k, as is the case in our application.

Next, the colour-coding solutions were extended to several biologically motivated extensions of the basic path-finding problem. These include: (1) constraining the set of proteins occurring in a path; (2) constraining the order of occurrence of the proteins in a path; and (3) finding pathway structures that are more general than simple paths. Due to the scarcely experimental populated protein interaction network, like the one we are considering, and due to the fact that little is known about the order of proteins in any signalling pathway in *P. falciparum*, we will consider only solving the first but modified version of the problems above, namely, given a set of proteins, constrain the maximal number of proteins occurring in a path. Note that this is slightly different to the first problem solved by Scott et al. (2006). They searched in their first biologically motivated problem as indicated above, pathway with a set of proteins, but in *P. falciparum*, known set of proteins that probably formulate signalling pathways are poorly known for a number of reasons (Doering, 1997; Koyama et al., 2009; Ward et al., 2004). The challenge to extract well defined sets using other eukaryotes is further complicated by the fact that about 60% of the *P. falciparum* proteins are hypothetical and share little or no sequence similarity with other eukaryotes (Koyama et al., 2009; Gardner et al., 2002). Therefore, for each given set that we formulate as shown below; we sought for pathways that contain a maximal number of proteins in that set. That means other proteins not in that set may be found in our predicted signalling pathways.

The paths computed are evaluated using two measures, namely weight *p*-value and the functional enrichment. Given a path with weight w, its weight *p*-value is defined as the percent of top-scoring paths in random networks (computed using the same algorithm that is applied to the real network) that have weight w or lower, where random networks are constructed by shuffling the

Table 1c
Extracted potential important signalling transduction pathways from the FIKK family proteins. Column one indicates the name of the signalling pathway, the second column shows minimum paths extracted, while optimizing the identified number of proteins in the pathway under consideration. The third column shows the weight *p*-value and column four detailed the products (from plasmodb) of the proteins in the identified potential signalling pathways.

Name/ Figure tag	Minimum paths	*p*-Value	Gene ID	Products
FIKK				
	PFA0130c → PFE1590w → MAL8P1.153 → PFA0215w	0.046	PFA0130c PFE1590w MAL8P1.153 PFA0220w	Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Transcription factor with AP2 domain(s), putative Ubiquitin carboxyl-terminal hydrolase, putative
	PFA0130c → PFE1590w → MAL8P1.153 → PF10_0075	0.039	PFA0130c PFE1590w MAL8P1.153 PF10_0075	Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Transcription factor with AP2 domain(s), putative Transcription factor with AP2 domain(s), putative
Fig. 2a	**PFA0130c → PFE1590w → MAL8P1.153 → PF11_0342**	**0.036**	**PFA0130c PFE1590w MAL8P1.153 PF11_0342**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 transcription factor with AP2 domain(s), putative conserved Plasmodium protein, unknown function**
	PFA0130c → PFE1590w → PF10_0232 → PF11_0506	0.036	PFA0130c PFE1590w PF10_0232 PF11_0506	Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Chromodomain-helicase-DNA-binding protein 1 homolog, putative Antigen 332, DBL-like protein
Fig. 1b	**PFA0130c → PFE1590w → PFF0590c → MAL8P1.153 → PFL1385c**	**0.046**	**PFA0130c PFE1590w PFF0590c MAL8P1.153 PFL1385c**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Homologue of human HSPC025 Transcription factor with AP2 domain(s), putative Merozoite Surface Protein 9, MSP-9**
Fig. 1a	**PFA0130c → PFE1590w → MAL8P1.153 → MAL8P1.23 → PF14_0463**	**0.039**	**PFA0130c PFE1590w MAL8P1.153 MAL8P1.23 PF14_0463**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Transcription factor with AP2 domain(s), putative Ubiquitin-protein ligase 1, putative Chloroquine resistance marker protein**
Fig. 2b	**PFA0130c → PFE1590w → PFE1605w → MAL8P1.153**	**0.036**	**PFA0130c PFE1590w PFE1605w MAL8P1.153**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Plasmodium exported protein (PHISTb), unknown function Transcription factor with AP2 domain(s), putative**
Fig. 2c	**PFA0130c → PFE1590w → PFF0220w → PFF0590c → MAL8P1.153**	**0.046**	**PFA0130c PFE1590w PFF0220w PFF0590c MAL8P1.153**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Conserved Plasmodium protein, unknown function Homologue of human HSPC025 Transcription factor with AP2 domain(s), putative**
Fig. 2d	**PFA0130c → PFE1590w → PFF1220w → MAL8P1.153**	**0.036**	**PFA0130c PFE1590w PFF1220w MAL8P1.153**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Conserved Plasmodium protein, unknown function transcription factor with AP2 domain(s), putative**
Fig. 2e	**PFA0130c → PFE1590w → PF07_0056 → MAL8P1.153 → MAL8P1.23**	**0.036**	**PFA0130c PFE1590w PF07_0056 MAL8P1.153 MAL8P1.23**	**Serine/Threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Conserved Plasmodium protein, unknown function Transcription factor with AP2 domain(s), putative Ubiquitin-protein ligase 1, putative**
	PFA0130c → PFE1590w → MAL8P1.153	0.007	PFA0130c PFE1590w MAL8P1.153	Serine/threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Transcription factor with AP2 domain(s), putative
	PFA0130c → PFE1590w → MAL8P1.153 → PF08_0034	0.041	PFA0130c PFE1590w MAL8P1.153 PF08_0034	Serine/threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Transcription factor with AP2 domain(s), putative histone acetyltransferase GCN5, putative
	PFA0130c → PFE1590w → MAL8P1.153 → MAL8P1.23	0.036	PFA0130c PFE1590w MAL8P1.153 MAL8P1.23	Serine/threonine protein kinase, FIKK family, putative Early transcribed membrane protein 5, ETRAMP5 Transcription factor with AP2 domain(s), putative Ubiquitin-protein ligase 1, putative

