Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Pyruvate kinase and glucose-6-phosphate dehydrogenase deficiencies and their association with malaria – population genetics and proteomic studies

Patrícia Isabel Pires Machado

Licenciada em Biologia pela Universidade de Évora

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor no Ramo de Ciências Biomédicas, Especialidade em Parasitologia, realizada sob orientação científica da Inv.ª Doutora Ana Paula Arez

Orientador: Inv.ª Doutora Ana Paula Arez
Unidade de Parasitologia Médica
Instituto de Higiene e Medicina Tropical

Co-Orientador: Prof. Catedrático Virgílio E. do Rosário
Unidade de Parasitologia Médica
Instituto de Higiene e Medicina Tropical

Comissão Tutorial: Inv.ª Doutora Leonor Gusmão
Instituto de Patologia e Imunologia Molecular da Universidade do Porto

O trabalho foi financiado pela Fundação para a Ciência e Tecnologia, através da bolsa de doutoramento ref. SFRH/BD/28236/2006 e dos projectos de investigação ref. POCI/SAU-ESP/55110/2004 e ref. PTDC/SAUMET/110323/2009.

ABRIL, 2013

Ao Xavier

Agradecimentos / Aknowledgments

A concretização deste trabalho só foi possível devido ao apoio de várias pessoas e instituições às quais quero agradecer, nomeadamente:

A Investigadora Doutora Ana Paula Arez, por me aceitar como sua estudante de doutoramento e, com isso, me ter aberto as portas do Mundo. Agradeço pela orientação e pelas discussões, mas especialmente pela confiança e por estimular a independência e sentido de responsabilidade, contribuindo para que crescesse como pessoa e cientista.

O Professor Doutor Virgílio E. do Rosário, pela co-orientação, pela partilha de conhecimento e pelo apoio, em particular durante a estadia em Maputo. Agradeço também pela sua energia e dinamismo que inspiram quem o rodeia.

A Investigadora Doutora Leonor Gusmão, por ter aceitado ser membro da minha Comissão Tutorial e por ser um verdadeiro exemplo de dedicação e rigor em Ciência, que eu tanto admiro. Obrigada por ter sempre tempo, mesmo quando está assoberbada de trabalho e solicitações.

O Doutor Licínio Manco, por me ter recebido no seu grupo "Genes, Populações e Doença" do Centro de Investigação em Antropologia e Saúde (CIAS) da Universidade de Coimbra. Agradeço pelas suas ideias, discussão de resultados e disponibilidade para colaborar activamente no presente trabalho. A sua colaboração foi verdadeiramente valiosa. Agradeço ainda por ter estabelecido o contacto com o Centro Hospitalar de Coimbra.

O Professor Doutor António Amorim, líder do grupo de Genética das Populações do Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), por me ter aberto as portas do seu grupo, pela partilha de conhecimentos, discussão de resultados e por espicaçar a inteligência dos seus estudantes. Aos restantes elementos deste grupo, agradeço por me terem ajudado em todas as dúvidas de bancada e por me terem proporcionado tão bons momentos no Porto. Um obrigado especial à Verónica Gomes, pelas nossas conversas e e todo o apoio na minha estadia no Porto; ao Rui Pereira, pela grande ajuda na leitura de resultados, por estar sempre presente quando preciso e pelo carinho com que sempre me brinda; à Mafalda Rocha e à Cíntia Alves, pela ajuda fundamental na preparação e corrida das amostras.

A Doutora Natércia Fernandes, da Faculdade de Medicina da Universidade Eduardo Mondlane, em Maputo, Moçambique, que estabeleceu a ponte entre Lisboa e Maputo, tratando da aprovação do projecto de trabalho no Comité Nacional de Bioética em Moçambique e me integrou na rotina hospitalar e laboratorial do Hospital Central de Maputo. A todos os elementos do Departamento de Bioquímica da Faculdade de Medicina da Universidade Eduardo Mondlane: o Dr. Sérgio Chibute, Director do Departamento, que sempre disponibilizou a sua ajuda quando solicitada; a Dra. Graça Salomé, pelo apoio no trabalho no Hospital e ajuda na realização dos questionários nas

enfermarias de Pediatria; e o Dr. Luis Sitoe, pelo apoio técnico no processamento de algumas amostras. A D. Violeta, responsável pelo Laboratório de Hematologia do Departamento de Pediatria, por me permitir acompanhar a rotina do laboratório (colheita de amostras, realização dos hemogramas, preparação de lâminas para diagnóstico de malária) e guardar religiosamente as amostras de sangue e as lâminas para o presente trabalho. O técnico deste mesmo laboratório, Sábado, pela partilha de experiências de vida, sonhos e ambições.

Ainda em Maputo, a Natacha, a Dida, a Antónia e a Juliana, por me terem feito sentir em casa e em família, tão longe que estava da minha casa e da minha família. Proporcionaram-me momentos extraordinários, que não vou esquecer nunca. Não sabia que havia pessoas assim, que abriam a porta de sua casa a uma desconhecida como se ela lá pertencesse e lá tivesse vivido sempre. À Filipa, Mosca, Jonhy e Zeca, duas palavras: Laurentina e 2M. Agradeço pelo companheirismo, pela festa, pela alegria de viver. Obrigada por terem tornado tudo tão fácil e tão bom!

A Doutora Letícia Ribeiro, Directora do Serviço de Hematologia do Centro Hospitalar de Coimbra, por ter disponibilizado a amostra de sangue com deficiência de PK e ter autorizado as minhas visitas ao Laboratório de Hematologia (de referência internacional para o diagnóstico das deficiências de PK e G6PD). Agradeço à Técnica Umbelina Rebelo, à Dra. Celeste Bento e ao Dr. Luís Relvas por partilharem os protocolos e procedimentos utilizados no diagnóstico de deficiências enzimáticas, imprescindíveis para a realização do rastreio em Moçambique.

A Dra. Isabel Albergaria e seus colaboradores, do Instituto Nacional de Saúde Dr. Ricardo Jorge, por também partilharem protocolos e procedimentos utilizados no diagnóstico de deficiências enzimáticas e por cederem bibliografia sobre o assunto, fundamental para o sucesso do trabalho em Moçambique.

Jerry Thomas, Jane Thomas-Oates, David Ashford and Ed Bergstrom, from Centre of Excellence in Mass Spectrometry, University of York, for all the support concerning Mass Spectrometry analyses. Special thanks to Marianne Loong and Ming Yang, for their company, sharing and support during my stay at York (I miss Asian food!).

A Doutora Fátima Nogueira, do Instituto de Higiene e Medicina Tropical (IHMT), por me ter iniciado nas culturas *in vitro* de *Plasmodium* e pelo apoio na preparação dos extractos proteicos. Agradeço particularmente por fazer de advogado do diabo na discussão de protocolos e resultados e me fazer parar e pensar sobre o propósito das coisas. O Doutor João Rodrigues, do IHMT, por estabelecer a ponte com a Universidade de York e pela sua disponibilidade na discussão de resultados e protocolos e ter sempre uma perspectiva positiva e uma palavra de incentivo perante o meu pessimismo.

A todos que comigo trabalharam no IHMT, nomeadamente, a Cristina Mendes e a Rute Félix, pela partilha diária de frustrações, dificuldades, sorrisos e lágrimas.

Estiveram sempre ao meu lado quando precisei. A Cláudia Gomes e a Mónica Guerra, as minhas parceiras "MixInfect", que tanto me ajudaram no trabalho de bancada. O Bruno Gomes, a Lara Borges, a Cláudia Istrate, a Ana Afonso e o Jorge Varanda, por tornarem os meus dias de laboratório mais felizes e por terem sempre uma palavra de coragem. A Celeste Figueiredo, pela eficiente e fundamental ajuda nos assuntos das papeladas e burocracias.

A Catarina Alves e a Dinora Lopes, pelo companheirismo e boa disposição de todos os dias. À Catarina agradeço também por toda a ajuda e apoio na fase final da escrita de tese, nomeadamente leitura de parte da mesma e ajuda na revisão bibliográfica e resumo. À Patrícia Salgueiro e à Ana Custódio estou muito grata por todo o apoio científico e pessoal. Agradeço especialmente por tão bem compreenderem a dificuldade na gestão dos vários papéis!

A Marta Machado pela grande prova de amizade. Acompanhou-me de dia e de noite na fase mais difícil da preparação da tese e fez directas mesmo quando eu sucumbi ao cansaço. Preparou figuras, tabelas, bibliografia, lista de abreviaturas, formatações. Sem a sua ajuda nunca teria conseguido. O que fez por mim não tem preço. Só posso retribuir na mesma moeda.

A todas as pessoas que colaboraram neste estudo disponibilizando a sua amostra de sangue.

O Instituto de Higiene e Medicina Tropical (IHMT) e o Centro de Malária e outras Doenças Tropicais (CMDT), por facultarem todas as condições necessárias para o desenvolvimento deste trabalho. Ainda a Unidade de Ensino e Investigação (UEI) de Parasitologia Médica, onde foi desenvolvida grande parte do trabalho experimental.

O Centro de Investigação em Antropologia e Saúde (CIAS), da Universidade de Coimbra, pela frutífera colaboração e por me ter permitido desenvolver a parte inicial do trabalho experimental, que incluiu a genotipagem de marcadores genéticos em amostras de DNA.

O Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), pela excelente colaboração e por me ter possibilitado fazer a maior parte do trabalho de genética populacional.

A Faculdade de Medicina da Universidade Eduardo Mondlane em Maputo, Moçambique, por me ter recebido e ter dado as condições necessárias para processamento das amostras de sangue colhidas no hospital. Também o Departamento de Pediatria e Banco de Sangue do Hospital Central de Maputo, por me terem aberto as portas e dado toda a liberdade para falar com crianças doentes, pais, dadores de sangue, médicos e técnicos de saúde.

The Centre of Excellence in Mass Spectrometry, from the University of York (United Kingdom), for receiving me as a temporary student, and give me all the technical support for Mass Spectrometry analysis.

A Fundação para a Ciência e Tecnologia, pela concessão da bolsa de doutoramento (ref. SFRH/BD/28236/2006) e fundos concedidos no âmbito dos projectos ref. POCI/SAU-ESP/55110/2004 e ref. PTDC/SAUMET/110323/2009, que tornaram possível a realização deste trabalho.

As minhas amigas de sempre, pelo apoio e amizade durante tantos anos. Por respeitarem e compreenderem as minhas ausências e os meus silêncios dos últimos tempos.

A minha família, que está sempre ao meu lado. Agradeço à minha madrinha pelo apoio extraordinário em todas as situações.

Os meus pais, por hoje, mais do que nunca, reconhecer o seu valor e o seu amor incondicional. Obrigada por tudo. A minha mãe é o meu maior exemplo de força, determinação e perseverança, características que tenho tentado reproduzir perante todos os desafios.

O Pedro, a pessoa que mais me tem apoiado nesta caminhada, por suportar o meu mau-humor, a minha impaciência, o meu desânimo e ser um companheiro e um pai fe-no-me-nal! Obrigada pelo profundo respeito pelas minhas decisões e liberdade. O Xavier, por… ter virado a minha vida do avesso! As palavras não chegam.

Resumo

Deficiência de piruvato cinase e deficiência de glucose-6-fosfato desidrogenase e a sua associação com a malária – estudos de genética populacional e de proteómica

Patrícia Isabel Pires Machado

PALAVRAS-CHAVE: Malária, polimorfismos genéticos humanos do glóbulo vermelho (GV), deficiência de piruvato cinase (PK), deficiência de glucose-6-fosfato desidrogenase (G6PD), marcas de selecção, proteómica, remodelação do glóbulo vermelho, fendas de Maurer.

A malária é reconhecida como uma das principais forças selectivas a actuar na história recente no genoma humano. Inúmeros polimorfismos genéticos têm sido descritos como protectores contra a gravidade da malária, como o alelo HbS (designado de traço falciforme) e o alelo G6PD A- (associado à deficiência de G6PD). Mais recentemente, também a deficiência de PK foi associada com a protecção contra a malária. Evidências desta associação foram obtidas em estudos com modelos de roedor e estudos *in vitro* utilizando GV humanos deficientes em PK. Até à data, não foram obtidos dados em populações humanas que revelem esta associação: ainda não foi identificada uma variante de PK com uma prevalência elevada em regiões endémicas de malária e não foram identificadas marcas de selecção na região do gene que codifica para a PK (gene *PKLR*). Além disso, os mecanismos subjacentes à protecção contra a malária por deficiências enzimáticas dos GV não estão bem esclarecidos.

Assim, os objectivos do presente estudo foram: investigar os polimorfismos genéticos humanos com associação com a malária em Cabo Verde; pesquisar marcas de selecção da malária na região do gene *PKLR* em populações Africanas; determinar a frequência da deficiência em PK e identificar uma eventual variante da enzima que possa estar sob selecção positiva em regiões endémicas de malária; avaliar o efeito das duas deficiências enzimáticas (PK e G6PD) na invasão e maturação do parasita em culturas *in vitro* de *Plasmodium* usando GV normais e deficientes; e analisar o perfil proteómico de GV infectados e não infectados, normais e com deficiência (em PK e G6PD), bem como de parasitas isolados de GV tanto deficientes como normais.

Em Cabo Verde (área epidémica), não foram identificadas marcas de selecção pela malária, através da análise dos vários polimorfismos. No entanto, quando a análise foi realizada em dois países endémicos (Angola e Moçambique), foram detectadas várias marcas de selecção: a genotipagem de microssatélites (STRs) e polimorfismos de base única (SNPs) localizados na vizinhança do gene *PKLR* revelou uma diferenciação consideravelmente maior entre as populações Africana e Europeia (Portuguesa), do que a diferenciação determinada aquando da utilização de marcadores genéticos neutros. Além disso, uma região genómica de maior amplitude apresentou um Desequilíbrio de Ligação (LD) significativo no grupo de malária não grave (e não no grupo de malária

grave), sugerindo que a malária poderá estar a exercer pressão selectiva sobre a região do genoma humano que envolve o gene *PKLR*.

No estudo que incidiu na determinação da prevalência da deficiência de PK no continente Africano (realizado em Moçambique), esta revelou-se elevada - 4,1% - sendo o valor mais elevado descrito até ao momento a nível mundial para esta enzimopatia. Na pesquisa de mutações que pudessem estar na causa deste fenótipo (baixa actividade de PK), foi identificada uma mutação não sinónima 829G>A (277Glu>Lys), significativamente associada à baixa actividade enzimática. Esta mutação foi também identificada em Angola, São Tomé e Príncipe e Guiné Equatorial, onde a frequência de portadores heterozigóticos foi entre 2,6 e 6,7% (valores que se encontram entre os mais elevados descritos globalmente para mutações associadas à deficiência em PK). Não foi possível concluir acerca da associação entre a deficiência de PK e o grau de severidade da malária e da associação entre o alelo 829A e a mesma, devido ao baixo número de amostras.

Os resultados dos ensaios de invasão/maturação do parasita sugeriram que, nos GV com deficiência de PK ou G6PD, a invasão (onde está envolvida a membrana do GV hospedeiro e o complexo apical do parasita) é mais relevante para a eventual protecção contra a malária do que a maturação. Os resultados da análise proteómica revelaram respostas diferentes por parte do parasita nas duas condições de crescimento (GV com deficiência de PK e GV com deficiência de G6PD). Esta resposta parece ser proporcional à gravidade da deficiência enzimática. Nos parasitas que cresceram em GV deficientes em G6PD (provenientes de um indivíduo assintomático), a principal alteração observada (relativamente às condições normais) foi o aumento do número de proteínas de choque térmico e chaperones, mostrando que os parasitas responderam às condições de stress oxidativo, aumentando a expressão de moléculas de protecção. Nos parasitas que cresceram em condições de *deficit* de PK (GV de indivíduo com crises hemolíticas regulares, dependente de transfusões sanguíneas), houve alteração da expressão de um maior número de proteínas (relativamente ao observado em condições normais), em que a maioria apresentou uma repressão da expressão. Os processos biológicos mais representados nesta resposta do parasita foram a digestão da hemoglobina e a troca de proteínas entre hospedeiro e parasita/remodelação da superfície do GV. Além disso, uma elevada percentagem destas proteínas com expressão alterada está relacionada com as fendas de Maurer, que desempenham um papel importante na patologia da infecção malárica. É colocada a hipótese de que a protecção contra a malária em GV deficientes em PK está relacionada com o processo de remodelação da membrana dos GV pelo parasita, o que pode condicionar a invasão por novos parasitas e a própria virulência da malária. Os resultados da análise do proteoma dos GV contribuirão para confirmar esta hipótese.

Abstract

Pyruvate kinase and glucose-6-phosphate dehydrogenase deficiencies and their association with malaria – population genetics and proteomic studies

Patrícia Isabel Pires Machado

KEYWORDS: Malaria, human red blood cell (RBC) genetic polymorphisms, pyruvate kinase (PK) deficiency, glucose-6-phosphate dehydrogenase (G6PD) deficiency, selection signatures, proteomics, RBC remodeling, Maurer's clefts.

Malaria has been recognized as the strongest known force for evolutionary selection in the recent history of the human genome. Several human genetic polymorphisms have been described as protective against malaria severity, as the HbS allele (sickle cell trait) and G6PD A- allele (causing G6PD deficiency). More recently, PK deficiency has also been described as protective against malaria. Evidences were obtained in murine models and *in vitro* studies using PK-deficient human RBC. Human population data has not been obtained so far: a high prevalent PK variant has yet to be identified in malaria endemic regions and selection signatures in the genome region around RBC PK-encoding gene (*PKLR*) have not been detected to date. Also, the mechanisms underlying malaria protection by RBC enzyme deficiencies are not clear.

So, the objectives of this study were: to investigate malaria associated genetic traits in Cape Verde; to look for selection signatures in the *PKLR* gene region in African populations; to determine PK deficiency frequency and identify a prevalent PK variant that could be under selection by malaria in endemic African regions; to assess parasite invasion and maturation of *Plasmodium falciparum* growing *in vitro* in PK and G6PD-deficient and normal RBC; and to analyze the proteomic profile of non-infected and infected PK and G6PD-deficient and normal RBC as well as of parasites isolated from both deficient and normal host cells.

In Cape Verde (epidemic area), no malaria selection signatures were found. However, when the analysis was performed in two malaria endemic countries (Angola and Mozambique), several selection marks were detected: data from Short Tandem Repeat (STR) and Single Nucleotide Polymorphic (SNP) loci spread along the *PKLR* gene region showed considerably higher differentiation between African and European (Portuguese) populations than that usually found for neutral markers, and a wider region showing strong Linkage Disequilibrium (LD) was found in the uncomplicated malaria group (and not in severe malaria group), suggesting that malaria may be shaping this genomic region in malaria countries. Additionally, when we performed the first study concerning the determination of PK deficiency prevalence in the African continent (in Mozambique), we were surprised with a high value: 4.1%. This was the higher frequency ever obtained for PK deficiency worldwide. Then, we looked for a mutation that could be in the origin of this phenotype and the missense mutation 829G>A

(277Glu>Lys) was significantly associated. When we did a research of this mutation in other African countries (Angola, Sao Tome and Principe and Equatorial Guinea), the heterozygous carrier frequency was 2.6-6.7%, which is also among the highest heterozygous frequencies associated to PK deficiency described so far. We could not conclude about the association of PK deficiency and allele 829A with malaria outcome due to low sample number.

Parasite invasion/maturation assays suggested that, in deficient RBC, the invasion step (or the cellular membranes) are more relevant for protection than maturation (the intracellular environment). Proteomic data from parasites growing in both G6PD and PK-deficient RBC revealed a distinct response from parasites growing in both deficient conditions, proportional to the phenotype severity. In parasites growing in G6PD-deficient RBC (asymptomatic individual), the main alteration was the increase of parasitic heat shock proteins and chaperones, showing that parasites are responding to oxidative stress conditions increasing the expression of protective molecules. In PK-deficient (transfusion-dependent individual with regular hemolytic crisis), a wider range of proteins displayed abundance alterations, the majority being down-expressed. The most represented biological processes in this response were hemoglobin digestion and protein trafficking/RBC remodeling. A high proportion of these altered proteins are related to Maurer's clefts, which play important roles in the pathology of malaria infection. We hypothesized that protection against malaria in PK-deficient RBC is associated with the RBC membrane remodeling process by the parasite, which may lead to a reduction in invasion by new parasites and malaria virulence itself. Data on the RBC proteome will contribute to confirm this hypothesis.

Abbreviations

ACTs	Artemisinin-based Combination Therapies
AFR	African
AHA	Acute Hemolytic Anemia
AI	Asymptomatic Infection
AI – INDELs	Ancestry Informative Insertion/Deletion polymorphisms
ANG	Angola
ATP	Adenosine Triphosphate
BIMCP	Bioko Island Malaria Control Project
bp	Base pairs
CA	Carbonic Anhydrases
cDNA	complementary Deoxyribonucleic Acid
CI	Confidence Intervals
DDT	Dichlorodiphenyltrichloroethane
DNA	Deoxyribonucleic Acid
EEA	European Economic Area
ESI	Electrospray Ionization
EU	European Union
FASP	Filter-Aided Sample Preparation Method
FST	Fixation Index
GMAP	Global Malaria Action Plan
GNI	Gross National Income
GO	Gene Ontology

GSH	Glutathione
G6P	Glucose-6-phosphate
G6PD	Glucose-6-phosphate Dehydrogenase
G6PDD	G6PD-Deficiency
G6PDN	G6PD-Normal
Hb	Hemoglobin
HBB	Beta Hemoglobin gene
HbS	Sickle Hemoglobin allele
HK	Hexokinase
HNSHA	Hereditary Nonspherocytic Hemolytic Anemia
HPLC	High Performance Liquid Chromatography
I	Infected
ILL	Illness Group
IPT	Intermittent Preventive Treatment
IRS	Indoor Residual Spraying
ITNs	Insecticide-Treated Nets
LC	Liquid Chromatography
LD	Linkage Disequilibrium
LLINs	Long-Lasting Insecticidal Nets
MALDI	Matrix-Assisted Laser Desorption/Ionization
mRNA	messenger Ribonucleic Acid
MDG	United Nations Millenium Development Goal
mtDNA	Mitochondrial Deoxyribonucleic Acid
MOZ	Mozambique
MS	Mass Spectrometry

NADP	Nicotinamide Adenine Dinucleotide Phosphate
NADPH	reduced form of Nicotinamide Adenine Dinucleotide Phosphate
n.d.	Not Determined
NI	No Infection/Infected
Ni-NTA	Nickel- Nitrilotriacetic Acid
OR	Odds Ratios
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PCR-RFLP	Polymerase Chain Reaction - Restriction Fragment Length Polymorphism
PEP	Phosphoenolpyruvate
PfCRT	*Plasmodium falciparum* Chloroquine Resistance Transporter
PfEMP1	*Plasmodium falciparum* Erythrocyte Membrane Protein 1
PfMDR	*Plasmodium falciparum* Multidrug Resistance Protein
PfP2	*Plasmodium falciparum* 60S ribosomal acidic protein P2
PK	Pyruvate Kinase
PKD	Pyruvate Kinase Deficiency/Deficient
PKN	Pyruvate Kinase Normal
PK-L	Pyruvate Kinase isoenzyme type L
PK-R	Pyruvate Kinase isoenzyme type R
PK-M2	Pyruvate Kinase isoenzyme type M2

PKLR	Pyruvate kinase, liver and RBC encoding gene
PNLP	Programa Nacional de Luta contra o Paludismo
PPP	Pentose Phosphate Pathway
PT-C	Portuguese healthy/control individuals
PT-PKD	Portuguese individuals with PK deficiency
PVM	Parasitophorous Vacuole Membrane
R	Ring-stage parasites
RBC	Red Blood Cell(s)
RBM	Roll Back Malaria
rDNA	ribosomal Deoxyribonucleic Acid
RDTs	Rapid Diagnostic Tests
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
ROS	Reactive Oxygen Species
S	Schizont-stage parasites
SBE	Single-base extension
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Electrophoresis
SH	Sulfhydryl
SISA	Simple Interactive Statistical Analysis software
SM	Severe Malaria
SNP	Single-Nucleotide Polymorphism
STR	Short Tandem Repeat
SSCP	Single Strand Conformational Polymorphism
TVN	Tubulovesicular Network

ToF	Time-of-Flight
UM	Uncomplicated Malaria
WHO	World Health Organization
2,3-DPG	2,3-diphosphoglycerate
6PGD	6-phosphoglyconate dehydrogenase

Amino acids

Three letter amino acid code	*Amino acid*
Ala	Alanine
Asp	Aspartic Acid
Glu	Glutamic Acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Lys	Lysine

Table of Contents

List of Figures

Chapter 1 – General Introduction

Chapter 2 - Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers

Chapter 3 - Malaria: looking for selection signatures in the human PKLR gene region

Chapter 4 – Pyruvate Kinase Deficiency in Sub-Saharan Africa: Identification of a Highly Frequent Missense Mutation (G829A; Glu277Lys) and Association with Malaria

Chapter 5 – Quantitative proteomics approach for the analysis of the human malaria parasite *Plasmodium falciparum* (trophozoite stage) and its erythrocyte host – a preliminary study

List of Tables

Chapter 1 -

General Introduction

1. Malaria

Human malaria is an infectious disease caused by five species of parasites of the genus *Plasmodium* (*Plasmodium falciparum, Plasmodium vivax, Plasmodium ovale, Plasmodium malariae* and *Plasmodium knowlesi*) and is transmitted by the bite of infected female mosquitoes of more than 30 species of the genus *Anopheles*. *Plasmodium falciparum* is the most deadly parasite species and predominates in Africa; *P. vivax* is less dangerous but more widespread, and the other three species are found much less frequently. Globally, an estimated 3.3 billion people were at risk of acquiring malaria in 2011 and the last records from 2010 revealed an estimated 219 million cases and 660 000 deaths in that year. The populations living in sub-Saharan Africa have the highest risk of get infected with *Plasmodium* and approximately 80% of cases and 90% of deaths are estimated to occur in the WHO African Region, with children less than five years of age and pregnant women most severely affected (WHO, 2012).

1.1. Global epidemiological data overview (from World Malaria Report 2012, WHO 2012)

In 2010, there were an estimated 219 million cases of malaria (range 154 - 289 million) and 660 000 deaths (range 610 000 - 971 000). Together, the Democratic Republic of the Congo and Nigeria account for over 40% of the estimated total of malaria deaths globally. In 2012, 104 countries with a worldwide distribution were endemic for malaria: 79 are classified as being in the malaria control phase, ten are in the pre-elimination phase and ten in the elimination phase. Another five countries without ongoing transmission are classified in the prevention of re-introduction phase. **Figure 1** shows categorization of countries as malaria free, controlling malaria (in malaria control phase) and eliminating malaria (including countries in pre and elimination phases) and **Fig. 2** shows categorization of countries according to whether human malaria is predominantly caused by *P. falciparum, P. vivax,* or both *P. falciparum* and *P. vivax* (the two most prevalent *Plasmodium* species worldwide). Countries in elimination phases, prevention of reintroduction and recently certified as malaria free are discriminated in supplementary **Table S1**.

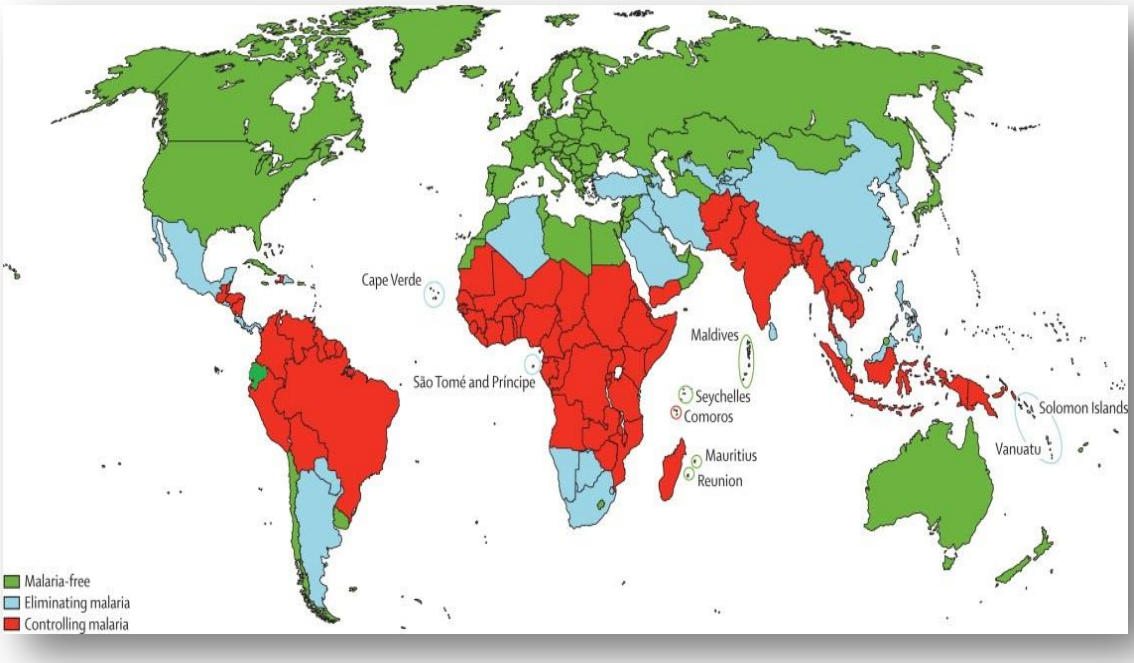


Fig. 1. World malaria distribution: categorization of countries as malaria free, eliminating malaria and controlling malaria (adapted from Feachem, et al., 2010 considering data from WHO, 2012).

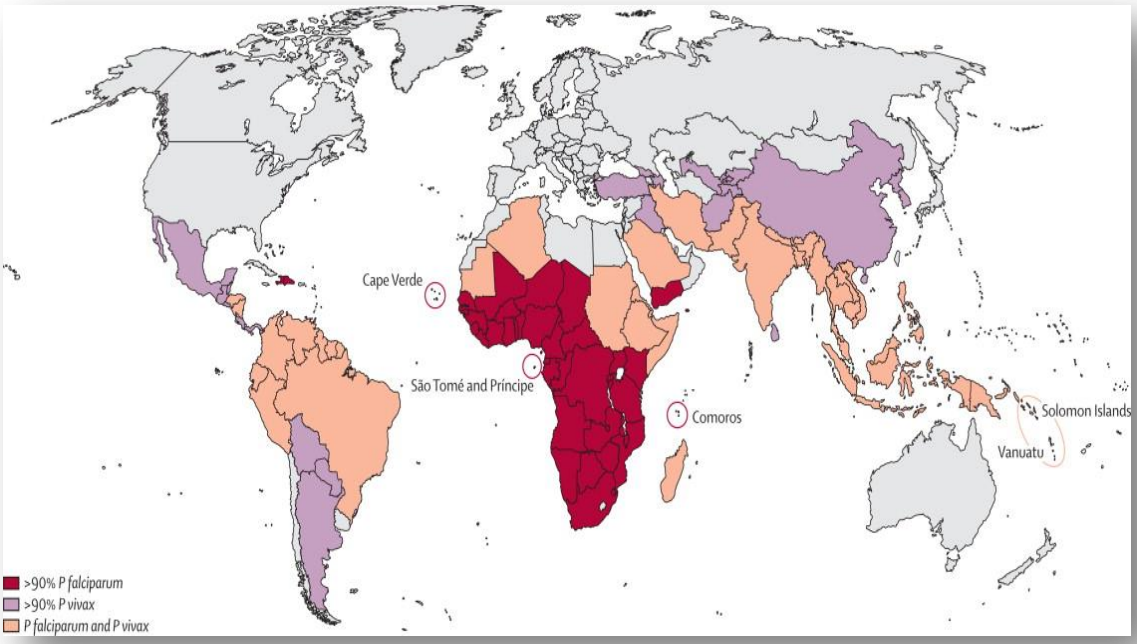


Fig. 2. World malaria distribution: categorization of countries according to whether human malaria is predominantly caused by *P. falciparum*, *P. vivax*, or both *P. falciparum* and *P. vivax* (from Feachem, et al., 2010).

Malaria is a preventable and treatable disease, since the currently recommended interventions are properly employed. These include: a) vector control through the use of insecticide-treated nets (ITNs), indoor residual spraying (IRS) and, in some specific settings, larval control; b) chemoprevention for the most vulnerable populations, particularly pregnant women and infants; c) confirmation of malaria diagnosis through microscopy or rapid diagnostic tests (RDTs) for every suspected case; and d) timely treatment with appropriate antimalarial medicines (according to the parasite species and drug resistance).

1.1.1. Vector control

By 2011, 32 countries in the African Region and 78 other countries worldwide had adopted the WHO recommendation to provide ITNs to all persons at risk for malaria. ITNs include both long-lasting insecticidal nets (LLINs) and conventional nets that are later treated with an insecticide. A total of 89 countries, including 39 in Africa, distribute ITNs free of charge. Every year, an estimated 150 million ITNs are needed to protect all populations at risk of malaria in sub-Saharan Africa. Between 2004 and 2010, the number of ITNs delivered annually by manufacturers to malaria-endemic countries in sub-Saharan Africa increased from 6 million to 145 million. The percentage of households owning at least one ITN in sub-Saharan Africa is estimated to have risen from 3% in 2000 to 53% in 2011, and remained at 53% in 2012. The proportion of the population sleeping under an ITN, representing the population directly protected, also increased from 2% in 2000 to 33% in 2011, and remained at 33% in 2012.

Indoor residual spraying remains a powerful vector control tool for reducing and interrupting malaria transmission. In 2011, 80 countries, including 38 in the African Region, recommended IRS for malaria control. In that year, 153 million people were protected by IRS worldwide, or 5% of the global population at risk. In the African Region, the proportion of the at-risk population that was protected rose from less than 5% in 2005 to 11% in 2010 and remained at that level in 2011, with 77 million people benefiting from the intervention.

Concerning larval control, WHO recommends larviciding only in settings where mosquito breeding sites are few, fixed, findable and easy to identify, map and treat. So,

in Africa, larviciding interventions are most likely to be appropriate in urban settings, and are unlikely to be cost effective in most rural settings where malaria mosquitoes breed in many small water sources.

Insecticide resistance is a major threat for vector control programmes. It has been detected in 64 countries with ongoing malaria transmission, affecting all major vectors species and all classes of insecticides. Pyrethroid resistance in Africa is one of the major reasons of concern, as this is the only class used on currently recommended LLINs. A substantial intensification of resistance monitoring is needed, using both bioassay susceptibility tests and genetic methods. Using the same insecticide for multiple successive IRS cycles is not recommended and in areas with high LLIN coverage, pyrethroids should not be used for IRS.

1.1.2. Chemoprevention

Intermittent preventive treatment (IPT) is recommended for population groups in areas of high transmission who are particularly vulnerable to *Plasmodium* infection and its consequences, particularly pregnant women and infants. In sub-Saharan Africa, an estimated 32 million pregnant women and a large portion of the estimated 28 million infants born each year would benefit from IPT. A total of 36 of 45 sub-Saharan African countries had adopted IPT for pregnant women as national policy by the end of 2011. In March 2012, WHO issued a recommendation on seasonal malaria chemoprevention for children aged 3‑59 months.

1.1.3. Diagnostic testing

Implementation of universal diagnostic testing in the public and private sectors would substantially reduce the global requirements for antimalarial treatment. In 2011, 41 of 44 countries with ongoing malaria transmission in the African Region and 46 of 55 countries in other WHO Regions reported having adopted a policy of providing parasitological diagnosis for all age groups. Malaria diagnostic testing is provided free of charge in the public sector in 84 countries around the world. The proportion of

suspected malaria cases receiving a diagnostic test in the public sector increased from 20% in 2005 to 47% in 2011 in the African Region and from 68% to 77% globally.

Most of the increase in testing in the African Region is attributable to an increase in the use of RDTs, which accounted for 40% of all cases tested in that region in 2011.

1.1.4. Treatment

Artemisinin-based combination therapies (ACTs) are recommended as the first-line treatment for malaria caused by *P. falciparum*: arthemeter plus lumefantrine, artesunate plus amodiaquine, artesunate plus mefloquine, artesunate plus sulfadoxine-pyrimethamine, or dihydroartemisinin plus piperaquine. The choice of the ACT should be based on the therapeutic efficacy in the country or area of intended use.

By 2011, 79 countries and territories had adopted ACTs as first-line treatment for *P. falciparum* malaria. *P. vivax* malaria should be treated with chloroquine where it is effective, or an appropriate ACT in areas where *P. vivax* is resistant to chloroquine. Treatment of *P. vivax* should be combined with a 14-day course of primaquine to prevent relapse. Severe malaria should be treated with injectable artesunate and followed by a complete course of an effective ACT as soon as the patient can take oral medications.

The number of ACT treatment courses delivered to the public and private sectors globally increased from 11 million in 2005 to 76 million in 2006, and reached 278 million in 2011. In the African Region in 2011, the total number of tests (both microscopy and RDTs) was less than half the number of ACTs distributed by national malaria control programmes, indicating that ACTs are given to many patients without confirmatory diagnostic testing.

1.1.5. Antimalarial drug resistance

Antimalarial drug resistance is a major public health problem which hinders the control of malaria. Resistance is occurring as a consequence of several factors,

including poor treatment policies, inadequate patient adherence to prescribed antimalarial regimens, and the widespread availability of artemisinin-based monotherapies and standard forms of the drug.

Parasite resistance to artemisinins has now been detected in four countries of the Greater Mekong subregion: Cambodia, Myanmar, Thailand and Viet Nam. Suspected artemisinin resistance is defined as an increase in parasite clearance time, as evidenced by ≥10% of cases with parasites detectable on day 3 after treatment with an ACT, whereas confirmed resistance is defined as treatment failure after treatment with an oral artemisinin-based monotherapy, with adequate antimalarial blood concentration, as evidenced by the persistence of parasites for seven days, or the presence of parasites at day 3 and recreduscence within 28-42 days. To date, neither the mechanism of artemisinin resistance, nor a molecular marker to screen for it, has been identified.

Despite the observed changes in parasite sensitivity to artemisinins, ACTs continue to cure patients provided that the partner drug is still efficacious. In Cambodia's Pailin province, resistance has been found to both components of multiple ACTs, and special provisions for directly observed therapy using a non-artemisinin-based combination (atovaquone-proguanil) have been put in place.

The World Health Organization recommends that oral artemisinin-based monotherapies should be progressively withdrawn from the market and replaced by ACTs. The number of countries which still allow the marketing of these products has decreased from 55 countries in 2008 to 16 countries in November 2012, of which nine are in the African Region.

1.1.6. Financing malaria control

The past decade has witnessed remarkable expansion in the financing and implementation of malaria control programmes. International disbursements for malaria control rose steeply from less than US$ 100 million in 2000 to US$ 1.71 billion in 2010 and were estimated to be US$ 1.66 billion in 2011 and US$ 1.84 billion in 2012. As funding has risen, international disbursements have been increasingly targeted to the African Region, to countries with the lowest gross national income (GNI) per capita,

and to countries with the highest malaria mortality rates. Domestic government funding for malaria control programmes also increased through 2005‑2011 and was estimated at US$ 625 million in 2011. While still falling short of the US$ 5.1 billion required to achieve universal coverage of malaria interventions, the financing provided for malaria control has enabled endemic countries to greatly increase access to malaria preventive interventions as well as diagnostic and treatment services.

Nevertheless, greater numbers of cases and deaths are estimated to have been averted between 2001 and 2010 in countries which had the highest malaria burdens in 2000. If the malaria incidence and mortality rates in 2000 had remained unchanged over the decade, 274 million more cases and 1.1 million more deaths would have occurred between 2001 and 2010. The majority of cases averted (52%) and lives saved (58%) are in the ten countries which had the highest estimated malaria burdens in 2000. Thus, malaria programmes have had their greatest impact where the burden is highest.

1.1.7. Malaria control and elimination

Malaria control is part of United Nations Millenium Development Goal (MDG) 6 (“Combat HIV/AIDS, malaria and other diseases”), Target 6C: “To have halted by 2015 and begun to reverse the incidence of malaria and other major diseases” (United Nations, 2012). In line with this, the Roll Back Malaria (RBM) partnership, the global coordinating body for fighting malaria, has created the Global Malaria Action Plan (GMAP) that, in 2011, has defined the following objectives: 1) Reduce global malaria deaths to near zero by end 2015; 2) Reduce global malaria cases by 75% by end 2015 (from 2000 levels); 3) Eliminate malaria by end 2015 in ten new countries (since 2008) and in the WHO European region (Roll Back Malaria, 2008).

Fifty countries are on track to reduce their malaria case incidence rates by 75%, however, these 50 countries account for only 3% (or 7 million) of the total estimated malaria cases worldwide. International targets for malaria will not be attained unless considerable progress is made in the 14 highest burden countries, which account for an estimated 80% of malaria deaths. Defeating malaria will require a high level of political commitment, strengthened regional cooperation, and the engagement of a number of sectors outside of health, including finance, education, defense, environment, mining,

industry and tourism. The fight against this disease needs to be integrated into the overall development agenda in all endemic countries.

1.2. Study areas

In this thesis, blood and DNA samples from five sub-Saharan African countries (Cape Verde, Mozambique, Angola, Republic of Equatorial Guinea and Democratic Republic of Sao Tome and Principe) and one European country (Portugal) were analyzed. A short description of the localization, geography and malaria epidemiological profile is provided below. A short overview on malaria recent cases in Europe is also presented.

1.2.1. Africa

Cape Verde (capital Praia, 14º55'15''N/23º30'30''W) is comprised of ten islands in the Atlantic Ocean, 500 km west of Senegal. Santiago is the largest island, where approximately half of the population resides. Malaria was almost eradicated between 1954 and 1970 and since 1973 autochthonous cases were only observed in this island (Alves, 1994). In Cape Verde, malaria has epidemic characteristics and is in pre-elimination phase since 2010. The incidence rate of confirmed indigenous malaria cases has decreased by 72% between 2000 and 2011. In 2011, 36 confirmed malaria cases and four deaths were recorded. The estimated percentage of population with IRS and antimalarial medicines coverage is currently 100% (WHO, 2012).