edges and weights of the original network, preserving vertex degrees. To evaluate the functional enrichment of a path *P*, its proteins are associated with known biological processes using Gene Ontology (GO) annotations (The Gene Ontology Consortium, 2000). It is then straightforward to compute the tendency of the proteins to have a common annotation using a method developed in Sharan et al. (2005). The scoring is done as follows: define a protein to be below a GO term *t*, if it is associated with *t* or any other term that is a descendant of *t* in the GO hierarchy. For each GO term *t* with at least one protein assigned to it, we compute a hypergeometric *p*-value based on the following quantities: (1) the number of proteins in *P* that are below *t*; (2) the total number of proteins below *t*; (3) the number of proteins in *P* that are below all parents of *t*; and (4) the total number of proteins below all parents of *t*. The *p*-value is further Bonferroni-corrected for multiple testing.

3. Results

We applied the methods above to search for minimum pathways in the *P. falciparum* interaction (weighted graph) network. Our findings can be found in Tables 2–8 of the Supplementary materials. A snapshot of these tables is presented in Table 1.

Table 1d
Vaid and Sharma (2006) and Vaid et al. (2008) motivated extracted potential important signalling transduction pathways. Column one indicates the name of the signalling pathway, the second column shows minimum paths extracted, while optimizing the identified number of proteins in the pathway under consideration. The third column shows the weight *p*-value and column four detailed the products (from plasmodb) of the proteins in the identified potential signalling pathways.

Name/Figure tag	Minimum paths	*p*-Value	Gene ID	Products
Phosphatidylinositol cycle				
Fig. 3b	**PFE0750c → PFL1385c → MAL13P1.256**	**0.005**	**PFE0750c PFL1385c MAL13P1.256**	**RNA recognition motif, putative Merozoite Surface Protein 9, MSP-9 Phosphatidylinositol transfer protein, putative**
Calcium modulated				
	PFB0540w → PFB0815w → PFF0675c → PF10_0345 → PF11_0111	0.044	PFB0540w PFB0815w PFF0675c PF10_0345 PF11_0111	Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 myosin E merozoite surface protein 3 asparagine-rich antigen
	PFB0540w → PFB0815w → PFF0220w → PFF0590c → PFL1385c	0.024	PFB0540w PFB0815w PFF0220w PFF0590c PFL1385c	Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 Conserved Plasmodium protein, unknown function Homologue of human HSPC025 Merozoite Surface Protein 9, MSP-9
	PFB0540w → PFB0815w → PFF1365c → MAL7P1.12	0.029	PFB0540w PFB0815w PFF1365c MAL7P1.12	Conserved Plasmodium protein, unknown function Calcium-dependent protein kinase 1 HECT-domain (ubiquitin-transferase), putative Erythrocyte membrane-associated antigen
Cell Cycle				
	PF10_0254 → PF10_0272 → PFL1385c	0.021	PF10_0254 PF10_0272 PFL1385c	Conserved plasmodium protein, unknown function 60S ribosomal protein L3, putative Merozoite Surface protein 9, MSP-9
cyclic nucleotide				
	PFC0435w → PFE0660c → PF10_0281 → PF11_0224	0.016	PFC0435w PFE0660c PF10_0281 PF11_0224	Conserved Plasmodium protein, unknown function Purine nucleotide phosphorylase, putative Merozoite TRAP-like protein, MTRAP Circumsporozoite-related antigen
Unknown signal transduction groups				
	PFA0125c → PFE0570w → PF11_0277 → PFL1385c	0.027	PFA0125c PFE0570w PF11_0277 PFL1385c	Erythrocyte binding antigen-181 RNA pseudouridylate synthase, putative Conserved Plasmodium protein, unknown function Merozoite Surface Protein 9, MSP-9

From the available literature (Doering, 1997; Koyama et al., 2009; Ward et al., 2004), we found the following identified classes of signal transduction pathways: cAMP dependent, cGMP dependent, MAP kinase, MAPK, phosphatidylinositol cycle, calcium signalling, protein phosphatases, calcium modulated protein kinase, cyclic nucleotide-dependent, CDK-like kinases, cell cycle kinases and the novel FIKK kinases. Searching the plasmodb database using their text option using keywords from the above specific identified pathways names, we found 43, 12, 7, 1, 25, 85, 2, 64, 234, 10, 620, or 36 *P. falciparum* genes in these classifications, respectively. Searching for the existence of these genes in the LaCount protein–protein interaction data, we found none from cAMP dependent, cGMP dependent, MAP kinase and MAPK signalling, but we found 9 from phosphatidylinositol cycle, 32 from calcium signalling, none from protein phosphatases, 28 from calcium modulated protein kinase, 55 from cyclic nucleotide-dependent, none from CDK-like kinases, 130 from cell cycle kinases and 8 from novel FIKK kinases proteins. To ensure we have exhaustively considered all putatively or annotated signal transduction pathways genes in *P. falciparum*, we further search plasmodb using the keyword 'signal transduction'. We found 1183 genes in this category and later filtered out all genes found earlier on using the known signalling pathways listed above. We also observed that there are many genes in the known pathways; which are not part of the filtered 1183 genes. We are then left with 940 genes that are not in any of the known signal transduction pathways. We call this group of genes "unknown" signalling pathways. Considering our modified first biological motivated problem; we evaluated the weight *p*-value and hypergeometric *p*-value (for functional enrichment) of each path extracted. Note that; following Scott et al. (2006); we extracted pathways with lengths less than or equal to 10 and considered only pathways whose weight and hypergeometric *p*-values were less than 0.05. We used only these criteria for now; since this is the very first time; this kind of analysis is being done on the only existing protein–protein interaction network (LaCount et al., 2005) for *P. falciparum*. We felt; it is important to be able to see explicitly all potential signalling pathways. Furthermore; the results obtained with these criteria have been biological proven to be reliable when applied to yeast protein–protein interaction network (Scott et al., 2006). We found minimum pathways for the genes in the known classifications above as given in Tables 2–7 of the supplementary data. For the "unknown"; we set *I* to be each gene in a sequence and found also in Table 8 (supplementary material); the listed pathways; whose weight and hypergeometric *p*-values are both less than 0.05.

To visualize the content of the tables diagrammatically, the highlighted (bold) ones in Tables 2–7 can be captured in the usual signalling pathways (Figs. 1 and 2). They are also highlighted in bold and tagged in Tables 1a–1d.

4. Discussion

Since our work is the first attempt to predict the main chains of signal transduction pathways in *P. falciparum*, we adhere strictly to the description of most signal pathways: "the proteins would transmit the signal from the membrane, where the signal is initiated, towards the nucleus by the activation of transcription factors, which in turn lead to transcription of the final effectors".

Table 1e
DomainSweep functional prediction for the proteins with unknown function in Tables 1a–1d above. m.p = membrane protein, n.p = nuclear protein, t.f = transcription factor, m.s.p = merozoite surface protein. The third column indicates putative hits that do not fulfil the criteria of a significant hit but have a score above a certain threshold. A significant hit has at least two hits of domains which are described in two protein family databases AND which are members of the same INTERPRO family/domain, or at least two motif hits or two block hits in correct order as described in an individual entry of the Prints or the Blocks database. We listed the first two as predicted from DomainSweep. Columns four and five indicate selected hits from and name of the specific domains or families, respectively.