Mozambique (capital Maputo, 25^{o}57'55''S/32^{o}35'21''E) is localized in south-eastern Africa with its east coast on the Indian Ocean. Malaria is endemic throughout the country in areas where the climate favors year-long transmission, with peak transmission observed after the rainy season (from December to April). Mozambique has achieved remarkable results in malaria control in recent years: in 2006, about 6.5 million cases were described; in 2011, only 1.8 million approximately were reported (3 086 deaths). This seems to be the result of the widespread of intervention strategies: in 2011, 36% of the population was protected by IRS and 46% by ITNs and 64% of all cases received an antimalarial medicine (ACTs) (Mabunda, et al., 2008; WHO, 2012).

Angola (capital Luanda, 8°50’8’’S/13°14’4’’E), Equatorial Guinea (capital Malabo, 3°45’7’’N/8°46’2’’E) and Sao Tome and Principe (capital Sao Tome, 0°20’10’’N/6°40’53’’E) are all in the western coast of Africa, bordered by the Atlantic Ocean.

In Angola, malaria still is a great public health problem with all population at high risk of infection, being the mainly cause of morbility and mortality in the country. Due to the successive wars, malaria vector control activities and operational studies have been interrupted for decades, with a consequent lack of basic information on malaria vectors. This lack of information plus the dearth of skilled malaria entomologists have been potential impediments to the goal of scaling up the use of IRS and ITNs as a major strategy for the control of malaria (Cuamba, et al., 2006). In 2011, only 4% of population was protected with IRS and about 40% with ITNs; 73% of cases were potentially treated with antimalarial medicines (ACTs), resulting in more than 2.5 million malaria cases in all population and 6 909 deaths. Angola reported slight decreases in malaria admissions and deaths since 2007, revealing that greater efforts are still needed to combat malaria in this region (WHO, 2012).

The Republic of Equatorial Guinea is located in Middle Africa and is constituted by an insular and a mainland region. The insular region consists of the islands of Bioko and Annobón. The capital Malabo is situated at Bioko island. The risk of get infected with malaria is high in all country. The ongoing Bioko Island Malaria Control Project (BIMCP) aims at reducing malaria transmission and eliminating malaria in this island. The first five year phase of the project began in 2004 and was extended by a second five year term starting in 2009. The mosquito vector suppression activities included twice-yearly IRS of insecticides on interior walls of all inhabited dwellings and in 2007 LLINs were distributed to all households to cover all sleeping areas. The results of these concerted efforts reduced malaria prevalence from 42% to 18% in children two to five years old between 2004 and 2008 (Overgaard, et al., 2012). Since 2009, however, the efforts seem to have slowed down: considering the all country, in 2009, 65% of the population was potentially protected by ITNs and 58% by IRS; in 2011, there is no available information on IRS coverage and only 1% of the population was reported to be covered by ITNs. The percentage of cases potentially treated with antimalarial medicines is described to be 30% in 2009 but only 8% in 2011. The number of malaria

cases in all population in 2009 was about 78 983 and in 2011 was near 33 830, but the number of deaths increased from 23 in 2009 to 52 in 2011 (WHO, 2012).

The Democratic Republic of Sao Tome and Principe consists of two islands, located about 140 kilometers apart and about 250 and 225 kilometers respectively, off the north-western coast of Gabon. The climate is tropical and the rainy season runs from October to May. The prevalence of malaria in Sao Tome and Principe before the 1980's was about 19% (Ceita, 1986), but a remarkable reduction has been achieved in the last decade: the number of confirmed malaria cases fell by 87% between 2000 and 2011 and the number of malaria admissions by 84%. However, recent years have seen a higher number of cases and admissions: the number of cases reported in 2011 (6 504) is the highest since 2005 and the number of malaria admissions is the highest since 2006. A strong association between interventions and their impact on malaria morbidity and mortality is seen in Sao Tome and Principe. Reported coverage with IRS, ITNs and antimalarial is 69%, 87% and 100%, respectively. However, the recent increase in malaria admissions despite maintaining high coverage of the interventions requires further investigation (WHO, 2012).

Cape Verde and Sao Tome and Principe are both on track to achieve ≥75% decrease in case incidence by 2015, reaching the goals defined in Global Malaria Action Plan. **Table S2** summarizes the epidemiological profile, intervention strategies and antimalarial policy from these five countries, whereas **Table S3** shows the intervention coverage estimation and reported malaria cases and deaths in the same countries in 2011, both as supplemental material.

1.2.2. Europe

The confirmed case rate of malaria reported by European Union/European Economic Area (EU/EEA) countries has remained stable in the last five years, fluctuating around one per 100 000 population. Almost all cases of malaria were imported; Greece is an exception with nearly 18% of indigenous cases. The highest rates of confirmed cases were reported by the United Kingdom, Luxembourg, Ireland and Belgium. In 2010, 6 759 confirmed cases of malaria were reported by 27 EU/EEA countries (does not include cases reported in French overseas territories). In that year,

Belgium, Greece and Spain reported locally acquired cases of malaria but only ten cases were confirmed as indigenous, eight from Greece and two from Spain (ECDC, 2012). For Spain this marked the first indigenous cases of malaria due to *P. vivax* since malaria was officially eradicated (Santa-Ollala, et al., 2010). Greece reported local transmission of malaria for the third year in a row: in the summer of 2009, a cluster of *P. vivax* malaria occurred in Lakonia, and in 2010, Greece recorded another eight cases, one of which was reported from Lakonia. In 2011, another malaria outbreak affected five districts, including Lakonia (Danis, et al., 2011). The seasonality and age distribution most likely reflect travel patterns to malaria endemic countries (ECDC, 2012).

In the past, malaria was endemic in Europe, but in the 1970s it was eliminated in most parts of the EU/EEA. However, cases of indigenous transmission of malaria have occasionally been reported over the last ten years (Armengaud, et al., 2008; Zoller, et al., 2009; Santa-Ollala, et al., 2010; Danis, et al., 2011). These reports indicate that local transmission of *P. falciparum* and *P. vivax* is still possible in the EU if mosquito vectors are present. This underlines the need for surveillance, preparedness and prevention in EU/EEA countries, including improved access to healthcare for seasonal workers (ECDC, 2012).

Portugal (capital Lisbon, $39^{o}30'$N, $8^{o}00'$W) is in south-western Europe. Malaria was endemic here until 1950's, when residual dichlorodiphenyltrichloroethane (DDT) spraying was introduced and followed by extensive detection of cases of malaria and their treatment. By 1958, the transmission of the infection (which has always been much below the one recorded in Africa) was interrupted in nearly all areas of the country and eradication was confirmed by WHO in 1973 (Bruce-Chwatt, 1977). The malaria vector *Anopheles atroparvus* still exists in Portugal and the current global warming can contribute to increase its density. This event, together with the increasing people exchanges with malaria endemic countries, may raise the risk of transmission in regions where malaria has been absent, as Portugal (Lage, 2010). During 2010, Portugal recorded 50 malaria imported cases (ECDC, 2012). Vigilance must be intensified and preventive measures must be put into practice.

2. The human malaria parasite, infection and disease

The human malaria parasite has a complex, multistage life cycle involving two hosts: the human host and a mosquito vector. Parasites develop their sexual life cycle and first asexual phase in the mosquito (sporogonic cycle); in man, they complete their asexual life cycle (schizogonic cycle), which can be divided in hepatic and erythrocytic, the latter being responsible for the malaria symptoms. *Plasmodium* life cycle is shown in **Fig. 3**.

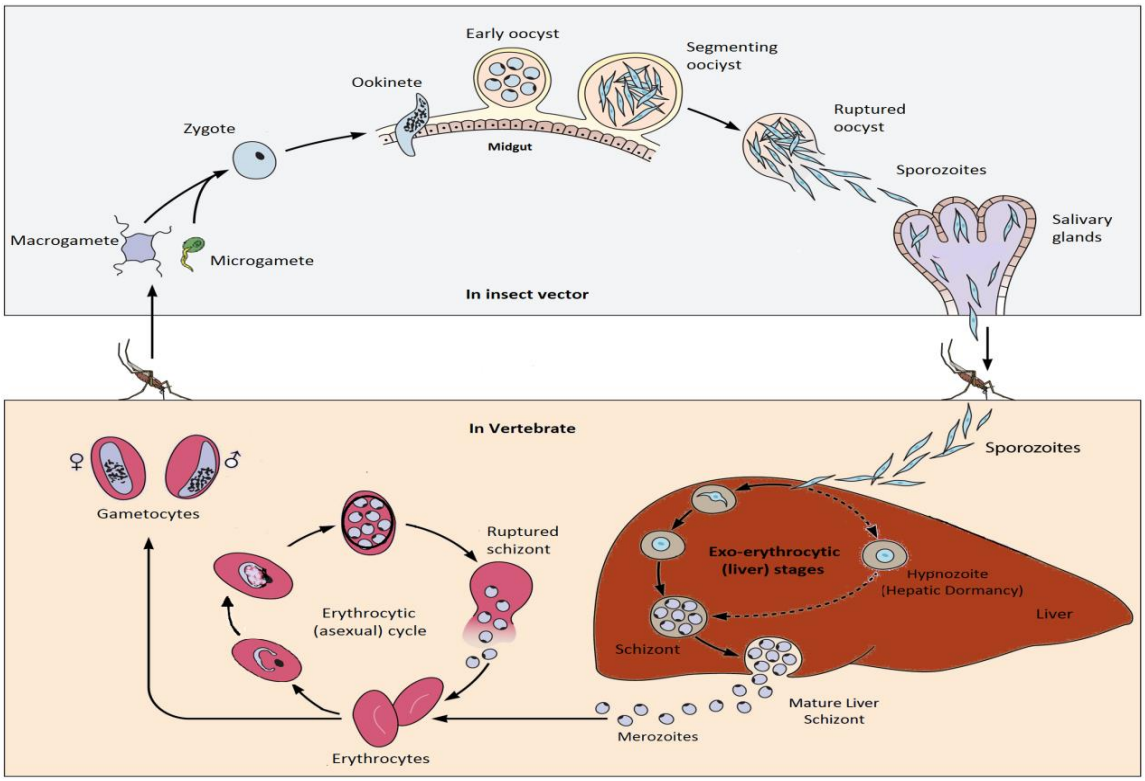


Fig. 3. *Plasmodium* life cycle (adapted from Tomé, 2013).

When a female *Anopheles* takes a blood meal in an infected person, gametocytes escape from the red blood cells (RBC) in the midgut of the mosquito to become free gametes, male and female. Then, fertilization occurs and a zygote is formed. This develops into the invasive ookinete, which bores into the stomach wall and becomes an oocyst, which grows and divides to produce thousands of invasive sporozoites. The mature cyst bursts and the free sporozoites migrate through the salivary glands. When

the mosquito feeds again, sporozoites are injected into the blood, causing malaria infection in the human host. The sporozoites that find a blood vessel, reach the liver, migrate into a few hepatocytes and then grow and divide to produce thousands of invasive merozoites. The infected liver cells burst, releasing merozoites into the blood. In *P. vivax* some sporozoites become hypnozoites, which lie dormant in liver cells, to develop months or years later and cause the illness to relapse. The occurrence of relapses indicating a dormant stage is also described in *P. ovale* but this has recently been questioned (Richter, et al., 2010). Merozoites invade RBC and become erythrocytic trophozoites. These grow originating schizonts and then divide into 8-16 new merozoites. When mature RBC bursts, merozoites are released and the cycle starts again. As the disease progresses, some merozoites develop into male or female gametocytes. These circulate but only develop further if they are taken up by a mosquito (Knell, et al., 1991).

The signs and symptoms of malaria typically begin 8–25 days following infection, however, symptoms may occur later in those who have taken antimalarial medications as prevention. Symptoms include febrile episodes with their tendency to regular periodic paroxysms (cyclical occurrence of sudden coldness followed by rigor and then fever and sweating), occurring every two days (tertian fever) in *P. vivax* and *P. ovale* infections, and every three days (quartan fever) for *P. malariae*. *Plasmodium falciparum* infection can cause recurrent fever every 36-48 hours or a less pronounced and almost continuous fever (Knell, et al., 1991; Carter and Mendis, 2002). *Plasmodium knowlesi* has an asexual cycle of about 24 hours, with an associated fever that typically occurs at the same frequency (quotidian fever) (Chin, et al., 1965; Jongwutiwes, et al., 2004; Cox-Singh, et al., 2008). Malaria also has many symptoms in common with other infectious illnesses, including body aches, headache and nausea, general weakness, and prostration. Untreated infections of malaria are characterized by enlargement of the spleen. In *P. falciparum* malaria, severe and life-threatening conditions commonly arise, characterized by dysfunction of vital organs, as the lungs, kidneys, liver, and, most notably, the brain during “cerebral malaria.” Severe anemia can also occur. These are the conditions which are associated with most of the mortality of acute malaria. Chronic infection with *P. malariae* can result in a nephrotic syndrome, and this, too, can eventually be fatal (Carter and Mendis, 2002). Human *P. knowlesi* infection has been

described to range from an asymptomatic to a rapidly fatal disease with severe hepato-renal dysfunction and acute respiratory distress syndrome (several references in Antinori, et al., 2013).

Repeated attacks of malaria due to any species of the parasites over several years severely debilitate body and mind. Cachexia, a wasting of body tissues, takes place, and splenic enlargement becomes a constant feature. Lethargic and with sunken and sallow features, spindly limbs, and hard swollen belly is the general description of the condition. In this state, the affected individual succumbs to diseases or other hardships that would scarcely threaten a person in reasonable health. Under the burden of chronic malaria, both the quality and duration of life are greatly reduced. An individual's experience of malaria at a particular time is, however, strongly governed by the type and degree of antimalarial immunity that he or she may have attained. The number of malaria inoculations experienced, and the intervals between them, are all important to the malaria immune status of an individual. Because of the time taken to achieve effective immunity to malaria under conditions of endemic infection, antimalarial immunity is often said to be “age dependent”. In the sense intended, however, it would be more accurate to say that it is “duration of exposure dependent”. There are, nevertheless, truly age-dependent aspects both to the attainment of immunity and to the pathologic responses to malaria infection. Very young children appear to have a poor capacity to acquire effective antimalarial immunity of any sort, while older children and adults may so do more readily. Infants and the very young are more prone to malaria anemia, while cerebral damage due to *P. falciparum* malaria predominates in slightly older children. Yet, other severe conditions, including renal, hepatic, and pulmonary failure, are most commonly seen in adults (Baird, et al., 1991; Baird, 1995; Carter and Mendis, 2002).

2.1. Origin and spread of human malaria parasites

The origin and evolution of *Plasmodium* parasites remains a highly debated subject, with much speculation and controversy (Liu, et al., 2010; Baron, Higgins and Dzik, 2011; Prugnolle, et al., 2011; Duval and Ariey, 2012). Malaria has probably been a human pathogen for the entire history of the species. Malaria parasites are very

remotely related to each other and their evolutionary divergence predates the origin of the hominids. Multiple switches between mammalian hosts are likely to explain the evolutionary history of human malarias (Ayala, Escalante and Rich, 1999; Joy, et al., 2003; Duval, et al., 2007; Garamszegi, 2009; Prugnolle, et al., 2011).

Early molecular phylogenetic studies showed that *P. falciparum* clustered with two avian parasites rather than with those infecting mammals, thus suggesting that *P. falciparum* was the result of a transfer from birds to humans (Waters, Higgins, and McCutchan, 1991; 1993). According to these studies, this transfer took place at the beginning of agricultural development, when the human habitat was settled about 10 000 years ago. However, this result was quickly questioned, due to the small number of ingroup taxa considered for the phylogenetic analyses and the use of 18S rDNA sequences, which have proved their weakness in studies on Haemosporidia phylogeny (Martinsen, Perkins and Schall, 2008). Subsequent analyses demonstrated that the closest sister taxon of *P. falciparum* was *P. reichenowi*, a parasite isolated from a chimpanzee. Escalante and Ayala (1994) suggested that these two parasites diverged at the time of the divergence between humans and chimpanzees. According to their results, *P. falciparum* did not directly originate from an avian malarial parasite. Nevertheless, the *P. falciparum/P. reichenowi* pair still was considered as a sister lineage of the parasites from birds and lizards.

Several other studies were performed with contradictory results and only recently the origin of *P. falciparum* seems to have been consistently established: Liu and collaborators (2010) analyzed the diversity of *Plasmodium* species in African great apes based on a very large collection of fecal samples from three subspecies of chimpanzees (*Pan troglodytes troglodytes*, *Pan troglodytes ellioti* and *Pan troglodytes schweinfurthii*), bonobos, and two subspecies of gorillas (*Gorilla gorilla gorilla* and *Gorilla gorilla graueri*), through the sequencing of mitochondrial, apicoplastic, and nuclear genes of *Plasmodium* isolates. This study confirmed the existence of a large diversity of *P. falciparum*–related parasites in gorillas but did not find any in natural populations of chimpanzees or bonobos, which suggested a likely gorilla origin for human *P. falciparum*, in opposition to all theories previously proposed. Based on these data, another study was performed indicating that *P. falciparum* probably first infected

ancestors of modern humans between 112 000 and 1 036 000 years ago (Baron, Higgins and Dzik, 2011).

Plasmodium vivax is morphologically identical to three other parasite species: *Plasmodium cynomolgi*, which infects monkeys of southern and southeastern Asia and West Pacific; *Plasmodium simium*, a parasite of the New World monkeys; and *Plasmodium schwetzi*, a parasite of chimpanzees in West and Central Africa (Carter and Mendis, 2002). In order to investigate the origin of present-day African *P. vivax*, a study was performed comparing the mitochondrial sequence diversity of parasites from Africa with those from other areas of the world. Mitochondrial genome sequencing revealed relatively little polymorphism within the African population compared to parasites from the rest of the world. This, combined with sequence similarity with parasites from India, suggested that the present day African *P. vivax* population in humans may have been introduced relatively recently from the Indian subcontinent. However, several evidences point to an African ancestral origin of this parasite (Culleton, et al., 2011).

Plasmodium malariae, in addition to infecting humans, is found in apparently indistinguishable form as a natural parasite of chimpanzees in West Africa and molecular genetic analysis has failed to distinguish *P. malariae* from *Plasmodium brasilianum* that infects New World monkeys in Central and South America (Carter and Mendis, 2002). Among the species infecting the great apes, *P. schwetzi* morphologically appears to be the closest relative to *P. ovale* (Duval and Ariey, 2012).

Plasmodium knowlesi shares a close phylogenetic relationship with *P. vivax* and morphological features that resemble those of both *P. falciparum* and *P. malariae*. Some Southeast Asian macaques species are the principal natural hosts of this parasite (Cox-Singh, 2012; Antinori, et al., 2013).

The impact of malaria is thought to have increased between 10 000 and 5 000 years ago when there were the beginnings of agriculture and consequently more human settlements. During this period, the numbers of both the human population and the mosquito vector increased, resulting in higher spread of malaria (Carter and Mendis, 2002). In adopting an agricultural way of life, human populations in sub-Saharan Africa changed from a low-density and mobile hunting and gathering life-style to communal living in settlements cleared in the tropical forest. This new, man-made environment

had two important consequences for the mosquito populations: the numbers and densities of humans began to increase under the new agricultural economy and the new life-style generated numerous small water collections close to the human habitations. Those who adopted agriculture thus transformed themselves into large, stable, and accessible sources of blood in the midst of abundant mosquito-breeding sites. The new situation provided a strong selective advantage to mosquito populations which became adapted to breed close to human habitation and to feed primarily on human blood. This led to the very high anthropophily of the vectors of African malaria and, in large part, their great vectorial efficiency (Livingstone, 1958; Colluzi, 1999). Agricultural village economies had also developed throughout the tropics and subtropics of Asia and the Middle East, however, malaria vectors have never acquired the same extraordinary preference for human blood as in Africa, probably because of the abundance of animal species in Asia whose domestication was achieved during the rise of agriculture (Carter and Mendis, 2002).

In most parts of the world, the anthropophilic index (the probability of a blood meal being on a human) of the vectors of malaria is much less than 50% and often less than 10 to 20%. By contrast, in sub-Saharan Africa, the vectors of human malaria usually have an anthropophilic index of 80 to almost 100%. This is probably the most important single factor responsible for the stability and intensity of malaria transmission in tropical Africa today (Bruce-Chwatt, Garrett-Jones and Weitz, 1966).

3. Malaria and human genetics

3.1. The imprint of malaria on the human genome

Such an ancient relationship between *Plasmodium* and the human species is expected to have profound effects on both parasite and human genomes. Infectious diseases are likely to have been major causes of mortality for much of human evolution, and, over time, changes in the environment, human demography (e.g. increasing population densities) and host-disease interactions have significantly altered the disease spectrum. Disease mortality and thus reproductive success has probably been influenced by an individual's genotype. Consequently, some aspects of modern patterns of

diversity have been determined by prehistoric diseases. The clearest examples are provided by malaria that, as above mentioned, has probably been a human pathogen for the entire history of the human species and even now affects about 220 million people each year and kills some 700 000 (Jobling, Hurles and Tyler–Smith, 2004). Malaria has actually been recognized as the strongest known force for evolutionary selection in the recent history of the human genome (Kwiatkowski, 2005) and the association between genetics and malaria susceptibility has gained a tremendous interest and relevance through the years, which is reflected by the number of papers published on the subject: since 2001, and considering only review publications, at least 15 papers are available (Craig, et al., 2001; Weatherall and Clegg, 2002; Kwiatkowski, 2005; Min-Oo and Gros, 2005; Williams, 2006a;b; Verra, Mangano and Modiano, 2009; Allison, 2009; Wellems, Hayton and Fairhurst, 2009; López, et al., 2010; Machado, et al., 2010; Hedrick, 2011; Moxon, Grau and Craig, 2011; Hedrick, 2012; Mohandas and An, 2012). Over the last decades, evidence has emerged revealing that genetic variants influence the onset, progression, severity and ultimate outcome of malaria infection in humans. The genetic component of susceptibility to malaria is complex and multigenic with a variety of genetic polymorphisms reported to influence both pathogenesis and host response to malaria. The most common and best characterized protective polymorphisms are those involving the RBC-specific structural proteins and enzymes. These polymorphisms include the variant hemoglobins, the thalassaemias, the Duffy antigen, variants of the RBC membrane and enzyme deficiencies as glucose-6-phosphate dehydrogenase (G6PD) deficiency. The alleles underlying these variants have reached very high frequencies in geographic regions where malaria is or was highly prevalent. More recently, pyruvate kinase (PK) deficiency has also been reported as protective against malaria in murine models and in studies performed *in vitro* with *P. falciparum* growing in PK-deficient RBC (Min-Oo, et al., 2003; Ayi, et al., 2008; Durand and Coetzer, 2008).

When the genetic basis of some RBC disorders was initially investigated, scientists found an unexpected paradox: the presence of high frequent deleterious mutations in some populations. Thalassemias (causing insufficient synthesis of α and β globin chains), for example, are very frequent around the shores of the Mediterranean sea, middle East, Africa and southeast Asia. Haldane, in 1949, then proposed the so

called “malaria hypothesis”, suggesting that a mutated allele reaches and maintains a high frequency, not because of an exceptionally high mutation rate, but because it is a consequence of a selective advantage against *P. falciparum* malaria, whose distribution overlaps the geographic distribution of thalassemia (Haldane, 1949).

Just a few years later, the "malaria hypothesis" was confirmed by Allison (1954), who found that the geographical distribution of the sickle-cell mutation in the beta hemoglobin gene (*HBB*) was correlated with malaria endemicity. Allison further noted that individuals who carried the sickle-cell trait (presenting only one HbS allele, causing the substitution of a glutamic acid for a valine, β6Glu>Val) were less easily parasitized than normal individuals, concluding that heterozygous carriers would have a selective advantage. Sickle-cell disease is a hereditary hemoglobin disorder caused by a mutation in both alleles of the *HBB* gene (HbSS individuals), that causes severe anemia and infections and lesions in vital organs reducing the life expectancy. Several evidences suggest the existence of an equilibrium between the elimination of the HbS allele, because of early death of homozygous individuals, and its preservation in heterozygous, due to the selective advantage against malaria. The HbS trait carriers seem, then, to be favored relatively to non-carriers and, as a consequence, HbS allele is positively selected. Globally, in Africa, the HbS allele can be found in a percentage between 5 and 40% (Weatherall and Clegg, 2001; 2002; Min-Oo and Gros, 2005).

Diseases are, by definition, disadvantageous, and genes leading to them will be selected against in the population. In the most extreme case, that of a fully-penetrant dominant disease or condition that prevents reproduction of affected individuals (e.g., because they die in childhood or are infertile), all mutations will produce affected individuals, who will then invariably fail to transmit the mutation. Therefore, all cases of the disease will be due to independent *de novo* mutations, and the incidence of the disease will equal the mutation rate. This incidence will be low, and mutations will probably occur with equal frequency in different populations, so the disease will be rare and have a relatively uniform geographical distribution. If, however, the phenotype is milder and individuals carrying the mutant allele reproduce, other factors including the strength of the selection and random genetic drift come into play, and the resulting incidence and distribution of the disease will be influenced by population processes, which include structure and history (e.g. founder events). Nonetheless, the default

expectation remains that the most disadvantageous individual mutations would not spread far, so diseases would be rare, found at similar frequencies in different populations, and originate from many different mutations. However, a few exceptional disorders are more frequent than would be expected. Factors influencing the frequency of diseases in individual populations include: mutation rate, mode of inheritance (dominant or recessive, autosomal or X-linked), selection, migration (including recent population movements), and past demography.

If susceptibility to a disease has some genetic basis, a search for the relevant gene(s) can be undertaken. Linkage analysis, haplotype analysis and association studies can be used to identify susceptible/protective alleles. However, care must be taken to determine whether any association discovered is due to true association with the disease or population structure, also referred to as population stratification (Jobling, Hurles and Tyler–Smith, 2004).

3.2. Red blood cell enzyme deficiencies and malaria

3.2.1. Glucose-6-phosphate dehydrogenase deficiency

Glucose-6-phosphate dehydrogenase (G6PD) deficiency was discovered in the 1950's when a minority of American soldiers developed acute hemolytic anemia upon exposure to antimalarial drugs (Alving, et al., 1956). It is an X-linked, hereditary genetic disorder caused by mutations in the *G6PD* gene, resulting in protein variants with different levels of enzyme activity, that are associated with a wide range of biochemical and clinical phenotypes (Cappellini and Fiorelli, 2008). It is the most common human enzymopathy, present in nearly 330 million people worldwide (Nkhoma, et al., 2009). Often, G6PD deficiency is referred to as favism, a disorder characterized by a hemolytic reaction to consumption of fava beans; however, this is misleading as not all people with G6PD deficiency will manifest a reaction to fava beans ingestion (Cappellini and Fiorelli, 2008).

3.2.1.1. Geographical distribution and prevalence of G6PD deficiency

The estimated global prevalence of G6PD deficiency is 4.9% (Nkhoma, et al., 2009). The highest prevalence is reported in Africa, southern Europe, the Middle East, Southeast Asia, and the central and southern Pacific islands; however, because of fairly recent migration, deficient alleles are nowadays quite prevalent in North and South America and in parts of northern Europe (Cappellini and Fiorelli, 2008). In Africa, the prevalence of G6PD deficiency has been reported as high as 28.1% in southwest Nigeria (May, et al., 2000), 22.5% in Congo (Bouanga, et al., 1998), 18% in Mozambique (Nieuwenhuis, et al., 1986), 15.7% in Mali (Duflo, et al., 1979), 13.0% in Uganda (Davis, et al., 2006), 9.0−15.5% in Gabon (Migot-Nabias, et al., 2000; Mombo, et al., 2003) and 10% in Angola (Miranda, 2006).

Establishing the prevalence of G6PD deficiency on a large scale has being controversial, since epidemiological studies based on enzyme activity screening have been imprecise and have not extended to global coverage and the frequency of G6PD deficiency can vary markedly, even over a small area. Moreover, X-linked disorders are usually thought to affect males only (and some studies just include males data to calculate G6PD deficiency frequencies), but in the case of G6PD deficiency, because of the high frequency of deficient alleles and the high incidence of consanguineous marriages, homozygous females have a relevant contribution to G6PD deficiency prevalence numbers. In addition, perhaps 10% of heterozygous females are also effectively G6PD-deficient due to unequal inactivation of their X-chromosomes. All these aspects contribute to an error underlying these estimations (WHO, 1989).

3.2.1.2. Function and structure of G6PD

Glucose-6-phosphate dehydrogenase catalyzes the first reaction in the pentose phosphate pathway (PPP): the oxidation of glucose-6-phosphate (G6P) to 6-phosphogluconolactone with the concomitant reduction of NADP to NADPH. The PPP is important in all cells for the production of reducing equivalents in the form of NADH (involved in protecting against toxicity of reactive oxygen species, ROS) and of pentose

sugars for the synthesis of nucleotides and nucleic acids. In RBC, the PPP has an even greater importance, since it is the only source of NADPH in these cells, as mitochondria are absent (Mason, Bautista and Gilsanz, 2007).

The amino acid sequence of G6PD has been highly conserved. Multiple sequence alignment shows amino acid sequence similarity throughout the protein but 3 highly conserved motifs. These are the peptide 198-RIDHYLGKE-206, the nucleotide–binding fingerprint, 38-GASGDLA-44 (consensus GxxGxxG/A), and the sequence 170-EKPFG-174 (consensus EKPxG) (Kotaka, et al., 2005). Biochemical evidence has shown that the 9 residue peptide is the site of G6P binding and catalysis (Camardella, et al., 1988; Lee, et al., 1992) and the nucleotide fingerprint is involved in NADP binding (Lee and Levy, 1992). The human G6PD is a tetramer; each monomer is composed of two domains and contains a single active site (Au, et al., 2000; Kotaka, et al., 2005).

3.2.1.3. Gene *G6PD* and genetics

The *G6PD* gene is localized in the q28 locus of the long arm of the X chromosome. It comprises 13 exons, spanning nearly 20 kb, encoding 515 amino acids (Mehta, Mason and Vulliamy, 2000). Females can thus be homozygous deficient or heterozygous deficient, whereas males are hemizygous deficient. Heterozygous-deficient women have a mixed population of RBC, owing to random inactivation of one of the two X chromosomes, known as lyonization. One of the RBC populations is G6PD deficient; the other has normal G6PD function (Lyon, 1961; Davidson, Nitowsky and Childs, 1963).

The G6PD locus is thought to be one of the most polymorphic loci among humans with almost 400 allelic variants reported (Beutler and Vulliamy, 2002). Most mutations underlying these variants are point mutations and small deletions that cause structural defects in the enzyme. The lack of severe mutations indicates that total G6PD deficiency is lethal. In most cases, mutations cause instability of the enzyme or altered activity, usually by decreased affinity of G6PD for its substrates, $NADP^+$ or G6P (Luzzatto, 2006). G6PD variants are classified according to their phenotypic effect: class 1, enzyme deficiency with chronic nonspherocytic hemolytic anemia; class 2, severe enzyme deficiency (<10% activity); class 3, moderate/mild enzyme deficiency

(10−60% activity); class 4, very mild or no enzyme deficiency (≥60−100% activity); class 5, increased enzyme activity. Variants from classes 2 and 3 are those that have reached appreciable gene frequencies (1-70%) in particular populations (Beutler,1996).

Different geographical areas have different sets of polymorphic variants. The Mediterranean variant (188Ser>Phe, caused by the substitution 563C>T) seems to be the most common deficient variant in the world and is widespread in the Mediterranean areas (Spain, Italy, Greece), the middle East and India (Vives-Corrons, et al., 1990; Kurdi-Haidar, et al., 1990), while the A- variant, formerly known as Betica (68Val> Met + 128Asn>Asp; caused by both 376A>G and 202 G>A) (Vulliamy, et al., 1988; Hirono and Beutler, 1988) accounts for the vast majority of G6PD deficiency in Africa. African populations also have a non-deficient variant G6PD A (126 Asn>Asp aused by 376A>G), the A- variant having arisen by a point mutation in the A allele (Beutler, 1989; Vulliamy, et al., 1991). Some polymorphic variants, as G6PD Union and G6PD Chatham have a wider distribution (Rovira, et al., 1994), while others are restricted to small populations such as tribal Indian groups (Kaeda, et al., 1995; Chalvam, et al., 2007). In China, a number of polymorphic variants are present each with a unique distribution throughout the country (Chiu, et al., 1991; Jiang, et al., 2006). The common African variant G6PD A- is usually a moderate/mild deficiency (10−15% of normal activity, hemizygous males). In contrast, the G6PD Mediterranean variant is more severe (< 1% of normal activity) (Beutler, 1996).

3.2.1.4. Clinical features of G6PD deficiency

The clinical manifestations of G6PD deficiency include neonatal jaundice, acute hemolytic anemia and chronic hemolytic anemia. Most people with a deficient G6PD allele never suffer any clinical manifestation and the sporadic variants causing chronic hemolysis are extremely rare, with a frequency of 1 in a million (Frank, 2005).

It is not clear why G6PD deficiency leads to an increased incidence of neonatal jaundice in both males and females (Weng, Chou and Lien, 2003). It seems that G6PD deficient neonates have an impaired ability to conjugate and clear bilirubin in the liver. Neonatal jaundice is more common in the more severe G6PD variants such as G6PD Mediterranean than in the milder variants such as G6PD A- (Mason, Bautista and

Gilsanz, 2007). Acute hemolytic anemia (AHA) manifests as acute episodes of intravascular hemolysis developing in a previously asymptomatic subject as a consequence of infection or the ingestion of certain drugs or fava beans (favism) (Mason, Bautista and Gilsanz, 2007). Infection is probably the most common cause of hemolysis in subjects with G6PD deficiency. Bacterial or viral infections have been reported as precipitants of AHA (Mehta, Mason and Vulliamy, 2000). The underlying mechanism is thought to relate to the release of oxidants by leukocytes during phagocytosis (Baehner, Nathan and Castle, 1971).

Divicine, isouramil, and convicine, which are thought to be the toxic constituents of fava beans, increase the activity of the PPP, promoting hemolysis in G6PD-deficient patients (Arese and de Flora, 1990), usually around 24h after the beans are eaten. Favism was noted to be present widely in Mediterranean countries, where it was originally noted, and also in the Middle East, the Far East, and North Africa, where the growth and consumption of fava beans was widespread (Kattamis, Kyriazakou and Chaidas, 1969). Favism is now widely believed to be most frequently associated with the Mediterranean variant of G6PD deficiency. Not all G6PD-deficient individuals undergo favism after ingestion of fava beans, and even the same individual can have an unpredictable response, suggesting that several factors affect the development of the disorder, including the health of the patient and the amount of fava beans ingested (Cappellini and Fiorelli, 2008).

There are several drugs that should be avoided or administered with caution in G6PD–deficient individuals due to the risk of drug-induced G6PD deficiency-related hemolysis. Primaquine is of special concern due to its use for the treatment of malaria (by the elimination of hypnozoites reservoirs of *P. vivax* and *P. ovale* and interruption of transmission since it has a potential gametocytocidal activity against the mature gametocytes of *P. falciparum*), in countries where the prevalence of G6PD deficiency is high (Beutler, et al., 2007).

Individuals who have inherited rare mutations (class 1 G6PD variants) have such a low enzyme activity that they suffer hemolytic anemia even in the absence of precipitating factors. Such variants have been described almost invariably in males within single kindred in many parts of the world. The severity of hemolysis shows great

variability with most patients presenting neonatal jaundice, often requiring exchange transfusion and splenomegaly (Beutler, Mathai and Smith, 1968).

The definitive diagnosis of G6PD deficiency is based on the estimation of enzyme activity, by quantitative spectrophotometric analysis of the rate of NADPH production from NADP. For rapid population screening, several semiquantitative methods have been applied, such as the fluorescent spot tests (Beutler, 1984). Molecular analysis is the only method by which a definitive diagnosis can be made of a female's status.

3.2.1.5. Pathophysiology of G6PD deficiency

In the RBC, the PPP is the only source of NADPH, which is essential to protect the RBC against the physiologically high levels of oxidative damage by maintaining a high level of reduced glutathione (GSH) in the cell to preserve a reducing environment. GSH protects the sulphydryl group in hemoglobin and in the RBC membrane from oxidation. In normal RBC the ratio between oxidized and reduced glutathione is 100:1. In the presence of oxidizing agents in the form of free radicals or peroxides the level of GSH drops and can be restored by the action of glutathione reductase which needs an adequate supply of NADPH. If NADPH concentrations cannot be maintained, as in G6PD deficiency, the GSH levels fall and oxidative damage occurs resulting ultimately in hemolysis (Pandolfi, et al., 1995; Mason, Bautista and Gilsanz, 2007; Stanton, 2012).

The exact mechanism whereby increased sensitivity to oxidative damage leads to hemolysis remains to be established. Most knowledge comes from favism, in which the compounds divicine and isouramil have a causal role in the irreversible oxidation of GSH and other protein-bound sulfhydryl (SH) groups, resulting in electrolyte imbalance, calcium homeostasis disorder, membrane cross-bonding and RBC phagocytosis (de Flora, et al., 1985; Turrini, et al., 1985). The recognition of deficient cells by macrophages may result from a modification of membrane carbohydrates. G6PD-deficient RBC have been shown to undergo glycoprotein modifications, which may lead to removal from circulation even in non-acute hemolysis (Horn, et al., 1995; Jain, 1998).

3.2.1.6. Glucose-6-phosphate dehydrohgenase deficiency and malaria

Several evidences have been accumulated associating G6PD deficiency to a malaria protective effect. The geographic co-distribution of G6PD deficiency and historical endemicity of malaria suggest that G6PD deficiency has risen in frequency through natural selection by malaria. This is supported by data from population and *in vitro* studies and also population genetics analyses identifying selection signatures for G6PD deficiency in the human genome. However, some of these data have been countered by other studies, meaning that this subject is controversial. Nevertheless, although some aspects remain to be elucidated, G6PD deficiency is widely accepted as protective against human malaria and provides one of the clearest examples of selection in the human genome. Concerning population studies, Ruwende and collaborators (1995), based on two large case-control studies of over 2 000 African children, showed that G6PD A- deficiency can reduce the risk of malaria infection by 46-58% in both heterozygous females and hemizygous males. In contrast, a few studies showed that only heterozygous females are protected against malaria. Bienzle and co-workers (1972), based on hospital samples, showed that infection rates in children were highest in hemizygous males and homozygous deficient females. The rates of infection were lowest in heterozygous females. Similar results based on hospital-based data were reported by others (Krutrachue, et al., 1962; Martin, et al., 1979). In this regard, it was suggested that hospital-based data may have an ascertainment bias as G6PD-deficient individuals with mild malaria are less likely to visit hospitals, as compared to G6PD-deficient individuals with severe malaria (Greene, 1993).

Then, if G6PD-deficient individuals are all protected against malaria, i.e., in selective advantage, deficient alleles would be expected to rapidly reach fixation in exposed populations (as it happened in the case of the Duffy O allele, where the near-fixation of the variant has occurred in African populations exposed to *P. vivax*). Although G6PD-deficient alleles are found at frequencies of up to 25% in some populations, these fall short of fixation, suggesting either that homozygous females are actually at disadvantage, or that the selective pressure varies over time or space (Jobling, Hurles and Tyler–Smith, 2004; Tripathy and Reddy, 2007).

Still considering the effect of X-linked inheritance but in studies performed *in vitro*, a study was carried out (Luzzatto, Usanga and Reddy, 1969) on differential parasitization of deficient and non-deficient RBC of the same individual in 20 heterozygous females. It was found that parasitization was 2-80 times greater in non-deficient that in deficient cells. Thus, both homozygous female and hemyzygous males should be protected.

Roth and co-workers (1983) cultured *P. falciparum* in blood samples from normal males and females, deficient hemizygous males and heterozygous females. Levels of parasitemia in hemizygous deficient males and heterozygous females were three times less than in normal controls and both hemizygous males and heterozygous females showed similar levels of parasitemia, suggesting that both hemizygous deficient males and heterozygous females are equally protected against malaria. In a different study, parasites growing in G6PD–deficient RBC only showed a reduction in multiplication rates when additional oxygen stress conditions were applied (Friedman, 1979).

Later, Usanga and Luzzatto (1985) described that the growth inhibition of *P. falciparum* in human G6PD-deficient RBC (both Mediterranean and A- variants) is overcome after two or three growth cycles. The parasite seems to undergo adaptive changes that gradually improve its ability to multiply in these deficient cells by producing its own G6PD enzyme (Usanga and Luzatto, 1985; Roth and Schulman, 1988). Cappadoro and coworkers (1998), contrarily, found that invasion and maturation of the parasite in both the first and second growth cycles were quantitatively indistinguishable in normal and deficient RBC (Mediterranean variant) and that G6PD mRNA was not significantly different in normal and deficient parasitized cells, claiming that preferential phagocytosis at an early stage of the schizogonic cycle is the most probable explanation for the protection conferred by this deficiency, instead of the intracellular oxidative stress itself.

A few studies have attempted to identify the signatures of selection for G6PD-deficient alleles in the human genome. Haplotype analysis of A- and Mediterranean mutations at G6PD locus indicated that they have evolved independently and have increased in frequency at a rate that is too rapid to be explained by genetic drift.

Moreover, they arose within the past 1 600 – 11 760 years, supporting the hypothesis that malaria has had a major impact on humans since the introduction of agriculture (Tishkoff, et al., 2001). A study from Verrelli and co-workers (2002) supported the previous results and found that the age of the A variant, which is also common in Africa, may not be consistent with the recent emergence of severe malaria and suggested that selection does not necessarily favor specific G6PD amino acid variants *per se* but enzyme deficiency in general is adaptive. Latter, an analysis of DNA sequence variation across the G6PD locus in humans, chimpanzees and other primates and estimates of linkage disequilibrium (LD) concluded that G6PD amino acid variants in humans have a recent increase in their frequency, whereas haplotype structure at G6PD locus in chimpanzees implies a history of several recombination events and very little overall LD. Amino acid variation is abundant in humans and our species has recently responded to malarial infection differently than our closest relative (Verreli, et al., 2006).

In a different study, it was observed that selection at *G6PD* gene has affected a region of >1.6 Mb of the human X chromosome, demonstrating that selection can have considerable effects on nucleotide variability over remarkably long genomic distances, even in African populations (Saunders, et al., 2005).

Genome wide data for haplotypes are available from projects like the International Hapmap project (http://www.hapmap.org) and, contrarily to expected, evidence for selection was found to be weak for *G6PD* (International HapMap consortium, 2005). This may be due to low single-nucleotide polymorphism (SNP) density at the Xq28 locus in the Hapmap data. Also, the tests used for detecting selection for the genome wide analysis have insufficient statistical power (Sabeti, et al., 2006).