Gene ID	Our prediction from predicted signalling pathways	Putative hits	Selected protein domains and families hits	The selected protein domains and families
PFB0540w	m.p.	GPCR, family 3, metabotropic glutamate receptor 3 Involucrin repeat	Ribosomal protein L35	PFAM A
PFF0220w	n.p.	Mycobacterial pentapeptide repeat Ribosomal protein VAR1	Anticodon nuclease activator family Bipartite nuclear lo	PFAM A PROSITE-PROFILES
PF14_0059	t.f.	Protein of unknown function DUF1754, eukaryotic Daxx protein	Transcription factor IIA, alpha/bet Transcription elongation factor Elf1	BLOCKS PFAM A
PFB0190c	m.p.	Sel1 repeat	Mitochondrial ribosomal protein (VAR1) Plasmodium histidine-rich protein (HRP) Putative stress-responsive nuclear e Bipartite nuclear lo IF-2: translation initiation factor I S8e: ribosomal protein S8.e	PFAM APROSITE-PROFILES TIGRFAMS
PFC0435w	t.p.	Botulinum neurotoxin	Mitochondrial ribosomal protein (VAR1) YL1 nuclear protein Bipartite nuclear lo ETRAMP: early transcribed membrane	PFAM APROSITE-PROFILES TIGRFAMS
PF10_0254	n.p.	Bipartite nuclear lo Asparagine-rich regi	Ribosomal protein S15 Transcription factor S-II (TFIIS), ce Heat shock factor binding protein 1 Plasmodium histidine-rich protein (HRP) Ribosomal protein S26e rho: transcription termination factor	PFAM A TIGRFAMS
MAL13P1.202	t.f.	Clostridium neurotoxin, translocation Phosphatidylinositol-4, 5-bisphosphate phosphodiesterase beta, conserved site	phage_rinA: phage transcriptional reg rho: transcription termination factor	TIGRFAMS
PFL1930w	t.f.	Uso1/p115 like vesicle tethering protein, head region	bZIP transcription factor Putative stress-responsive nuclear en Asparagine-rich regi Bipartite nuclear lo	PFAM APROSITE-PROFILES
PF14_0257	t.f.	Transcription factor IIA, alpha/beta subunit Translation initiation factor eIF3 subunit	P21_Cbot: transcriptional regulator	TIGRFAMS
PFD0090c	t.f.	Exported protein, PHISTa/c, conserved d omain, Plasmodium Basic helix–loop–helix, Nulp1-type	Myb-like DNA-binding domain	PFAM A
PF11_0342	m.s.p.	P60-like TAFII55 protein, conserved region	Merozoite surface protein (SPAM)	PFAM A
PFE1605w	n.p.	Apoptosis regulator, Bcl-2 relatedANTIGEN SURFACE MALARIA	Nuclear factor I protein pre-N-termin Putative stress-responsive nuclear en	PFAM APFAM A
PFF1220w	n.p.	Botulinum neurotoxin Asparagine-rich regi Lysine-rich region p	Mitochondrial ribosomal protein (VAR1)	PFAM A
PF07_0056	t.f.	Subtilin biosynthesis protein SpaC Putative 5-3 exonuclease	Poxvirus Late Transcription Factor VL	PFAM A
PF11_0277	t.f.	Autophagy-related protein 6 Uso1/p115 like vesicle tethering protein, head region	ribosomal protein L29	TIGRFAMS

We thus extracted, for example, from the tables of the supplementary data, the following signalling pathways, namely, calcium modulated, calcium signalling, cell cycle kinases, cyclic nucleotide, phosphatidylinositol cycle, FIKK in *P. falciparum*. They are highlighted (in red) in the tables in the supplementary materials and reproduced here in Tables 1a–1c. The biological validation of these pathways will certainly be useful and attractive for designing new strategies against malaria.

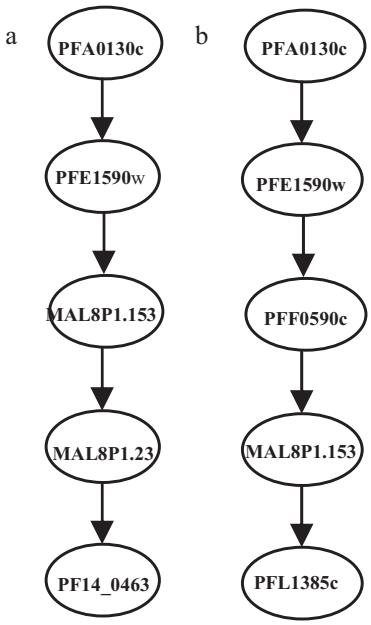

Fig. 1. Potential vital signalling pathways from the FIKK family proteins as extracted into Table 1c. (a) Potential chloroquine resistance signalling pathway and (b) potential signalling pathway that may have signal the start of the invasion process of Red Blood Cell (RBC) by the merozoites.

Vaid and Sharma (2006) reported the first signalling pathway in *P. falciparum*, which involves activation of protein kinase B-like enzyme (PfPKB) by calcium/Calmodulin (CaM). This is depicted in Fig. 3(a) as given in Fig. 7 of Vaid and Sharma (2006), but it has not been characterized in term of the genes responsible. In their study, they also noted that PfPKB is expressed mainly in the schizont/merozoite stages of *P. falciparum*, and the calcium necessary for PfPKB activation by CaM is dependent on the activation of phospholipase C (PLC). Therefore, the PfPKB pathway is regulated by CaM and phospholipase C-mediated calcium release. The erythrocyte invasion is a multistep process, which involves the interaction between the merozoite and the erythrocyte followed by reorientation of the merozoite, which leads to the formation of a tight junction between the merozoite apical end and the erythrocyte membrane (Soldati et al., 2004). Vaid et al. (2008) carried out a further study, which showed that the PfPKB pathway is important for erythrocyte invasion. Their study demonstrated

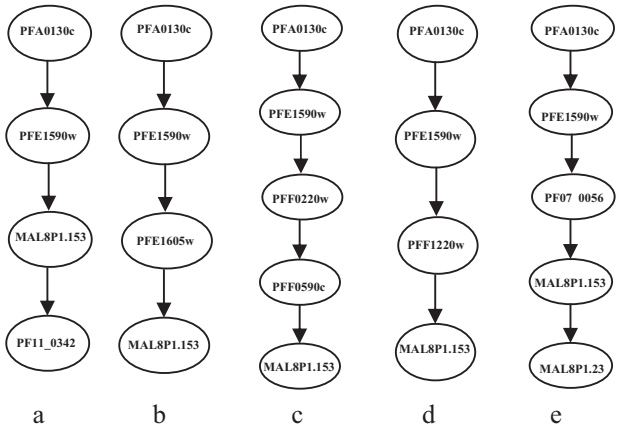

Fig. 2. Hypothetical functional predictions from some predicted signalling pathways from the FIKK family proteins as extracted into Table 1a. (a) P11_0342 was predicted to be a Merozoite Surface Protein, (b) PFE1605w, (c) PFF0220w and (d) PFF1220w as nucleus proteins and (e) PF07_0056 as a transcription factor.

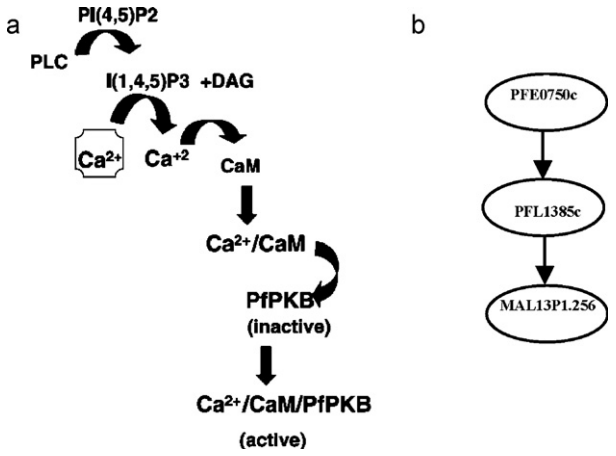

Fig. 3. (a) The Ca^{2+}/Calmodulin-PfPKB signalling pathway as biologically dissected by Vaid and Sharma (2006). (b) The potential corresponding Ca^{2+}/Calmodulin-PfPKB signalling pathway of Vaid and Sharma (2006) from the protein–protein interaction data of LaCount et al. (2005).

that PLC-mediated control of calcium release is important for merozoite invasion and that CaM may be involved in invasion, due to the localization of CaM at the apical end of the merozoites. It was shown in their previous work that PfPKB is one of the very few CaM targets to be identified in *P. falciparum*, so it then follows that the PfPKB pathway may be important for invasion.

Using these findings, we search from Tables 6 and 8 of the supplementary materials for pathways that contain the combination of: phospholipase C/CaM/PfPKB (protein kinase B-like enzyme). We found only the first entry in Table 1d, as depicted in Fig. 3(b). We hypothesize this pathway as the Vaid and Sharma (2006) Ca^{2+}/Calmodulin-PfPKB signalling pathway characterized

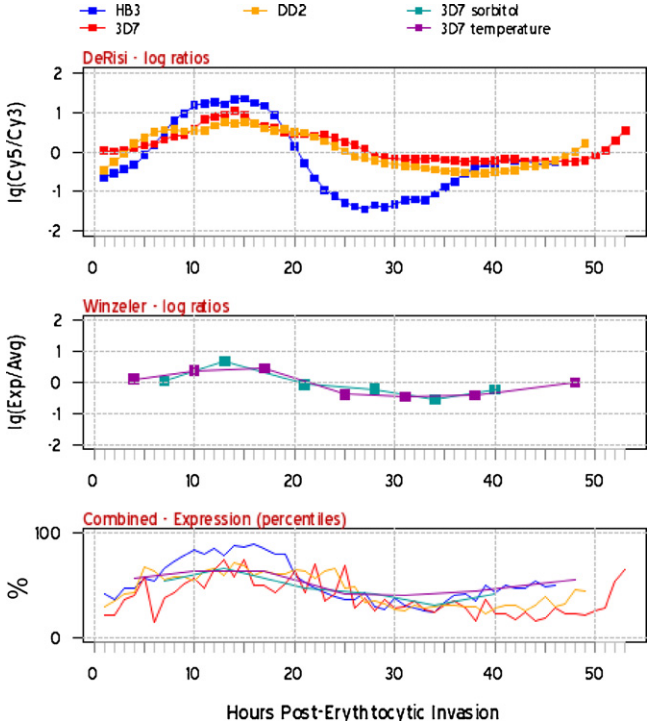

Fig. 4. A stage specific expression profile data for PFA0130c as obtained from plasmodb shows that this serine protease protein is highly expressed at the ring stage for all the different cultures (HB3, 3D7, DD2 in Bozdech et al., 2003 and two 3D7 (sorbitol and temperature) in Le Roch et al., 2003) of the parasite used in experiments.