3.2.2. Pyruvate kinase deficiency

Pyruvate kinase deficiency is an inherited metabolic disorder of the enzyme PK, which can be caused by a variety of mutations leading to lowered production, activity or stability of the enzyme. It is the most frequent enzyme abnormality of the glycolytic pathway and the second most common cause of hereditary non-spherocytic hemolytic

anemia, after G6PD deficiency (Zanella and Bianchi, 2000; Zanella, et al., 2007). The first case was detected in 1961 (Valentine, Tanaka and Miwa, 1961), and since then more than 500 affected families have been identified, but many more remain unreported in the absence of usual clinical or molecular features (Zanella and Bianchi, 2000). Pyruvate kinase deficiency is classically described as being transmitted as an autosomal recessive trait with clinical symptoms only occurring in compound heterozygotes with two mutant alleles and in homozygotes. However, inheritance as dominant trait has also been reported (Etiemble, et al., 1984).

3.2.2.1. Geographical distribution and prevalence of PK deficiency

Pyruvate kinase deficiency has a worldwide geographical distribution and it has been recognized as highly frequent in the Old Order Amish deme from Pennsylvania (Muir, et al., 1984) and Ohio (Kanno, et al., 1994) due to the high level of inbreeding in this population group. Establishing the actual prevalence of this pathology has been extremely difficult and confusing (Carey, et al., 2000) due to the methods employed. Disease prevalence estimates based on the numbers of affected patients are expected to be substantially lower than estimates based on the prevalence of heterozygotes in the population: prenatal or neonatal mortality lowers the frequency with which a disease is found in the population at large and, additionally, the errors in diagnosis are not infrequent (Beutler and Gelbart, 2000a). Pyruvate kinase deficiency has an estimated prevalence of 1:20 000 in the general white population as assessed by gene frequency studies (Beutler and Gelbart, 2000b) and 0.1%-3.12% in Asian region based in PK activity measurements (Abu-Melha, et al., 1991; Feng, Tsang and Mak, 1993; Yavarian, et al., 2008). Data from the African region was not available so far. Heterozygote frequencies are around 1-2% in most population studies, ranging from 0.2% to 6% (Fung, Keung and Chung, 1969; Beutler and Gelbart, 2000b; Yavarian, et al., 2008, Berghout, et al., 2012).

3.2.2.2. Function and structure of PK

Pyruvate kinase catalyzes the last step of glycolysis: the conversion of phosphoenolpyruvate (PEP) to pyruvate, coupled to the synthesis of one adenosine triphosphate (ATP) molecule. Glycolysis is the metabolic pathway that converts glucose into pyruvate and the free energy released in this process is used to form the high energy compounds ATP and NADPH. Pyruvate kinase plays a central role in cellular metabolism since PK is one of the major regulatory enzymes of glycolysis and the product of the reaction, pyruvate, feeds into a number of metabolic pathways (Kayne, 1973). Four PK isozymes are present in mammalian tissues (Hall and Cottam, 1978): L-type (in liver mainly) and R-type (in RBC), that are both encoded by *PKLR* gene on chromosome 1 (Satoh, et al., 1988) and under the control of two tissue-specific promoters (Noguchi, et al., 1987); and M_1-type (in skeletal muscle, heart and brain) and M_2-type (mainly in early fetal and proliferating tissues), which are encoded by the *PKM* gene on chromosome 15 (Tani, et al., 1988) and produced by alternative DNA splicing (Noguchi, et al., 1987).

The three-dimensional structure of human R-type PK has been determined (Valentini, et al., 2002), revealing the typical four-domain subunit architecture found in all PK of known three-dimensional structure. Each subunit consists of four domains: the A (residues 85-159 and 263-431) and C domains (residues 432-574), together with the small N-terminal domain (residues 57-84) form the main body of the subunit; the B domain (residues 57-84) is loosely packed to the rest of the molecule. The active site resides between A and B domains, whereas the allosteric site is located in a pocket of the C domain.

3.2.2.3. Gene *PKLR* and genetics

The *PKLR* gene is over 9.5 kb and is located in the locus q21 of chromosome 1. The cDNA is 2060 bp long and codes for 574 amino acids. The codifying region is split into 12 exons, 10 of which are common to the two isoforms, while exons 1 and 2 are specific for the RBC and the hepatic enzyme respectively (Noguchi, et al., 1987). The *PKLR* gene is highly polymorphic with more than 190 mutations described to date and several polymorphisms, most of them in non-coding regions (Zanella, et al., 2007;

Berghout, et al., 2012). Most mutations are missense (69%), splicing and stop codon (13% and 5% respectively), whereas small deletions, insertions and frameshift mutations are rare (Zanella, et al., 2007). Most mutations have only been found once, but there is a clear accumulation of some mutations with a strong ethnic and regional background. In the Eastern hemisphere, the mutation 1468T seems to be the most common (Beutler and Gelbart, 2000), whereas in the Western hemisphere, mutations 1529A and 1456T occur more frequently. The 1529A mutation seems to predominate in the USA (41.6%) (Baronciani and Beutler, 1995) and Northern European areas (41%) (Lenzner, et al., 1997). Mutation 1456T is probably the most common in Southern Europe (32% in Spain, 29% in Italy and Portugal), where, contrarily, mutation 1529A is rare (Zanella, et al., 1997; Zarza, et al., 1998; Manco, et al., 1999). The prevalence of PK deficiency in Africa is unknown but the 1456T allele was found in Afro-American individuals (Beutler and Gelbart, 2000) and 1614T allele was identified in Sao Tome and Principe (Manco, et al., 2009) at a low frequency. More recently, three additional mutations (277Glu>Lys, 295Ala>Ile and 507His>His) were identified (one allele only each) in populations from sub-Saharan regions (Berghout, et al., 2012).

3.2.2.4. Clinical features of PK deficiency

Although abnormalities in *PKLR* gene may result in alterations of both RBC and liver enzyme, clinical symptoms are confined to RBC, since the hepatic deficiency is usually compensated by the persistent enzyme synthesis in hepatocytes (Nakashima, et al., 1977). In RBC this does not happen because as enucleated cells, new protein synthesis does not occur. Clinical manifestations of PK deficiency comprise anemia of variable severity, ranging from very mild or fully compensated anemia detected only in adulthood and by chance, to life-threatening neonatal anemia and jaundice necessitating exchange transfusion and subsequent continuous transfusion therapy (Zanella, et al., 2007). Hydrops foetalis and death in the neonatal period have also been reported in rare cases (Ferreira, et al., 2000; Fermo, et al., 2005; Pissard, et al., 2006). Slight-to-moderate splenomegaly and splenectomy are also common in these patients, resulting in stabilization of hemoglobin at a slightly higher level. Hematological features also

include reticulocytosis, but this is not proportional to the severity of hemolysis (Mentzer, et al., 1971).

Since hematological features of PK deficiency are not distinctive from other hemolytic anemias, the diagnosis ultimately depends upon the determination of enzyme activity and DNA testing. Most anemic homozygotes or compound heterozygotes patients have about 5-40% of the normal level of PK activity (Zanella and Bianchi, 2000), however, in some patients, hemolytic anemia may be associated with normal or even increased enzyme activity (Lestas, Kay and Bellingham, 1987; Colombo, Zanella and Sirchia, 1988).

Patients with identical genotype may be differently affected, even within the same family. The variability of clinical expression could depend on possible individual differences in metabolomic or proteolytic activity that may diversely modulate the basic effect of the mutation, and on the ability to compensate for the enzyme deficiency by overexpressing isozymes or using alternative pathways (Zanella, et al., 2007). The compensatory persistence of PK-M_2 in mature RBC has been described in some severely affected patients (Kanno, et al., 1994; Lenzner, et al., 1997).

3.2.2.5. Pathophysiology of PK deficiency

The key abnormalities in PK deficiency are ATP depletion, although not constant, and increased content of 2,3-DPG. It is believed that ATP depletion, through the impairment of some vital ATP-dependent reactions, initiates a series of events leading to hemolysis. ATP-depleted cells lose large amounts of potassium and water, becoming dehydrated and rigid. Then, stasis, acidosis and hypoxia, by further inhibiting the glycolytic activity, contribute to the entrapment and premature destruction of the poorly deformable RBC in the microcirculation of the reticulo-endothelial system, particularly in the spleen, liver and bone marrow. However, there is no constant relationship between the metabolic impairment and the severity of hemolysis, and ATP depletion cannot explain the hemolysis in PK variants that result in a normal or increased ATP content (several references in Zanella and Bianchi, 2000). Alterations of the pattern of RBC intermediate metabolites other than ATP can contribute to hemolysis, at least in some cases. The elevated 2,3-DPG level may also contribute to the

hemolytic process by further impairing the glycolytic flux through the inhibition of hexokinase (HK) (Rijsen and Staal, 1977). The 2,3-DPG is also an inhibitor of G6PD and 6-phosphoglyconate dehydrogenase (6PGD) (Tomoda, et al., 1983), causing the impairment of PPP activity and further contributing to hemolysis.

Cell destruction appears to be brought about mostly by the phagocytosis of metabolic unable cells, the surface of which is recognized by the phagocytic cells. Several abnormalities of PK deficient RBC membranes have actually been reported: membranes from PK deficient cells are denser than normal (Allen, et al., 1983), display a more precocious than normal membrane glycoprotein self-digestion during *in vitro* incubation at 37ºC and are much more susceptible than normal to the cytotoxic activity of mouse macrophages (Zanella, et al., 1979).

3.2.2.6. Pyruvate kinase deficiency and malaria

The first report associating PK deficiency with malaria was published ten years ago by Gros and his team in a mouse model (Min-Oo, et al., 2003). In this study, it was observed that two congenic recombinant strains of mice were protected against *Plasmodium chabaudi* infection and the 269T>A mutation (90Ile>Asn) was identified in the *PKLR* gene as underlying this protection. A strong association was detected between homozygosity for 269T>A and decreased parasitemia and survival to infection. The 269T>A mutation has also been described in a human case of PK deficiency (in this case, the association with malaria was not ascertained) (van Solinge, et al., 1997). The result initially obtained by Min-Oo and collaborators (2003) was then explored by the same team, looking for the phenotypic expression of the loss-of-function 269A allele and the correlation between enzyme activity, extent of hemolytic anemia and protection against malaria, always in murine models (Min-Oo, et al., 2004; Fortier, et al., 2005; Min-Oo, et al., 2007). A second variant (338Gly>Asp) was identified with a more severe phenotype and they concluded that the degree of protection was associated with the severity of the PK deficiency. Additionally, these studies suggested that increased phagocytosis of sterile and *P. chabaudi* infected deficient RBC might decisively contribute to reduce parasitemia and increase survival to infection.

Later, the association between PK deficiency and malaria was investigated through *in vitro* experiments. Gros and collaborators (Ayi, et al., 2008) and Durand and Coetzer (2008) performed two independent studies to compare the growth of *P. falciparum* in normal and PK-deficient human RBC. A significant reduction in the invasion of RBC by parasites during three consecutive growth cycles was observed in the homozygous deficient cells. In heterozygous, no significant effect was observed. For both homozygous and heterozygous, no significant differences were detected in parasite intracellular maturation in RBC from deficient and control normal cells. Enhanced phagocytosis of ring-parasitezed RBC was also detected.

The mechanisms by which PK deficiency affects the ability of *Plasmodium* to replicate inside deficient RBC are not clarified, but may involve the following possibilities: a) greater membrane rigidity affecting parasite invasion (ATP-depleted cells lose large amounts of potassium and water, becoming dehydrated and rigid); b) altered membrane properties resulting in shortened half-lives of non-infected and infected RBC through increased phagocytosis; c) greater abundance of metabolic intermediates, such as 2,3-DPG, and of oxidative species, resulting in less hospitable intracellular environment; d) altered ratio of reticulocytosis to mature RBC in circulating blood affecting replication of *Plasmodium* species preferring mature red cells as host; and e) impairment of intra RBC parasite glucose metabolism (Roth, 1990; Zanella and Bianchi, 2000; Min-Oo, et al., 2003).

Trying to understand the molecular basis of protection conferred by PK deficiency, Gros and his group (Ayi, et al., 2009) examined the ATP levels in PK-deficient RBC and observed that there was a correlation between ATP levels and both inhibition of parasite invasion and enhancement of phagocytosis of RBC infected with ring-stage parasites. They also observed that parasites invading PK-deficient RBC respond to low intraerythrocytic ATP levels by means of a parallel increase in parasite-derived ATP via up-regulation of *P. falciparum* specific PK. Based on these results and others a model was suggested in this study for PK deficiency protection against malaria: together with the reduction in ATP production, there is an increase in 2,3-DPG in PK-deficient cells, that contribute to the maintenance of GSH in the reduced state and, as a consequence, excessive amounts of free radicals may be generated that transform oxyhemoglobin to methemoglobin and, ultimately, to hemichromes, contributing to

mechanical destabilization of the PK-deficient RBC membrane and disruption of the cell membrane cytoskeletal protein network, namely the spectrin-actin band 4.1 complex, with consequent band 3 aggregation and impairment of parasite invasion.

4. Aims and thesis structure

The role of PK deficiency in malaria protection in humans is not clear. Up to now, evidence for this protection came from murine models (a significant association was detected between PK deficiency and decreased parasitemia and survival to malaria infection) and from *in vitro* studies using PK-deficient human RBC (a significant reduction in the invasion of RBC by parasites in homozygous PK-deficient RBC and enhanced phagocytosis of ring-parasitized PK-deficient RBC were observed). Human population data is clearly missing: a high prevalent PK variant has yet to be identified in malaria endemic regions and selection signatures in the *PKLR* genome region have not been detected so far. Moreover, proteomic data on *Plasmodium* infection is very scarce: the total proteome from normal RBC infected with *Plasmodium* has not been characterized; similarly, the proteome from PK-deficient and G6PD-deficient RBC infected and non-infected with *Plasmodium* have not been studied and the proteome of the parasite itself growing in PK-deficient and G6PD-deficient RBC has not yet been investigated. These proteomic data would bring key information on infection dynamics and mechanisms underlying protection against malaria.

The main objective of this thesis was then, to investigate the association between PK deficiency and malaria in humans. In a general way, the present work intended to contribute to the knowledge of human genetic factors associated to malaria protection as well as identify the underlying protective mechanisms in order to potentially use them as targets of therapeutic intervention. It also aimed at contributing to the knowledge of RBC enzyme deficiencies overall.

The specific objectives were:

1. To investigate malaria associated genetic traits (mainly *PKLR* and *G6PD* polymorphisms) in Cape Verde that could explain the low morbidity from malaria in the archipelago.

2. To look for malaria selection signatures in the *PKLR* gene region in African populations.
3. To determine PK deficiency frequency and identify a prevalent PK variant that could be under selection by malaria in endemic African countries.
4. To assess parasite invasion and maturation of *P. falciparum* growing *in vitro* in PK and G6PD-deficient and normal RBC.
5. To analyze the proteomic profile of non-infected and infected PK and G6PD-deficient and normal RBC as well as of parasites isolated from both deficient and normal host cells.

G6PD deficiency is widely accepted as protective against human malaria and provides one of the clearest examples of selection in the human genome. So, this enzymopathy was mainly used in the present work as a control to the experiments carried out for PK deficiency.

In order to address these specific objectives, the dissertation is organized in seven chapters. Specifically, Chapter 1 corresponds to the General Introduction and it is followed by four results chapters. Chapters 2, 3 and 4 concern the objectives 1, 2 and 3, respectively, and are all published as research papers. Chapter 5 refers to objectives 4 and 5 and is presented as a paper in preparation. In Chapter 6, a General Discussion is provided, including an integrated overview of the results from previous chapters and pointing out possible future avenues of research and some key limitations to this study. Finally, Chapter 7 is a brief conclusion of the study.

References

Abu-Melha, A.M., Ahmed, M.A., Knox-Macaulay, H., Al-Sowayan, S.A. and el-Yahia A., 1991. Erythrocyte pyruvate kinase deficiency in newborns of eastern Saudi Arabia. *Acta haematologica*, 85(4), pp.192-94.

Allen, D.W., Groat, J.D., Finkel, B., Rank, B.H., Wood, P.A. and Eaton, J.W., 1983. Increased adsorption of cytoplasmic proteins to the erythrocyte membrane in ATP-depleted normal and pyruvate kinase-deficient mature cells and reticulocytes. *American journal of hematology*, 14(1), pp.11-25.

Allison, A.C., 1954. Protection afforded by sickle-cell trait against subtertian malareal infection. *British medical journal*, 1(4857), pp.290-94.

Allison, A.C., 2009. Genetic control of resistance to human malaria. *Current opinion in immunology*, 21(5), pp.499-505.

Alves, J., 1994. Programme National de Lutte contre le paludisme. Plan d'Action 1994-98. Ministère de la Santé. République du Cap-Vert.

Alving, A.S., Carson, P.E., Flanagan, C.L. and Ickes, C.E., 1956. Enzymatic deficiency in primaquine-sensitive erythrocytes. *Science*, 124(3220), pp.484-85.

Antinori, S., Galimberti, L., Milazzo, L. and Corbellino, M., 2013. *Plasmodium knowlesi*: the emerging zoonotic malaria parasite. *Acta tropica*, 125(2), pp.191-201.

Arese, P. and De Flora, A., 1990. Pathophysiology of hemolysis in glucose-6-phosphate dehydrogenase deficiency. *Seminars in hematology*, 27(1), pp.1-40.

Armengaud, A., Legros, F., D'Ortenzio, E., Quatresous, I., Barre, H., Houze, S., Valayer, P., Fanton, Y. and Schaffner, F., 2008. A case of autochthonous *Plasmodium vivax* malaria, Corsica, August 2006. *Travel medicine and infectious disease*, 6(1-2), pp.36-40.

Au, S.W., Gover, S., Lam, V.M. and Adams, M.J., 2000. Human glucose-6-phosphate dehydrogenase: the crystal structure reveals a structural NADP(+) molecule and provides insights into enzyme deficiency. *Structure*, 8(3):293-303.

Ayala, F.J., Escalante, A.A. and Rich, S.M., 1999. Evolution of *Plasmodium* and the recent origin of the world populations of *Plasmodium falciparum*. *Parassitologia*, 41(1-3), pp.55-68.

Ayi, K., Liles, W.C., Gros, P. and Kain, K.C., 2009. Adenosine triphosphate depletion of erythrocytes simulates the phenotype associated with pyruvate kinase deficiency and confers protection against *Plasmodium falciparum in vitro*. *The Journal of infectious diseases*, 200(8), pp.1289-99.

Ayi, K., Min-Oo, G., Serghides, L., Crockett, M., Kirby-Allen, M., Quirt, I., Gros, P., Kain, K.C., 2008. Pyruvate kinase deficiency and malaria. *The New England journal of medicine*, 358(17), pp.1805-10.

Baehner, R.L., Nathan, D.G. and Castle, W.B., 1971. Oxidant injury of caucasian glucose-6-phosphate dehydrogenase-deficient red blood cells by phagocytosing leukocytes during infection. *The Journal of clinical investigation*, 50(12), pp.2466-73.

Baird, J.K., 1995. Host age as a determinant of naturally acquired immunity to *Plasmodium falciparum*. *Parasitology today*, 11(3), pp.105-11.

Baird, J.K., Jones T.R., Danudirgo, E.W., Annis, B.A., Bangs, M.J., Basri, H., Purnomo., Masbar, S., 1991. Age-dependent acquired protection against Plasmodium falciparum in people having two years exposure to hyperendemic malaria. *The American journal of tropical medicine and hygiene*, 45(1), pp.65-76.

Baron, J.M., Higgins, J.M. and Dzik, W.H., 2011. A revised timeline for the origin of *Plasmodium falciparum* as a human pathogen. *Journal of molecular evolution*, 73(5-6), pp.297-304.

Baronciani, L. and Beutler, E., 1995. Molecular study of pyruvate kinase deficient patients with hereditary nonspherocytic hemolytic anemia. *The Journal of clinical investigation*, 95(4), pp.1702-709.

Berghout, J., Higgins, S., Loucoubar, C., Sakuntabhai, A., Kain, K.C. and Gros, P., 2012. Genetic diversity in human erythrocyte pyruvate kinase. *Genes and immunity*, 13(1), pp.98-102.

Beutler, E., 1984. *Red Cell Metabolism: A Manual of Biochemical Methods*. 3rd sub ed. Grune & Stratton, Philadelphia, PA.

Beutler E., 1989. Glucose-6-phosphate dehydrogenase: new perspectives. *Blood*, 73(6), pp.1397-401.

Beutler, E., 1996. G6PD: population genetics and clinical manifestations. *Blood reviews*, 10(1), pp.45-52.

Beutler, E., Duparc, S. and G6PD Deficiency Working Group., 2007. Glucose-6-phosphate dehydrogenase deficiency and antimalarial drug development. *The American journal of tropical medicine and hygiene*, 77(4), pp.779-89.

Beutler, E. and Gelbart, T., 2000a, PK deficiency prevalence and the limitations of a population-based survey. *Blood*, 96(12), pp. 4005-06.

Beutler, E. and Gelbart, T., 2000b. Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population. *Blood*, 95(11), pp.3585-88.

Beutler, E., Mathai, C.K. and Smith, J.E., 1968. Biochemical variants of glucose-6-phosphate dehydrogenase giving rise to congenital nonspherocytic hemolytic disease. *Blood*, 31(2), pp.131-50.

Beutler, E. and Vulliamy, T.J., 2002. Hematologically important mutations: glucose-6-phosphate dehydrogenase. *Blood cells, molecules & diseases*, 28(2), pp.93-103.

Bienzle, U., Ayeni, O., Lucas, A.O. and Luzzatto, L., 1972. Glucose-6-phosphate dehydrogenase and malaria. Greater resistance of females heterozygous for enzyme deficiency and of males with non-deficient variant. *Lancet*, 1(7742), pp.07-10.

Bouanga, J.C., Mouélé, R., Préhu, C., Wajcman, H., Feingold, J. and Galactéros, F., 1998. Glucose-6-phosphate dehydrogenase deficiency and homozygous sickle cell disease in Congo. *Human heredity*, 48(4), pp.192-7.

Bruce-Chwatt, L.J., 1977. Malaria eradication in Portugal. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 71(3), pp.232-40.

Bruce-Chwatt, L.J., Garrett-Jones, C. and Weitz, B., 1966. Ten years' study (1955-64) of host selection by anopheline mosquitos. *Bulletin of the World Health Organization*, 35(3), pp.405-39.

Camardella, L., Caruso, C., Rutigliano, B., Romano, M., Di Prisco, G. and Descalzi-Cancedda, F., 1988. Human erythrocyte glucose-6-phosphate dehydrogenase. Identification of a reactive lysyl residue labelled with pyridoxal 5'-phosphate. *European journal of biochemistry*, 171(3), pp.485-9.

Cappadoro, M., Giribaldi, G., O'Brien, E., Turrini, F., Mannu, F., Ulliers, D., Simula, G., Luzzatto, L. and Arese, P., 1988. Early phagocytosis of glucose-6-phosphate dehydrogenase (G6PD)-deficient erythrocytes parasitized by *Plasmodium falciparum* may explain malaria protection in G6PD deficiency. *Blood*, 92(7), pp.2527-34.

Cappellini, M.D. and Fiorelli, G., 2008. Glucose-6-phosphate dehydrogenase deficiency. *Lancet*, 371(9606), pp.64-74.

Carey, P.J., Chandler, J., Hendrick, A., Reid, M.M., Saunders, P.W., Tinegate, H., Taylor, P.R. and West, N., 2000. Prevalence of pyruvate kinase deficiency in northern European population in the north of England. Northern Region Haematologists Group. *Blood*, 96(12), pp.4005–06.

Carter, R. and Mendis, K.N., 2002. Evolutionary and historical aspects of the burden of malaria. *Clinical microbiology reviews*, 15(4), pp.564-94.

Ceita, J.G.V., 1986. Malaria in São Tomé and Príncipe, In: A. Buck. (ed) *Proceedings of the Conference on Malaria in Africa*. American Institute of Biological Sciences/USAID: Washington, DC. pp 142–155.

Chalvam, R., Mukherjee, M.B., Colah, R.B., Mohanty, D. and Ghosh, K., 2007. G6PD Namoru (208 T--> C) is the major polymorphic variant in the tribal populations in southern India. *British journal of haematology*, 136(3), pp.512-13.

Chin, W., Contacos, P.G., Coatney, R.G. and Kimbal, H.R., 1965. A naturally acquired quotidian-type malaria in man transferable to monkeys. Science 149 (3686), pp.865.

Chiu, D.T., Zuo, L., Chen, E., Chao, L., Louie, E., Lubin, B., Liu, T.Z., Du, C.S., 1991. Two commonly occurring nucleotide base substitutions in Chinese G6PD variants. *Biochemical and biophysical research communications*, 180(2), pp.988-93.

Colombo, M.B., Zanella, A. and Sirchia, G., 1988. 2,3-Diphosphoglycerate and 3-phosphoglycerate in red cell pyruvate kinase deficiency. *British journal of haematology*, 69(3), pp.423-24.

Coluzzi, M., 1999. The clay feet of the malaria giant and its African roots: hypothesis and inferences about origin, spread and control of *Plasmodium falciparum* *Parassitologia* 41(1-3), pp.277-83.

Cox-Singh, J., 2012. Zoonotic malaria: *Plasmodium knowlesi*, an emerging pathogen. *Current opinion in infectious diseases*, 25(5), pp.530-36.

Cox-Singh, J., Davis, T.M., Lee, K.S., Shamsul, S.S., Matusop, A., Ratnam, S., Rahman, H.A., Conway, D.J. and Singh, B., 2008. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 46(2), pp.165-71.

Craig, A., Hastings, I., Pain, A. and Roberts, D.J., 2001. Genetics and malaria--more questions than answers. *Trends in parasitology*, 17(2), pp.55-56.

Cuamba, N., Choi, K.S. and Townson, H., 2006. Malaria vectors in Angola: distribution of species and molecular forms of the *Anopheles gambiae* complex, their pyrethroid insecticide knockdown resistance (kdr) status and *Plasmodium falciparum* sporozoite rates. *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/5/1/2> [Accessed 19 December 2012].

Culleton, R., Coban, C., Zeyrek, F.Y., Cravo, P., Kaneko, A., Randrianarivelojosia, M., Andrianaranjaka, V., Kano, S., Farnert, A., Arez, A.P., Sharp, P.M., Carter, R. and Tanabe, K., 2011. The origins of African *Plasmodium vivax*; insights from mitochondrial genome sequencing. *Plos One*, [online] Available at: <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0029137> [Accessed 12 December 2012].

Danis, K., Baka, A., Lenglet, A., Van Bortel, W., Terzaki, I., Tseroni, M., Detsis, M., Papanikolaou, E., Balaska, A., Gewehr, S., Dougas, G., Sideroglou, T., Economopoulou, A., Vakalis, N., Tsiodras, S., Bonovas, S. and Kremastinou, J., 2011. Autochthonous *Plasmodium vivax* malaria in Greece, 2011, *Euro surveillance: bulletin européen sur les maladies transmissibles*, [online] Available at: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19993> [Accessed 17 December 2012].

Davidson, R.G., Nitowsky, H.M., Childs, B., 1963. Demonstration of two populations of cells in the human female heterozygous for glucose-6-phosphate dehydrogenase variants. *Proceedings of the National Academy of Sciences of the United States of America*, 50(3), pp.481–485.

Davis, J.C., Clark, T.D., Kemble, S.K., Talemwa, N., Njama-Meya, D., Staedke, S.G. and Dorsey, G., 2006. Longitudinal study of urban malaria in a cohort of Ugandan children: description of study site, census and recruitment, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/5/1/18> [Accessed 16 December 2012].

De Flora, A., Benatti, U., Guida, L., Forteleoni, G. and Meloni, T., 1985. Favism: disordered erythrocyte calcium homeostasis. *Blood*, 66(2), pp.294-97.

Duflo, B., Diallo, A., Toure, K. and Soula, G., 1979. Glucose-6-phosphate dehydrogenase deficiency in Mali. Epidemiology and pathological aspects. *Bulletin de la Société de pathologie exotique et de ses filiales*, 72(3), pp.258-64.

Durand, P.M. and Coetzer, T.L., 2008. Pyruvate kinase deficiency protects against malaria in humans. *Haematologica*, 93(6), pp.939-40.

Duval, L. and Ariey, F., 2012. Ape *Plasmodium* parasites as a source of human outbreaks. *Clinical microbiology and infection*, 18(6), pp.528-32.

Duval, L., Robert, V., Csorba, G., Hassanin, A., Randrianarivelojosia, M., Walston, J., Nhim, T., Goodman, S.M. and Ariey, F., 2007. Multiple host-switching of Haemosporidia parasites in bats, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/6/1/157> [Accessed 20 December 2012].

Escalante, A.A. and Ayala, F.J., 1994. Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), pp.11373-77.

Etiemble, J., Picat, C., Dhermy, D., Buc, H.A., Morin, M. and Boivin, P., 1984. Erythrocytic pyruvate kinase deficiency and hemolytic anemia inherited as a dominant trait. *American journal of hematology*, 17(3), pp.251-60.

European Centre for Disease Prevention and Control (ECDC), 2012. *Annual epidemiological report reporting on 2010 surveillance data and 2011 epidemic intelligence data*, [online] Available at: <http://www.ecdc.europa.eu/en/publications/Publications/Forms/ECDC_DispForm.aspx ?ID=767> [Accessed 20 December 2012].

Feachem, R.G., Phillips, A.A., Hwang, J., Cotter, C., Wielgosz, B., Greenwood, B.M., Sabot, O., Rodriguez, M.H., Abeyasinghe, R.R., Ghebreyesus, T.A. and Snow, R.W., 2010. Shrinking the malaria map: progress and prospects. *Lancet*. 376(9752), pp.1566-78.

Feng, C.S., Tsang, S.S. and Mak, Y.T., 1993. Prevalence of pyruvate kinase deficiency among the Chinese: determination by the quantitative assay. *American journal of hematology*, 43(4), pp.271-73.

Fermo, E., Bianchi, P., Chiarelli, L.R., Cotton, F., Vercellati, C., Writzl, K., Baker, K., Hann, I., Rodwell, R., Valentini, G. and Zanella, A., 2005. Red cell pyruvate kinase deficiency: 17 new mutations of the PK-LR gene. *British journal of haematology*, 129(6), pp.839-46.

Ferreira, P., Morais, L., Costa, R., Resende, C., Dias, C.P., Araújo, F., Costa, E., Barbot, J. and Vilarinho, A., 2000. Hydrops fetalis associated with erythrocyte pyruvate kinase deficiency. *European journal of pediatrics*, 159(7), pp.481-82.

Fortier, A., Min-Oo, G., Forbes, J., Lam-Yuk-Tseung, S. and Gros, P., 2005. Single gene effects in mouse models of host: pathogen interactions. *Journal of leukocyte biology*, 77(6):868-77.

Frank, J.E., 2005. Diagnosis and management of G6PD deficiency. *American family physician*, 72(7), pp.1277-82.

Friedman, M.J., 1979. Oxidant damage mediates variant red cell resistance to malaria. *Nature*, 280(5719), pp.245-47.

Fung, R.H., Keung, Y.K. and Chung, G.S., 1969. Screening of pyruvate kinase deficiency and G6PD deficiency in Chinese newborn in Hong Kong. *Archives of disease in childhood*, 44(235), pp.373-76.

Garamszegi, L.Z., 2009. Patterns of co-speciation and host switching in primate malaria parasites, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/8/1/110> [Accessed 20 December 2012].

Greene, L.S., 1993. G6PD Deficiency as protection against *falciparum* malaria: An Epidemiologic critique of population and experimental studies. *Yearbook of Physical Anthropology*, 36, pp.153–178.

Haldane, J.B.S., 1949. The rate of mutation of human genes. *Hereditas*, 35(S1), pp.267–273.

Hall, E.R. and Cottam, G.L., 1978. Isozymes of pyruvate kinase in vertebrates: their physical, chemical, kinetic and immunological properties. *The International journal of biochemistry*, 9(11), pp.785-93.

Hedrick, P.W., 2011. Population genetics of malaria resistance in humans. *Heredity*, 107(4), pp.283-304.

Hedrick, P.W., 2012. Resistance to malaria in humans: the impact of strong, recent selection, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/11/1/349> [Accessed 19 December 2012]

Hirono, A. and Beutler E., 1988. Molecular cloning and nucleotide sequence of cDNA for human glucose-6-phosphate dehydrogenase variant A(-). *Proceedings of the National Academy of Sciences of the United States of America*, 85(11):3951-54.

Horn, S., Bashan, N., Peleg, N. and Gopas, J., 1995. Membrane glycoprotein modifications of G6PD deficient red blood cells. *European journal of clinical investigation*, 25(1), pp. 32-38.

Jain, S.K., 1998. Glutathione and glucose-6-phosphate dehydrogenase deficiency can increase protein glycosylation. *Free radical biology & medicine*, 24(1), pp.197-201.

Jiang, W., Yu, G., Liu, P., Geng, Q., Chen, L., Lin, Q., Ren, X., Ye, W., He, Y., Guo, Y., Duan, S., Wen, J., Li, H., Qi, Y., Jiang, C., Zheng, Y., Liu, C., Si, E., Zhang, Q., Tian, Q. and Du, C., 2006. Structure and function of glucose-6-phosphate dehydrogenase-deficient variants in Chinese population. *Human genetics*, 119(5), pp.463-78.

Jobling, M.A., Hurles, M.E. and Tyler–Smith, C. 2004. *Human Evolutionary Genetics: Origins, Peoples & Disease.* New York: Garland Science.

Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Sata, T. and Kanbara, H., 2004. Naturally acquired Plasmodium knowlesi malaria in human, Thailand. *Emerging infectious diseases*, 10(12), pp.2211-13.

Joy, D.A., Feng, .X, Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., Beerli, P. and Su, X.Z., 2003. Early origin and recent expansion of *Plasmodium falciparum. Science*, 300(5617), pp.318-21.

Kaeda, J.S., Chhotray, G.P., Ranjit, M.R., Bautista, J.M., Reddy, P.H., Stevens, D., Naidu, J.M., Britt, R.P., Vulliamy, T.J., Luzzatto, L. and Mason, P.J., 1995. A new glucose-6-phosphate dehydrogenase variant, G6PD Orissa (44 Ala-->Gly), is the major polymorphic variant in tribal populations in India. *American journal of human genetics*, 57(6), pp.1335-41.

Kanno, H., Wei, D.C., Chan, L.C., Mizoguchi, H., Ando, M., Nakahata, T., Narisawa, K., Fujii, H. and Miwa, S., 1994. Hereditary hemolytic anemia caused by diverse point mutations of pyruvate kinase gene found in Japan and Hong Kong. *Blood*, 84(10), pp.3505-09.

Kattamis, C.A., Kyriazakou, M. and Chaidas, S., 1969. Favism: clinical and biochemical data. *Journal of medical genetics,* 6(1), pp.34-41.

Kayne, F.J., 1973. Pyruvate kinase, In: P.D. Boyer, ed. 1973. *The Enzymes.* New York: Academic press Inc, pp.353-382

Knell, A.J.,1991. *Malaria : a publication of the tropical programme of the Wellcome Trust.* Knell, A.J. ed. Oxford: Oxford University Press.

Kotaka, M., Gover, S., Vandeputte-Rutten, L., Au, S.W., Lam, V.M., Adams, M.J., 2005. Structural studies of glucose-6-phosphate and NADP+ binding to human glucose-6-phosphate dehydrogenase. *Acta crystallographica. Section D, Biological crystallography*. 61(Pt 5), pp.495-504.

Kruatrachue, M., Charoenlarp, P., Chongsuphajaisiddhi, T. and Harinasuta, C., 1962. Erythrocyte glucose-6-phosphate dehydrogenase and malaria in Thailand. *Lancet*, 2(7267), pp.1183-86.

Kurdi-Haidar, B., Mason, P.J., Berrebi, A., Ankra-Badu, G., al-Ali, A., Oppenheim, A. and Luzzatto, L., 1990. Origin and spread of the glucose-6-phosphate dehydrogenase variant (G6PD-Mediterranean) in the Middle East. *American journal of human genetics*, 47(6), pp.1013–1019.

Kwiatkowski, D.P., 2005. How malaria has affected the human genome and what human genetics can teach us about malaria. *American journal of human genetics*, 77(2), pp.171-92.

Lage, S., 2010. *Teremos malária em Portugal*?. [online] Ciência Hoje. Available at: http://www.cienciahoje.pt/index.php?oid=43820&op=all [Accessed 16 December, 2012].

Lee, W.T. and Levy, H.R., 1992. Lysine-21 of Leuconostoc mesenteroides glucose 6-phosphate dehydrogenase participates in substrate binding through charge-charge interaction. *Protein science: a publication of the Protein Society*, 1(3), pp.329-34.

Lenzner, C., Nürnberg, P., Jacobasch, G., Gerth, C. and Thiele, B.J., 1997. Molecular analysis of 29 pyruvate kinase-deficient patients from central Europe with hereditary hemolytic anemia. *Blood*, 89(5), pp.1793-99.

Lestas, A.N., Kay, L.A. and Bellingham, A.J., 1987. Red cell 3-phosphoglycerate level as a diagnostic aid in pyruvate kinase deficiency. *British journal of haematology*, 67(4), pp.485-88.

Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson JD, Keele BF, Ndjango JB, Sanz CM, Morgan DB, Locatelli, S., Gonder, M.K., Kranzusch, P.J., Walsh, P.D., Delaporte, E., Mpoudi-Ngole, E., Georgiev, A.V., Muller, M.N., Shaw, G.M., Peeters, M., Sharp, P.M., Rayner, J.C. and Hahn, B.H., 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature*, 467(7314), pp.420-25.

Livingstone, F.B., 1958. Anthropological implications of sickle cell gene distribution in West Africa. *American Anthropologist*, 60(3), pp.533-62

López, C., Saravia, C., Gomez, A., Hoebeke, J. and Patarroyo, M.A., 2010. Mechanisms of genetically-based resistance to malaria. *Gene*, 467(1-2), pp.1-12.

Luzzatto, L., 2006. Glucose 6-phosphate dehydrogenase deficiency: from genotype to phenotype. *Haematologica*, 91(10), pp.1303–1306.

Luzzatto, L., Usanga, F.A. and Reddy, S., 1969. Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites. *Science*, 164(3881), pp.839-42.

Lyon, M.F., 1961. Gene action in the X-chromosome of the mouse (*Mus musculus L.*). *Nature*, 190, pp.372-73.

Mabunda, S., Casimiro, S., Quinto, L. and Alonso, P., 2008. A countrywide malaria survey in Mozambique. I. Plasmodium falciparum infection in children in different epidemiological settings. *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/7/1/216> [Accessed 20 December 2012].

Machado, P., Mendes, C., do Rosário, V.E. and Arez, A.P., 2010. A contribuição dos polimorfismos humanos do eritrócito na proteção contra a malária. *Revista Pan-Amazônica de Saúde*, 1(4), pp.85-96.

Manco, L. and Ribeiro, M.L., 2009. Novel human pathological mutations. Gene symbol: PKLR. Disease: pyruvate kinase deficiency. *Human genetics*, 125(3), pp.343.

Manco, L., Ribeiro, M.L., Almeida, H., Freitas, O., Abade, A. and Tamagnini, G., 1999. PK-LR gene mutations in pyruvate kinase deficient Portuguese patients. *British journal of haematology*, 105(3), pp.591-95.

Martin, S.K., Miller, L.H., Alling, D., Okoye, V.C., Esan, G.J., Osunkoya, B.O. and Deane, M., 1979. Severe malaria and glucose-6-phosphate-dehydrogenase deficiency: a reappraisal of the malaria/G-6-P.D. hypothesis. *Lancet*, 1(8115), pp.524-26.

Martinsen, E.S., Perkins, S.L. and Schall, J.J., 2008. A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Molecular phylogenetics and evolution*, 47(1), pp.261-73.

Mason, P.J., Bautista, J.M. and Gilsanz, F., 2007. G6PD deficiency: the genotype-phenotype association. *Blood reviews*, 21(5):267-83.

May, J., Meyer, C.G., Grossterlinden, L., Ademowo, O.G., Mockenhaupt, F.P., Olumese, P.E., Falusi, A.G., Luzzatto, L. and Bienzle, U., 2000. Red cell glucose-6-phosphate dehydrogenase status and pyruvate kinase activity in a Nigerian population. *Tropical medicine & international health*, 5(2), pp.119-23.

Mehta, A., Mason, P.J. and Vulliamy, T.J., 2000. Glucose-6-phosphate dehydrogenase deficiency. *Baillière's best practice & research. Clinical haematology*, 13(1), pp.21-38.

Mentzer, W.C.Jr., Baehner, R.L., Schmidt-Schönbein, H., Robinson, S.H. and Nathan, D.G., 1971. Selective reticulocyte destruction in erythrocyte pyruvate kinase deficiency. *The Journal of clinical investigation*, 50(3), pp.688-99.

Migot-Nabias, F., Mombo, L.E., Luty, A.J., Dubois, B., Nabias, R., Bisseye, C., Millet, P., Lu, C.Y. and Deloron, P., 2000. Human genetic factors related to susceptibility to mild malaria in Gabon. *Genes and immunity*, 1(7), pp.435-41.

Min-Oo, G., Fortin, A., Tam, M.F., Gros, P., Stevenson, M.M., 2004. Phenotypic expression of pyruvate kinase deficiency and protection against malaria in a mouse model. *Genes and immunity*, 5(3), pp.168-75.

Min-Oo, G., Fortin, A., Tam, M.F., Nantel, A., Stevenson, M.M. and Gros, P., 2003. Pyruvate kinase deficiency in mice protects against malaria. *Nature genetics*, 35(4), pp.357-62.

Min-Oo, G. and Gros, P., 2005. Erythrocyte variants and the nature of their malaria protective effect. *Cellular microbiology*, 7(6), pp.753-63.

Min-Oo, G., Tam, M., Stevenson, M.M. and Gros, P., 2007. Pyruvate kinase deficiency: correlation between enzyme activity, extent of hemolytic anemia and protection against malaria in independent mouse mutants. *Blood cells, molecules & diseases*, 39(1), pp.63-69.

Miranda, J., 2006. *Estudo de determinantes genéticos de susceptibilidade/resistência em crianças com diferentes níveis clínicos de gravidade de malária (Luanda, Angola).* MSc. Instituto de Higiene e Medicina Tropical- Universidade Nova de Lisboa.

Mohandas, N. and An, X., 2012. Malaria and human red blood cells. *Medical microbiology and immunology*, 201(4), pp.593-98

Mombo, L.E., Ntoumi, F., Bisseye, C., Ossari, S., Lu, C.Y., Nagel, R.L. and Krishnamoorthy, R., 2003. Human genetic polymorphisms and asymptomatic *Plasmodium falciparum* malaria in Gabonese schoolchildren. *The American journal of tropical medicine and hygiene*, 68(2), pp.186-90.

Moxon, C.A., Grau, G.E. and Craig, A.G., 2011. Malaria: modification of the red blood cell and consequences in the human host. *British journal of haematology*, 154(6), pp.670-679.

Muir, W.A., Beutler, E. and Wasson, C., 1984. Erythrocyte pyruvate kinase deficiency in the Ohio Amish: origin and characterization of the mutant enzyme. *American journal of human genetics*, 36(3), pp.634-9.

Nakashima, K., Miwa, S., Fujii, H., Shinohara, K., Yamauchi, K., Tsuji, Y. and Yanai, M., 1977. Characterization of pyruvate kinase from the liver of a patient with aberrant erythrocyte pyruvate kinase, PK Nagasaki. *The Journal of laboratory and clinical medicine*, 90(6), pp.1012-20.

Nieuwenhuis, F., Wolf, B., Bomba, A. and De Graaf, P., 1986. Haematological study in Cabo Delgado province, Mozambique; sickle cell trait and G6PD deficiency. *Tropical and geographical medicine*, 38(2), pp.183-87.

Nkhoma, E.T., Poole, C., Vannappagari, V., Hall, S.A. and Beutler, E., 2009. The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood cells, molecules & diseases*, 42(3), pp.267-78.

Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. and Tanaka, T., 1987. The L- and R-type isozymes of rat pyruvate kinase are produced from a single gene by use of different promoters. *The Journal of biological chemistry*, 262(29), pp.14366-71.

Overgaard, H.J., Reddy, V.P., Abaga, S., Matias, A., Reddy, M.R., Kulkarni, V., Schwabe, C., Segura, L., Kleinschmidt, I. and Slotman, M.A., 2012. Malaria transmission after five

years of vector control on Bioko Island, Equatorial Guinea. *Parasites & vectors*, [online] Available at: <http://www.parasitesandvectors.com/content/5/1/253> [Accessed 19 December 2012].

Pandolfi, P.P, Sonati, F., Rivi, R., Mason, P., Grosveld, F. and Luzzatto, L., 1995. Targeted disruption of the housekeeping gene encoding glucose 6-phosphate dehydrogenase (G6PD): G6PD is dispensable for pentose synthesis but essential for defense against oxidative stress. *The EMBO Journal,* 14(21), pp. 5209–5215.

Pissard, S., Max-Audit, I., Skopinski, L., Vasson, A., Vivien, P., Bimet, C., Goossens, M., Galacteros, F. and Wajcman, H., 2006. Pyruvate kinase deficiency in France: a 3-year study reveals 27 new mutations. *British journal of haematology,* 133(6), pp.683-89.

Prugnolle, F., Durand, P., Ollomo, B., Duval, L., Ariey, F., Arnathau, C., Gonzalez, J.P., Leroy, E. and Renaud, F., 2011. A fresh look at the origin of Plasmodium falciparum, the most malignant malaria agent. *PLoS pathogens*, [online] Available at: <http://www.plos pathogens.org/article/info%3Adoi%2F10.1371%2Fjournal.ppat.1001283> [Accessed 18 December 2012].

Richter, J., Franken, G., Mehlhorn, H., Labisch, A. and Häussinger, D., 2010. What is the evidence for the existence of *Plasmodium ovale* hypnozoites?. *Parasitology research,* 107(6), pp.1285-90.

Rijksen, G. and Staal, G.E., 1977. Regulation of human erythrocyte hexokinase. The influence of glycolytic intermediates and inorganic phosphate. *Biochimica et biophysica acta*, 485(1), pp.75-86.

Roll Back Malaria, 2008. *Global malaria action plan.* [pdf] Geneva: Roll Back Malaria Partnership. Available at:<http://rbm.who.int/gmap/gmap.pdf> [Accessed 14 December 2012].

Roth, E.Jr., 1990. *Plasmodium falciparum* carbohydrate metabolism: a connection between host cell and parasite. *Blood Cells,* 16(2-3), pp.453-60.

Roth, E.F.Jr., Raventos-Suarez, C., Rinaldi, A. and Nagel, R.L., 1983. Glucose-6-phosphate dehydrogenase deficiency inhibits in vitro growth of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America,* 80(1), pp.298-99.

Roth, E.Jr. and Schulman, S., 1988. The adaptation of *Plasmodium falciparum* to oxidative stress in G6PD deficient human erythrocytes. *British journal of haematology,* 70(3), pp.363-67.

Rovira, A., Vulliamy, T.J., Pujades, A., Luzzatto, L. and Corrons, J.L., 1994. The glucose-6-phosphate dehydrogenase (G6PD) deficient variant G6PD Union (454 Arg-->Cys) has a worldwide distribution possibly due to recurrent mutation. *Human molecular genetics*, 3(5), pp.833-35.

Ruwende, C., Khoo, S.C., Snow, R.W., Yates, S.N., Kwiatkowski D., Gupta, S., Warn, P., Allsopp, C.E., Gilbert, S.C., Peschu, N., Newboldá, C.I., Greenwoodá, B.M., Marshá, K. and Hill, A.V.S., 1995. Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. *Nature,* 376(6537), pp.246-49.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D. and Lander, E.S., 2006. Positive natural selection in the human lineage. *Science*, 312(5780), pp.1614-20.

Santa-Olalla, P., Vazquez-Torres, M.C., Latorre-Fandos, E., Mairal-Claver, P., Cortina-Solano, P., Puy-Azón, A., Adiego Sancho, B., Leitmeyer, K., Lucientes-Curdi, J. and Sierra-Moros, M.J., 2010. First autochthonous malaria case due to *Plasmodium vivax* since eradication, Spain, October 2010, *Euro surveillance: bulletin européen sur les maladies transmissibles*, [online] Available at: <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19684> [Accessed 17 December 2012].

Satoh, H., Tani, K., Yoshida, M.C., Sasaki, M., Miwa, S. and Fujii, H., 1988. The human liver-type pyruvate kinase (PKL) gene is on chromosome 1 at band q21. *Cytogenetics and cell genetics, 47(3)*, pp.132-33.

Saunders, M.A., Slatkin, M., Garner, C., Hammer, M.F. and Nachman, M.W., 2005. The extent of linkage disequilibrium caused by selection on G6PD in humans. *Genetics*, 171(3), pp.1219-29.

Stanton, R.C., 2012. Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB Life*, 64(5), pp.362-69.

Tani, K., Yoshida, M.C., Satoh, H., Mitamura, K., Noguchi, T., Tanaka, T., Fujii, H. and Miwa, S., 1988. Human M2-type pyruvate kinase: cDNA cloning, chromosomal assignment and expression in hepatoma. *Gene*, 73(2), pp.509-16.

Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E.H., Williams, S.M. and Clark, A.G., 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*, 293(5529), pp.455-62.

Tomé, César., 2013. Life cycle of *Plasmodium* parasites. Mapping Ignorance. [blog] 3 January. Available at: <http://mappingignorance.org/2013/01/03/the-flick-of-a-switch-controls-the-fate-of-human-parasites/parasites_figure2/> [Accessed 26 January 2013].

Tomoda, A., Lachant, N.A., Noble, N.A. and Tanaka, K.R., 1983. Inhibition of the pentose phosphate shunt by 2,3-diphosphoglycerate in erythrocyte pyruvate kinase deficiency. *British journal of haematology*, 54(3), pp.475-84.

Tripathy, V. and Reddy, B.M., 2007. Present status of understanding on the G6PD deficiency and natural selection. *Journal of postgraduate medicine*, 53(3), pp.193-202.

Turrini, F., Naitana, A., Mannuzzu, L., Pescarmona, G. and Arese, P., 1985. Increased red cell calcium, decreased calcium adenosine triphosphatase, and altered membrane proteins during fava bean hemolysis in glucose-6-phosphate dehydrogenase-deficient (Mediterranean variant) individuals. *Blood*, 66(2), pp.302-05.

United Nations, 2012. *Millennium Development Goals Report 2012*. [pdf] New York: United Nations. Available at: < http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2012/ English2012.pdf#page=44> [Accessed 15 December 2012].

Usanga, E.A. and Luzzatto, L., 1985. Adaptation of *Plasmodium falciparum* to glucose 6-phosphate dehydrogenase-deficient host red cells by production of parasite-encoded enzyme. *Nature*, 313(6005), pp.793-5.

Valentini, G., Chiarelli, L.R., Fortin, R., Dolzan, M., Galizzi, A., Abraham, D.J., Wang, C., Bianchi, P., Zanella, A. & Mattevi, A. (2002) Structure and function of human erythrocyte pyruvate kinase. Molecular basis of nonspherocytic hemolytic anemia. *Journal of Biological Chemistry*, 277(26), pp.23807–23814.

Valentine, W.N., Tanaka, K.R. and Miwa, S., 1961. A specific erythrocyte glycolytic enzyme defect (pyruvate kinase) in three subjects with congenital non-spherocytic hemolytic anemia. *Transactions of the Association of American Physicians,* 74, pp.100-110.

Van Solinge, W.W., Kraaijenhagen, R.J., Rijksen G. and Nielsen, F.C., 1997. Novel mutations in the human red cell type pyruvate kinase gene: two promoter mutations in cis, a splice site mutation, a nonsense- and three missense mutations. *Blood,* 90(10), pp.1197-1197.

Verra, F., Mangano, V.D. and Modiano, D., 2009. Genetics of susceptibility to *Plasmodium falciparum*: from classical malaria resistance genes towards genome-wide association studies. *Parasite immunology,* 31(5), pp.234-53.

Verrelli, B.C., McDonald, J.H., Argyropoulos, G., Destro-Bisol, G., Froment, A., Drousiotou, A., Lefranc, G., Helal, A.N., Loiselet, J., Tishkoff, S.A., 2002. Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *American journal of human genetics*, 71(5), pp.1112-28.

Verrelli, B.C., Tishkoff, S.A., Stone, A.C. and Touchman, J.W., 2006. Contrasting histories of G6PD molecular evolution and malarial resistance in humans and chimpanzees. *Molecular biology and evolution,* 23(8), pp.1592-601.

Vives-Corrons, J.L., Kuhl, W., Pujades, M.A. and Beutler, E., 1990. Molecular genetics of the glucose-6-phosphate dehydrogenase (G6PD) Mediterranean variant and description of a new G6PD mutant, G6PD Andalus1361A. *American journal of human genetics*, 47(3), pp.575-79.

Vulliamy, T.J., D'Urso, M., Battistuzzi, G., Estrada, M., Foulkes, N.S., Martini, G., Calabro, V., Poggi, V., Giordano, R., Town, M., Luzzatto, L. and Persico, M.G., 1988. Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. *Proceedings of the National Academy of Sciences of the United States of America*, 85(14):5171-75.

Vulliamy, T.J., Othman, A., Town, M., Nathwani, A., Falusi, A.G., Mason, P.J. and Luzzatto, L., 1991. Polymorphic sites in the African population detected by sequence analysis of the glucose-6-phosphate dehydrogenase gene outline the evolution of the variants A and A-. *Proceedings of the National Academy of Sciences of the United States of America*, 88(19), pp.8568-71.

Waters, A.P., Higgins, D.G. and McCutchan, T.F., 1991. *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proceedings of the National Academy of Sciences of the United States of America,* 88(8), pp.3140-44.

Waters, A.P., Higgins, D.G. and McCutchan, T.F., 1993. Evolutionary relatedness of some primate models of *Plasmodium. Molecular biology and evolution*, 10(4), pp.914-23.

Weatherall, D.J. and Clegg, J.B., 2001. Inherited haemoglobin disorders: an increasing global health problem. *Bulletin of the World Health Organization*, 79(8), pp.704-12.

Weatherall, D.J. and Clegg, J.B., 2002. Genetic variability in response to infection: malaria and after. *Genes and immunity*, 3(6), pp.331-37.

Wellems, T.E., Hayton, K. and Fairhurst, R.M., 2009. The impact of malaria parasitism: from corpuscles to communities. *The Journal of clinical investigation*, 119(9), pp.2496-505.

Weng, Y.H., Chou, Y.H. and Lien, R.I., 2003. Hyperbilirubinemia in healthy neonates with glucose-6-phosphate dehydrogenase deficiency. *Early human development*, 71(2), pp.129-36.

Williams, T.N., 2006a. Human red blood cell polymorphisms and malaria. *Current opinion in microbiology*, 9(4), pp.388-94.

Williams, T.N., 2006b. Red blood cell defects and malaria. *Molecular and biochemical parasitology*, 149(2), pp.121-27.

World Health Organization (WHO) working group, 1989. Glucose-6-phosphate dehydrogenase deficiency. *Bulletin of the World Health Organization*, 67(6), pp. 601-611.

World Health Organization (WHO), 2012. *World Malaria Report.* [pdf] Geneva: WHO Press. Available at: http://www.who.int/malaria/publications/world_malaria_report_2012/en/> [Accessed 8 December 2012].

Yavarian, M., Karimi, M., Shahriary, M. and Afrasiabi, A.R., 2008. Prevalence of pyruvate kinase deficiency among the south Iranian population: quantitative assay and molecular analysis. *Blood cells, molecules & diseases*, 40(3), pp.308-11.

Zanella, A. and Bianchi, P., 2000. Red cell pyruvate kinase deficiency: from genetics to clinical manifestations. *Baillière's best practice & research. Clinical haematology*. 13(1):57-81.

Zanella, A., Bianchi, P., Baronciani, L., Zappa, M., Bredi, E., Vercellati, C., Alfinito, F., Pelissero, G. and Sirchia, G., 1997. Molecular characterization of PK-LR gene in pyruvate kinase-deficient Italian patients, *Blood*, 89(10), pp.3847-52.

Zanella, A., Brovelli, A., Mantovani, A., Izzo, C., Rebulla, P. and Balduini, C., 1979. Membrane abnormalities of pyruvate kinase deficient red cells. *British journal of haematology*, 42(1), pp.101-08.

Zanella, A., Fermo, E., Bianchi, P., Chiarelli, L.R. and Valentini, G., 2007. Pyruvate kinase deficiency: the genotype-phenotype association. *Blood reviews*, 21(4), pp.217-31.

Zarza, R., Alvarez, R., Pujades, A., Nomdedeu, B., Carrera, A., Estella, J., Remacha, A., Sánchez, J.M., Morey, M., Cortes, T., Pérez Lungmus, G., Bureo, E. and Vives Corrons, J.L., 1998. Molecular characterization of the PK-LR gene in pyruvate kinase deficient Spanish patients. Red Cell Pathology Group of the Spanish Society of Haematology (AEHH). *British journal of haematology*, 103(2), pp.377-82.

Zoller, T., Naucke, T.J., May, J., Hoffmeister, B., Flick, H., Williams, C.J., Frank, C., Bergmann, F., Suttorp, N. and Mockenhaupt, F.P., 2009. Malaria transmission in non-

endemic areas: case report, review of the literature and implications for public health management, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/8/1/71> [Accessed 20 December 2012].

Chapter 2 –

Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers

This chapter was published as a research paper:

Alves, J., Machado, P., Silva, J., Gonçalves, N., Ribeiro, L., Faustino, P., do Rosário, V.E., Manco, L., Gusmão, L., Amorim, A. and Arez, A.P., 2010. Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers. *Blood cells, molecules & diseases*, 44(1):62-8.

Blood Cells, Molecules, and Diseases 44 (2010) 62–68

Contents lists available at ScienceDirect

Blood Cells, Molecules, and Diseases

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

Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers

Joana Alves [a,b], Patrícia Machado [a], João Silva [a], Nilza Gonçalves [a], Letícia Ribeiro [c], Paula Faustino [d], Virgílio Estólio do Rosário [a], Licínio Manco [e], Leonor Gusmão [f], António Amorim [f,g], Ana Paula Arez [a,*]

[a] Centre for Malaria and Tropical Diseases, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, 100, 1349-008 Lisbon, Portugal
[b] Ministry of Health, Palácio do Governo, CP 47, Cabo Verde
[c] Hematology Department, Centro Hospitalar de Coimbra, Portugal
[d] Genetics Department, Instituto Nacional de Saúde Dr Ricardo Jorge, Lisbon, Portugal
[e] Centre of Research in Anthropology in Health/Department of Anthropology, Universidade de Coimbra, Portugal
[f] Institute of Molecular Pathology and Immunology of University of Porto (IPATIMUP), Oporto, Portugal
[g] Faculty of Sciences, Universidade do Porto, Portugal

ARTICLE INFO

Article history:
Submitted 31 July 2009
Available online 17 October 2009

(Communicated by Sir D. Weatherall, F.R.S., 17 September 2009)

Keywords:
Hemoglobin S
Glucose-6-Phophate-dehydrogenase
Pyruvate Kinase
Cabo Verde
Malaria

ABSTRACT

Malaria has occurred in the Cabo Verde archipelago with epidemic characteristics since its colonization. Nowadays, it occurs in Santiago Island alone and though prophylaxis is not recommended by the World Health Organization, studies have highlight the prospect of malaria becoming a serious public health problem as a result of the presence of antimalarial drug resistance associated with mutations in the parasite populations and underscore the need for tighter surveillance.

Despite the presumptive weak immune status of the population, severe symptoms of malaria are not observed and many people present a subclinical course of the disease. No data on the prevalence of sickle-cell trait and red cell glucose-6-phosphate dehydrogenase deficiency (two classical genetic factors associated with resistance to severe malaria) were available for the Cabo Verde archipelago and, therefore, we studied the low morbidity from malaria in relation to the particular genetic characteristics of the human host population. We also included the analysis of the pyruvate kinase deficiency associated gene, reported as putatively associated with resistance to the disease.

Allelic frequencies of the polymorphisms examined are closer to European than to African populations and no malaria selection signatures were found. No association was found between the analyzed human factors and infection but one result is of high interest: a linkage disequilibrium test revealed an association of distant loci in the PKLR gene and adjacent regions, only in non-infected individuals. This could mean a more conserved gene region selected in association to protection against the infection and/or the disease.

Introduction

According to de Meira et al. [1] epidemic malaria is known to have occurred in the Cabo Verde archipelago since the remote past . Malaria should have been introduced in the archipelago during its colonization in the XV century. Records from 1507 report that the old Portuguese sailing ships (*caravelas*) from the spice route were not allowed in Cabo Verde ports because of the fear of getting malaria [2]. In 1952, da Costa Monteiro [3] reported malaria as the most serious public health problem in the archipelago, Santiago being the most affected island.

Cabo Verde is comprised of 10 islands in the Atlantic Ocean, 500 km west of Senegal. Santiago is the largest island, where approximately half of the population resides (capital: Praia). Malaria was almost eradicated between 1954 and 1970 and since 1973 autochthonous cases are only observed in this island [4]. The World Health Organization (WHO) [5] considers there to be a limited risk of malaria between September and November. There is no recommendation for prophylaxis but recent studies highlight the prospect of malaria recurring as a serious public health problem in Cabo Verde and underscores the need for a closer and continuous surveillance. The population is considered to be non-immune or semi-immune and irregular outbreaks occur. An outbreak in 1995–1996 in the St. Catarina district was followed by parasitological and molecular analysis during 1 year [6]. Studies indicated that malaria is maintained as asymptomatic and sub-patent infections and that the majority of the circulating parasite populations harbor chloroquine-resistant mutations [7].

In the previous two studies, no complicated malaria cases were found in spite of high parasitaemias. Most of patent parasitaemias

* Corresponding author. Fax: +351 213622458.
E-mail address: aparez@ihmt.unl.pt (A.P. Arez).

1079-9796/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bcmd.2009.09.008

were above the range of 1000–10,000 parasites/μl, usually considered a cut-off level for malaria attacks [8]. However, individuals of all ages presented no more than mild symptoms such as fever, headache, nausea and general malaise. This population seemed not to develop severe symptoms of malaria despite its presumptive weak immune status and many persons exhibit a subclinical course. The low morbidity associated with malaria infections in this island may be related to factors of both parasites and host, which may control the severity of the malaria infection.

In the localized outbreak in St. Catarina district [6], we suggest that the genetically homogeneous circulating parasite population could have been a weakly virulent parasite. However, when different localities were studied [7] and *Plasmodium falciparum* heterogeneity was observed this hypothesis proved untenable. Therefore, no evidence is available regarding the contribution of parasite factors to the low morbidity observed in the island.

The establishment of clinical symptoms could be attenuated due to premunition, already described for other areas of unstable and low-level transmission of malaria [9,10,11]. Also, differences in clinical consequences of infection with *P. falciparum* as consequence of host factors have already been demonstrated [12,13] and the most common and best characterized protective polymorphisms are those involving the erythrocyte-specific proteins and enzymes, such as hemoglobin (Hb) and glucose-6-phosphate dehydrogenase (G6PD) variants.

Questioning if the observed low morbidity in Santiago Island could be a consequence of particular characteristics of the host population and since no data on the frequency of these human genetic polymorphisms are available for the Cabo Verde archipelago we studied the prevalence of HbS allele responsible for the sickle-cell trait (heterozygosity for the HbS mutation in β-globin gene, Hb β globin) and the prevalence of G6PD variants, two classical genetic factors strongly associated to resistance against human severe malaria.

Further, both may have a crucial importance in the control and management of malaria cases. Malaria can be one of the major causes of hospitalization and death in patients with sickle cell anemia and as a result, antimalarial prophylaxis is included in the standard management of patients with the disease. However, with the spread of chloroquine resistance there is an on-going debate on which drugs should now be used [14]. Concerning G6PD deficiency, the epidemic conditions of *P. falciparum* malaria justify the use of primaquine as a gametocidal drug but this drug presents potentially fatal side effects in G6PD-deficient individuals [15].

In sub-Saharan Africa, X-linked G6PD is essentially a tri-allelic polymorphism: G6PDB, the most common allele associated to normal enzymatic activity; G6PDA, which results in approximately 85% of the normal enzymatic activity and the G6PDA$^-$ deficiency allele, which implies only around 12% of normal enzymatic activity with a range of 5–25% in sub-Saharan Africa [16,17]. However, considering the history of Cabo Verde settlement and the reported high European contribution, [18] we also searched for the G6PD Mediterranean (Med) variant, the most common in countries surrounding the Mediterranean Sea [19]. This variant is associated with 3% of normal enzyme activity and usually ranges in frequency from 2% to 20% in Europe [20].

More recently, pyruvate kinase (PK) deficiency was associated with resistance to the disease in rodent models [21] and humans [22,23]. Up to now, elevated frequencies of pyruvate kinase liver and red cells (PKLR)-deficient alleles have not been recorded in areas endemic for malaria, although a systematic analysis has not been done. The information about the frequency of PK deficiency in African populations is clearly limited [24,25]. We, therefore, included its analysis in this study. The PKLR gene (1q21) encodes for either PK-L (in liver) or PK-R (in red cells), according to the use of tissue-specific promoters (leading to structural differences in the protein N-terminal region). The coding region is split into 12 exons, 10 of which are shared by the 2 isoforms, while exons 1 and 2 are specific for the erythrocyte and hepatic isozyme, respectively. About 180 mutations associated with PK-deficiency and 8 polymorphic sites have been reported in the PKLR gene [26].

Materials and methods

Study area and Isolates

Biological material–DNA samples obtained from blood–was already available for this analysis. Samples were collected in localities from different Districts of Santiago Island (Praia—south, St Catarina—west, St Cruz—east and Tarrafal—north) in 1995–1996 [6], 1998–2000 and 2003 [7]. From a total of 1056 available samples, a sub-sample of 257 unrelated individuals was used for the present study (99 individuals from Praia, 23 from St Cruz, 119 from St Catarina and 16 from Tarrafal).

Individual data such as gender, age, and malaria history were available. Further, given that each individual was well characterized for *Plasmodium*-infection (species and genotype) and clinical status (most of them asymptomatic and a few with mild symptoms), two groups were defined: 64 infected (I—presence of infection at least once during the collections period) and 188 non-infected (NI—absence of infection throughout the collection period); infection status was uncertain in 5 individuals.

For the analysis of PK polymorphisms, two additional groups were also analyzed–80 adult healthy Portuguese individuals–PT-C (DNA prepared from finger-prick blood samples collected in 2006 at Health Centre of Coruche, Portugal as described in [27]) and 21 Portuguese individuals with hereditary nonspherocytic hemolytic anemia (HNSHA) caused by PK-deficiency—PT-PKD (DNA prepared from venous blood samples). These PK-deficient individuals were previously diagnosed by PK enzyme assay and molecular genetic analysis [28,29].

The investigation was approved by the Ministry of Health of Cabo Verde and by the Ethical Committee at institutions involved in the study. Each person (or parent) was informed of the nature and aims of the study and told that participation was voluntary.

Detection of hemoglobin S allele (HbS)

The mutation at c.6 of the β globin gene was detected using an adaptation of the technique described by Waterfall and Cobb [30] and the homozygous HbSS status was confirmed by a PCR-RFLP technique (details as Supplementary Material).

Detection of glucose-6-phosphate dehydrogenase polymorphisms

Mutations in the G6PD gene were detected by a PCR-RFLP method as described in Tishkoff et al. [20] (details as Supplementary Material). The possible nine genotypes were grouped as follows: hemizygous males G6PDB and G6PDA, homozygous females G6PDBB and G6PDAA and heterozygous females G6PDBA (variants with a putative normal enzyme activity) as g6pd$^+$; hemizygous males G6PDA$^-$ and homozygous females G6PDA$^-$A$^-$ (putative deficient variants) as g6pd$^-$ and heterozygous females G6PDBA$^-$ and G6PDAA$^-$ (variants with a putative intermediate enzyme activity) as g6pd$^\pm$ [31].

Detection of pyruvate kinase polymorphisms

Analysis of PKLR gene was done by two approaches: (1) typing of polymorphic loci and searching for relevant mutations associated to PK-deficiency previously described and (2) search for new microsatellite regions—short tandem repeats (STRs) in the gene and adjacent regions. In total, a PKLR gene spanning region of 95 kb was analyzed.

Analysis of binary polymorphisms

Two mutations were investigated: 269T>A (90Ile>Asn) at exon 3, the mutation identified in mice as associated to malaria protection [21], and already described in PK-deficient individuals [32] (technical details as Supplementary Material) and 1456C>T (486Arg>Trp) at exon 11, the most common mutation responsible for PK deficiency in humans from Portugal and some Sub-Saharan regions [33,34,35]. Also, two polymorphisms were analyzed: the single nucleotide polymorphism (SNP) 1705A/C at exon 12 [35,36,] and the T10/19 repeat at intron 10 [37], common polymorphic sites in São Tomé e Príncipe [24].

Analysis of STRs

After searching for STRs in the PKLR gene (accession nr AY316591) and downstream/upstream adjacent regions (accession nr AL713999), 4 loci were chosen for analysis: 2 inside (intron 3-IVS3 and intron 11-IVS11) and 2 downstream the gene (25 kb—locus PKA and 65 kb —PKV). IVS11 was the only one already described as polymorphic [38] (see Supplementary Material for amplification conditions and analysis of PCR products).

Statistical analysis

Pearson χ^2 test was used for comparison of populations from different districts and malaria I and NI groups. Additionally, PK polymorphisms were also compared with the two Portuguese groups, PT-C and PT-PKD. Allelic frequencies and selection signatures were investigated (genetic diversity, Hardy–Weinberg equilibrium deviation and linkage disequilibrium) with Arlequin 3.11. for Windows [39]. For all tests, a significance level of 0.05 was considered.

Results

Hemoglobin polymorphisms

The β globin genotype was successfully defined for a total of 217 individuals (84%). From these, 92% were HbAA, 7% HbAS and 1% HbSS. HbS allele was only found in Praia (11% of HbAS) and St Catarina (4% of HbAS and 3% of HbSS) districts with a very low frequency (0.05).

I and NI individuals distributed similarly among HbAA and HbAS genotypes [21% I and 79% NI in the HbAA group (unknown infection status of 3 individuals) and 23% I and 77% NI in the HbAS group (unknown infection status of 1 individual)]. All 3 HbSS individuals were I.

Glucose-6-phosphate dehydrogenase polymorphisms

G6PD genotype was measured in a total of 176 (68%) individuals, 77 males and 99 females. Seventy-four percent of males presented G6PDB genotype, 25% G6PDA and 1% G6PDA$^-$; 61% of females presented G6PDBB genotype, 29% G6PDBA, 6% G6PDAA and 4% G6PDAA$^-$ (no genotypes G6PDBA$^-$ and G6PDA$^-$A$^-$ were found).

In the total population, allelic frequencies were $f(B)=0.95$, $f(A)=0.04$ and $f(A^-)=0.008$, respectively but A$^-$ allele was only found in Praia and Tarrafal districts, being much more frequent in the latter—0.019 and 0.115, respectively ($P=0.007$), which reflected the presence of 3 G6PDAA$^-$ genotypes. G6PDMed variant was not found.

Ninety-seven percent of individuals were G6PD$^+$, 2% were G6PD$^\pm$ and 1% were G6PD$^-$. Normal condition seems to be equally prevalent in both genders (99% G6PD$^+$ in males and 96% in females); 1% and 0% of G6PD$^-$ in males and females, respectively and 4% of G6PD$^\pm$ in females.

Among A$^-$ carriers, all except one G6PDAA$^-$ female were NI. I and NI individuals distributed similarly among G6PD$^+$ and G6PD$^\pm$ groups [33% I and 64% NI in the G6PD$^+$ group (unknown infection status of 5 individuals) and 25% I and 75% NI in the G6PD$^\pm$ group]. The only G6PD$^-$individual was NI.

Pyruvate kinase polymorphisms

Binary polymorphisms

The 269T>A (exon 3) and 1456C>T (exon 11) mutations were screened with success in 253 (98%) and 255 (98%) individuals respectively and mutated alleles were not found.

Polymorphisms 1705A/C (exon 12) and T10/19 (intron 10) were accomplished in a total of 200 individuals (78%). Regarding 1705A/C, 19.5% were of AA genotype, 33% CC and 47.5% AC. Allelic frequencies were $f(A)=0.43$ and $f(C)=0.57$. Regarding (T)10/19, 27% were of 10/10 genotype, 20.5% 19/19 and 52.5% 10/19. Allelic frequencies were determined to be $fT(10)=0.53$ and $fT(19)=0.47$. The analysis of possible haplotypes revealed that 1705C/(T)10 exhibited a frequency of 0.52 and 1705A/(T)19 a frequency of 0.42. The other two, 1705A/(T)10 and 1705C/(T)19, presented very low frequencies (0.01 and 0.05, respectively).

F_{ST} values were calculated for all pairs of districts and only St Catarina and St Cruz revealed significant differences ($P=0.045\pm0.02$). Concerning both 1705A/C and (T)10/19 allelic frequencies, while St Catarina follows the general trend [$f(A)=0.41$ and $f(C)=0.59$; $f(T)10=0.54$ and $f(T)19=0.46$], in St Cruz values are inverted [$f(A)=0.57$ and $f(C)=0.43$; $f(T)10=0.39$ and $f(T)19=0.61$]. Haplotype frequencies were also different in St Cruz—on the opposite to the general population, 1705A/(T)19 was the predominant haplotype with a frequency of 0.57, followed by 1705C/(T)10 with 0.39; 1705C/(T)19 was present with a frequency of 0.05 and 1705A/(T)10 was absent.

In total population, no significant differences were found between I and NI. However, when districts were compared separately, certain differences were found in St Catarina as regards locus (T)10/19—the group of I individuals showed a significantly higher heterozygosity than expected ($P=0.009$) and allelic frequencies were inverted comparing to the general trend [$fT(10)=0.48$ and $fT(19)=0.52$] in the NI. Regarding haplotypes, in the NI group, both 1705C/(T)10 and 1705A/(T)19 showed similar frequencies (0.47 and 0.45, respectively) and 1705C/(T)19 showed higher frequency than in the other groups (0.07).

STRs

The 4 STR loci in the PKLR gene and downstream adjacent region–IVS3 (intron 3), IVS11 (intron 11), PKA (25 kb downstream) and PKV (65 kb downstream) (Fig. 1)–were screened in 252 individuals (98%).

All STRs were confirmed to be polymorphic with variable number of repeats—the number of (ATT) repeats in the IVS11 locus varied between 7 and 18, the number of (AAAT) repeats in the PKA locus varied between 6 and 21 and the number of (TTTA) repeats in the PKV locus varied between 8 and 13. The IVS3 locus is the most polymorphic with 8 repeat regions and it is interrupted. The consensus sequence determined, allele classification, etc. are presented as Supplementary Material. The number of repeats in this locus varied between 27 and 43.2, which were nomenclature as alleles 1 to 26.

In the overall population of Cabo Verde, IVS3 locus presented the greatest diversity indices with the larger allele number and expected heterozygosity (Table 1). Observed heterozygosity was according to Hardy–Weinberg expected frequencies for all loci, except for IVS3, which it is significantly below the expected ($P=0.000$). All pairs of loci revealed a marked Linkage Disequilibrium (LD) ($P=0.000$), i.e. a significant LD for a ≈75 kbp spanning region (IVS3 was not considered as it was not in Hardy–Weinberg equilibrium).

When districts were compared and F_{ST} values calculated, significant values were obtained for all pairs including St Cruz (vs. Praia—0.012, $P=0.018$; vs. St Catarina—0.015, $P=0.009$; vs. Tarrafal—0.012, $P=0.045$). All the other three revealed no differences between each

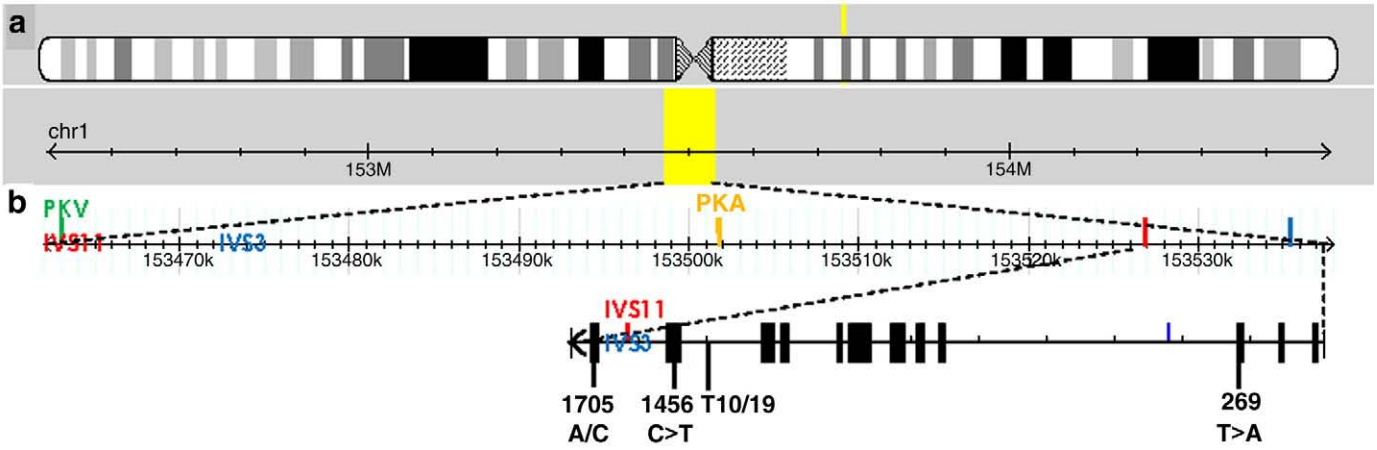


Fig. 1. The 95-kbp fragment analyzed, including PKLR gene and flanking regions. (a) Localization in chromosome 1q21; (b) localization of all mutations and polymorphisms (269T>A, 1456C>T, 1705A/C, (T)10/19, PKV, PKA, IVS11 and IVS3) genotyped in the present study.

other. No conspicuous differences seemed to exist regarding allelic frequencies or inferred haplotypes except that it is the only district when IVS3 observed heterozygosity was according to Hardy–Weinberg expected frequencies.

Regarding the studied Portuguese groups—PT-C and PT-PKD, IVS3 locus also presented the greatest diversity indices with the larger allele number and expected heterozygosity (Table 2). Observed heterozygosity was according to Hardy–Weinberg expected frequencies for all loci in the PT-C but not in PT-PKD. In this one, both IVS3 and IVS11 were significantly below the expected ($P = 0.000$ and $P = 0.002$, respectively). Again excluding IVS3 from the analysis, PT-C only showed LD for the closer loci (PKV/PKA and PKA/IVS11), while PT-PKD just had LD for PKV/IVS11 but since this latter, as IVS3, was not in Hardy–Weinberg equilibrium, we may say that no LD was observed between loci in this group.

When F_{ST} values were calculated for the two Portuguese groups, a significant value was obtained (0.025, $P = 0.009$). When those were calculated for all the studied populations pairs, significant values ($P = 0.000$) were obtained for all: CV-Total vs. PT-C—0.068 and vs. PT-PKD—0.111; CV-St Cruz vs. PT-C—0.111 and vs. PT-PKD—0.170; CV-I-St Catarina vs. PT-C—0.076 and vs. PT-PKD—0.122; CV-NI-St Catarina vs. PT-C—0.076 and vs. PT-PKD—0.124.

When groups of I and NI were analyzed separately, lower number of alleles was observed in I for all loci (IVS3: I—20, NI—25; IVS11: I—9, NI—11; PKA:I—10, NI—11) except for PKV (5 alleles in both groups) but this may be due to the smaller sample size of the I group (I—128, NI—376 alleles). Calculation of F_{ST} revealed no significant differences between the groups, both presenting the same most frequent alleles for all loci and no specific haplotypes being associated to any of them.

Yet, LD analysis revealed different results. While in the NI, as in overall population, a marked LD was observed between all pair of loci, in the I this effect was not found between the most distant loci—IVS11 and PKV. This could also be related with the smaller sample size of the I group, as it also happened in those districts with smaller sample size when were analyzed separately (St Cruz—46 alleles and Tarrafal—32 alleles). However, when I and NI from St Catarina, which have similar sample sizes (I—112 alleles and NI—118), were compared, the same was observed—a marked LD between all pair of loci in the overall population and NI alone and no linkage between IVS11 and PKV in the I. Besides, I and NI from St Catarina revealed no significant differences between them but IVS3 observed heterozygosity was according to Hardy–Weinberg expected frequencies in the I group.

Table 1
Diversity indices for the studied short tandem repeats in the Cabo Verde population.

Loci	Number of alleles	Heterozygosity		
		Observed	Expected	*P*-value
IVS3	26	0.825	0.927	0.000
IVS11	11	0.873	0.850	0.458
PKA	11	0.766	0.804	0.256
PKV	6	0.619	0.640	0.404

Discussion

The study of malaria epidemiology is crucial for control, especially in countries like Cabo Verde where mosquito vectors are in close proximity to susceptible host populations and tourists. In Cabo Verde, we are addressing the three biological entities involved in the complex malaria life-cycle doing both parasitological [6,7] and entomological studies (on-going). The present study addresses some human host genetic polymorphisms in association to malaria.

Sickle cell disease affects millions of people worldwide and it is most common among people whose ancestors come from sub-Saharan Africa, India, Saudi Arabia and Mediterranean countries. Frequencies of the heterozygous state for the sickle cell gene (HbAS) range from 2% to 38% in sub-Saharan Africa where HbS allele frequencies frequently exceed 25% [14,16,40]. Sickle-cell trait is the best described host-specific factor shown to confer strong protection against *P. falciparum* in numerous studies over the course of the last 50 years [41,42,43,44].

Deficient G6PD alleles are distributed worldwide with a global prevalence of deficiency of 4.9% and an estimate of nearly 330 million people carrying a deficiency-associated mutation in the G6PD gene

Table 2
Diversity indices for the studied short tandem repeats in the Portuguese groups.

Loci	PT-C				PT-PKD			
	Number of alleles	Heterozygosity			Number of alleles	Heterozygosity		
		Obs	Exp	*P*		Obs	Exp	*P*
IVS3	19	0.913	0.906	0.389	11	0.524	0.792	0.000
IVS11	9	0.738	0.682	0.636	5	0.476	0.708	0.002
PKA	8	0.488	0.512	0.162	4	0.143	0.139	1.000
PKV	5	0.588	0.601	0.697	4	0.476	0.580	0.294

PT-C: Portuguese healthy individuals; PT-PKD: Portuguese individuals with hereditary nonspherocytic hemolytic anemia (HNSHA) caused by PK-deficiency; Obs: Observed Heterozygosity; Exp: Expected Heterozygosity; *P*: *P*-value.

[45]. The highest prevalence is reported in Africa, southern Europe, the Middle East, Southeast Asia, and the central and southern Pacific islands; however, because of fairly recent migration, deficient alleles are nowadays quite prevalent in North and South America and in parts of northern Europe [19]. In most areas of high prevalence of G6PD deficiency, several polymorphic alleles are found but tropical regions of Africa are one exception, where the variant G6PD A^- accounts for about 90% of G6PD deficiency with frequencies of 5–25% [16,17]. The coincident worldwide distribution of malaria and mutated G6PD alleles made "The malaria/G6PD hypothesis" generally accepted [46]. Further evidence of protection against severe *P falciparum* malaria comes from both epidemiological studies [47] as well as from *in vitro* work [48,49].

PK deficiency along with G6PD deficiency, are the two most frequent enzyme disorders causing chronic hemolytic anemia worldwide. In families with no consanguinity, PK-deficient individuals are usually compound heterozygotes and prevalence of heterozygous individuals is estimated to be 1–2% in most studies [50]. The highest frequencies of the PK deficiency associated alleles are found in Europe and Asia with a prevalence ranging from 1% to 3.6% [26,33]. As these regions were historically endemic for malaria, it could have been responsible for maintaining this frequency or the ~180 mutations resulting in PK-deficiency are simply the product of random variation or other population genetic phenomena. However, in Africa, although the prevalence of PK deficiency is not known, the perception exists that it is rare, which may reflect a lack of testing [23]. If the marked *in vitro* protective effect of homozygosity for PK deficiency against malaria translates into the field (further supported by the murine model data), the argument that malaria has maintained the polymorphic frequency of the abnormal alleles may be plausible. In addition, the large number of PKLR mutations *per se* also suggests that these have been maintained by a selective force [23].