partially in terms of the genes responsible. This partial characterization, we believe, is due to the scarcely experimental populated protein interaction network underlining our present computational platform. It is important to note that the potential corresponding Ca^{2+}/Calmodulin-PfPKB signalling pathway of Vaid and Sharma (2006) extracted by us is the only pathway that involves a merozoite surface protein among the identified phosphatidylinositol cycle proteins. Further experiments using this set of genes could lead to complete characterization of the pathway in terms of the genes responsible.

We also use the keywords "merozoite" and "erythrocyte" to search all entries of Tables 2–7, and to avoid a trivial result, both of them to search Table 8, we found all the other entries in Table 1d. They are highlighted (in blue) in the tables of the supplementary data. Again, we believe, the biological validation of these pathways will certainly be useful and attractive for designing new strategies against malaria. One interesting thing about the pathways extracted in Table 1d is that the proteins of unknown function in Tables 1a–1c are also the proteins of unknown function in Table 1d, except for PF11_0277. Although this is not the aim of this study, but it is worth to mention that the gene PFA0125c (Table 8), which encodes the protein "erythrocyte binding antigen-181" maybe an important "choke point" in *P. falciparum*. This has not been mentioned in the analysis by LaCount et al. (2005).

Several interesting hypotheses are in particular obtained from the FIKK protein family as shown in Figs. 1a and b and 2. It has been noted by Ward et al. (2004) and Schneider and Mercereau-Puijalon (2005) that among all the *P. falciparum* protein kinases that have been identified, the FIKK protein family is particularly noteworthy. Koyama et al. (2009) suggested that the FIKK kinases may have a role in parasite-induced signalling events because members of this family are exported into the erythrocytes where they are found associated with the Maurer's clefts, and one of the paralogs, R45, is transported to the host cell membrane. This hypothesis is also reflected in our results as we predicted a signal transduction pathway from the FIKK family (see Fig. 1a) that ends up on a chloroquine resistance marker protein, PF14_0463, which indicates that interference with FIKK proteins might reverse *P. falciparum* from resistant to sensitive phenotype.

The Maurer's clefts are established by the parasite within its host cell and play an essential role in directing proteins from the parasite to the erythrocyte surface. Presently, they are appreciated as a novel type of secretory organelle. They play an important role in the export of protein from the parasite across the cytoplasm of the host cell to the erythrocyte surface. This is remarkable since erythrocytes lack secretory organelles found in other eukaryotic cells. As a result, the parasite cannot rely on the host cell for its proteins need and therefore must establish a de-novo secretory system in the host cell cytoplasm, in a compartment outside of its own confines (Frischknecht and Lanzer, 2008). The signal pathways in Fig. 1a assign FIKK proteins to this pathway as enabling the resistance of the parasite by excreting chloroquine via an efflux process (Krogstad et al., 1987, 1988).

With respect to the Red Blood Cell (henceforth, RBC), Miller et al. (2002) noted that what remains completely unknown is which merozoite surface molecules recognize the RBC surface and then signal the start of the invasion process. There was a hypothesis that suggested that RBC invasion requires the cleavage of a surface protein on the RBC by an unknown parasite serine protease. It has been noted that understanding this pathway will give insight into the parasite virulence and will facilitate rational vaccine design against merozoites invasion (Miller et al., 2002). The signalling pathway predicted and depicted in Fig. 1b suggests the transduction pathway of that process. The serine protease protein among the proteins involved in this pathway is PFA0130c. From Le Roch et al. (2003) and Bozdech et al. (2003) respectively, it is known that the 48-h *P. falciparum* intraerythrocytic developmental cycle (IDC) initiates with merozoite invasion of RBCs and is followed by the formation of the parasitophorous vacuole (PV) during the ring stage. The parasite then enters a highly metabolic maturation phase, the trophozoite stage, prior to parasite replication. In the schizont stage, the cell prepares for reinvasion of new RBCs by replicating and dividing to form up to 32 new merozoites. The ring stage, immediately after the merozoite invasion, happens between the 1st hour to the 24th hour, the trophozoite stage begins from the 8th hour to the 33rd hour, while the schizont stage picked up from the 24th hour to the 48th hour. A stage specific expression profile data (see Fig. 4) for PFA0130c as obtained from plasmodb shows that this serine protease protein is highly expressed at the ring stage for all the different cultures of the parasite used in experiments. These hypotheses need of course to be experimentally validated.