In the present study of the β globin chain of Hb, 6% of HbAS individuals and a frequency 0.05 of HbS allele are low values for a sub-Saharan region. Also G6PD deficiency associated mutations occurs in a very low frequency in this population (0.6%). Concerning PKLR polymorphisms, frequencies of alleles or haplotypes also differ from those described for African populations. Allelic frequencies of polymorphism 1705A/C [f(A) = 43%, f(C) = 57%] are closer to the European populations [f(A) ~ 29%, f(C) ~ 71%] than to Saharawi population from North Africa [f(A) ~ 62%, f(C) ~ 38%] or sub-Saharan populations [f(A) ~ 67%, f(C) ~ 33%] [25]. Allelic frequencies of the repeat (T)10/19 [f(10) = 53%, f(C) = 47%] are closer to the Portuguese population [f(10) ~ 78%, f(19) ~ 22%] than to São Tomé e Príncipe (Gulf of Guinea, West Africa) [f(10) ~ 36%, f(19) ~ 64%] [24]. Allelic frequencies of all these polymorphisms seem always be closer to the European, particularly to the Portuguese populations. The most frequent haplotypes 1705C/(T)10 and 1705A/(T)19, were the only two observed in Portugal and Central Europe [37]. However, the other two, 1705A/(T)10 and 1705C/(T)19 also occurred in low frequencies. As in São Tomé e Príncipe [24], there is a strong but not total association for combinations among these two biallelic systems.

Such low frequencies of traits HbS and G6PDMED are somehow unexpected. It could be due to the already well known importance of Caucasian admixture in the population of Cabo Verde [18] but these traits are quite prevalent in the Mediterranean region, an endemic region for malaria in the past. Further, Santiago Island should have had less contribution from Caucasians as demonstrated before in previous studies with mtDNA [51], Y-chromosome lineages [52] and autosomal STR [53].

Nevertheless, particular settlements with a strong African contribution to the genetic composition of the population seemed to persist as it may be the case of St Cruz. This district located in the east coast of the island showed F_{ST} values significantly different with all other studied populations both considering loci 1705A/C and (T)10/19 or STRs analysis. Moreover, allelic frequencies of the first two loci [f(A) = 57%, f(C) = 43%; fT(10) = 39%, fT(19) = 61%] were closer to the Saharawi population from North Africa and São Tomé e Príncipe (see above). Although we do not have historical reports about St Cruz or its capital Pedra Badejo (former Port of São Tiago), it is commonly said that the escaping slaves (Cabo Verde became an important provisioning station for slaves headed for the Americas) used to hide in this area, from where they could escape to the Island of Maio. This could justify such a stronger African contribution for the genetic background of this population but this should be further analyzed with more balanced sample sizes.

In the present study, no malaria related clinical data were available but regarding the infection status no association seems to occur with either the Hb β globin or the G6PD genotype. Also no haplotype or polymorphism of PKLR gene was associated to infected or non-infected individuals. Nevertheless some striking results related with PKLR analysis deserve a special remark. A linkage disequilibrium test revealed an association of distant loci only in non-infected individuals. This could mean a more conserved gene region in these individuals, which could happen if it would confer any protection against the infection and/or disease. Further, other peculiarities were found in the two groups. Infected individuals from St Catarina showed a significantly higher heterozygosity than expected in the locus T10/19 and on the opposite, it was the only group where IVS3 observed heterozygosity was within Hardy–Weinberg expected frequencies. Non-infected individuals from this district showed inverted allelic frequencies of the locus T10/19 comparing to the general trend and haplotypes 1705C/T10 and 1705A/T19 presented similar frequencies and 1705C/T19 showed higher frequency than in other studied groups. Further studies are needed to assess if these findings have a real biological meaning or are simply sampling artifacts.

Concluding remarks

This was the first study where data on sickle cell trait and G6PD deficiency frequencies were obtained for Cabo Verde human populations.

In this study no association was found between the analyzed human genetic factors and infection status of individuals. Three main reasons may have contributed for this: (1) the role of erythrocyte polymorphisms are usually associated and much easier demonstrated in severe than in mild or asymptomatic cases [54], (2) the cross-sectional sampling makes the infected/non-infected classification a faint case definition for an association study and 3) selective pressure of malaria, even if it had occurred, could never had a strong effect in this area due to its epidemic character.

Nonetheless, the finding of a very low frequency of G6PD deficiency associated alleles (A^- and MED) have important implications for the malaria control strategies defined by the National Program to Fight against Malaria (*Programa Nacional de Luta contra o Paludismo*, PNLP) viewing that it is recommended by WHO [55] that primaquine (potentially lethal in G6PD-deficient individuals) should be added to the drug regimen to block transmission in epidemic conditions such as Cabo Verde.

Regarding the PKLR gene, responsible for PK deficiency, recently reported as conferring protection against malaria in rodent and *in vitro* models, this study has not shown any clear association with malaria infection. Selective advantage afforded individuals protection from severe life-threatening complications of malaria and did not necessarily decrease their susceptibility to infection. Further, pyruvate kinase deficiency is a heterogeneous condition and most of the clinical phenotypes are mild or moderate in severity [26]. This suggests that the reproductive cost of PK deficiency was not limiting and mutations/polymorphisms would be spread in apparently healthy individuals.

Nevertheless, this is, to our knowledge, the first genetic population study about this putative association and results such as the region in linkage identified in the non-infected group deserve further investigation. Also, to further assess the assumption of a

protective effect of PK deficiency, further studies are being performed in other African populations from malaria highly endemic areas with well-defined malaria clinical cases (different severity level), well-characterized *Plasmodium*-infection and Hb β globin and G6PD status (to control for negative epistasis) and immediate enzymatic activity dosage at collection.

Acknowledgments

We are grateful to the population of Santiago Island, Cabo Verde who accepted to collaborate in this study. We thank the Health Delegates and technicians of Health Care Units of St Cruz, Tarrafal, St Catarina (especially Ana Veiga, Aníbal Monteiro, Antonino Monteiro and Edna Semedo) and Praia (especially Ernesto Cabral), Jorge de Pina (National Program against Malaria, Cabo Verde), Encarnação Horta and Marta Remédios (Institute of Hygiene and Tropical Medicine, Portugal) for technical assistance. We are also grateful to Doutor João Pinto for his participation in some sampling periods.

This study was supported by "Financiamento Programático do Laboratório Associado CMDT.LA/IHMT", POCI—Programa Operacional Ciência e Inovação 2010 (IPATIMUP) and POCI/SAU-ESP/55110/2004 (FCT/MCTES, Portugal). J. Alves and A.P. Arez were funded by FCT/MCTES Portugal (SFRH/BD/153451/2005 and SFRH/BPD/1624/2000—until 2007, respectively).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bcmd.2009.09.008.

References

[1] M.T.V. de Meira, T.S. Simões, J.F. Pinto Nogueira, Observações sobre sezonismo nas ilhas do Sal, Boa Vista e S. Nicolau (Cabo Verde), An. Inst. Med. Trop. (Lisb) IV (1947) 213-238.
[2] F.J.C. Cambournac, H. Santa Rita Vieira, M.A. Coutinho, et al., Note sur l'éradication du paludisme dans l'île de Santiago (Republique du Cap-Vert), An. Inst. Hig. Med. Trop. (Lisb) 10 (1984) 23–34.
[3] M. da Costa Monteiro, O sezonismo em Cabo Verde, An. Inst. Med. Trop. (Lisb) IX (1952) 461-483.
[4] J. Alves, Programme National de Lutte contre le paludisme. Plan d'Action 1994-98. Ministère de la Santé. République du Cap-Vert, 1994.
[5] WHO, International travel and health. Vaccination requirements and malaria situation. World Health Organization, 2005, http://whqlibdoc.who.int/publications/2005/9241580364_country_list.pdf (accessed 30 Jul 2009).
[6] A.P. Arez, G. Snounou, J. Pinto, et al., A clonal *Plasmodium falciparum* population in an isolated outbreak of malaria in the Republic of Cabo Verde, Parasitology 118 (1999) 347–355.
[7] J. Alves, A.L. Roque, P. Cravo, et al., Epidemiological characterization of *Plasmodium falciparum* in the Republic of Cabo Verde: implications for potential large-scale re-emergence of malaria, Malar. J. 5 (2006) 32.
[8] B. Carme, M.P. Hayette, A. Mbitsi, H. Moudzeo, J.C. Bouquety, *Plasmodium falciparum* index and level of parasitemia: diagnostic and prognostic value in the Congo, Ann. Soc. Belg. Med. Trop. 75 (1995) 33–41.
[9] I.M. Elhassan, L. Hviid, P.H. Jakobsen, et al., High proportion of subclinical *Plasmodium falciparum* infections in an area of seasonal and unstable malaria in Sudan, Am. J. Trop. Med. Hyg. 53 (1995) 78–83.
[10] J.M. Gonzalez, V. Olano, J. Vergara, et al., Unstable, low-level transmission of malaria on the Colombian Pacific Coast, Ann. Trop. Med. Parasitol. 91 (1997) 349–358.
[11] F.M. Leoratti, R.R. Durlacher, M.V. Lacerda, et al., Pattern of humoral immune response to *Plasmodium falciparum* blood stages in individuals presenting different clinical expressions of malaria, Malar. J. 7 (2008) 186.
[12] A.V. Hill, The genomics and genetics of human infectious disease susceptibility, Annu. Rev. Genomics Hum. Genet. 2 (2001) 373–400.
[13] F. Verra, V.D. Mangano, D. Modiano, Genetics of susceptibility to *Plasmodium falciparum*: from classical malaria resistance genes towards genome-wide association studies, Parasite Immunol. 31 (2009) 234–253.
[14] J. Makani, T.N. Williams, K. Marsh, Sickle cell disease in Africa: burden and research priorities, Ann. Trop. Med. Parasitol. 101 (2007) 3–14.
[15] M.J. Bouma, M. Goris, T. Akhtar, et al., Prevalence and clinical presentation of glucose-6-phosphate dehydrogenase deficiency in Pakistani Pathan and Afghan refugee communities in Pakistan; implications for the use of primaquine in regional malaria control programmes, Trans. R. Soc. Trop. Med. Hyg. 89 (1995) 62–64.
[16] A. Enevold, L.S. Vestergaard, J. Lusingu, et al., Rapid screening for glucose-6-phosphate dehydrogenase deficiency and haemoglobin polymorphisms in Africa by a simple high-throughput SSOP-ELISA method, Malar. J. 4 (2005) e61.
[17] M.A. Saunders, M. Slatkin, C. Garner, et al., The extent of linkage disequilibrium caused by selection on G6PD in humans, Genetics 171 (2005) 1219–1229.
[18] H. Spínola, J. Bruges-Armas, D. Middleton, A. Brehm, HLA polymorphisms in Cabo Verde and Guiné-Bissau inferred from sequence-based typing, Hum. Immunol. 66 (2005) 1082–1092.
[19] M.D. Cappellini, G. Fiorelli, Glucose-6-phosphate dehydrogenase deficiency, Lancet 371 (2008) 64–74.
[20] S.A. Tishkoff, R. Varkonyi, N. Cahinhinan, et al., Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance, Science 293 (2001) 455–462.
[21] G. Min-Oo, A. Fortin, M.F. Tam, et al., Pyruvate kinase deficiency in mice protects against malaria, Nat. Genet. 35 (2003) 357–362.
[22] K. Ayi, G. Min-Oo, L. Serghideset al, Pyruvate kinase deficiency and malaria, N. Engl. J. Med. 358 (2008) 1805–1810.
[23] P.M. Durand, T.L. Coetzer, Pyruvate kinase deficiency protects against malaria in humans, Haematologica 93 (2008) 939–940.
[24] L. Manco, A.L. Oliveira, C. Gomes, et al., Population genetics of four PKLR intragenic polymorphisms in Portugal and São Tomé e Princípe (Gulf of Guinea), Hum. Biol. 73 (2001) 467–474.
[25] E. Mateu, A. Perez-Lezaun, R. Martinez-Arias, et al., PKLR-GBA region shows almost complete linkage disequilibrium over 70 kb in a set of worldwide populations, Hum. Genet. 110 (2002) 532–544.
[26] A. Zanella, E. Fermo, P. Bianchi, et al., Pyruvate kinase deficiency: the genotype-phenotype association, Blood Rev. 21 (2007) 217–231.
[27] C. Alves, V. Gomes, M.J. Prata, et al., Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR YFiler) in Portugal, Forensic Sci. Int. 171 (2007) 250–255.
[28] L. Manco, M.L. Ribeiro, H. Almeida, et al., PK-LR gene mutations in pyruvate kinase deficient Portuguese patients, Br. J. Haematol. 105 (1999) 591–595.
[29] L. Manco, M.L. Ribeiro, V. Máximo, et al., A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency, Br. J. Haematol. 110 (2000) 993–997.
[30] C.M. Waterfall, B.D. Cobb, Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis, Nucleic Acids Res. 29 (2001) e119.
[31] J. May, C.G. Meyer, L. Gro terlinden, et al., Red cell glucose-6-phosphate dehydrogenase status and pyruvate kinase activity in a Nigerian population, Trop. Med. Int. Health 5 (2000) 119–123.
[32] W.W. van Solinge, R. van Wijk, R.J. Kraaijenhagen, et al., Novel mutations in the human red cell type pyruvate kinase gene: two promoter mutations in cis, a splice site mutation, a nonsense- and three missense mutations, Blood 90 (1997) 1197.
[33] E. Beutler, T. Gelbart, Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population, Blood 95 (2000) 3585–3588.
[34] L. Manco, A. Abade, Pyruvate kinase deficiency: prevalence of the 1456 C→T mutation in the Portuguese population, Clin. Genet. 60 (2001) 472–473.
[35] C. Lenzner, P. Nurnberg, B.-J. Thiele, et al., Mutations in the pyruvate kinase L gene in patients with hereditary hemolytic anemia, Blood 83 (1994) 2817–2822.
[36] H. Kanno, H. Fujii, A. Hirono, et al., Identical point mutations of the R-type pyruvate kinase (PK) cDNA found in unrelated PK variants associated with hereditary hemolytic anemia, Blood 79 (1992) 1347–1350.
[37] C. Lenzner, P. Nurnberg, G. Jacobasch, B.-J. Thiele, Complete genomic sequence of the human PK-L/R-gene includes four intragenic polymorphisms defining different haplotype backgrounds of normal and mutant Pk-genes, DNA Seq. 8 (1997) 45–53.
[38] C. Lenzner, G. Jacobash, A. Reis, et al., Trinucleotide repeat polymorphism at the PKLR locus, Hum. Mol. Genet. 3 (1994) 523.
[39] L. Excoffier, G. Laval, S. Schneider, Arlequin ver. 3.0: An integrated software package for population genetics data analysis, Evol. Bioinform. Online 1 (2005) 47–50.
[40] B. Modell, M. Darlison, Global epidemiology of haemoglobin disorders and derived service indicators, Bull. WHO 86 (2008) 480–487.
[41] D.A. Hassan, A.P. Arez, H.S. Mohamed, et al., Reduced sequestration of *P. falciparum* infected erythrocytes among sickle cell trait malaria patients: an *in vivo* evidence among a West African tribe in Sudan, Ann. Trop. Med. Parasitol. 102 (2008) 743–748.
[42] M. Aidoo, D.J. Terlouw, M.S. Kolczak, et al., Protective effects of the sickle cell gene against malaria morbidity and mortality, Lancet 359 (2002) 1311–1312.
[43] J.R. Aluoch, Higher resistance to *Plasmodium falciparum* infection in patients with homozygous sickle cell disease in western Kenya, Trop. Med. Int. Health 2 (1997) 568–571.
[44] J. Miranda, N. Gonçalves, I. Picanço, et al., Sickle-cell trait and red cell glucose-6-phosphate dehydrogenase status and malaria morbidity in Angola, Proc. 5th Eur. Congr. Trop. Med. Int. Health (2007) 65–68.
[45] E.T. Nkhoma, C. Poole, V. Vannappagari, et al., The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis, Blood Cells Mol. Dis. 42 (2009) 267–278.
[46] L. Luzzatto, U. Bienzle, The malaria/G6PD hypothesis, Lancet 1 (1979) 1183–1184.
[47] C. Ruwende, S.C. Khoo, R.W. Snow, et al., Natural selection of hemi and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria, Nature 376 (1995) 246–249.
[48] M. Cappadoro, G. Giribaldi, E. O'Brien, et al., Early phagocytosis of Glucose-6-Phosphate Dehydrogenase (G6PD)-deficient erythrocytes parasitized by *Plasmodium falciparum* may explain malaria protection in G6PD deficiency, Blood 92 (1998) 2527–2534.
[49] E.F. Roth Jr., M.C. Calvin, I. Max-Audit, et al., The enzymes of the glycolytic pathway in erythrocytes infected with *Plasmodium falciparum* malaria parasites, Blood 72 (1988) 1922–1925.

[50] I. Max-Audit, Anémie hémolytique due à un déficit en pyruvate kinase du globule rouge. Orphanet. http://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=FR&Expert=766.0, 2001 (accessed 31 Jul 2009).
[51] A. Brehm, L. Pereira, H.J. Bandelt, A. Amorim, Mithocondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade, Ann. Hum. Genet. 66 (2002) 49–60.
[52] R. Gonçalves, A. Rosa, A. Freitas, et al., Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers, Hum. Genet. 113 (2003) 467–472.
[53] A.T. Fernandes, R. Velosa, J. Jesus, et al., Genetic differentiation of the Cabo Verde archipelago population analysed by STR polymorphisms, Ann. Hum. Genet. 67 (2003) 340.
[54] F. Migot-Nabias, S. Pelleau, L. Watier, et al., Red blood cell polymorphisms in relation to *Plasmodium falciparum* asymptomatic parasite densities and morbidity in Senegal, Microbes Infect. 8 (2006) 2352–2358.
[55] WHO, Malaria epidemics: prediction, preparedness and control. WHO Expert Committee on Malaria. Twentieth report, http://apps.who.int/malaria/docs/ecr20_7.htm, 1998 (accessed 03 Jul 2009).

Chapter 3 –

Malaria: looking for selection signatures in the human *PKLR* gene region

This chapter was published as a research paper:

Machado, P., Pereira, R., Rocha, A.M., Manco, L., Fernandes, N., Miranda, J., Ribeiro, L., do Rosário, V.E., Amorim, A., Gusmão, L. and Arez, A.P., 2010**.** Malaria: looking for selection signatures in the human PKLR gene region. *British journal of haematology,* 149(5), pp.775-84.

Malaria: looking for selection signatures in the human *PKLR* gene region

Patrícia Machado,[1] Rui Pereira,[2,3] Ana Mafalda Rocha,[2] Licínio Manco,[4,5] Natércia Fernandes,[6] Juliana Miranda,[7] Letícia Ribeiro,[5] Virgílio E. do Rosário,[1] António Amorim,[2,8] Leonor Gusmão[2] and Ana Paula Arez[1]

[1]Centre for Malaria and Tropical Diseases, Malaria Unit, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, [2]Institute of Molecular Pathology and Immunology of University of Porto (IPATIMUP), Oporto, Portugal, [3]Institute of Legal Medicine, Universidade de Santiago de Compostela, Spain, [4]Research Centre for Anthropology and Health Department of Anthropology, Universidade de Coimbra, [5]Haematology Department, Centro Hospitalar de Coimbra, Portugal, [6]Central Hospital of Maputo and Faculty of Medicine, Universidade Eduardo Mondlane, Mozambique, [7]Paediatric Hospital David Bernardino, Luanda, Angola, and [8]Faculty of Sciences, Universidade do Porto, Portugal

Received 11 November 2009; accepted for publication 12 February 2010

Correspondence: Dr Patrícia Machado, Centre for Malaria and Tropical Diseases, Malaria Unit, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira, 100, 1349-008 Lisbon, Portugal.
E-mail: pmachado@ihmt.unl.pt

Summary

The genetic component of susceptibility to malaria is both complex and multigenic and the better-known protective polymorphisms are those involving erythrocyte-specific structural proteins and enzymes. *In vivo* and *in vitro* data have suggested that pyruvate kinase deficiency, which causes a nonspherocytic haemolytic anaemia, could be protective against malaria severity in humans, but this hypothesis remains to be tested. In the present study, we conducted a combined analysis of Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs) in the pyruvate kinase-encoding gene (*PKLR*) and adjacent regions (chromosome 1q21) to look for malaria selective signatures in two sub-Saharan African populations from Angola and Mozambique, in several groups with different malaria infection outcome. A European population from Portugal, including a control and a pyruvate kinase-deficient group, was used for comparison. Data from STR and SNP loci spread along the *PKLR* gene region showed a considerably higher differentiation between African and Portuguese populations than that usually found for neutral markers. In addition, a wider region showing strong linkage disequilibrium was found in an uncomplicated malaria group, and a haplotype was found to be associated with this clinical group. Altogether, this data suggests that malaria selective pressure is acting in this genomic region.

Keywords: Human malaria, selection signatures, pyruvate kinase-deficiency, *PKLR,* molecular markers.

According to the World Malaria Report 2008 (World Health Organization, WHO, 2008), 109 countries are currently endemic for malaria, 45 of which are within the African region, and 247 million malaria cases were estimated among the 3·3 billion people at risk in 2006. These cases resulted in nearly a million deaths, mostly of children under 5 years old. Despite this disastrous picture, the current combination of tools and methods to combat malaria, including long-lasting insecticidal nets and artemisinin-based combination therapy (ACT), supported by indoor residual spraying of insecticide and intermittent preventive treatment in pregnancy, is leading to a significant reduction of cases in some countries, such as Gambia (Ceesay *et al*, 2008), Kenya (O'Meara *et al*, 2008) and São Tomé and Príncipe (unpublished observations). However, both *Anopheles* mosquito and *Plasmodium* parasite have developed resistance to insecticides (Anto *et al*, 2009) and new drugs (Noedl *et al*, 2008), which clearly shows that the fight against the disease continues to be a difficult challenge.

Malaria has been reported as one of the strongest known forces for evolutionary selection in the recent history of the

First published online 4 April 2010
doi:10.1111/j.1365-2141.2010.08165.x

human genome. The genetic component of susceptibility to malaria is complex and multigenic, with a variety of genetic polymorphisms reported to influence both pathogenesis and host response to infection (Kwiatkowski, 2005; Min-Oo & Gros, 2005; Williams, 2006). The identification of these variants might, therefore, help to improve the development of therapeutic and disease-prevention strategies.

The most common and best characterised malaria protective polymorphisms are those involving erythrocyte-specific structural proteins and enzymes, such as sickle cell disease and glucose-6-phosphate dehydrogenase (G6PD)-deficiency. More recently, pyruvate kinase (PK)-deficiency has also been reported as protective against malaria in murine models (Min-Oo *et al*, 2003) and two studies have reported the *in vitro* culturing of *P. falciparum* in PK-deficient blood with a significant decrease in parasite replication (Ayi *et al*, 2008; Durand & Coetzer, 2008). However, the possibility that PK-deficiency may affect susceptibility to malaria in humans remains to be confirmed.

Apart from results in murine models and *in vitro* cultures, there is no population data supporting a positive association between PK-deficiency and malaria protection. Given the differences in selection pressure that mice and humans have been exposed to over tens of millions of years, the major susceptibility genes in the two species are unlikely to be the same (Hill, 1998), and the possibility that any crucial insufficiency of the erythrocytes, besides PK-deficiency, may influence the development of the parasite make clear the need to perform additional studies to clarify this question. Moreover, until now, contrary to G6PD-deficiency or sickle cell disease, elevated frequencies of PK-deficiency have not been recorded in malaria endemic areas; however, a systematic analysis has never been done and even the information about the frequency of PK-deficiency in African populations is clearly limited (Manco *et al*, 2001; Mateu *et al*, 2002).

The first study including a population genetic approach concerning the possible association between the *PKLR* gene (PK-encoding gene) and malaria was carried out at the Island of Santiago, Cabo Verde (Alves *et al*, 2010). Although no association was then found between any *PKLR* polymorphism and infection status, a strong linkage between distant loci in the gene and adjacent regions was reported only in non-infected individuals. This linkage could mean that there is a more conserved gene region that is selected if protective against the infection and/or disease. The present study aimed to further analyse this previous preliminary result by looking at the *PKLR* gene and adjacent regions in individuals belonging to different population groups (from Angola and Mozambique, both malaria endemic countries, and from Portugal, a country with no malaria transmission) and to different malaria status (asymptomatic infection, mild and severe malaria), with the goal of identifying potential selection signatures in this genomic region imprinted by malaria.

Material and methods

Study areas

Angola and Mozambique are both sub-Saharan countries. Angola (capital Luanda, 8°50′ 18″S, 13°14′ 4″E) is localised in south-western Africa and is bordered by the Atlantic Ocean to the west; Mozambique (capital Maputo, 25°57′ 55″S, 32°35′ 21″E) is in south-eastern Africa with its east coast on the Indian Ocean. Both have a tropical climate with two seasons, one wet and warm from September to May, and the other dry and cold from June to August. Malaria, predominantly caused by *Plasmodium falciparum*, is endemic (Cuamba *et al*, 2006; Mabunda *et al*, 2008). Portugal (39°30′N, 8°00′W) is in south-western Europe. Malaria transmission was interrupted in nearly all parts of the country by 1958 and eradication was confirmed by WHO in 1973 (Bruce-Chwatt, 1977).

Sampling

A total of 417 DNA samples were analysed in this study. There were 316 collected from both uninfected and infected non-related children with a different malaria outcome: 166 from Luanda, Angola (ANG) [44 with severe malaria, 43 with uncomplicated malaria, 37 from asymptomatic infected individuals and 42 from healthy aparasitaemic individuals (uninfected)] and 150 from Maputo, Mozambique (MOZ) (51 with severe malaria and 99 with uncomplicated malaria). The pooling of all samples from Angola (ANG) and Mozambique (MOZ) constituted the African group (AFR). Two groups from Portugal were also analysed: there were 80 samples from healthy individuals (control Portuguese group, PT-C) (described in Alves *et al*, 2007) and 21 belonging to individuals with PK-deficiency (PT-PKD) (described in Manco *et al*, 1999, 2000).

Malaria outcome was defined as follows: (i) Severe malaria (SM): slide positive for blood-stage asexual *P. falciparum* at any parasite density, fever (axillary temperature ≥37·5°C), haemoglobin level of Hb≤50 g/l and/or other symptoms, such as coma, prostration or convulsions; (ii) Uncomplicated malaria (UM): slide positive for blood-stage asexual *P. falciparum* at any parasite density, fever (axillary temperature ≥37·5°C) and haemoglobin level of Hb>50 g/l; and (iii) Asymptomatic infection (AI): slide positive for blood-stage asexual *P. falciparum* at any parasite density in the absence of fever or other symptoms of clinical illness. The additional group of uninfected children (NI) was defined as slide negative and the absence of fever or other symptoms of clinical illness. Slide negativity was afterwards confirmed by Polymerase Chain Reaction (PCR). The illness group (ILL) comprised all the individuals expressing clinical disease: SM plus UM.

Blood collection and DNA extraction

Blood sample collections by finger-prick were carried out in Angola in August 2005 and in Mozambique during 2006 from children aged 3 months to 15 years who reported to the Emergency Services of the Paediatric Hospital David Bernardino, Luanda (Angola) or to the Paediatric Emergency Services of Central Hospital of Maputo, Health Centre of Bagamoyo or Health Centre of Boane (Mozambique). The blood was drawn after the clinician examination (malaria was considered to be the primary diagnosis if *Plasmodium* parasites were found in the peripheral blood and if other likely causes of the clinical presentation could be excluded at the admission) but before the administration of any anti-malarial therapeutics and/or blood transfusion. The registration of symptoms, axillary temperature, haemoglobin level and history of malaria was done for all individuals.

The investigation was approved by both the Ministry of Public Health of Angola and Mozambique and by the local Ethical Committees at the institutions involved in the study. Each individual and parent/tutor of the children was informed of the nature and aims of the study and told that participation was voluntary; informed consents were obtained from all individuals.

DNA was extracted using standard phenol-chloroform or chelex procedures from peripheral blood. In the case of infected individuals, human and *Plasmodium* DNA were extracted simultaneously.

Genotyping

A section of chromosome 1q21, including the *PKLR* gene and adjacent regions, with a total length of ≈ 95 Kb, was genotyped for 4 Short Tandem Repeats (STRs) and 15 Single Nucleotide Polymorphisms (SNPs). Samples were also genotyped for 32 Ancestry Informative Insertion/Deletion polymorphisms (AI-INDELs) distributed throughout the genome. The localization of polymorphisms in chromosome 1 is represented in Fig 1.

STRs

The STRs used were IVS3 (in intron 3), IVS11 (intron 11), PKA (≈ 25 kb upstream from the *PKLR* gene) and PKV (≈ 65 kb upstream from the gene) and were genotyped after multiplex PCR as described in Alves *et al* (2010).

SNPs

SNPs localised in a region closer to *PKLR* than the above-mentioned STRs were genotyped using a SNaPshot (Applied Biosystems, Foster City, CA, USA) multiplex reaction.

The DNA sequence of chromosome 1q21, including the *PKLR* gene and flanking regions, was screened for SNPs in the HapMap database (http://hapmap.ncbi.nlm.nih.gov/). A total of 13 SNPs were selected in a region of 40,970 bp that spanned the *PKLR* gene (chr1:153515199..153556169; data source: HapMap Data Rel 22/phaseII Apr07, on NCBI B36 assembly, dbSNP b126), starting at 18 334 bp upstream and extending to 11 055 bp downstream of the gene. All the SNPs described for the *PKLR* gene were selected for genotyping, except rs3020781, which had amplification difficulties. SNPs outside of the gene that showed variation in the reference African population (Yoruba, Nigeria), with a minor allele frequency above 15% and distances between contiguous SNPs greater than 1600 bp, were included in the study.

Two additional mutations were investigated in the *PKLR* gene: 1456C>T, because it is the most common mutation in South Europe, namely in Portugal (Manco & Abade, 2001) and the only one described in PK-deficient Afro-American individuals (Beutler & Gelbart, 2000), and 1614A>T, identified in São Tome and Príncipe (Manco *et al*, 2009).

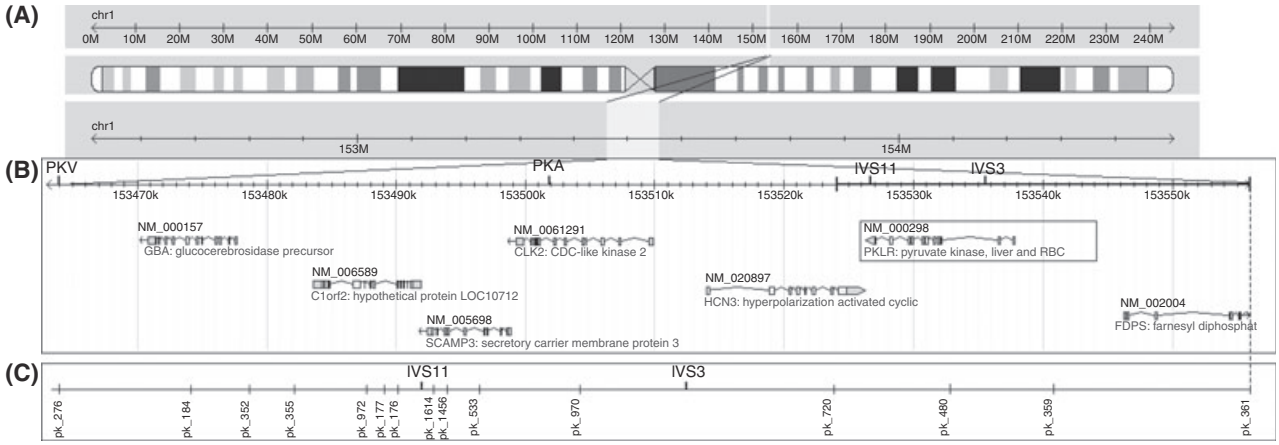


Fig 1. The 95 kb fragment analysed in this study, including *PKLR* gene. (A) Localization in chromosome 1q21; (B) The 4 STR loci (PKV, PKA, IVS11 and IVS3) genotyped in the present study and genes near *PKLR*; (C) The 15 SNP loci analysed spread along a region closer to the gene *PKLR*. *Adapted from* http://www.hapmap.org.

Primers were designed for the flanking regions of each of the 15 SNPs in the GenBank database sequence AY316591 with Primer 3 software v.0.4.0 (Rozen & Skaletsky, 2000; primer sequences in Table SI). Primers were first tested in singleplex and then multiplex reactions were carried out according to Goios *et al*, 2008, using the Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany).

For each SNP, an SBE-Primer was designed with Primer 3 software (Table SII). Amplified products were purified with ExoSAP-IT (Amersham Biosciences, Uppsala, Sweden) and SNaPshot reactions were then performed using the SNaPshot Multiplex Kit (Applied Biosystems) in a reaction volume of 5 μl with primer concentrations as indicated, under the following conditions: 96°C for 10 s, 55°C for 5 s, and 60°C for 30 s, repeated for 27 cycles. The final products were purified with SAP (Amersham Biosciences) and run in an abi prism 3130 Genetic Analyzer. Allele assignment was performed using GeneMapper 4.0 (Applied Biosystems).

Ancestry informative INDELs

The high levels of genetic substructure in Africa, even within small geographic regions, require the determination of individual ancestry and proper correction for substructure in association studies (Campbell & Tishkoff, 2008). To look into the structure of our African groups and to investigate if our PT-PKD group could have a relevant African genetic component, which would suggest that PK-deficiency could be frequent in that region, 32 INDEL polymorphic regions localised throughout the genome were genotyped as described in Santos *et al* (2010). In this work, we used only a subset of the original assay, comprising the INDELs that are especially informative of African and European ancestry. An additional reference Portuguese group (PT-REF) that was previously typed for these INDEL loci (Santos *et al*, 2010) was also used in this analysis.

Statistical analysis

Analysis was performed by comparing population groups (ANG, MOZ, PT-C, PT-PKD) and malaria status groups (SM, UM, AM, NI, ILL). STR and SNP results were explored with Arlequin 3.1 (Excoffier *et al*, 2005): determination of the allele frequencies, expected and observed heterozygosity and population pairwise F_{ST} values, Hardy–Weinberg equilibrium tests, Linkage Disequilibrium (LD) tests, haplotype frequency estimation and analysis of molecular variance (amova). When there were multiple tests, Bonferroni's correction was applied, dividing 0·05 by the number of tests to obtain the actual cut-off for significance. The allelic association of SNPs and STRs with malaria status groups was assessed by a Pearson's 2×2 contingency table chi-square test using Simple Interactive Statistical Analysis (SISA, http://www.quantitativeskills.com/sisa/). Odds ratios (OR) and 95% confidence intervals (CI) were estimated using SISA. Allelic richness with rarefaction of private alleles was calculated with HP-Rare (Kalinowski, 2005). Bayesian clustering analysis as implemented by Structure 2.2 (Pritchard *et al*, 2000) was used to infer population substructure/ancestry from the INDEL data set, without prior information on sampling groups, under the admixture model with correlated allele frequencies. Ten independent runs with 10^5 burn-in steps and 10^5 interactions were done for each value of K ($K = 1$ to 5 clusters). For INDELs, ARLEQUIN 3.1 (Excoffier *et al*, 2005) was also used for F_{ST} calculations.

Results

STRs

The allele frequencies for the four STR loci found in ANG, MOZ, PT-C and PT-PKD are shown in Table SIII. The IVS3 locus presented the greatest diversity indices in all groups, with the highest number of alleles and expected heterozygosity. In both African groups, the observed genotype frequencies were according to Hardy-Weinberg expectations for all loci except for IVS3, which revealed a heterozygosity significantly below the expected ($P \leq 0{\cdot}000$). In Portuguese groups, all loci were in Hardy-Weinberg equilibrium in the control PT-C ($P = 0{\cdot}378$ for IVS3) but not in the PT-PKD group, which showed a strong deviation from the expected values for IVS3 ($P \leq 0{\cdot}000$) and IVS11 ($P = 0{\cdot}006$).

When F_{ST} values were calculated, no significant differentiation was obtained for the pair ANG vs. MOZ ($F_{ST} = 0{\cdot}002$; $P = 0{\cdot}189$). When Portuguese groups were compared, significant values were obtained, as expected: F_{ST}(PT-C vs. PT-PKD) = 0·025; $P \leq 0{\cdot}000$. Since no differentiation was found between Angola and Mozambique, a single group was formed for all of the African samples (AFR) and it was compared to Portuguese groups to investigate if African and Portuguese PK-deficient individuals were genetically closer in this genomic region than African and Portuguese controls. If so, we could hypothesise that PK-deficiency could be frequent in Africa (because of some kind of selective advantage conferred by the disease). The F_{ST} values obtained were as follows: F_{ST}(AFR vs. PT-C) = 0·102 and F_{ST}(AFR vs. PT-PKD) = 0·153 ($P \leq 0{\cdot}000$ for both tests).

No significant differentiation was found between the several malaria status groups, whether considering each of the four STR loci separately or all together. As F_{ST} was not significant when comparing ANG and MOZ, UM and SM, samples from both countries were pooled into two larger groups, but still no significant values were found between these groups. No STR or SNP allele was associated with any malaria status group ($P > 0{\cdot}05$) and OR values were non-significant for all groups. Moreover, when STR allelic private richness was calculated (considering 42 genes for all groups as PT-PKD only included 21 samples), private alleles were not identified, supporting the previous result. However, allele 16 of locus IVS11 ($\chi^2 = 10{\cdot}918$; $P < 0{\cdot}001$ and OR = 6·200 with 95% CI 1·858–20·685) and allele 36·2 of locus IVS3 ($\chi^2 = 13{\cdot}265$; $P < 0{\cdot}001$

and OR = 5·961 with 95% CI 2·072–17·154) were significantly associated only with PT-PKD. These two specific alleles were not associated with any particular malaria status group.

The African groups ANG and MOZ showed a marked LD for all pairs of loci ($P \leq 0{\cdot}000$). Conversely, the group PT-C only showed LD for the closer loci (PKV/PKA and PKA/IVS11), while the PT-PKD group only showed LD for PKV/IVS11. However, when the African malaria status groups were analysed separately, only UM sets from both Angola and Mozambique had significant results for all pairs of loci ($P \leq 0{\cdot}008$), i.e. significant LD for a region spanning $\approx$ 75 Kb (IVS3 was not considered for this test as it was not in Hardy-Weinberg equilibrium). Furthermore, when UM samples from Angola and Mozambique were pooled in one single larger group, the previous result was reinforced: $P \leq 0{\cdot}000$ for all LD tests between locus pairs. Therefore, we searched for a haplotype (PKV/PKA/IVS11/IVS3) that could be associated with this larger UM group and 9/11/13/34 revealed this association, although it was borderline ($\chi^2 = 5{\cdot}898$, $P = 0{\cdot}015$; OR = 5·267; 95% CI: 1·188–23·355).

The population groups studied all revealed a large number of low frequency inferred haplotypes. The most common haplotypes were: in ANG, 10/14/12/38, 11/12/15/35, 11/11/17/35 and 10/13/12/34, with an approximate frequency of 3% each; in MOZ, haplotype 9/11/13/34 was prominent (6·3%, from which 5·5% were in UM) and four additional haplotypes were also frequent ($\approx$ 3%): 10/13/14/35, 11/9/17/37·2, 10/13/12/35 and 10/14/12/38; in PT-C, the most frequent haplotype was 9/9/14/40·2 (5·6%), followed by 10/9/14/38·2, 10/9/14/39·2 and 9/9/14/37·2 (about 4%); and in PT-PKD, the most frequent haplotypes were 10/9/14/38·2 (23·8%), 9/9/15/36·2 (19·0%) and 9/9/16/38·2 (11·9%). These last two were not detected in PT-C and 9/9/15/36·2 was exclusively found in PT-PKD.

An AMOVA that considered these four loci for comparison in the follow three populations, Africa (NI, AM, UM and SM from Angola, UM and SM from Mozambique), Portugal – control (PT-C) and Portugal – PK-deficiency (PT-PKD), resulted in a significant percentage of variation between the three populations (10·92%, $P \leq 0{\cdot}000$) and within each group (88·97%, $P \leq 0{\cdot}000$). A non-significant value was obtained between groups within each population (0·12%, $P = 0{\cdot}512$).

SNPs

Overall, 15 SNPs were analysed in this study: 13 were identified in the HapMap database and two were mutations previously described to be associated with PK-deficiency. These mutations were not identified in any of the African groups studied or in the control Portuguese individuals. Mutation 1456C>T was identified in eight Portuguese PK-deficient individuals, two of whom were homozygous for the T allele (Manco *et al*, 1999, 2000). The allele frequencies found in the studied population groups are shown in Table SIV.

No significant differentiation was found between ANG and MOZ or between PT-C and PT-PKD, whether considering all 13 loci simultaneously or separately. A significant differentiation was found between African and Portuguese groups: F_{ST}(AFR vs. PT-C) = 0·239, F_{ST}(AFR vs. PT-PKD) = 0·341, $P \leq 0{\cdot}000$ for both tests.

Comparing NI, AI, SM and UM from Angola and Mozambique, F_{ST} values were not significant for any pairs of groups tested. Given that there were no differences between the two African populations, UM and SM from both countries were pooled into larger groups for comparison, but still no differences were found. The same result was obtained when these groups were compared to NI and AI.

The observed heterozygosity was according to the Hardy-Weinberg expected frequencies in all population groups but, strikingly, when performing an analysis on the malaria status groups from Angola, all loci in UM and SM that were localised in exon 12 (pk_177, pk_176 and pk_972) or downstream (pk_276, pk_184, pk_352 and pk_355) had a deviation from Hardy-Weinberg equilibrium ($P < 0{\cdot}050$) with an excess of heterozygotes (as seen in Fig 2). However, when Bonferroni's correction was applied ($P < 0{\cdot}004$ for significance), none of these results were statistically significant. However, when individuals of SM and UM were combined into the single ILL group, the deviation was significant even under Bonferroni's correction. These results were not obtained for the Mozambican groups, where the observed heterozygosity was similar to expectation.

African populations showed higher haplotype diversity than the Portuguese. The five main inferred haplotypes (pk_276/pk_184/.../pk_361, ordered as in Fig 1) were identified in both ANG and MOZ and also observed in the malaria status groups from each country. No specific haplotype was associated with any group. In PT-C, two main haplotypes, already identified in the African groups, were observed: G/G/T/C/G/A/G/T/C/G/A/C/A/T/A (frequency of 76%) and A/A/C/G/A/G/T/T/C/C/A/G/C/C/C (frequency of 18%). In PT-PKD, two main haplotypes were identified: one was the most common in PT-C, whereas the other was exclusive to this group, because of the mutation 1456T (G/G/T/C/G/A/G/T/T/G/A/C/A/T/A), which was in complete LD with all adjacent loci (Fig 3). When we looked for selective sweeps in African groups in this genomic segment, they were not found: in a general way, the expected heterozygosity in loci from ANG and MOZ was higher but followed the trend observed in PT-C and PT-PKD (Fig 2).

Similarly to AMOVA using the STRs, AMOVA using all of the SNP loci resulted in significant percentages of variation between the populations [Africa (NI, AM, UM and SM from Angola, UM and SM from Mozambique), Portugal – control (PT-C) and Portugal – PK-deficiency (PT-PKD)] and within each group (25·47%, $P \leq 0{\cdot}000$ and 74·52%, $P \leq 0{\cdot}000$, respectively). The percentage of variation between groups within each population was not significant (≤0·00%, $P = 0{\cdot}481$).

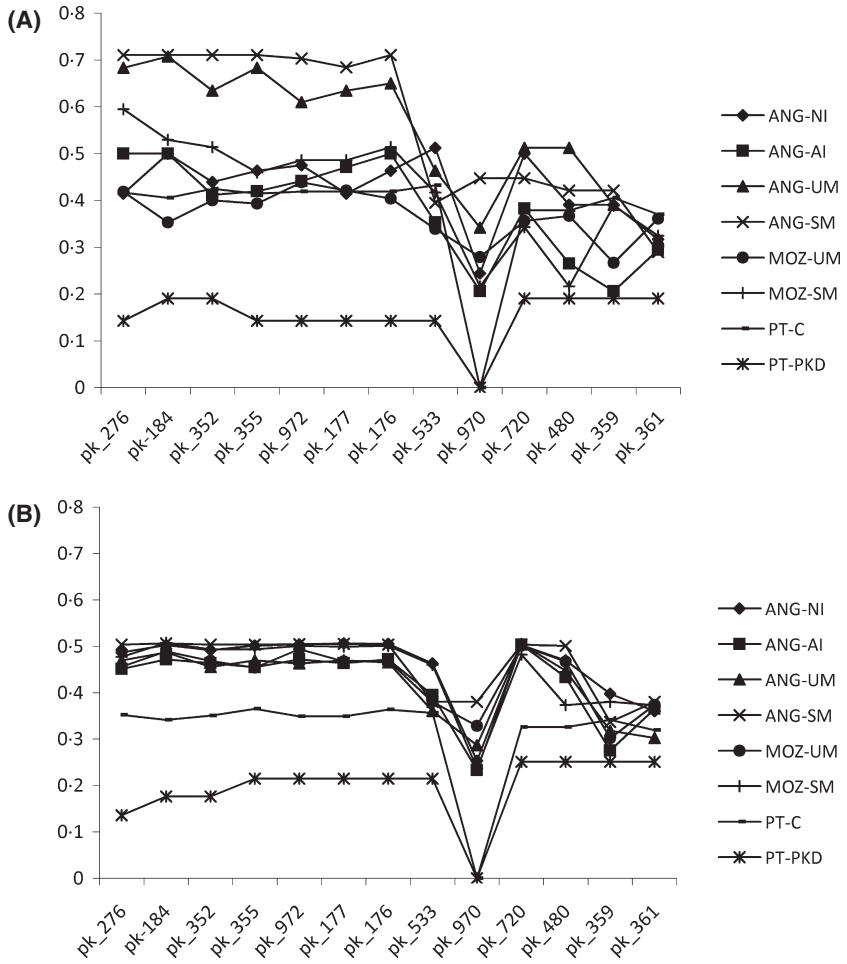


Fig 2. Observed (A) and expected (B) heterozygosity of the SNP loci in Portuguese groups and malaria status groups from both Angola and Mozambique. ANG-UM and ANG-SM revealed a heterozygote excess for all loci included between pk_276 and pk_176. ANG-NI: Angola – non-infected; ANG-AI: Angola – asymptomatic infection; ANG-UM: Angola – uncomplicated malaria; ANG-SM: Angola – severe malaria; MOZ-UM: Mozambique – uncomplicated malaria; MOZ-SM: Mozambique – severe malaria; PT-C: Portugal – control group; PT-PKD: Portugal – PK-deficiency group.

A combined analysis was performed using all STR and SNP loci, and the results supported those reported above: significant F_{ST} values were obtained when African groups were compared to Portuguese groups. A significant differentiation was also obtained between the two Portuguese groups, PT-C and PT-PKD.

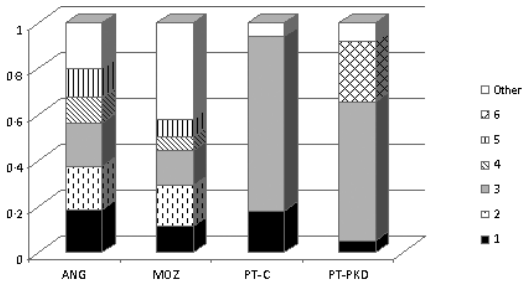
Ancestry informative INDELs

The structure of African and Portuguese (PT-PKD and PT-REF) groups was examined through the genotyping of 32 INDELs. K = 2 was, undoubtedly, the most likely number of clusters, corresponding to the African and Portuguese samples. Even when K = 3 to K = 5 were tested, the division between African and Portuguese clusters was obvious (Fig 4). A clear differentiation was achieved between African and PT-REF (F_{ST} = 0·392; $P \leq 0{\cdot}000$) and African and PT-PKD (F_{ST} = 0·423; $P \leq 0{\cdot}000$) groups. MOZ and ANG could be slightly differentiated (F_{ST} = 0·003; $P \leq 0{\cdot}000$) by genetic distance analysis but not when using Structure 2.2 software, even when only the two African groups were considered (data not shown). No differentiation was achieved between PT-REF and PT-PKD, or between malaria status groups within MOZ or within ANG under any circumstance.

Discussion

A combined analysis with STR and SNP data was used to search for malaria selection signatures in the *PKLR* gene region. Two different approaches were performed: inter-population analysis, opposing two populations from malaria endemic regions (Angola and Mozambique) to a Portuguese population with no malaria, and an intra-population analysis, comparing malaria status groups within populations.

STR and SNP allelic frequencies in ANG and MOZ were similar and quite different from PT-C and PT-PKD, reflecting



Haplotype	pk_276	pk_184	pk_352	pk_355	pk_972	pk_177	pk_176	pk_1614	pk_1456	pk_533	pk_970	pk_720	pk_480	pk_359	pk_361
	G/A	**G/A**	**T/C**	**C/G**	**G/A**	**A/G**	**G/T**	**T>A**	**C>T**	**G/C**	**A/G**	**C/G**	**A/C**	**T/C**	**A/C**
1	A	A	C	G	A	G	T	T	C	C	A	G	C	C	C
2	A	A	C	G	A	G	T	T	C	C	A	C	A	C	A
3	G	G	T	C	G	A	G	T	C	G	A	C	A	T	A
4	A	A	C	G	A	G	T	T	C	C	A	G	A	C	A
5	G	G	T	C	G	A	G	T	C	C	G	G	C	C	A
6	G	G	T	C	G	A	G	T	**T**	G	A	C	A	T	A

Fig 3. Estimated frequencies of inferred haplotypes in the studied population groups. ANG: Angola; MOZ: Mozambique; PT-C: control Portuguese; PT-PKD: Portuguese with PK-deficiency. The segment between pk_276 and pk_176 was extremely conserved in all haplotypes with only two possible allelic combinations, indicated by different greys in the lower panel.

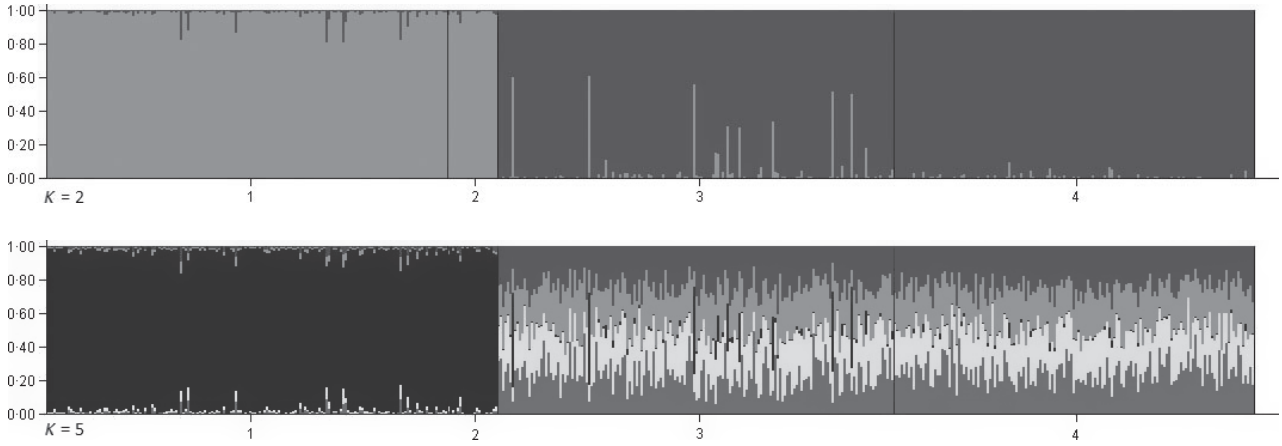


Fig 4. Estimated population structure determined with STRUCTURE 2.2. (no prior information of sampling groups, under the admixture model with correlated allele frequencies; ten independent runs with 10^5 burn-in steps and 10^5 interactions). Each bar represents a single individual and is partitioned into K different grey-shaded segments that represent the individual's estimated coefficients of ancestry. K = 2 is the most suitable division, with clusters corresponding to the Portuguese (mainly light grey) and African (mainly dark grey) samples. 1- PT-REF [reference group from Portugal (Santos *et al*, 2010)]; 2- PT-PKD (individuals with PK-deficiency from Portugal) 3- ANG (Angola); 4-MOZ (Mozambique).

structural differences. In fact, when sample structure was tested using ancestry informative INDEL markers, two clusters were clearly formed: one with all ANG and MOZ samples and one including all PT-PKD and PT-REF samples.

F_{ST} among human populations from major geographical regions, based on more than 370 STRs, was estimated to be 0·05 (Rosenberg *et al*, 2002), and it was estimated to be 0·10 when based on 600,000 SNPs (Li *et al*, 2008). Moreover, an AMOVA using the same STR loci (Rosenberg *et al*, 2002) showed 3·6% to 5·2% variation between major regions of the world and 3·1% variation between populations within Africa. In this study, F_{ST} values obtained between African and Portuguese groups were considerably higher, varying between 0·102 and 0·153 for STRs and between 0·239 and 0·341 for SNPs. In addition, an AMOVA for STR loci had a significant outcome of 10·92% variation between Africans and Portuguese, whereas variation between groups within each population was 0·12%. In a typical multilocus sample, it is reasonable to assume that all autosomal loci have experienced the same demographic history and the same rates and patterns of migration. Loci showing unusually large amounts of differentiation may indicate regions of the genome that have been subject to diversifying selection (Holsinger & Weir, 2009) of which malaria could have been the cause. The AMOVA results show that, whereas variation between Africa and

Portugal more than doubled in this study, the opposite occurred in the degree of variation between groups within populations, suggesting that some (selective) force is homogenising this genomic fragment in African regions and, at the same time, extending the differences between Africa and other global areas. Curiously, the F_{ST} value for Africans versus. PT-PKD was higher than for Africans versus. PT-C, suggesting that, even if PK-deficiency is frequent in sub-Saharan Africa, mutations should be different from those found in the Portuguese.

Concerning the Portuguese groups, differentiation was only significant when STR data was used, which may be explained by the different molecular resolution of SNPs and STRs: in humans, the average nucleotide mutation rate is assumed to be $2{\cdot}5 \times 10^{-8}$ and the STR mutation rate has been estimated to be 10^{-2}–10^{-5} per generation (Tishkoff & Verrelli, 2003). Thus, SNPs are best used for inferring human evolutionary history over longer time scales and STRs can be used to trace recent demographic events (Agrafioti & Stumpf, 2007). Therefore, we can presume that Portuguese PK-deficiency variants have emerged recently, which is supported by the lower diversity found within this group.

No differentiation was ever obtained between malaria status groups, either using SNPs or STRs, although insufficient sampling of each group may be influencing this result. Of all the STR loci, IVS3 in the *PKLR* gene was the only one with frequencies that were out of Hardy-Weinberg equilibrium in the African groups, with a significant excess of homozygotes. This had already been observed in a previous study with African samples from Cabo Verde (Alves *et al*, 2010). Conversely, as expected, the control group PT-C, had a heterozygosity that was similar to that expected. These data suggest that IVS3 homozygosity is being promoted in some manner. Possible causes for the Hardy-Weinberg equilibrium deviation include admixture and substructure or non-random mating patterns. However, as this deviation was observed in several African populations, it is possible that it is caused by the impact of selection pressures from environmental conditions (e.g. infectious diseases like malaria). IVS3 is in intron 3, a critical functional location as it is where the splicing of exon 2 occurs for the production of PKL mRNA, and as it is not a simple polymorphic locus (it includes eight contiguous variation regions), it should be carefully analysed.

The LD test for the STRs showed a significant LD along the entire studied region for UM. This is interesting as suggests an association between this conserved genomic block and a mild malaria outcome. Moreover, this LD emphasises the result previously found in Cabo Verde, where an LD test revealed an association of these same loci but in non-infected individuals (Alves *et al*, 2010). Additionally, this LD outcome is not expected under neutrality, which also supports our results: several datasets show differences in haplotype structure between African and non-African samples, where blocks are significantly smaller in African samples and extend longer and are less diverse in non-Africans (Tishkoff & Verrelli, 2003). Reinforcing the LD result, a haplotype was identified as associated with this group: 9/11/13/34. This association must be further analysed since it is not robust ($P = 0{\cdot}015$), but we believe that insufficient sampling may be the cause for this deficiency, as this association was identified only when UM and SM samples from both Angola and Mozambique were pooled together in a larger group.

The LD test for the SNPs had a significant result in all groups and populations for all pairs of SNP loci in exon 12 and upstream (between loci pk_276 and pk_176). Curiously, the ILL group from Angola had a significant SNP heterozygote excess exactly in the same region. Three of these loci are located in exon 12 of *PKLR*, and the remaining are in the *HCN3* gene. This gene, coding for a hyperpolarisation-activated cyclic nucleotide-gated potassium channel 3, is a voltage-gated channel performing ionic, potassium and sodium transport (Uniprot database/Swiss-Prot Q9P1Z3) and is highly expressed in early erythroid cells (Su *et al*, 2004), which produce mature erythrocytes. Heterozygosity in this genomic fragment seems to be associated with clinical malaria in Angola but not in Mozambique, suggesting that, additionally to malaria, some geographic factor may be involved in this scenario.

Five main inferred SNP haplotypes were identified in ANG and MOZ and only two in PT-C (contained within those five) and two in PT-PKD. These results were expected as African populations are older and have maintained a larger N whereas non-African populations have experienced a bottleneck event during the expansion of modern humans out of Africa within the past 100 000 years (Tishkoff & Verrelli, 2003). The high mutation rate of STRs explains why the same STR haplotype diversity is present in both African and non-African regions. Haplotype 6 was exclusive to PT-PKD, differing only from haplotype 3 (the most common in PT-C) at the pk_1456 locus. As a result of its strong LD, the segment between pk_276 and pk_176 was extremely well-conserved in all haplotypes, with only two possible allelic combinations. The remaining segment revealed strong recombination. Neither of the two mutations that were potentially associated with PK-deficiency in Africa (as indicated in previous reports) were identified in our African samples.

Previous studies have also examined this particular genomic fragment, seeking other disease-associated variants. Multiple studies in populations from diverse origins have shown linkage of type 2 diabetes (T2D) to chromosome 1q over a broad region and the *PKLR* gene arises as the first candidate (Wang *et al*, 2002, 2009; Das & Elbein, 2007). A search for prevalence of T2D in the African continent revealed that Afro-Americans have a two-fold increase in risk for T2D compared to other populations in the United States, but its prevalence is lower in Africa (1–2%) than among people of African descendant in industrialised nations (11–13%) (Rotimi *et al*, 2004). In addition, this region includes the *GBA* gene, coding for the housekeeping enzyme beta-glucocerebrosidase, which has mutations causing Gaucher disease; however, especially high frequencies of this disease have not been detected in Africa

(Goldblatt & Beighton, 1979). Therefore, the probability that these diseases would be selectively acting on this genomic region is lower than it is for malaria, denying the possibility of relevant selective confounding factors.

In summary, in this study, several results were obtained supporting the hypothesis that malaria is acting as a selective force in the *PKLR* gene region. Firstly, F_{ST} values between African and Portuguese populations using STR and SNP data from this specific fragment were considerably higher than those found using STR and SNP neutral markers, and the same was observed with AMOVA, revealing that this genomic section is under selection; secondly, the LD block included a more extensive region in the mild malaria group and a haplotype was found to be associated with this clinical group, suggesting that this conserved genomic block is associated with some protection against malaria severity. Thus, the output of this work, using human population data, seems to be in agreement with the results previously obtained with murine models and *in vitro Plasmodium* culturing. For future work, a larger number of samples from malaria status sets should be used and locus IVS3 should be carefully analysed. A more extensive field work with deeper phenotype discrimination and identification of PK abnormal alleles is currently under way.

Acknowledgements

We thank all individuals and parents/tutors of children who participated in this study and to all health technicians working at Emergency Services of the Paediatric Hospital David Bernardino (Luanda, Angola), Paediatrics Department of Central Hospital of Maputo, Health Centres of Bagamoyo and Boane (Maputo, Mozambique) for all technical support.

This study was supported by 'Financiamento Programático do Laboratório Associado CMDT.LA/IHMT' and POCI/SAU-ESP/55110/2004 (Fundação para a Ciência e Tecnologia/Ministério da Ciência, Tecnologia e Ensino Superior, FCT/MCTES, Portugal). P. Machado, R. Pereira and A. P. Arez were funded by FCT/MCTES Portugal (SFRH/BD/28236/2006, SFRH/BD/30039/2006 and SFRH/BPD/1624/2000—until 2007, respectively).

References

Agrafioti, I. & Stumpf, M.P. (2007) SNPSTR: a database of compound microsatellite-SNP markers. *Nucleic Acids Research*, **35** (Database issue), D71–D75.

Alves, C., Gomes, V., Prata, M.J., Amorim, A. & Gusmão, L. (2007) Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFlSTR YFiler) in Portugal. *Forensic Science International*, **171,** 250–255.

Alves, J., Machado, P., Silva, J., Gonçalves, N., Ribeiro, L., Faustino, P., do Rosário, V.E., Manco, L., Gusmão, L., Amorim, A. & Arez, A.P. (2010) Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers. *Blood Cells, Molecules and Diseases*, **44,** 62–68.

Anto, F., Asoala, V., Anyorigiya, T., Oduro, A., Adjuik, M., Owusu-Agyei, S., Dery, D., Bimi, L. & Hodgson, A. (2009) Insecticide resistance profiles for malaria vectors in the Kassena-Nankana district of Ghana. *Malaria Journal*, **8,** 81.

Ayi, K., Min-Oo, G., Serghides, L., Crockett, M., Kirby-Allen, M., Quirt, I., Gros, P. & Kain, K.C. (2008) Pyruvate kinase deficiency and malaria. *The New England Journal of Medicine*, **358,** 1805–1810.

Beutler, E. & Gelbart, T. (2000) Estimating the prevalence of pyruvate kinase deficiency from gene frequency in the general white population. *Blood*, **95,** 3585–3588.

Bruce-Chwatt, L.J. (1977) Malaria eradication in Portugal. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **71,** 232–240.

Campbell, M.C. & Tishkoff, S.A. (2008) African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annual Review of Genomics and Human Genetics*, **9,** 403–433.

Ceesay, S.J., Casals-Pascual, C., Erskine, J., Anya, S.E., Duah, N.O., Fulford, A.J.C., Sesay, S.S.S., Abubakar, I., Dunyo, S., Sey, O., Palmer, A., Fofana, M., Corrah, T., Bojang, K.A., Whittle, H.C., Greenwood, B.M. & Conway, D.J. (2008) Changes in malaria indices between 1999 and 2007 in The Gambia: a retrospective analysis. *Lancet*, **372,** 1545–1554.

Cuamba, N., Choi, K.S. & Townson, H. (2006) Malaria vectors in Angola: distribution of species and molecular forms of the Anopheles gambiae complex, their pyrethroid insecticide knockdown resistance (kdr) status and Plasmodium falciparum sporozoite rates. *Malaria Journal*, **5,** 2.

Das, S.K. & Elbein, S.C. (2007) The search for type 2 diabetes susceptibility loci: the chromosome 1q story. *Current Diabetes Reports*, **7,** 154–164.

Durand, P.M. & Coetzer, T.L. (2008) Pyruvate kinase deficiency protects against malaria in humans. *Haematologica*, **93,** 939–940.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1,** 47–50.

Goios, A., Gusmão, L., Rocha, A.M., Fonseca, A., Pereira, L., Bogue, M. & Amorim, A. (2008) Identification of mouse inbred strains through mitochondrial DNA single-nucleotide extension. *Electrophoresis*, **29,** 4795–4802.

Goldblatt, J. & Beighton, P. (1979) Gaucher disease in the Afrikaner population of South Africa. *South African Medical Journal*, **55,** 209–210.

Hill, A.V. (1998) Host genetics of infectious diseases: old and new approaches converge. *Emerging Infectious Diseases*, **4,** 695–697.

Holsinger, K.E. & Weir, B.S. (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, **10,** 639–650.

Kalinowski, S.T. (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. *Molecular Ecology Notes*, **5,** 187–189.

Kwiatkowski, D.P. (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *The American Journal of Human Genetics*, **77,** 171–192.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L. & Myers, R.M. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, **319,** 1100–1104.

Mabunda, S., Casimiro, S., Quinto, L. & Alonso, P. (2008) A country-wide malaria survey in Mozambique. I. Plasmodium falciparum infection in children in different epidemiological settings. *Malaria Journal*, **7,** 216.

Manco, L. & Abade, A. (2001) Pyruvate kinase deficiency: prevalence of the 1456C–>T mutation in the Portuguese population. *Clinical Genetics*, **60,** 472–473.

Manco, L., Ribeiro, M.L., Almeida, H., Freitas, O., Abade, A. & Tamagnini, G. (1999) PK-LR gene mutations in pyruvate kinase deficient Portuguese patients. *British Journal of Haematology*, **105,** 591–595.

Manco, L., Ribeiro, M.L., Máximo, V., Almeida, H., Costa, A., Freitas, O., Barbot, J., Abade, A. & Tamagnini, G. (2000) A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency. *British Journal of Haematology*, **110,** 993–997.

Manco, L., Oliveira, A.L., Gomes, C., Granjo, A., Trovoada, M., Ribeiro, M.L., Abade, A. & Amorim, A. (2001) Population genetics of four PKLR intragenic polymorphisms in Portugal and São Tomé e Princípe (Gulf of Guinea). *Human Biology*, **73,** 467–474.

Manco, L., Trovoada, M.J. & Ribeiro, M.L. (2009) Novel Human Pathological Mutations. Gene Symbol: PKLR Disease: pyruvate kinase deficiency. *Human Genetics*, **125,** 340.

Mateu, E., Perez-Lezaun, A., Martinez-Arias, R., Andres, A., Vallés, M., Bertranpetit, J. & Calafell, F. (2002) PKLR-GBA region shows almost complete linkage disequilibrium over 70 kb in a set of worldwide populations. *Human Genetics*, **110,** 532–544.

Min-Oo, G. & Gros, P. (2005) Erythrocyte variants and the nature of their malaria protective effect. *Cellular Microbiology*, **7,** 753–763.

Min-Oo, G., Fortin, A., Tam, M.F., Nantel, A., Stevenson, M.M. & Gros, P. (2003) Pyruvate kinase deficiency in mice protects against malaria. *Nature Genetics*, **35,** 357–362.

Noedl, H., Se, Y., Schaecher, K., Smith, B.L., Socheat, D. & Fukuda, M.M. (2008) Evidence of artemisinin-resistant malaria in western Cambodia. *The New England Journal of Medicine*, **359,** 2619–2620.

O'Meara, W.P., Bejon, P., Mwangi, T.W., Okiro, E.A., Peshu, N., Snow, R.W., Newton, C.R. & Marsh, K. (2008) Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet*, **372,** 1555–1562.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155,** 945–959.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. & Feldman, M.W. (2002) Genetic structure of human populations. *Science*, **298,** 2381–2385.

Rotimi, C.N., Chen, G., Adeyemo, A.A., Furbert-Harris, P., Parish-Gause, D., Zhou, J., Berg, K., Adegoke, O., Amoah, A., Owusu, S., Acheampong, J., Agyenim-Boateng, K., Eghan, Jr, B.A., Oli, J., Okafor, G., Ofoegbu, E., Osotimehin, B., Abbiyesuku, F., Johnson, T., Rufus, T., Fasanmade, O., Kittles, R., Daniel, H., Chen, Y., Dunston, G. & Collins, F.S. (2004) A genome-wide search for type 2 diabetes susceptibility genes in West Africans: the Africa America Diabetes Mellitus (AADM) Study. *Diabetes*, **53,** 838–841. Erratum in: *Diabetes*, **53**, 1404.

Rozen, S. & Skaletsky, H.J. (2000) Primer 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (ed. by S. Krawetz & S. Misener), pp. 365–386. Humana Press Inc, Totowa, New Jersey, USA.

Santos, N.P., Ribeiro-Rodrigues, E.M., Ribeiro-dos-Santos, A.K., Pereira, R., Gusmão, L., Amorim, A., Gerreiro, J.F., Zago, M.A., Matte, C., Hutz, M.H. & Santos, S.E. (2010) Assessing individual interethnic admixture and population substructure using a 48 insertion-deletion ancestry-informative marker panel. *Human Mutation*, **31,** 184–190.

Su, A., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R. & Hogenesch, J.B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, **101,** 6062–6067.

Tishkoff, S.A. & Verrelli, B.C. (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual Review of Genomics and Human Genetics*, **4,** 293–340.

Wang, H., Chu, W., Das, S.K., Ren, Q., Hasstedt, S.J. & Elbein, S.C. (2002) Liver pyruvate kinase polymorphisms are associated with type 2 diabetes in northern European Caucasians. *Diabetes*, **51,** 2861–2865.

Wang, H., Hays, N.P., Das, S.K., Craig, R.L., Chu, W.S., Sharma, N. & Elbein, S.C. (2009) Phenotypic and molecular evaluation of a chromosome 1q region with linkage and association to type 2 diabetes in humans. *Journal of Clinical Endocrinology & Metabolism*, **94,** 1401–1408.

WHO. (2008) *World Malaria Report 2008*. http://apps.who.int/malaria/wmr2008/malaria2008.pdf

Williams, T.N. (2006) Red blood cell defects and malaria. *Molecular and Biochemical Parasitology*, **149,** 121–127.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table SI. SNP loci selected for analysis (ordered according to localization), allelic frequencies and primers used for multiplex PCR.

Table SII. Single Base Extension (SBE) primers used for SNaPshot reaction.

Table SIII. STR loci allele frequencies found in Angola (ANG), Mozambique (MOZ), control Portuguese (PT-C) and PK-deficient Portuguese (PT-PKD).

Table SIV. SNP loci allelic frequencies observed in Angola, Mozambique and Portuguese groups.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Chapter 4 –

Pyruvate kinase deficiency in sub-Saharan Africa: identification of a highly frequent missense mutation (G829A;Glu277Lys) and association with malaria

This chapter was published as a research paper:

Machado, P., Manco, L., Gomes, C., Mendes, C., Fernandes, N., Salomé, G., Sitoe, L., Chibute, S., Langa, J., Ribeiro, L., Miranda, J., Cano, J., Pinto, J., Amorim, A., do Rosário, V.E. and Arez, A.P., 2012. Pyruvate kinase deficiency in sub-Saharan Africa: identification of a highly frequent missense mutation (G829A;Glu277Lys) and association with malaria. *PLoS One*, 7(10):e47071.

OPEN ACCESS Freely available online

PLOS ONE

Pyruvate Kinase Deficiency in Sub-Saharan Africa: Identification of a Highly Frequent Missense Mutation (G829A;Glu277Lys) and Association with Malaria

Patrícia Machado[1], Licínio Manco[2], Cláudia Gomes[1], Cristina Mendes[1], Natércia Fernandes[3], Graça Salomé[3], Luis Sitoe[3], Sérgio Chibute[3], José Langa[4], Letícia Ribeiro[5], Juliana Miranda[6], Jorge Cano[7], João Pinto[1], António Amorim[8,9], Virgílio E. do Rosário[1], Ana Paula Arez[1]*

1 Centro de Malária e outras Doenças Tropicais, Unidade de Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal, **2** Centro de Investigação em Antropologia e Saúde (CIAS), Universidade de Coimbra, Coimbra, Portugal, **3** Faculdade de Medicina da Universidade Eduardo Mondlane, Maputo, Mozambique, **4** Banco de Sangue do Hospital Central de Maputo, Maputo, Mozambique, **5** Departmento de Hematologia, Centro Hospitalar de Coimbra, Coimbra, Portugal, **6** Hospital Pediátrico David Bernardino, Luanda, Angola, **7** Centro Nacional de Medicina Tropical, Instituto de Salud Carlos III, Madrid, Spain, **8** Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP), Porto, Portugal, **9** Faculdade de Ciências da Universidade do Porto, Porto, Portugal

Abstract

Background: Pyruvate kinase (PK) deficiency, causing hemolytic anemia, has been associated to malaria protection and its prevalence in sub-Saharan Africa is not known so far. This work shows the results of a study undertaken to determine PK deficiency occurrence in some sub-Saharan African countries, as well as finding a prevalent PK variant underlying this deficiency.

Materials and Methods: Blood samples of individuals from four malaria endemic countries (Mozambique, Angola, Equatorial Guinea and Sao Tome and Principe) were analyzed in order to determine PK deficiency occurrence and detect any possible high frequent PK variant mutation. The association between this mutation and malaria was ascertained through association studies involving sample groups from individuals showing different malaria infection and outcome status.

Results: The percentage of individuals showing a reduced PK activity in Maputo was 4.1% and the missense mutation G829A (Glu277Lys) in the PKLR gene (only identified in three individuals worldwide to date) was identified in a high frequency. Heterozygous carrier frequency was between 6.7% and 2.6%. A significant association was not detected between either PK reduced activity or allele 829A frequency and malaria infection and outcome, although the variant was more frequent among individuals with uncomplicated malaria.

Conclusions: This was the first study on the occurrence of PK deficiency in several areas of Africa. A common PKLR mutation G829A (Glu277Lys) was identified. A global geographical co-distribution between malaria and high frequency of PK deficiency seems to occur suggesting that malaria may be a selective force raising the frequency of this 277Lys variant.

Citation: Machado P, Manco L, Gomes C, Mendes C, Fernandes N, et al. (2012) Pyruvate Kinase Deficiency in Sub-Saharan Africa: Identification of a Highly Frequent Missense Mutation (G829A;Glu277Lys) and Association with Malaria. PLoS ONE 7(10): e47071. doi:10.1371/journal.pone.0047071

Editor: Georges Snounou, Université Pierre et Marie Curie, France

Received July 3, 2012; **Accepted** September 7, 2012; **Published** October 17, 2012

Funding: This study was supported by PEst-OE/SAU/LA0018/2011 - Proj. Estratégico LA0018 2011/2012 (http://cmdt.ihmt.unl.pt/index.php/pt/) and PTDC/SAU-MET/110323/2009, "Fundacão para a Ciência e Tecnologia/Ministério da Educação e Ciência", FCT/MEC (http://alfa.fct.mctes.pt/index.phtml.pt), Portugal. PM holds a FCT grant (SFRH/BD/28236/2006). IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Education and Science, and is partially supported by Fundação para a Ciência e a Tecnologia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: aparez@ihmt.unl.pt

Introduction

Infectious diseases have been one of the major causes of mortality during most of human evolution. For many diseases, mortality and hence reproductive success are influenced by certain individual genotype. Consequently, some aspects of modern patterns of human genetic diversity should have been determined by diseases dating from prehistoric times [1]. The clearest example are provided by malaria, which even now affects 500 million people each year and kills some two million. The selective pressure that malaria has imposed to human populations has been reflected in dozens of molecular variants described as protective against the infection and disease [2–4]. Of these, the most well studied and widely accepted are probably the sickle cell allele (hemoglobin HbS allele), α and β thalassemias and glucose-6-phosphate (G6PD) deficiency (alleles A and A-), all showing an extensive overlap of geographical distribution and exceptionally high frequencies in malaria endemic regions.

Pyruvate kinase (PK) deficiency, caused by mutations in the pyruvate kinase, liver and RBC (PKLR) gene (chromosome 1q21) is one of the most recently described erythrocyte abnormalities associated to malaria. Evidences of its protective effect were

obtained both in murine models [5] and in *Plasmodium falciparum in vitro* cultures using human PK-deficient blood [6,7]. Also, population studies showed that a selective pressure is shaping the PKLR genomic region in individuals from malaria endemic countries (Cape Verde, Angola and Mozambique), being malaria infection the most likely driving force [8,9].

PK catalyzes the conversion of phosphoenolpyruvate (PEP) into pyruvate with the synthesis of ATP in the last step of glycolysis. PEP and pyruvate are involved in a great deal of energetic and biosynthetic pathways and the regulation of PK activity has proven to be of great importance for the entire cellular metabolism [10]. PK deficiency, worldwide distributed, is the most common enzyme abnormality in the erythrocyte glycolytic pathway causing hereditary chronic nonspherocytic hemolytic anemia. It is transmitted as an autossomal recessive trait and clinical symptoms usually occur in homozygotes and in compound heterozygotes for two mutant alleles. The clinical phenotype is heterogeneous, ranging from a mild chronic hemolytic anemia to a severe anemia presenting at birth and requiring exchange transfusion [11].

High frequencies of PK deficiency have not yet been recorded in malaria endemic areas but a systematic analysis has never been performed. Considering the previous knowledge of co-distribution between malaria endemicity and protective polymorphisms, we questioned if a PK variant could be exceptionally prevalent in malaria endemic areas. Therefore, the aims of the present study were: i) to determine PK deficiency occurrence in sub-Saharan African countries, ii) to assess frequency of PK variants underlying this deficiency, iii) to investigate possible associations between PK deficiency and malaria infection.

Materials and Methods

Sampling

This study is based on the molecular analysis of six sets of blood samples collected in four sub-Saharan African areas – Mozambique, Angola, Equatorial Guinea and Sao Tome and Principe (see Figure 1) - and in a malaria non-endemic area – Portugal (Europe).

In this study, 296 unrelated whole blood samples from individuals who attended to the Central Hospital of Maputo (Mozambique) between September and December 2008 were analyzed: 144 from children (6 months to 14 years-old) who presented to the Emergency Services of the Pediatric Department with some kind of complaint, and 152 from healthy blood donor adults (16 to 65 years-old) who presented to the Blood Bank. In order to increase the sample size of the set with a malaria outcome characterization, an additional group of 151 DNA samples extracted from blood samples collected from 3 months to 15 years-old children in Mozambique [9] was also genotyped.

In the Pediatric Department, blood was collected by venous puncture after the clinician examination but before the administration of any anti-malarial drug and/or blood transfusion. The registration of symptoms, axillary temperature and hemoglobin level was done for all individuals. Children who had received a blood transfusion in the last six months were excluded from the study. Anemic and *Plasmodium* infection status were considered at collection time. In the Blood Bank, the blood samples were randomly collected from blood donors. In the admission, a solubility test for rapid detection of hemoglobin S (adapted from Loh [12]) was performed in order to exclude allele S carriers. After blood collection in a tube, a blood spot in a filter paper was prepared from each sample for later subsequent DNA extraction by a standard phenol-chloroform method.

In addition to these samples from Mozambique, a set of 343 DNA samples from malaria-infected and non-infected unrelated individuals, which were already available from other studies, were also analyzed: 164 from Angola [9], 38 from Equatorial Guinea [13] and 67 from Sao Tome and Principe [14]. Finally, 74 samples from non-infected Portuguese individuals from all age groups were used as control samples [8]. Overall, 790 samples were analyzed.

Ethics statement

Regarding the survey in Mozambique, the human isolates collection was approved by local Ethical Committee (Comité Nacional de Bioética para a Saúde, Health Ministry of Mozambique, IRB 00002657, ref. 226/CNBS/08) and IHMT (Conselho de Ética do Instituto de Higiene e Medicina Tropical, CEIHMT, 14-2011-PN). A detailed work plan, questionnaires and informed consent forms were submitted to the Ethical Committees of the participant institutions in the study, which approved the survey. Each individual and parent/tutor of the children was informed of the nature and aims of the study and was told that participation was voluntary; written informed consent was obtained from each person (or parent/tutor). Blood sample collection followed strict requirements set by the Ethical Committees: blood samples from children who attended to the Pediatric Department were the remaining volume of the samples previously collected for the medical diagnosis; in the Blood Bank, during the blood donation, a small volume was put aside in a tube. In this way, no extra blood collection was needed and the patient, blood donor and the routine health services were not significantly disturbed. All ethical aspects related with the other sets of samples collected in previous studies, are described in the respective reports [8,9,13,14].

Plasmodium infection and malaria outcome groups

In the Central Hospital of Maputo, the rapid test OptiMAL-IT (DiaMed, Switzerland) was used for malaria diagnosis in all the patients with suspicion of malaria infection, and a blood smear was prepared for microscopic visualization to confirm diagnosis; later, all samples were amplified by Polymerase Chain Reaction (PCR), using *Plasmodium* species specific primers [15].

Malaria outcome was defined as follows: (i) Severe Malaria (SM): positive PCR for any species of *Plasmodium*, fever (i.e. axillary temperature $\geq$37,5°C), hemoglobin level of Hb$\leq$5 g/dL and/or any of these symptoms: coma, prostration or convulsions; (ii) Uncomplicated Malaria (UM): positive PCR for any *Plasmodium* species, fever and hemoglobin level of Hb$>$5 g/dL; and (iii) Asymptomatic Infection (AI): positive PCR for any *Plasmodium* species in the absence of fever (i.e. axillary temperature $<$37,5°C) or other symptoms of clinical illness; (iv) No infection (NI): negative PCR and absence of fever or other symptoms of clinical illness.

Based on malaria infection and symptoms data, the 144 samples from the Pediatric Department of Central Hospital of Maputo collected in 2008 were organized in the following malaria outcome groups: SM (18 samples); UM (27 samples) and NI (99 samples). The 152 samples from the Blood Bank were organized in the following groups: AI (4 samples) and NI (148 samples). Outcome groups were also defined using the same criteria for the set of isolates from Angola (43 SM, 43 UM, 37 AI and 41 NI) and for the set of isolates previously collected in Mozambique (52 SM, 97 UM and 2 NI), both described in Machado *et al.* [9]. In total, we had 611 samples with malaria infection and outcome characterization - 459 samples from children (113 SM, 167 UM, 37 AI and 142 NI) and 152 samples from adults (4 AI and 148 NI).

Figure 1. Geographic location of the countries Mozambique, Angola, Sao Tome and Principe, Equatorial Guinea (Africa), Pakistan (Asia) and Portugal (Europe).
doi:10.1371/journal.pone.0047071.g001

Determination of PK activity

PK activity was measured in lyzed erythrocytes from all the 296 fresh blood samples (after plasma and buffy coat strict removal) collected in Mozambique in 2008, with an enzymatic assay adapted from Beutler [16], according to the instructions of the kit "Determination of pyruvate kinase (EC 2.7.1.40) in erythrocytes hemolysate or serum/heparinized plasma" (Instruchemie, The Netherlands). The enzymatic reactions were running at room temperature. A PK-deficient and a normal control were used in each assay to validate the activity values and to classify the samples within the following phenotypes: normal, intermediate or deficient activity.

Identification of a PK variant underlying PK-reduced activity

Samples with a PK activity value less than or equal to 75% of the normal control sample activity were analyzed by the Single Strand Conformational Polymorphism (SSCP) method (described in Manco *et al.* [17]) in order to find a mutation associated with this phenotype. The promoter region and eleven exons of the PKLR gene were amplified with specific primers (see Table S1, supporting information) and run in an acrylamide-bisacrylamide gel (10%), together with a wild-type amplicon, to detect differences in migration patterns caused by an alteration in DNA chain composition (exon 2 was not analyzed since it is specific for the hepatic isoenzyme). The amplification conditions were: initial denaturation at 94°C for 5 minutes, followed by 35 cycles of 94°C for 45 seconds, a specific annealing temperature for 45 seconds (see Table S1), and 72°C for 1 minute, with a final extension at 72°C for 5 minutes. The samples with a different migration pattern were further analyzed by automatic DNA sequencing (Macrogen Inc., Korea). The exon 7, in which a mutation was identified, was then amplified in all samples from all groups by PCR with the specific primers and conditions indicated in Table S1 and the amplicons were sequenced (Macrogen Inc., Korea).

Statistical analysis

The association between alleles and malaria outcome groups was assessed by Pearson's chi-square tests and Fisher's exact test, this latter considered when there were a few cases in each comparison group (less than five), using the Simple Interactive Statistical Analysis software (SISA) [18]. Odds ratios (OR) and 95% confidence intervals (CI) were also estimated using SISA. Arlequin 3.1 software [19] was used to determine allele frequencies, population pairwise F_{ST} (to test for differentiation between populations), expected and observed values of heterozygosity and to perform Hardy–Weinberg equilibrium tests. Prediction of the possible impact of the amino acid substitution on the structure and function of the human PK protein was performed with the Polyphen software [20]. Finally, PyMol software [21] was used for the 3D structure simulation of the wild type and mutant variants.