The popular description of most signalling pathways is: "the proteins would transmit the signal from the membrane, where the signal is initiated, towards the nucleus by activation of transcription factors, which in turn lead to transcription of the final effectors". We applied this to suggest the functions of some genes as depicted in the proposed signal transduction pathways of Fig. 2. From their position in Fig. 2a–e, we hypothesize that PF11_0342 is a putative Merozoite Surface Protein, PFE1605w, PFF022w and PFF1220 are nuclear proteins, and PF07_0056 is a transcription factor. For the other proteins in Tables 1a–1d, we predicted their functions as described in the second column of Table 1e.

In an attempt to corroborate our prediction above, we used the DomainSweep software of del Val et al. (2007) to predict the function of these proteins as listed in Tables 1a–d. The result of DomainSweep on these proteins is shown in Table 1e.

From Table 1e, two results are conveyed, one, DomainSweep may be able to play a vital role in the re-annotation effort on-going for *P. falciparum* proteins and two, the information extracted by our work (apart from providing information about potential signalling pathways) can be used to collaborate the results of DomainSweep.

For the genes with "unkown" classification, the question is which type of cellular response is transmitted by the predicted signal transduction pathways. The answer to this question will give insight into a number of other signalling pathways and help us to understand better how the malaria parasite reacts and responds to its environments.

5. Conclusion and future work

In this work, we have been able for the first time to mine signal transduction pathways from the most deadly malaria parasite, *P. falciparum*. We have been able to use these results to suggest important hypotheses that can help to explain the mechanisms that signal chloroquine resistance process by the malaria via an efflux process, and which signals start the invasion process of RBC by the merozoites. One of our predicted pathways may also have provided the Vaid and Sharma (2006) Ca^{2+}/Calmodulin-PfPKB signalling pathway characterized in terms of the genes responsible. The PfPKB pathway has been shown to be important to the erythrocyte invasion (Vaid et al., 2008). We have also been able to use our results to predict functionality for some proteins.

Our present work has given lead to several future studies. To further address the problem of data scarcity (in particular with regard to the protein–protein interaction information available for the malaria parasite), we need to develop techniques to deal with missing edges, i.e. protein–protein interaction that have never been observed but exist in reality. One way to do this is to integrate transcription factors into the derived network, resulting into what has been called an integrated cellular weighted network of transcription–regulation and protein–protein interaction (Yeger-

Lotem et al., 2004). For the malaria parasite *P. falciparum*, only about a third of the number of transcription-associated proteins (TAPs) usually found in the genome of a free-living eukaryote is presently known (Coulson et al., 2004). To mine them on a large scale, a complete but reliable TAP database in *P. falciparum* is underway by Ewejobi et al. (in preparation).

Presently, from plasmodb, we have 137 metabolic pathways covering 2521 genes for *P. falciparum*. This is just about half of the annotated genes of this organism. Therefore, we will in the future consider the application here to the *P. falciparum* metabolic network. This will definitely be useful to predict the classification of more genes into metabolic pathways.

Author's contributions

JO, BB, RE and EA put up the general concepts and design of the study. JO and IE carried out the implementation of these concepts and design under the direction of EA. JO, IE, BB and EA carried out the results analysis. JO, IE, BB, RE and EA drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was funded by the Covenant University Staff Development Program and by the German Academics Exchange Program (DAAD). We are grateful to Douglas LaCount, Karine LeRoch and Zbynek Bozdech for making their experimental data available via the web. We are also grateful to Shalom Chinedu and Kehinde Okonjo (both from Covenant University) for useful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2010.11.006.

References

Bebek, G., Yang, J., 2007. PathFinder: mining signal transduction pathway segments from protein–protein interaction networks. BMC Bioinformatics 8, 335.

Ben Mamoun, C., Gluzman, L.Y., Hott, C., 2001. Coordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite plasmodium falciparum revealed by microarray analysis. Mol. Microbiol. 39 (1), 26–36.

Bozdech, Z., et al., 2003. The transcriptome of the intraerythrocytic development cycle of plasmodium falciparum. PLoS Biol. 1 (1), E5.

Breman, J.G., Alilio, M.S., Mills, A., 2004. Conquering the intolerable burden of malaria: what's new, what's needed: a summary. Am. J. Trop. Med. Hyg. 71 (Suppl. 2), 1–15.

Coulson, R.M.R., Hall, N., Ouzounis, C.A., 2004. Comparative genomics of transcriptional control in the human malaria parasite *P. falciparum*. Genome Res. 14, 1548–1554.

del Val, C., Ernst, P., Falkenhahn, M., Fladerer, C., Glatting, K.H., Suhai, S., Hotz-Wagenblatt, A., 2007. ProtSweep, 2Dsweep and DomainSweep: protein analysis suite at DKFZ. Nucleic Acids Res. 35 (Web Server issue), W444–W450.