Results

PK deficiency screening in Maputo, Mozambique

Ninety-eight from the 144 samples collected in the Pediatric Department (68%) in Mozambique in 2008 were from children with a hemoglobin concentration <9 g/dL (considered anemic) and 41 samples (28.5%) were infected with *P. falciparum*. Nineteen of the infected individuals were also anemic. Four (2.6%) of the 152 samples from the adult blood donors in Blood Bank showed an asymptomatic infection with *P. falciparum* (see Table 1)

From the 296 samples set, 12 (4.1%) presented PK activity values between 39% and 75% of the normal control activity (established in an average of 3.2 U/g Hb) (see Table 2): 8 from the Blood Bank (5.3%) and 4 from the Pediatrics (2.8%). They were all classified as intermediate activity phenotype. From the 98 samples with a hemoglobin level <9 g/dl (Pediatric Department), only 3 (3.1%) had a PK reduced activity.

Identification of a PK variant underlying PK-reduced activity

A migration pattern alteration was observed in the amplicon of exon 7 of 5 out of 12 samples with low activity (41.7%) by SSCP (see Figure 2): 4 from blood donors and 1 from Pediatrics. Sequencing of these 5 amplicons revealed a G>A substitution in all of them, being in homozygosis (A/A) in one sample. This is a non-synonymous mutation located in the nucleotide 829 of the PK mRNA sequence originating an alteration of the amino acid 277 of the PK protein: a glutamic acid (Glu, coded by GAG) is replaced by a lysine (Lys, coded by AAG). When this mutation was searched in all the other 284 samples with normal activity, it was detected in heterozygosis in 16 samples: 7 from children and 9 from blood donors. Overall, 21 samples (7.1%) had the 829A allele that displayed a frequency of 3.7%.

No association was found between the 829A allele and anemia (2.7–9 g/dL Hb). Conversely, a strong association was found between the allele 829A and PK deficient activity: $\chi^2 = 14.38$ ($P<0.00$), OR = 5.58 (95% CI: 2.07–15.03). Of the 6 samples with the lowest PK activity values (between 39% and 47% of the normal activity), 5 had the mutation. All the 6 other samples with an activity between 47% and 75% of the normal activity were wild type.

As visualized in the 3D PK structure simulation (see Figure 3), this 277 residue is exposed, showing a peripheral position. The prediction of the substitution Glu277Lys effect on the structure and function of the human protein PK was "Possibly Damaging" (score of 0.90) supporting the previous OR result and suggesting that this mutation is likely to be non-functional.

Searching the mutation G829A in other African malaria endemic areas

The mutation G829A was found in the other three African countries, always in heterozygosis: in 11 samples from Angola (6.7%), 1 sample from Equatorial Guinea (2.6%) and 2 samples from Sao Tome and Principe (3.0%). Allele 829A frequencies were 3.4%, 1.3% and 1.5%, respectively. In the Mozambican group from 2005, the frequency of individuals heterozygous for 829A was 5.3%, giving an allele frequency of 2.6%. The mutation was not found in the control group from Portugal. Considering all the Mozambican 447 samples, a frequency of carrier individuals of 5.8% and 829A allele frequency of 3.0% were estimated.

The observed genotype frequencies (829GG, 829AG and 829AA) were according to Hardy-Weinberg expectations for all populations ($P = 0.40$ in Mozambique; $P = 1.00$ in Angola, Equatorial Guinea and Sao Tome and Principe). Estimates of F_{ST} were non-significant between all pairs of African populations ($F_{ST} \leq 0.00$ for all) ($P = 1.00$ for Mozambique vs. Angola; $P = 0.50$ for Mozambique vs. Equatorial Guinea; $P = 0.30$ for Mozambique vs. Sao Tome and Principe; $P = 0.51$ for Angola vs. Equatorial Guinea; $P = 0.35$ for Angola vs. Sao Tome and Principe; and $P = 1.00$ for Equatorial Guinea vs. Sao Tome and Principe).

Association among PK-reduced activity, the mutation G829A and malaria infection/outcome

Six-hundred and eleven DNA samples belonging to individuals characterized for their infection and malaria disease outcome status were analyzed: 459 samples from children (113 SM, 167 UM, 37 AI and 142 NI) from Angola and Mozambique and 152 samples from adults (4 AI and 148 NI), from Mozambique. No significant differentiation between samples from Angola and Mozambique were observed, so all samples together were considered for this analysis.

Allele 829A frequencies were as follows (see Table 3): in children, 3.1% in SM, 3.3% in UM, 2.7% in AI and 2.5% in NI; in adults 4.4% in NI. In terms of malaria infection in children, allele A frequencies were 3.2% in infected and 2.5% in non-infected. In adults, this analysis in terms of infection was not considered due to the low number of infected individuals. Although the mutation frequency was higher in uncomplicated (UM) than in severe malaria (SM) group, no significant association was observed between 829A allele and disease outcome ($\chi^2 = 0.02$, $P = 1.00$; OR = 1.07, 95% CI: 0.41–2.80). No significant association was found either between 829A allele and infection ($\chi^2 = 0.33$, $P = 0.57$; OR = 1.29, 95% CI: 0.54–3.08) or between PK deficient activity (low enzyme activity) and infection ($P = 0.30$), though 11 from the 12 samples with PK reduced activity were non-infected.

Table 1. PK activity, anemia and *Plasmodium* infection status in the sample set from Maputo, Mozambique (2008).

	Pediatrics	Blood Bank	Total
Age Group	Children (6 months–14 years old); with some complaint	Adults (16–65 years old); healthy blood donors	6 months–65 years old
Nr of samples	144	152	296
Low PK activity (39–75% of control)	4 (2.8%)	8 (5.3%)	12 (4.1%)
Anemia (Hb<9 g/dL)	98 (68.1%)	n.d.	n.d.
***Plasmodium* infection**	41 (28.5%)	4 (2.6%)	45 (15.2%)
Anemia+Infection	19 (13.2%)	n.d.	n.d.

n.d.: not determined.

doi:10.1371/journal.pone.0047071.t001

Table 2. Samples with a reduced PK activity (between 39 and 75% of the normal control) and respective infection status and malaria outcome and 829 locus genotype.

				PK Activity U/g Hb					
#	Sample	Assay	Activity	Average	Control N	Average/Control N	Control DEF	Inf/Malaria outcome	829G/A
1	**BS_128**	1	1.69	**1.69**	3.48	**0.49**	0.85	NI	GG
2	**BS_176**	1	1.88						
	BS_176	2	1.93	**1.91**	3.48	**0.55**	0.85	NI	GG
3	**BS_197**	1	1.56						
	BS_197	2	1.34	**1.45**	3.48	**0.42**	0.85	NI	**GA**
4	**BS_199**	1	1.73						
	BS_199	2	0.99	**1.36**	3.48	**0.39**	0.85	NI	**GA**
5	**BS_212**	1	1.85						
	BS_212	2	1.43	**1.64**	3.48	**0.47**	0.85	NI	**GA**
6	**BS_220**	1	1.35						
	BS_220	2	1.52	**1.44**	3.48	**0.41**	0.85	NI	GG
7	**BS_230**	1	1.46						
	BS_230	2	1.59	**1.53**	3.48	**0.44**	0.85	NI	**AA**
8	**BS_327**	1	1.74						
	BS_327	2	1.96	**1.85**	3.48	**0.53**	0.85	NI	GG
9	**N_1159**	1	1.93						
	N_1159	2	2.27	**2.10**	2.91	**0.72**	0.73	NI	GG
10	**N_1391**	1	2.19	**2.19**	2.91	**0.75**	0.73	NI	GG
11	**N_1464**	1	1.69	**1.69**	2.91	**0.58**	0.73	NI	GG
12	**O_2341**	1	1.35	**1.35**	2.91	**0.46**	0.73	SM	**GA**

BS: samples collected in the Blood Bank; O and N: samples collected in the Department of Pediatrics; Inf/Malaria outcome: infection status and malaria outcome; 829G/A: 829 genotype; NI: non-infected; SM: severe malaria.
doi:10.1371/journal.pone.0047071.t002

Discussion

This is the first study aimed at determining PK deficiency occurrence as well as at studying a potential widespread PKLR mutation in the African continent.

Figure 2. SSCP results showing a migration pattern alteration in the exon 7 amplicons caused by the G829A substitution (10% acrylamide-bisacrylamide gel) - samples at the extremes (wild type isolate in the middle).
doi:10.1371/journal.pone.0047071.g002

In the first instance, PK deficiency was studied in samples from Maputo, Mozambique, measuring PK activity in anemic individuals, as this is described as a symptom of the disease. However, anemia was neither associated to PK reduced activity nor 829A allele. The overall prevalence rate of PK reduced activity was 4.1% in the study population (5.3% from blood donors and 2.8% from children). Although children samples were, most of them, clinical cases with a considerable anemic status, a higher PK deficiency prevalence was not found in these samples and no association was detected between PK low activity and anemia. In this regard, a study carried in 2002 revealed that 74% of the children under five and 50% of the women in reproductive age from Mozambique was anemic [22], showing that anemia is not a proper indicator of erythrocyte deficiencies in developing countries.

The missense mutation G829A (Glu277Lys) was identified in 41.7% of Mozambican PK deficient isolates with a strong association with reduced activity phenotype. This mutation was then searched in additional Mozambican samples and other sub-Saharan regions and the 829A allele was detected in all of them at allele frequencies between 1.3% (in Equatorial Guinea) and 3.4% (in Angola). The allele 829A was not present in the Portuguese samples. Although two African groups could be established according to these frequencies (Angola and Mozambique with higher frequencies vs. Equatorial Guinea and Sao Tome and Principe with lower frequencies), F_{ST} values were not significantly different between them. These differences may be explained by sample size bias (447 samples from Mozambique and 164 from

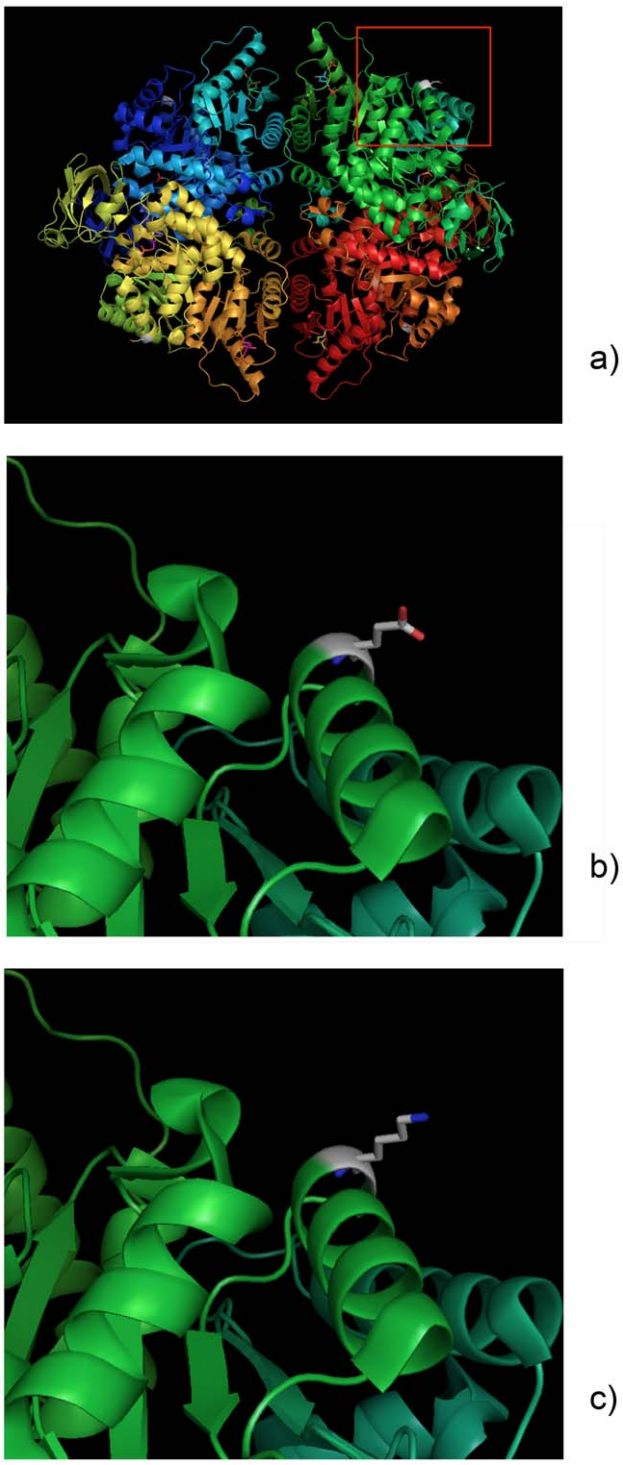


Figure 3. Location of the amino acid 277 in the PK protein and simulation of the 3D wild type 277Glu and mutant 277Lys PK variants structure with the software PyMol. a) Peripheral position of the amino acid 277 (domain A); b) Wild type variant 277Glu; c) Mutant variant 277Lys.
doi:10.1371/journal.pone.0047071.g003

Angola were processed against 38 from Equatorial Guinea and 64 from Sao Tome and Principe) or design bias (isolates from Mozambique and Angola were obtained in hospital-based studies, whereas the others were collected in households by active search). In addition, genetic substructure among geographic regions cannot be excluded as a hypothesis for this disparity. Differences in malaria selective pressure are not a probable cause, since it has probably been similar in all these regions in the past.

Prevalence of PK deficiency seems to vary greatly among ethnic groups and geographic regions, as well as the mutations in the PKLR gene. Some authors have estimated a prevalence of 1:20 000 in the general white population [23]. In Europe, an incidence of 3.3 per million has been reported in the north of England [24], and a prevalence of 0.24% and 1.1% have been described in Spain [25] and Turkey [26], respectively. In Asia, the frequency of PK deficiency among the Hong Kong Chinese population was <0.1% [27] whilst among the south Iranian population was 1.9% [28]. In Saudi Arabia, a prevalence of 3.12% was registered in newborns [29]. These studies were all based in PK activity measurements. The estimated mutant allele frequencies of common variants generally vary between 0.2 and 0.8% [23] with the highest heterozygous prevalence described so far in Saudi Arabia (6%) [28,30] and Hong Kong (3.4%) [31]. However, these last allele frequencies were not calculated from mutation genotyping but only estimated from the Beutler's screening qualitative procedure and enzyme assay [16], which result in less reliable estimates of heterozygosity. Moreover, consanguinity is extremely high in Saudi Arabia, exceeding 80% in some regions [29], which tends to bias the results.

The PK deficiency recorded in Mozambique (4.1%) and 829GA heterozygous prevalence (2.6–6.7%) determined from unrelated individuals from sub-Saharan populations is, to our knowledge, the highest estimated worldwide so far. We initially hypothesized that this would be the result of a strong malaria pressure, but a significant association between both PK low activity and 829A and malaria infection and outcome was not found. However, only 12 samples were available for testing a possible effect of low enzyme activity on severity of malaria and 20 samples for testing a possible effect of 829A allele meaning that larger numbers are required to formally conclude. Moreover, since this was a cross-sectional study, infection and malaria outcome groups were established according to a malaria phenotype in a specific time point (the collection day), that may not accurately reflect the true individual phenotype. Nevertheless, there was higher mutation prevalence in the uncomplicated malaria group supporting that further analysis is essential to complete the present study.

The Glu277Lys mutation here identified has been previously reported in the PKLR mutation database [32] and has recently been described [30] in only two individuals: one from the Mandenka ethnic group (one of the largest ethnic groups in West Africa) and other from the Brahui ethnic group from Pakistan, showing that is also present in Middle East. Since the haplotypes that include this mutation in these two individuals are different, it was suggested that it has arisen separately. In Pakistan, as in sub-Saharan countries, malaria continues to be a major public health problem. Both *P. falciparum* and *Plasmodium vivax* are widely distributed and the estimated number of annual malaria episodes in this country is 1.5 million [33].

The simulation of this Glu277Lys substitution on the human PK protein suggested that this mutation is likely to be non-functional. This residue is extremely well conserved and the result complies with the prediction from SIFT from a previous work [30]. Probably, the charge change (Glu is negatively whereas Lys is positively charged) at an exposed site alters the enzyme action. Considering this result together with the knowledge about PK deficiency that clinical symptoms usually occur in homozygotes for a mutant PKLR allele, it was surprising to find that the 829AA genotype belonged to a healthy blood donor without anemia

Table 3. Allele 829A frequencies in infection and malaria outcome groups.

	CHILDREN[1]			ADULTS[2]		
Infection/Clinical group	**Samples**	**829A carriers**	**829A frequency**	**Samples**	**829A carriers**	**829A frequency**
SM	113	7 (6.2%)	3.1%	0	0 (0%)	0 (0%)
UM	167	11 (6.6%)	3.3%	0	0 (0%)	0 (0%)
AI	37	2 (5.4%)	2.7%	4	0 (0%)	0 (0%)
NI	142	7 (4.9%)	2.5%	148	13[3] (8.8%)	4.7%
INF (SM+UM+AI)	317	20 (6.3%)	3.2%	4	0 (0%)	0 (0%)
TOTAL	459	27 (5.9%)	2.9%	152	13 (8.6%)	4.6%

[1]Samples from children well characterized for infection and malaria outcome status from Maputo, Mozambique (collected within this study and in a previous one [9]) and from Angola (collected previously [9]) who attended to the Pediatrics Department.
[2]Samples from adult blood donors from Maputo, Mozambique (collected within this study).
[3]Including one 829AA homozygote (the only one identified in the study).
SM: severe malaria; UM: uncomplicated malaria; AI: asymptomatic infection; NI: non-infected; INF: infected.
doi:10.1371/journal.pone.0047071.t003

symptoms, with a PK activity of 0.44 with regard to the normal control. In this case we were expecting an activity similar to the deficient control sample (0.8 U/g Hb). However, the results obtained regarding PK activity must carefully be considered since the range of values obtained in Mozambique was narrow, far below the values expected with the use of the kit and generally obtained in other labs (about 3.7–8.2 U/g Hb at 25°C and about 7.4–16.4 U/g Hb at 37°C), with a thin gap between normal and reduced activity. This can be explained by the lower room temperature in the lab (about 20°C), when compared to those generally maintained in this procedure (25°C or 37°C). Yet, the procedure was efficient since it was possible to identify samples with reduced activity. Actually, there was no direct relation between the genotype and phenotype: although a significant association between 829A and a reduction in the enzyme activity was found out (and the samples with the lowest activity were those ones with the 829A allele), the phenotype of allele A carriers was highly variable with a large number of individuals within normal PK activity range. A previous study emphasizes the difficulty in predicting the consequences of mutations simply from the location and the nature of the target residues [10]: the clinical manifestations of a genetic disease reflect the interactions of a variety of physiological and environmental factors, including genetic background, concomitant functional polymorphisms of other enzymes, posttranslational or epigenetic modifications, ineffective erythropoiesis and differences in splenic function, and do not solely depend on the molecular properties of the altered molecule.

To conclude, a geographical co-distribution between malaria and PK-deficiency seems to occur: the Middle East and sub-Saharan Africa are the regions with the highest PK deficiency prevalence described so far, as determined in the present study. These are regions with a strong malaria pressure, suggesting that malaria may be an agent of contribute to the selection of PK deficiency variants in these regions. Conversely, the prevalence of PK deficiency is extremely low in the general white populations. Moreover, some of the genes that confer resistance to malaria are among the most variable genes in the human genome [4] and this is the case for PKLR gene, which presents more than 180 mutations and 8 polymorphic sites [11].

Additional studies with a larger sampling effort including longitudinal malaria clinical history characterization and a search of the variant 277Lys in other malaria endemic regions will be conducted to clarify the results in this survey.

Supporting Information

Table S1 List of primers and annealing temperatures (a.t.) used in the amplification of *PKLR* promoter (Prom) and coding regions by PCR.
(DOCX)

Acknowledgments

Authors would like to express their gratitude to Dra. Umbelina Rebelo, Dra. Celeste Bento and Dr. Luís Relvas from the Hematology Department, Centro Hospitalar de Coimbra, as well as to Dra. Isabel Abergaria and the technicians from the Clinical Chemistry Lab, Instituto Nacional de Saúde Dr. Ricardo Jorge (Portugal) for all their help concerning the methodologies and protocols for PK assays. The authors also want to thank to João Rodrigues for doing the 3D structure simulation of the PK variants with PyMol software. Deep appreciation for the contribution of D. Violeta and Sábado from the Pediatric Lab and all the technicians from the Blood Bank (Central Hospital of Maputo, Mozambique) for collecting blood samples and to all volunteers that agreed in participate in the present study. Very special thanks to Natacha, Antónia, Dida and Juliana, Filipa and Pedro for their unconditional support during the stay at Mozambique.

Author Contributions

Conceived and designed the experiments: APA. Performed the experiments: PM CG CM LM. Analyzed the data: PM APA. Contributed reagents/materials/analysis tools: APA LM AA. Wrote the paper: PM APA. Did the field work at Mozambique (2008): PM GS JL LS NF SC. Processed the biological material and data collection in Mozambique, Angola, Sao Tome and Principe, Equatorial Guinea and Portugal, respectively: NF JM JP JC AA. Contributed with a critical review of the paper: AA CM JC JP LM LR SC VdR.

References

1. Jobling MA, Hurles ME, Tyler-Smith C (2004) Human evolutionary genetics: origins, peoples and disease. Garland Science, New York.
2. Verra F, Mangano VD, Modiano D (2009) Genetics of susceptibility to *Plasmodium falciparum*: from classical malaria resistance genes towards genome-wide association studies. Parasite Immunol 31: 234–253.
3. Allison AC (2009) Genetic control of resistance to human malaria. Curr Opin Immunol 21: 499–505.
4. Hedrick PW (2011) Population genetics of malaria resistance in humans. Heredity 107: 283–304.
5. Min-Oo G, Fortin A, Tam MF, Nantel A, Stevenson MM, et al. (2003) Pyruvate kinase deficiency in mice protects against malaria. Nat Genet 35: 357–362.

6. Ayi K, Min-Oo G, Serghides L, Crockett M, Kirby-Allen M, et al. (2008) Pyruvate kinase deficiency and malaria. N Engl J Med 358: 1805–1810.
7. Durand PM, Coetzer TL (2008) Pyruvate kinase deficiency protects against malaria in humans. Haematologica 93: 939–940.
8. Alves J, Machado P, Silva J, Gonçalves N, Ribeiro L, et al. (2010) Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers. Blood Cells Mol Dis 44: 62–68.
9. Machado P, Pereira R, Rocha AM, Manco L, Fernandes N, et al. (2010) Malaria: looking for selection signatures in the human PKLR gene region. Br J Haematol 149: 775–784.
10. Valentini G, Chiarelli LR, Fortin R, Dolzan M, Galizzi A, et al. (2002) Structure and function of human erythrocyte pyruvate kinase. Molecular basis of nonspherocytic hemolytic anemia. J Biol Chem 277: 23807–23814.
11. Zanella A, Fermo E, Bianchi P, Chiarelli LR, Valentini G (2007) Pyruvate kinase deficiency: the genotype-phenotype association. Blood Rev 21: 217–231.
12. Loh WP (1968) A new solubility test for rapid detection of haemoglobin. J Indiana State Med Assoc 61: 1651–1652.
13. Mendes C, Dias F, Figueiredo J, Mora VG, Cano J, et al. (2011) Duffy negative antigen is no longer a barrier to *Plasmodium vivax* - molecular evidences from the African West Coast (Angola and Equatorial Guinea). PLoS Negl Trop Dis 5: e1192.
14. Pinto J, Sousa CA, Gil V, Ferreira C, Gonçalves L, et al. (2000) Malaria in São Tomé and Príncipe: parasite prevalences and vector densities. Acta Trop 76: 185–193.
15. Snounou G, Viriyakosol S, Jarra W, Thaithong S, Brown KN (1993) Identification of the four human malaria parasite species in field samples by the polymerase chain reaction and detection of a high prevalence of mixed infections. Mol Biochem Parasitol 58: 283–292.
16. Beutler E (1984) Red Cell Metabolism: A Manual of Biochemical Methods. Grune & Stratton, Philadelphia, PA.
17. Manco L, Ribeiro ML, Almeida H, Freitas O, Abade A, et al. (1999) PK-LR gene mutations in pyruvate kinase deficient Portuguese patients. Br J Haematol 105: 591–595.
18. Simple Interactive Statistical Analysis software, SISA. Available: http://www.quantitativeskills.com/sisa/. Accessed 2012 Jun 29.
19. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47–50. Available: http://cmpg.unibe.ch/software/arlequin3/. Accessed 2012 Jun 29.
20. Polyphen software. Available: http://genetics.bwh.harvard.edu/php/. Accessed 2012 9 Aug 9.
21. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. Available: http://www.pymol.org/. Accessed 2012 Jun 29.
22. Ministério da Saúde, Direcção Geral da Saúde, República de Moçambique (2002) Moçambique: Investir na Nutrição é Reduzir a Pobreza. Análise das Consequências dos Problemas Nutricionais nas Crianças e Mulheres. Maputo.
23. Beutler E, Gelbart T (2000) Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population. Blood 95: 3585–3588.
24. Carey PJ, Chandler J, Hendrick A, Reid MM, Saunders PW, et al. (2000) Prevalence of pyruvate kinase deficiency in northern European population in the north of England. Northern Region Haematologists Group. Blood 96: 4005–4006.
25. García SC, Moragón AC, López-Fernández ME (1979) Frequency of glutathione reductase, pyruvate kinase and glucose-6-phosphate dehydrogenase deficiency in a Spanish population. Hum Hered 29: 310–313.
26. Akin H, Baykal-Erkiliç A, Aksu A, Yücel G, Gümüşlü S (1997) Prevalence of erythrocyte pyruvate kinase deficiency and normal values of enzyme in a Turkish population. Hum Hered 47: 42–46.
27. Feng CS, Tsang SS, Mak YT (1993) Prevalence of pyruvate kinase deficiency among the Chinese: determination by the quantitative assay. Am J Hematol 43: 271–273.
28. Yavarian M, Karimi M, Shahriary M, Afrasiabi AR (2008) Prevalence of pyruvate kinase deficiency among the south Iranian population: quantitative assay and molecular analysis. Blood Cells Mol Dis 40: 308–311.
29. Abu-Melha AM, Ahmed MA, Knox-Macaulay H, Al-Sowayan SA, el-Yahia A (1991) Erythrocyte pyruvate kinase deficiency in newborns of eastern Saudi Arabia. Acta Haematol 85: 192–194.
30. Berghout J, Higgins S, Loucoubar C, Sakuntabhai A, Kain KC, et al. (2012) Genetic diversity in human erythrocyte pyruvate kinase. Genes Immun 13: 98–102.
31. Fung RH, Keung YK, Chung GS (1969) Screening of pyruvate kinase deficiency and G6PD deficiency in Chinese newborn in Hong Kong. Arch Dis Child 44: 373–376.
32. University Medical Center. (2007) PKLR Mutation Database. Laboratory for Red Blood Cell Research: Ultrecht. Available: http://www.pklrmutationdatabase.com/. Accessed 2012 Jun 29.
33. WHO EMRO (2011) World Health Organization, Regional Office of the Eastern Mediterranean, Epidemiological Situation, Country Profiles. Available: http://www.emro.who.int/rbm/CountryProfiles-pak.htm. Accessed 2012 Jun 29.

Chapter 5 –

Quantitative proteomics approach for the analysis of the human malaria parasite *Plasmodium falciparum* (trophozoite stage) and its red blood cell host – a preliminary study

This chapter is a paper in preparation:

Machado, P., Nogueira, F., Rodrigues, J., Manco, L., Ribeiro, L., Bergstrom, E., Ashford D., Vitorino, R., Thomas-Oates, J., Thomas, J., Arez, A.P., 2013. *In prep*.

Quantitative proteomics approach for the analysis of the human malaria parasite *Plasmodium falciparum* (trophozoite stage) and its red blood cell host – a preliminary study

Patrícia Machado[1], Fátima Nogueira[1], João Rodrigues[1], Licínio Manco[2], Letícia Ribeiro[3], Ed Bergstrom[4], David Ashford[4], Rui Vitorino[5], Jane Thomas-Oates[4], Jerry Thomas[4], Ana Paula Arez[1]

[1]Instituto de Higiene e Medicina Tropical, Unidade de Parasitologia Médica, Rua da Junqueira, 100, 1349-008 Lisboa, Portugal

[2]Centro de Investigação em Antropologia e Saúde (CIAS), Universidade de Coimbra, Coimbra, Portugal

[3]Departmento de Hematologia, Centro Hospitalar de Coimbra, Coimbra, Portugal

[4]Centre of Excellence in Mass Spectrometry, University of York, York, United Kingdom

[5]Centro de Espectrometria de Massa, Universidade de Aveiro, Aveiro, Portugal

ABSTRACT

In the last years, we have provided some data supporting the association between malaria and PK deficiency in humans, which resulted from human population studies. Proteomic information from *Plasmodium* infection is scarce and there are no studies characterizing the total proteome of infected red blood cells (RBC). Moreover, the proteome of both PK and G6PD-deficient RBC and from parasites growing in these cells have not been characterized. Considering all these, we performed a proteomic study in which we intended to detect the relative abundance of proteins from both PK- and G6PD-deficient RBC, as also from *Plasmodium* parasites infecting these cells. These would retrieve key information about malaria dynamics but also about enzyme deficiencies causing important hemolytic anemias. Up to now, only results from the parasite proteome (trophozoyte stage) are available. In parasites growing in G6PD-deficient RBC there was an over-expression of defensive molecules against oxidative stress (heat shock proteins and chaperones); in parasites growing in PK-deficient RBC (severe phenotype) a general protein under-expression was observed, with the proteins involved in hemoglobin catabolism and trafficking/RBC remodelling being the most affected. The influence of these alterations in the protective mechanisms against malaria are discussed.

INTRODUCTION

The malaria parasite has a complex and multistage life cycle developing in two hosts: the humans and the *Anopheles* mosquito. In humans, the parasite develops asexually, resulting in proliferation in the blood stream within the red blood cells (RBC), going through ring, trophozoite, schizont and merozoite stages, or it develops into a male or female gametocyte (the sexual precursor forms) that, after ingestion by the mosquito during a blood meal, develop into mature gametes that fertilize to form zygotes. Repeated periodic cycles of parasitic development occur within the RBC (48h in the case of *Plasmodium falciparum*) causing the clinical symptoms of the disease.

The completion of the *P. falciparum* 3D7 genome sequencing (Gardner, et al., 2002) and the significant advances in mass spectrometry (MS) techniques over the past decade have provided the basis for proteomics studies on malaria. During the last five years, these experiments mostly enumerated proteins but today quantitative measurements are performed in practically all studies (quantitative MS proteomics reviewed in Bantscheff, et al., 2012). Such proteome surveys are bringing to light the substantial role of regulatory processes occurring after mRNA is made (posttranscriptional, translational and degradation regulation) in the determination of protein concentrations, contributing at least as much as transcription itself (Vogel and Marcotte, 2012).

Ten years ago, two studies were simultaneously published analyzing the proteome of several *P. falciparum* stages by high-accuracy MS (Florens, et al., 2002; Lasonder, et al., 2002) and since then, an increasing number of *Plasmodium* MS proteomic investigations have been performed (Nirmalan, Sims and Hyde, 2004; Hall, et al., 2005; Gelhaus, et al., 2005; Acharya, et al., 2009; Smit, et al., 2010). *Plasmodium falciparum* has a 23-megabase nuclear genome organized in 14 chromosomes, with 5 268 protein-encoding genes identified. About 60% (3 208 hypothetical proteins) of those predicted proteins did not have sufficient similarity to proteins in other organisms to justify provision of functional assignments. Thus, almost two-thirds of the proteins appear to be unique to this organism (Gardner, et al., 2002). Ten years later, most *Plasmodium* proteins remain with unknown function, confirming this hypothesis (Oehring, et al., 2012) and showing that this is a really peculiar organism. The number

of proteins detected to date in the proteome analysis of *Plasmodium* asexual forms (sporozoites, merozoites, trophozoites, schizonts and gametocytes) is about 2 500 (including hypothetical proteins, with or without known function). Just over half were found in one-stage only, suggesting that stage-specific specialization is substantial, and only 6% were common to sporozoites, merozoites, trophozoites and gametocytes (Florens, et al., 2002). Proteome investigations in *Plasmodium* growing in specific conditions have focused on protein expression under drug treatment (Prieto, et al., 2008; Briolant, et al., 2010) and on specific malaria pathways, as those involved in invasion (Kuss, et al., 2012). In this latter study, it was observed that the malaria parasite is able to adapt to variations in the host cell environment by posttranscriptional regulation, emphasizing the importance of proteomic studies for the knowledge of the biology of the parasite.

The RBC proteome has also been explored (RBC proteomics reviewed in D'Alessandro, Righetti and Zolla, 2010). Mature RBC have a life span of approximately 120 days and are optimally adapted for oxygen and carbon dioxide as well as for proton transport. They consist of a plasma membrane that envelopes a concentrated (33%) solution of proteins of which hemoglobin constitutes approximately 98% of the global proteome. The absence of nucleus and the loss of cytoplasmic organelles allow the RBC passing through narrow capillaries, with a concomitant drastic shape change, to properly accomplish its most important biological tasks (Roux-Dalvai, et al., 2008).

Very recently, RBC proteome analysis has been extended to infection with *Plasmodium* in order to detect changes induced by the parasite. Sicard, et al. (2011) detected the activation of a PAK-MEK signaling pathway in infected RBC that may be involved in the regulation of ion transport or membrane mechanical properties. Fontaine, et al. (2012) described host protein modifications following *P. falciparum* infection at the RBC membrane level, namely of cytoskeletal proteins, which were up-represented (band 4.1, spectrin, adducin and dematin). Several interactions between parasite-encoded proteins and cytoskeletal host proteins have been described and may explain the increased infected RBC plasma membrane permeability and rigidity. A different approach was followed by Ray, et al. (2012) that analyzed the alterations in the human serum proteome as a consequence of infection by *P. falciparum* and *P. vivax*. Functional pathway analysis revealed the modulation of different vital physiological

pathways, including acute phase response signaling, chemokine and cytokine signaling, complement cascades and blood coagulation.

Mature RBC, with no nucleus, mitochondria or ribosomes cannot make oxidative phosphorylation or protein synthesis. However, these cells need an active metabolism to keep the integrity of membrane and maintenance of functional status of hemoglobin. The enzymes of RBC allow meeting these tasks by supporting two important metabolic pathways: glycolysis and the pentose phosphate pathway. An enzymatic deficiency in these patways may affect the production of ATP or NADPH with alteration in the membrane and cell removal (Jacobasch and Rapoport, 1996; Cappadoro, et al., 1998; Ayi, et al., 2009). The most frequent RBC enzymatic disorder worldwide is the glucose-6-phosphate dehydrogenase (G6PD) deficiency (G6PDD), followed by pyruvate kinase (PK) deficiency (PKD) and polymorphisms in these enzymes have been associated to malaria protection. In this respect, a single proteomics report is available trying to explain the protection conferred by the G6PD A- African variant: Méndez, et al. (2011) analyzed the major oxidative changes occurring in the host membrane proteins during the erythrocytic development of *P. falciparum* by redox proteomics. Fifteen carbonylated membrane proteins were exclusively identified in infected G6PD A- RBC revealing a selective oxidation of host proteins upon malarial infection. As a result, three pathways in the RBC were oxidatively damaged in G6PD A-: traffic/assembly of exported parasite proteins in RBC cytoskeleton and surface, oxidative stress defense proteins, and stress response proteins. The identification of hemichromes (denatured hemoglobins) associated with membrane proteins also supported a role for oxidative modifications in protection against malaria by G6PD variants.

In this study, we intended to perform a comprehensive proteomic analysis of malaria infection and so we looked to the infected RBC under several perspectives. We tried to define a quantitative proteomic profile of non-infected and infected RBC (healthy, PKD and G6PDD), as well as of parasites growing in these different environments, to know the effect of these enzyme disorders on parasite development as well as the changes occurring in the RBC upon infection. The combination of proteome data from the parasite and the host cell will shed new light on: the parasite requirements

for development; the mechanisms responsible for the lower susceptibility of enzyme-deficient RBC to malaria; and the host-parasite interactions.

This is the first time that parasite and host proteins were extracted from the same cell cultures allowing a cause-effect reliable comparison between both protein expression profiles. This is also the first time that the proteomes of *Plasmodium* growing in G6PDD and PKD conditions as also of PKD RBC were studied – a step forward in the comprehension of infection dynamics and enzyme deficiencies.

METHODS

1. Individuals

Three individuals originated from Portugal voluntarily participated in this study donating their blood (all $0Rh^+$): one with PKD, other with G6PDD (both previously diagnosed and genotyped for mutations) and a healthy control [normal activity of both PK (PKN) and G6PD (G6PDN)]. The characteristics of case individuals are described in Table 1. G6PDD individual is asymptomatic whereas the PKD individual has a severe clinical phenotype, with 2-3 severe hemolytic crises every year, needing blood transfusions. He is splenectomised and present high reticulocyte counts (30–40%) (previously studied in Manco, et. al., 1999; Manco, et al., 2002). The last blood transfusion occurred 10 months before the blood collection for this study.

Blood samples were collected by intravenous puncture in vacutainer tubes containing K_2EDTA for both invasion/maturation and proteomic assays. White blood cells were removed by three cycles of centrifugation and washing of the blood samples with sterile saline solution (NaCl 0.9% w/v) and final hematocrit was adjusted to 50%. Washed RBC were stored at 4ºC and used to initiate the experiments in a maximum period of three days after collection.

Table 1. Characteristics of case individuals with PKD and G6PDD.

Subject	Gender	Age (years)	Percentage of control activity (%)[1]	Mutations[2]	Effect	Symptoms
PK-deficient (PKD)	M	14	18.0	IVS10(+1)G>C IVS10(+1)G>C	Splicing mutation	Transfusion-dependent
G6PD-deficient (G6PDD)	M	29	4.2	202G>A 376A>G	Val>Met Asn>Asp	Asymptomatic

[1]determined by the protocol described in Beutler, 1984.

[2]identified by PCR-RFLP and automatic sequencing.

2. *Plasmodium falciparum in vitro* cultures

Plasmodium falciparum 3D7 were maintained in continuous culture in healthy RBC at 5% hematocrit, at 37ºC, 5% CO_2, 5% O_2 and 90% N_2, as described (Trager and Jensen, 1976). Human serum was replaced by 0.5% AlbuMAXII (Invitrogen) in the culture medium. Prior to initiate the assays, cultures were synchronized twice with D-sorbitol (Lambros and Vanderberg, 1979).

3. Invasion and maturation assays

These assays were performed with 3 ml-synchronized cultures in 25 cm^2 flasks with an initial 5% hematocrit and parasitemia of 0.7% of schizonts: 21 µl of healthy RBC infected with schizonts (100% hematocrit, 5% parasitemia) were mixed with 258 µl of non-infected healthy (PKN and G6PDN), PKD or G6PDD washed RBC (hematocrit 50%) and culture medium was added up to 3 ml. Along the assays, new RBC were never added to the cultures. The experiments concerning PKD (denominated PK assay) and G6PDD (denominated G6PD assay) were performed independently and each had its own controls (although it corresponded exactly to the same blood from the same donor): the control from PK assay was termed PKN and the control from the G6PD assay was termed G6PDN. Each assay was performed in duplicate, meaning: in

PK assay, two 3 ml cultures in PKN and two 3 ml cultures in PKD; and in G6PD assay, two 3 ml cultures in G6PDN and two 3 ml cultures in G6PDD.

Parasitemias were determined by direct counting of parasites in Giemsa stained RBC smears in an optical microscope. Invasion levels were measured as the percentage of rings after 24, 72 and 120 hours of incubation, and maturation levels were measured as the percentage of schizonts after 48, 96 and 144 hours (as in Ayi, et al., 2008).

Moreover, invasion was evaluated calculating (adapted from Ayi, et al., 2004):

- the ratio between ring parasitemia at 24 hours and initial schizont parasitemia (first cycle of invasion),

- the ratio between ring parasitemia at 72 hours and schizont parasitemia at 48 hours (second cycle of invasion), and

- the ratio between ring parasitemia at 120 hours and schizont parasitemia at 96 hours (third cycle of invasion).

Similarly, maturation levels were measured determining:

- the ratio between schizont parasitemia at 48 hours and ring parasitemia at 24 hours (first cycle of maturation),

- between schizont parasitemia at 96 hours and ring parasitemia at 72 hours (second cycle of maturation) and

- between schizont parasitemia at 144 hours and ring parasitemia at 120 hours (third cycle of maturation).

3.1. Statistical analysis

Statistical analysis was performed with the software GraphPad Prism version 6.00 for Windows (http://www.graphpad.com/). Wilcoxon signed rank test, a paired difference test to compare two matched samples, was used to search for significant differences in *P. falciparum* growth in normal and deficient RBC. A significance level of 0.05 was considered.

4. Proteomics

Briefly, a MS proteomic experiment follows the next steps: extracts preparation, digestion into peptides, peptide separation (mostly by capillary High Performance Liquid Chromatography, HPLC), sample ionization (by Electrospray Ionization, ESI or matrix-assisted laser desorption/ionization, MALDI), MS and data analysis (Steen and Mann, 2004). These procedures are described below and the strategy is represented in **Fig. 1**.