Deng, M., Sun, F., Chen, T., 2003. Assessment of the reliability of protein–protein interactions and protein function prediction. PSB 140–151.

Doering, C.D., 1997. Signal transduction in malaria parasites. Parasitol. Today 13 (8), 307–313.

Ewejobi, I., Makolo, A., Bulashevska, S., Brors, B., Eils, R., Llinas, M., Adebiyi, E. The malaria parasite transcription factors and their DNA binding sites using a holistic probabilistic framework, in preparation.

Frischknecht, F., Lanzer, M., 2008. The *Plasmodium falciparum* Maurer's clefts in 3D. Mol. Microbiol. 6 (4), 687–691.

Gallup, J.L., Sachs, J.D., 2001. The economic burden of malaria. Am. J. Trop. Med. Hyg. 64 (Suppl. 1–2), 85–96.

Gangman, Y., Sze, H.-S., Thon, M.R., 2007. Identifying clusters of functionally related genes in genomes. Bioinformatics 23 (9), 1053–1060.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., Barrell, B., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419 (6906), 498–511.

Koyama, F.C., Chakrabarti, D., Garcia, C.R.S., 2009. Molecular machinery of signal transduction and cell cycle regulation in *Plasmodium*. Mol. Biochem. Parasitol. 165, 1–7.

Krogstad, D.J., Gluzman, I.Y., Kyle, D.E., Odunola, A.M.J., Martin, S.K., Milhous, W.K., Schlesinger, P.H., 1987. Efflux of chloroquine from *Plasmodium falciparum*: mechanism of chloroquine resistance. Science 238, 1283–1285.

Krogstad, D.J., Schlesinger, P.H., Herwaldt, B.L., 1988. Antimalarial agents: mechanism of chloroquine resistance. Antimicrob. Agents Chemother. 799–801.

LaCount, D.J., Marissa, V., Rakesh, C., Amit, P., Russell, B., Jay, R., Hesselberth, Lori, W., Schoenfeld, Irene, O., Sudhir, S., Cornelia, K., Stanley, F., Robert, E., Hughes, 2005. A protein interaction network of the malaria parasite *Plasmodium falciparum*. Nature 438, 103–107.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Kathleen Moch, J., David Haynes, J., De la Vega, P., Holder, A.A., Batalov, S., Carucci, D.J., Winzeler, E.A., 2003. Discovery of gene function by expression profiling of the malaria parasite fife cycle. Science 301 (5639), 1503–1508.

Miller, L.H., Baruch, D.I., Marsh, K., Doumbo, O.K., 2002. The pathogenic basis of malaria. Nature 415, 673–679.

Oyelade, J.O., Adebiyi, E.F., Yah, S.C., Olaseinde, G.I., 2008. Computational identification of functional related gene in malaria parasites. Poster proceeding. ISMB.

Schneider, A.G., Mercereau-Puijalon, O., 2005. A new Apicomplexa-specific protein kinase family: multiple members in *Plasmodium falciparum*, all with an export signature. BMC Genomics 6 (1), 30.

Scott, J., Ideker, T., Karp, R.M., Sharan, R., 2006. Efficient algorithms for detecting signaling pathways in protein interaction networks. J. Comput. Biol. 13 (2), 133–144.

Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Sittler, T., Karp, R.M., Ideker, T., 2005. Conserved patterns of protein interaction in multiple species. PNAS 102 (6), 1974–1979.

Soldati, D., Foth, B.J., Cowman, A.F., 2004. Molecular and functional aspects of parasite invasion. Trends Parasitol. 20, 567–574.

The Gene Ontology Consortium, 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

Vaid, A., Sharma, P., 2006. PfPKB, a protein kinase B-like enzyme from *Plasmodium falciparum*. II. Identification of calcium/calmodulin as its upstream activator and dissection of a novel signalling pathway. J. Biol. Chem. 281 (37), 27126–27133.

Vaid, A., Thomas, D.C., Sharma, P., 2008. Role of Ca^{2+}/Calmodulin-PfPKB signalling pathway in erthrocycte invasion by *Plasmodium falciparum*. J. Biol. Chem. 283 (9), 5589–5597.

Ward, P., Equinet, L., Packer, J., Doerig, C., 2004. Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. BMC Genomics 5, 79.

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U., Margalit, H., 2004. Network motifs in integrated cellular networks of transcription – regulation and protein interaction. PNAS 101 (16), 5934–5939.

Zhou, Y., Young, J.A., Santrosyan, A., Chen, K., Frank Yan, S., Winzeler, E., 2005. In-silico gene function prediction using ontology-based pattern identification. Bioinformatics 21 (7), 1237–1245.