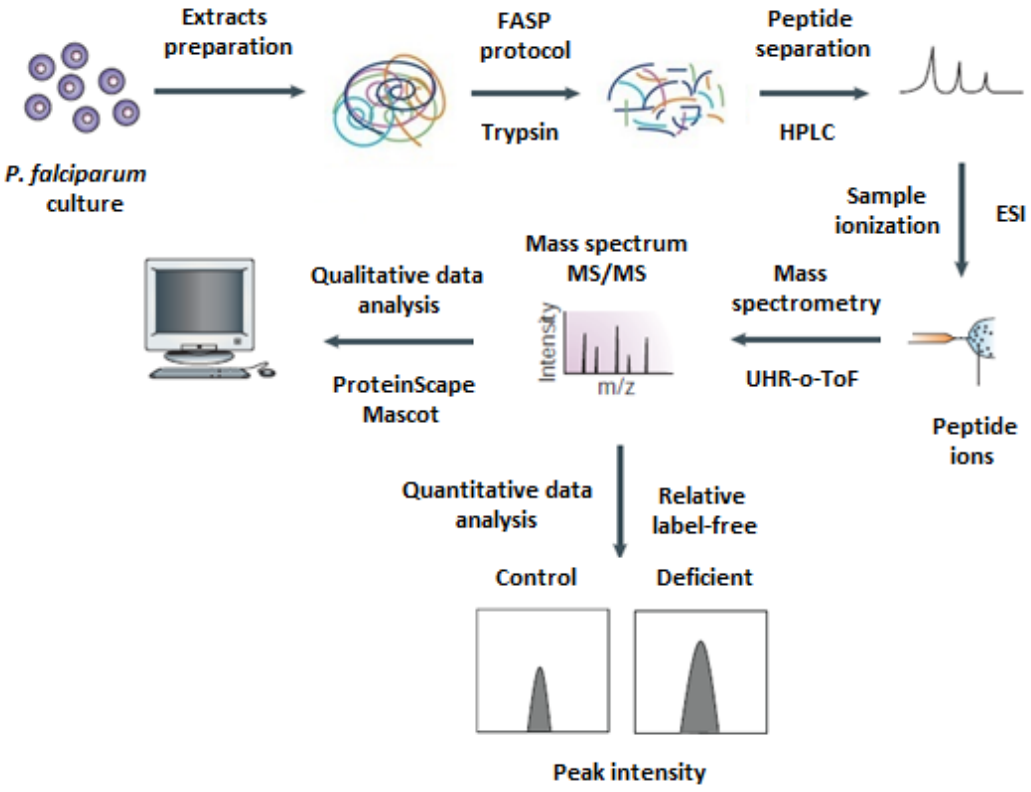


Fig. 1. The MS proteomic strategy followed in the present study. *Plasmodium falciparum* (18h trophozoite stage) and RBC extracts were prepared from *in vitro* cultures after lysis. Proteins were digested into peptides with trypsin and prepared for MS following the FASP prototocol. The generated peptide mixture was separated by HPLC and ionized by ESI and analyzed by a UHR-o-ToF mass spectrometer. Finally, the peptide-sequencing data that were obtained from the mass spectra were searched against human and *P. falciparum* protein databases using MASCOT and protein abundance determined in a relative and label free manner comparing peak intensities. FASP: filter-aided sample preparation method (Wisniewski,

et al., 2009); HPLC: high-performance liquid chromatography; ESI: electrospary ionization; UHR-o-ToF: ultra high resolution–orthogonal–time of flight (adapted from Steen and Mann, 2004).

4.1. Parasite growth

The parasite growth was performed in 15 ml-synchronized cultures in 75 cm^2 flasks with an initial 5% hematocrit and parasitemia of 0.7% of schizonts: 105 µl of healthy RBC infected with schizonts (100% hematocrit, 5% parasitemia) were mixed with 1290 µl of non-infected (PKN and G6PDN), PKD or G6PDD washed RBC (hematocrit 50%) and culture medium was added up to 15 ml. When the cultures reached a parasitemia of 5-10%, these were divided into two flasks, adjusting the hematocrit to 5% with the type of RBC to be tested (healthy, PKD or G6PDD). The PK and G6PD assays were performed independently and in duplicate. Each had its own controls (PKN and G6PDN). Non-parasitized (NI) PKN, PKD, G6PDN and G6PDD controls were also kept in culture under the same conditions as parasitized RBC. So, totally, 16 cultures (in 32 flasks, because each was divided into two flasks when initially reached the 5-10% parasitemia, as mentioned above) were maintained for protein extraction purposes (PK assay: 2 PKD, 2 PKD_NI, 2 PKN, 2 PKN_NI; G6PD assay: 2 G6PDD, 2 G6PDD_NI; 2 G6PDN, 2 G6PDN_NI). The extracts were prepared one cycle after cultures synchronization (with D-sorbitol), with a parasitemia of about 15% of young trophozoites (approximately 18 hours post-invasion).

4.2. Protein extracts preparation

No previous studies describing the extraction of proteins from both parasites and RBC from the same *Plasmodium* culture were found, so the followed procedure was adapted from available protocols in order to obtain the higher achievable quantity of each fraction (parasite, RBC cytoplasm and RBC membranes) but with the lower contamination among fractions as possible.

The cultures were transferred to 15 ml tubes, centrifuged at 2500 xg and the medium discarded. The packed RBC were lysed with a hypotonic lysis buffer [ice-cold

5 mM sodium phosphate pH8 with a protease inhibitor cocktail (Roche)] and the infected RBC were centrifuged at 18 000 xg for 20 min at 4ºC to separate the RBC fraction from the parasites. The upper reddish phase (RBC) was then transferred to a new tube.

4.2.1. Red blood cells

The reddish fraction was centrifuged at 18 000 xg for 20 min at 4ºC and the upper (cytoplasm) and lower (membrane ghosts) phases put in different tubes.

4.2.1.1. Membrane ghosts

Ghosts preparation was adapted from Pasini, et al., 2006. Initially they were washed with lysis buffer until the supernatant becomes colorless (at least five times). Each washing consisted in the addition of 10xV lysis buffer, mixing, centrifugation at 10 000 xg, 10 min at 4ºC and removal of the supernatant. More stringent washings (at 20 000 xg, 10 min at 4ºC) were then followed until the ghosts got yellowish. Pellets were stored at -80ºC.

4.2.1.2. Cytoplasm

The cytoplasmic fraction was centrifuged at 50 000 xg at 4ºC for 30 min and the supernatant transferred for a new tube. Then, two protocols were tested to remove hemoglobin: the Ni-NTA (nickel-nitrilotriacetic acid) Super Flow, from Qiagen (as reported in Ringrose, et al., 2008), that uses a nickel-charged resin with affinity for hemoglobin, and the HemoVoid - Hemoglobin Reagent Depletion Kit (Biotech Support Group), that derives from a silica-based library of individual mixed-mode ligand combinations (ionic, hydrophobic, aromatic, polymer). In the first method, after the supernatant has passed through the resin, the resin was washed with imidazole 5 mM solution and then with imidazole 10 mM solution. To elute hemoglobin, a 100 mM solution was used. Imidazole binds to Ni-NTA resin and competes with hemoglobin: at low concentrations inhibits non-specific binding and at higher concentrations elutes hemoglobin. The most successful method was applied to all cytoplasmic samples.

4.2.2. Parasite

Pellet was washed three times with cold PBS (centrifugations at 9 000 xg, 10 min at 4ºC) and the parasites lysed with lysis buffer [PBS, 0.1% Triton X-100 and protease inhibitor cocktail (Roche)] and three cycles of freeze-thawing (-70ºC – 37ºC), followed by a centrifugation at 9 000 xg, 10 min at 4ºC. The supernatant (parasite extract) was transferred to a new tube. The most efficient protocol in removing hemoglobin from RBC cytoplasmic fraction was tested in hemoglobin removal from the parasite extracts from one of the two experiments (G6PD assay).

4.3. Proteins quantification and visualization

Protein concentrations were determined using the colorimetric Bradford assay in a Nanodrop 1000 spectrophotometer (Thermo Scientific), according to the manufacturer's instructions. A calibration curve was assembled from measuring prediluted BSA standards. Proteins were separated by SDS-PAGE (Laemmli, 1970) in 12.5% acrylamide: bisacrylamide 37.5:1 gels (using the Mini-PROTEAN system, BioRad) or in precast SDS-polyacrylamide gel (NuPAGE Novex, 4-12% Bis-Tris Gel, Invitrogen) and stained with Coomassie Blue Brilliant R250 reagent.

4.4. Mass Spectrometry

Since our aim was the identification of peptides and subsequently definition of a global protein profile of our samples, a label-free shotgun proteomics approach was followed (revised in Matzke, et al., 2012), meaning that there was no predefined peptides of interest and that the protein quantification was determined in a label-free manner. Only parasite extracts were analyzed by MS so far; the analysis of the RBC fractions is still ongoing.

4.4.1. Protein samples preparation

After proteins extraction, these were prepared for MS using the filter-aided sample preparation (FASP) method (Wisniewski, et. al., 2009), in which trypsin enzyme

was used to cleave the proteins into peptides. The peptides were dried by the Speed-Vac system and eluted in 20 µl of Elga water.

4.4.2. Qualitative and quantitative mass spectrometry

After trypsinization (FASP protocol), samples were analyzed by nano-ESI-LC MS using a nano Acquiry Ultra Performance LC coupled to an UHR (ulta high resolution)-o-ToF mass spectrometer (maXis, Bruker). In this technique, peptides are separated by capillary nano-high performance liquid chromatography (nano-HPLC), ionized by ESI, and the generated ions are then separated according to their mass-to-charge (*m/z*) ratio. The MS then proceeded to obtain primary structure (sequence) information about these peptides coupling two stages of MS (MS/MS).

In a first instance, only qualitative data (peptides identification) was acquired and only one MS/MS run (technical replicate) was performed for each parasite sample (PKN1, PKN2, PKD1, PKD2, G6PDN1, G6PDN2, G6PDD1, G6PDD2). To get proteins quantitation, new MS data had to be acquired (new runs) and because of cost and time restrictions, control replicates were pooled together (PKN1+PKN2 and G6PDN1+G6PDN2). Each of the six samples ran three times (technical replicates). Protein quantification was label-free (peptides were not tagged and peptide peak intensities were used as a surrogate for abundance) and relative (presented as relative to control sample).

The bioinformatics platform ProteinScape (Bruker) was used for the storage and processing of MS data, including search results and quantitative data. Peptides identification was performed with the software MASCOT (version 2.3.02) against SwissProt (www.uniprot.org) and PlasmoDB (plasmodb.org) databases. Search parameters allowed for one missed tryptic cleavage site, the carbamidomethylation of cysteine and the possible oxidation of methionine. All identified proteins had a MASCOT score greater than 20, considering a $p< 0.05$ as significance level. Identifications were considered valid when they contained at least two peptide sequences per protein. The higher the score (calculated based on the correlation between the MS/MS spectrum and a theoretical one) of a candidate protein, the higher the confidence in the identification.

4.4.3. BioInformatic analysis

The proteins identified by MS were classified according to function with PANTHER (Protein Analysis Through Evolutionary Relationships) software classification system (www.pantherdb.com), which was designed to classify proteins (and their genes) according to:

a) Family and subfamily (families considered groups of evolutionarily related proteins; subfamilies group related proteins that also have the same function);

b) Molecular function (the function of the protein by itself or with directly interacting proteins at a biochemical level, e.g. a protein kinase);

c) Biological process (the function of the protein in the context of a larger network of proteins that interact to accomplish a process at the level of the cell or organism, e.g. mitosis);

d) Pathway (similar to biological process, but a pathway also explicitly specifies the relationships between the interacting molecules).

Details of the methods can be found in Thomas, et al., 2003 and Mi, et al., 2005. PANTHER classification is based on Gene Ontology (GO) project (http://www.geneontology.org), that standardizes the representation of gene and gene product attributes across species and databases.

An integrated analysis of the identified proteins in each experiment was performed with Cytoscape v2.8.3 (http://www.cytoscape.org/), regarding protein-protein interactions and biological processes.

Proteins non-classified by PANTHER and Cytoscape were manually investigated in several databases in order to get a functional profile for all proteins (PlasmDB, plasmodb.org; UniProt, www.uniprot.org; Malaria Parasite metabolic Pathways, http://priweb.cc.huji.ac.il/malaria/).

RESULTS AND DISCUSSION

Note: *The results presented as supplementary material are indicated with an "S" preceding the numeration.*

1. *Plasmodium falciparum* invasion and maturation assays

Parasites grew in both PKD and PKN RBC and their morphology (both ring and schizont stages) were similar (**Fig. 2 and 3**). In PKD cultures, reticulocytes were observed, as expected (high reticulocyte counts in this individual described previously in Manco, et al., 1999 and Manco, et al., 2002).

For six of the eight cultures from both PK and G6PD assays, the peak of parasitemia was reached 72h after inoculation. Cultures PKD1 and G6PD1 were the exception (maximum parasitemia reached 48h later, at 120h). After this, parasitemia dropped until total hemolysis, at 168-216h (**Fig. S1 and S2**).

In PK assays, the growth pattern of the parasites was similar in both PKN RBC and the same was observed for parasites growing in PKD RBC. Generally, parasitemias were always higher in PKN until 120h after inoculation, but after this time, parasites in PKD RBC achieved higher parasitemias.

Similarly, in G6PD assays, the parasitemias were higher in G6PDN in the beginning of the assays (until 96h after inoculation). After this time, there was no correspondence between the two cultures of each type of RBC. The culture G6PDN2 achieved parasitemias similar to those in PKN RBC but all the other grew slightly.

An increase in parasitemia reflects the invasion of RBC by new parasites, while the decrease reveals the maturation period during which some parasites die. This is showed in the maturation and invasion data.

No substantial differences were observed in gametocyte parasitemias.

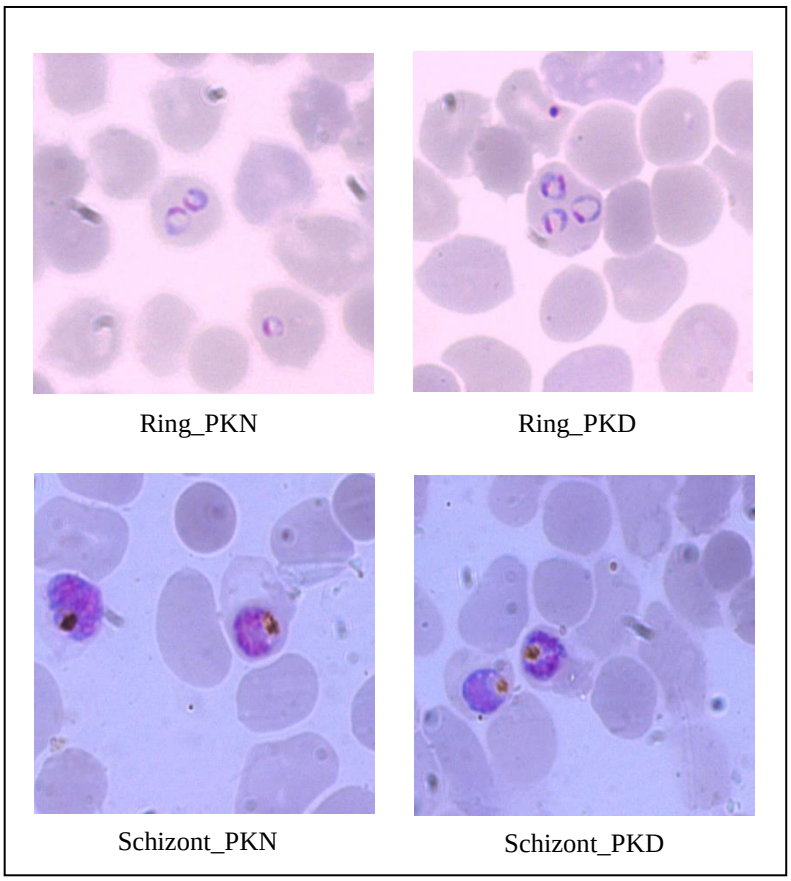


Fig. 2. Pyruvate kinase assay: *P. falciparum* 3D7 (ring and schizont stages) growing in normal (PKN) and PK-deficient (PKD) RBC, observed in Giemsa stained smears with an optical microscope. Amp: 1000x.

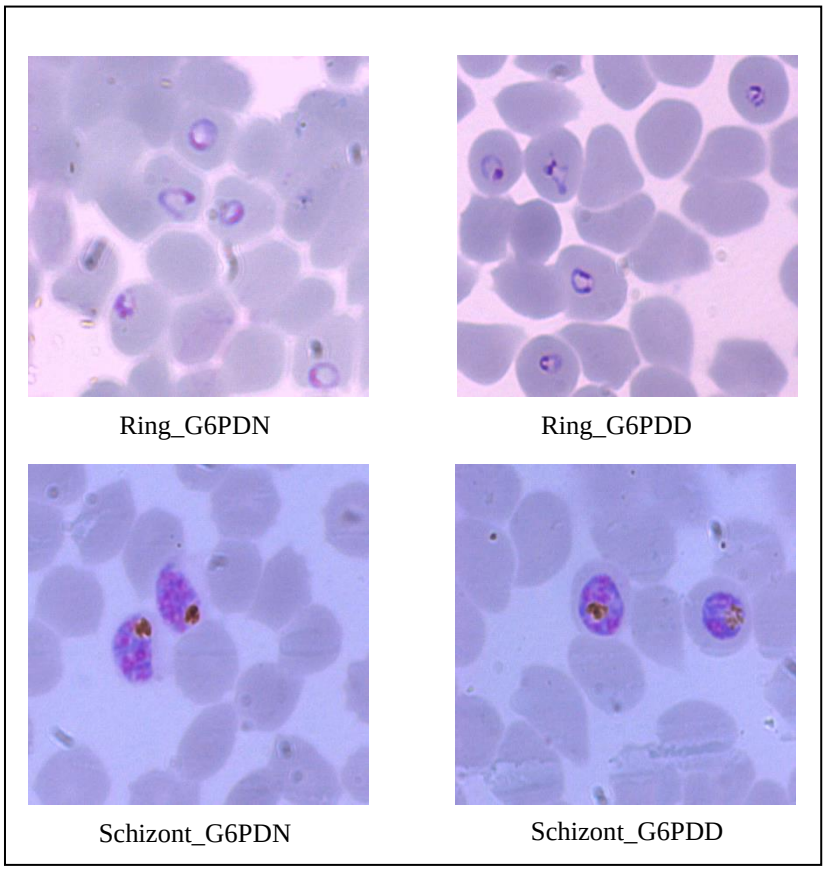


Fig. 3. Glucose-6-phosphate dehydrogenase assay: *P. falciparum* 3D7 (ring and schizont stages) growing in normal (G6PDN) and G6PD-deficient (G6PDD) RBC, observed in Giemsa stained smears with an optical microscope. Amp: 1000x.

Parasites invasion and maturation were assessed by two different ways: by ring and shizont parasitemias, respectively (making possible to compare with results in Ayi, et al., 2008), and calculating the ratios between the ring parasitemia and the schizont parasitemia 24h before (invasion) and between the schizont parasitemia and the ring parasitemia 24h before (maturation). These results are shown in **Fig. 4-7** and **Tables S1-S4**.

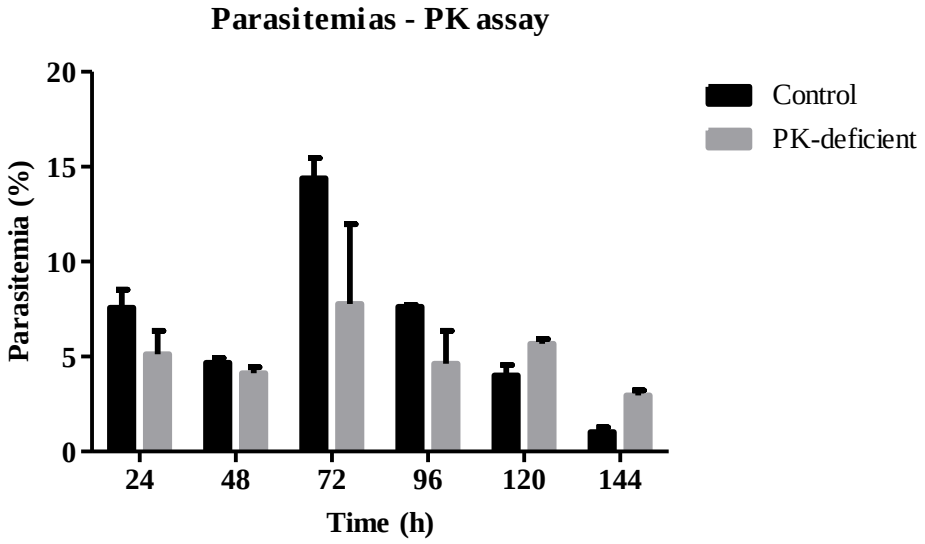


Fig. 4. Percentage of ring (24h, 72h and 120h after inoculation) and schizont parasitemias (48h, 96h and 144h after inoculation) of *P. falciparum* in three growing cyles in control (PKN) and PK-deficient (PKD) RBC. The results are the combination of mean values obtained in two replicates.

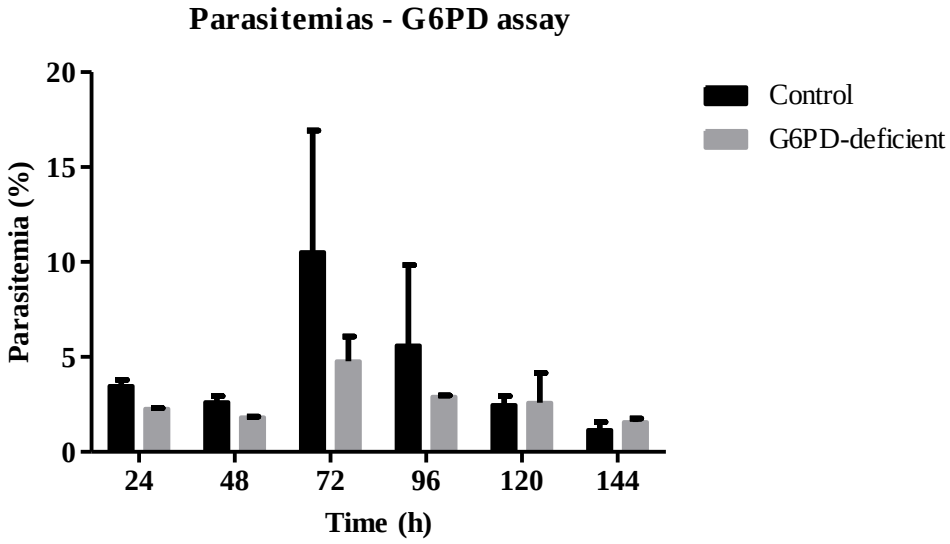


Fig. 5. Percentage of ring (24h, 72h and 120h after inoculation) and schizont parasitemias (48h, 96h and 144h after noculation) of *P. falciparum* in three growing cyles in control (G6PDN) and G6PD-deficient (G6PDD) RBC. The results are the combination of mean values obtained in two replicates.

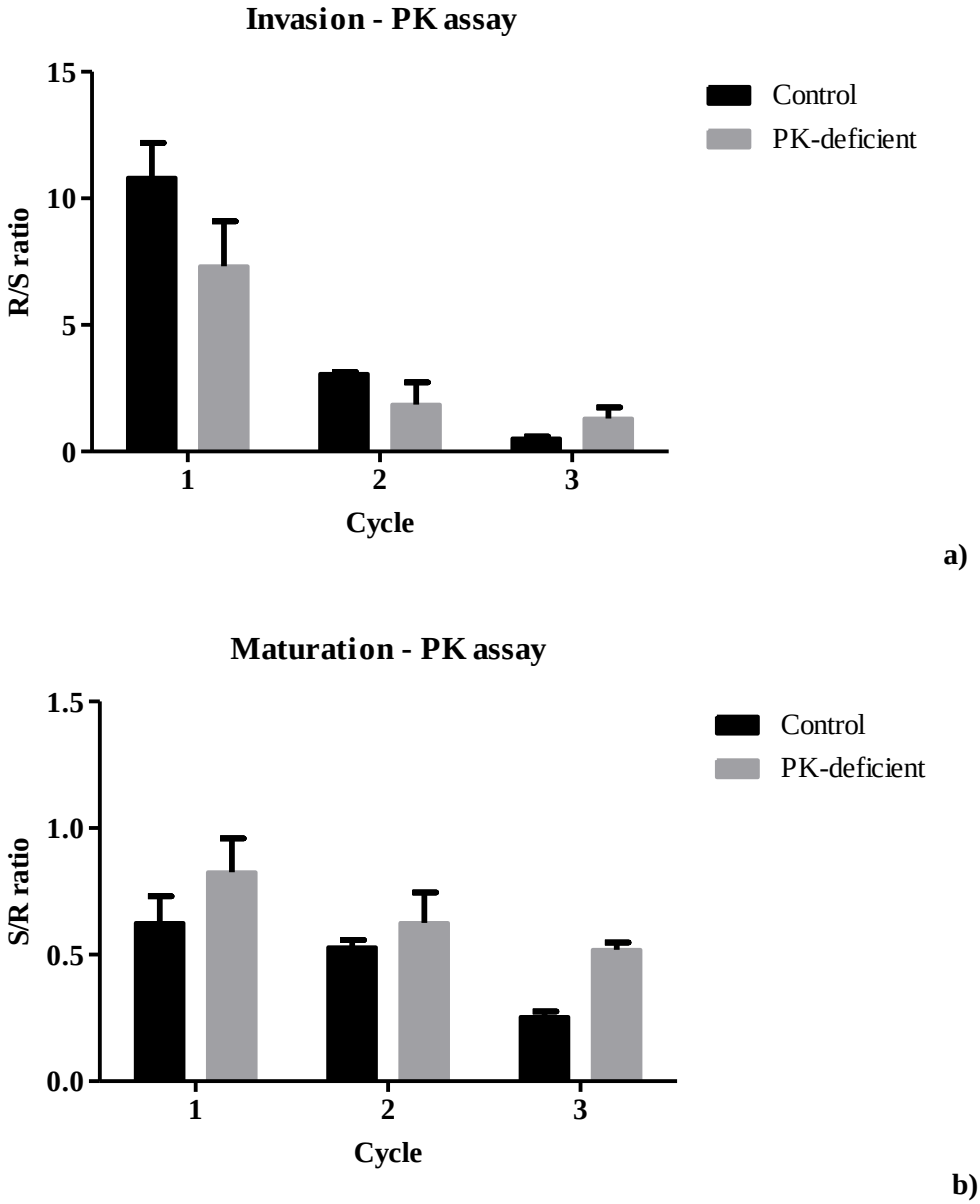
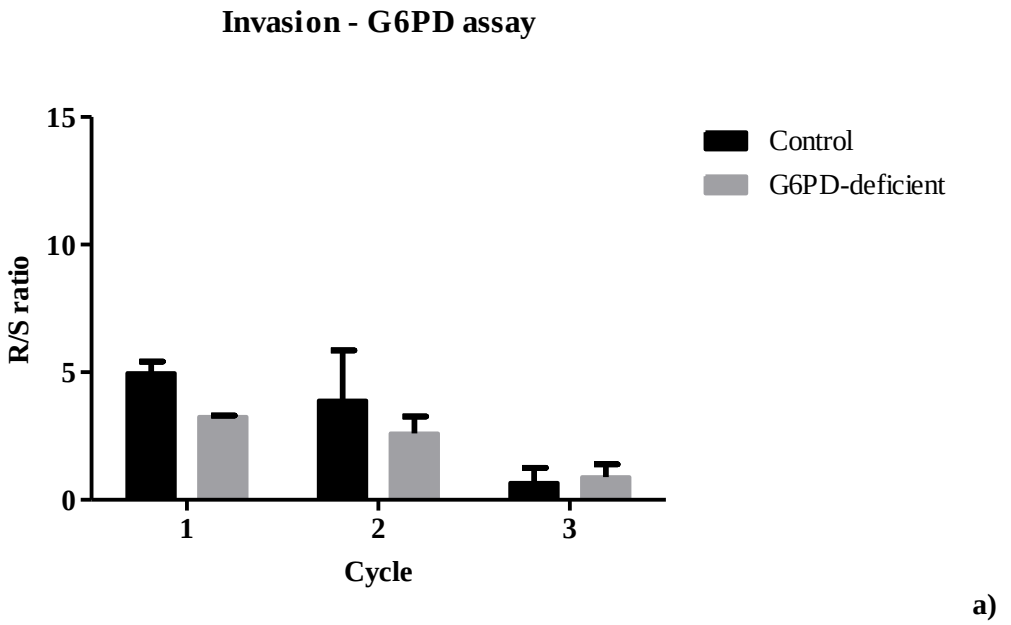
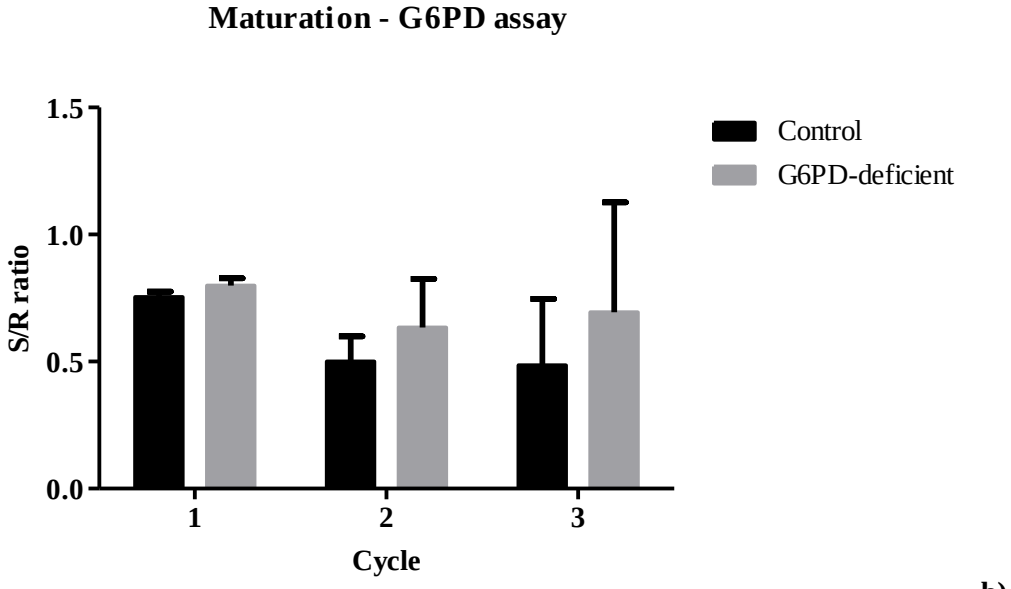


Fig. 6. Invasion and maturation ratios of *P.falciparum* in three growing cyles in control (PKN) and PK-deficient (PKD) RBC. The results are the combination of mean values obtained in two replicates. **a) Invasion:** cycle 1- ratio between the percentage of ring-stage parasites (R) at 24h and schizont-stage parasites (S) at 0h; cycle 2– ratio between R at 72h and S at 48h; cycle 3- ratio between R at 120h and S at 96h. **b) Maturation:** cycle 1- ratio between S at 48h and R at 24h; cycle 2- ratio between S at 96h and R at 72h; cycle 3- ratio between S at 144h and R at 120h.



a)



b)

Fig. 7. Invasion and maturation ratios of *P. falciparum* in three growing cyles in control (G6PDN) and G6PD-deficient (G6PDD) RBC. The results are the combination of mean values obtained in two replicates. **a) Invasion:** cycle 1- ratio between the percentage of ring-stage parasites (R) at 24h and schizont-stage parasites (S) at 0h; cycle 2– ratio between R at 72h and S at 48h; cycle 3- ratio between R at 120h and S at 96h. **b) Maturation:** cycle 1- ratio between S at 48h and R at 24h; cycle 2- ratio between S at 96h and R at 72h; cycle 3– ratio between S at 144h and R at 120h.

For both PK and G6PD assays, based on parasitemias only, invasion and maturation were both always higher in normal RBC, except in the third cycle of invasion and maturation. When invasion and maturation were assessed by ratios, the results were similar in invasion (in the third cycle, invasion superior in both PKD and G6PDD RBC) but not in maturation: maturation was higher in deficient RBC (PKD and G6PDD) in the three cycles. However, none of these differences were statistical significant, with the insufficient number of replicates probably contributing to this result (**Tables S1-S4**).

The disparity of the results is explained by the way they were obtained: the invasion ratios show the number of parasites that have invaded new RBC, originated from the schizonts measured 24h before; the maturation ratios are the number of parasites that develop into schizonts from the rings measured in the day before. Ratios give an idea of continuity, whereas the other assessment method is based on the number of parasites at a single moment. **Figures 6 and 7** show how parasitemias increase and decrease over 24h, so we can see that there are more parasites invading normal than deficient RBC (**Fig. 6a and 7a**) but more parasites are dying during its maturation in healthy RBC than in deficient ones (**Fig. 6b and 7b**). These results indicate that invasion is more relevant for parasite growth impairment in enzyme-deficient conditions than maturation. The protection mechanism related to these polymorphisms may be associated to a less efficient invasion of the RBC, instead of a more difficult development in the deficient environment, suggesting that membranes of deficient RBC that are about to be invaded may be the key for protection. Another possibility, is the emergence of some defect in the apical complex of new merozoites (that have developed inside deficient RBC), that may be hampering their invasion.

Moreover, some kind of selection seems to occur in the invasion step, limiting the ring parasitemia in deficient RBC in the first cycles, but once the parasites have invaded the deficient cells, these are more able to complete its erythrocytic cycle than the parasites that have grown in a normal environment. In the third invasion cycle, the parasites remaining after two "selective cycles" seem to be more competent to efficiently invade deficient RBC (higher invasion ratios). However, a "selective" mechanism is unlikely to occur in a *Plasmodium* clone (3D7) so quickly (three cycles). Besides, we cannot ignore that normal cultures have experienced a more severe

hemolysis and nutrient depletion (because of previous higher parasitemias) which may contribute for the lower third cycle invasion ratio in normal cells.

These results corroborate previous ones obtained in G6PDD RBC: Luzzatto, Usanga, and Reddy (1969) described an impaired growth in heterozygous females and Roth, et al. (1983) in hemizgotic males. Later, Usanga and Luzzatto (1985) reported that the growth inhibition of *P. falciparum* in human G6PDD RBC (both Mediterranean and A- variants) is overcome after two or three growth cycles, in agreement with our observations. The parasite seems to undergo adaptative changes that gradually improve its ability to multiply in these deficient cells by producing its own G6PD enzyme (Usanga and Luzatto, 1985; Roth and Schulman, 1988). Cappadoro and coworkers (1998), contrarily, found that invasion and maturation of the parasite in both the first and second growth cycles were quantitatively indistinguishable in normal and deficient RBC (Mediterranean variant) and that G6PD mRNA was not significantly different in normal and deficient parasitized cells, claiming that preferential phagocytosis at an early stage of the schizogonic cycle is the most probable explanation for the protection conferred by this deficiency, instead of the intracellular oxidative stress itself.

The interest in PK deficiency and its association with malaria is more recent and only two studies have been published regarding *P. falciparum in vitro* growing in PKD RBC (Durand and Coetzer, 2008; Ayi, et al., 2008), although from individuals with a different genotype from the individual from this study, which may be relevant if the phenotype is different. Durand and Coetzer (2008) used RBC from a homozygous and a heterozygous for the missense 1529G>A mutation. RBC from the homozygous PKD patient demonstrated a dramatic resistance to *P. falciparum* infection. The parasitemia in the heterozygote was slightly lower than the control but there was no statistically significant difference between them. Ayi, et al. (2008) used RBC from heterozygous and homozygous individuals for the loss-of-function mutation 1269G>A, and also from a homozygous subject for a single-base deletion at nucleotide position 823 of *PKLR*. In this study, invasion and maturation were assessed as the ring and schizont parasitemias at 24, 72 and 120h and at 48h, 96h and 144h, respectively (as also performed in the present study). There was a significant reduction in the invasion of RBC by *P. falciparum* parasites during three consecutive growth cycles in the homozygous subjects. In subjects carrying heterozygous mutations in *PKLR*, no significant effect was

observed. For both homozygous and heterozygous, no significant differences were detected in intracellular maturation between RBC from deficient subjects and those from control, however, as mentioned above, maturation was determined through schizont parasitemias. Interestingly, when we carefully looked for these data it was obvious that in the experiment with homozygous mutant cells more parasites died during its maturation in healthy RBC than in deficient ones, as in our study. This was not clear in the heterozygous mutant RBC experiment.

These results in PK experiments point to an adaptative response similar to that previously described for parasites growing in G6PDD RBC. Actually, a pyruvate kinase of parasitic origin has been described (Oelshlegel, Sander, and Brewer, 1975) and seems to be involved in this process: it has been shown that ATP levels are reduced in PKD RBC and there is a correlation between ATP levels and both inhibition of parasite invasion and enhancement of phagocytosis of RBC infected with ring-stage parasites. Moreover, the proportion of parasites that successfully invade PKD RBC appear to meet their ATP requirements for intracellular maturation by up-regulating their parasite specific pyruvate kinase (mRNA levels 8 to 13-fold increased) (Ayi, et al., 2009). Based on these results and others, a model is suggested by Ayi et al. (2009) for PK deficiency protection against malaria: together with the reduction in ATP production, there is an increase in 2,3-diphosphoglycerate (2,3-DPG) in PKD cells, that contribute to the maintenance of glutathione in the reduced state and, as a consequence, excessive amounts of free radicals may be generated that transform oxyhemoglobin to methemoglobin and, ultimately, to hemichromes, contributing to mechanical destabilization of the PKD RBC membrane and disruption of the cell membrane cytoskeletal protein network, namely, the spectrin-actin band 4.1 complex, with consequent band 3 aggregation and impairment of parasite invasion.

The proteomic analysis will help to clarify these protection mechanisms, namely if there is an increase in *P. falciparum* G6PD and PK expression when growing in cells deficient in these enzymes, and if there is relevant alterations in the RBC membrane proteins.

In the present study, no statistical differences were observed neither in invasion nor maturation, but only two replicates were performed in each assay, which

dramatically reduces the statistical power of the analysis. However, similarly to the results obtained in the up mentioned studies, we could observe that invasion was clearly higher in normal cells in the first and second replication cycles. Unfortunately, neither invasion nor maturation ratios were calculated in the studies from Durand and Coetzer (2008) and Ayi, et al. (2008), so we could compare with our results.

2. Proteomics

2.1. Protein extracts preparation

The preparation of protein extracts was hampered by numerous technical constraints, namely the lack of protocols describing the extraction of proteins from both parasites and RBC from the same cell culture and the high dynamic range of protein concentrations in blood component proteomes.

The high-abundant protein hemoglobin (Hb) completely masks low-abundance species, so, one of the greatest challenges in this task, was the removal of this protein, together with the adaptation of protocols to obtain extracts with enough quality for MS. For example, most of the described procedures for *Plasmodium* proteins extraction (e.g. Nirmalan, Sims, and Hyde, 2004; Southworth, Hyde and Sims, 2011) use saponin solution (0.05%) for release the parasites from RBC. However, the use of detergents may break some of the molecular interactions between protein and lipids and may differentially remove associated membrane proteins (Pasini, et al., 2010). Therefore, in the present study, a hypotonic phosphate lyses buffer was employed since it is believed to have minimal effects on RBC membrane protein equilibrium, in which we were also interested.

We were able to get protein extracts from both *Plasmodium* and RBC (membrane and cytoplasmic fractions from infected and non-infected cells) and the quantities and concentrations obtained are shown in **Tables S5-S7**. **Figures S3-S5** show the protein extracts separated by SDS-PAGE, from *Plasmodium* (**S3**) and membrane of RBC (**S4 and S5**). The identification of some abundant membrane proteins were predicted (shown in **Fig. S3 and S4**) considering their molecular weight and comparing with previous results from Delobel, et al., 2012: spectrin α (281 kDa), spectrin β (246

kDa), band 3 (102 kDa) and β-actin (42 kDa). We strongly expected to identify at least band 3 and spectrins since the transmembrane protein band 3 occurs at one million copies per cell (comprising 30% of the membrane proteome) and spectrin tetramer occurs at 100,000 copies per cell, comprising 75% of the cytoskeleton (Pasini, et al., 2010). *Plasmodium* extracts were contaminated with human proteins (as expected, since parasite proteins are much less abundant): at least spectrins and Hb (about 15 kDa band) were observed (Hb not present in **Fig. S3** because the gel fairly ran but clear in **Fig. S10**, lane B). Due to the amphipathic nature of Hb, a portion associates with the RBC membrane during lysis (Pasini, et al., 2010) but repeated washes at low temperature (4ºC) in hypotonic phosphate buffer significantly reduced Hb contents of membrane ghosts (no Hb band detected in gels from **Fig. S4 and S5**).

In cytoplasmic fractions, the Hb band was clearly identified as an intense band; carbonic-anhydrase (CA) and catalase were also recognized (**Fig. S7, lane B**). Hemoglobin is an iron-containing metalloprotein highly adapted to the specific function of oxygen transport in the RBC. Two α chains plus two β chains constitute HbA, which in normal adult life comprises about 97% of the total Hb; α chains combine with δ chains to constitute HbA-2, which with HbF (fetal Hb) makes up the remaining 3% of adult Hb (NCBI EntrezGene, Gene ID: 3039). The α chain is composed of 141 amino acids and has a molecular weight of 15 126 Da; the β chain has 146 amino acids and a molecular weight of 15 866 Da (Hill, et al., 1962). Apart from the Hbs, CA represents the principal protein constituent of RBC (Rickli, et al., 1964). Carbonic anhydrases are a large family of zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide. CA1 (about 29 kDa) is closely linked to CA2 and CA3 genes on chromosome 8, and it encodes a cytosolic protein which is found at the highest level in RBC (NCBI EntrezGene, gene ID: 759). Catalase (about 59 kDa) is a key antioxidant enzyme that plays a critical role in protecting cells against the toxic effects of hydrogen peroxide, removing over half of the hydrogen peroxide generated in normal human RBC (Kirkman and Gaetani, 1984).

It was not possible to detect differences between the bands pattern of extracts from normal and deficient RBC.

2.2. Hemoglobin removal

Alvarez-Llamas, et al. (2009) described the RBC proteome analysis as "enormously difficult" due to the high content of Hb. Hemoglobin depletion has been pointed as a crucial step for RBC proteome analysis in numerous studies (Prabakaran, et al., 2007; Roux-Dalvai, et al., 2008; Pasini, et al., 2010) and even today, as in 2008, "no available approach exists for the specific depletion of Hb together with the CA1, which accounts for approximately 97% and 1% of the RBC proteome, respectively" (Ringrose, et al., 2008).

Two reagents were initially tested for Hb removal: the Ni-NTA Super Flow (Qiagen), and the HemoVoid - Hb Reagent Depletion Kit (Biotech Support Group). The first has been optimized by Qiagen for 6xHis-tagged proteins purification but Ringrose and his team (2008) explored, successfully, its affinity for Hb. When this protocol was used in the present study, a clear reduction in Hb was observed but, contrarily to the results from Ringrose, the same pattern of SDS-PAGE bands was obtained (**Fig. S6**). On the other hand, Hemovoid not only removed most Hb as also allowed the detection of more proteins (**Fig. S7**). So, Hemovoid was used for Hb removal from all cytoplasmic extracts (**Fig. S8 and S9** show the SDS-PAGE results and **Table S7** the quantification of these extracts). Then, the reagent was tested for Hb removal in parasite extracts (only in G6PD assay samples), but this resulted in an unacceptable fraction lost of parasite proteins (**Fig. S10 and Table S8**), suggesting that the reagent is not suitable for parasite extracts. In the absence of a worthwhile alternative method to specifically remove Hb from *Plasmodium* extracts, parasite fractions were analyzed by MS without Hb removal.

2.3. Mass spectrometry

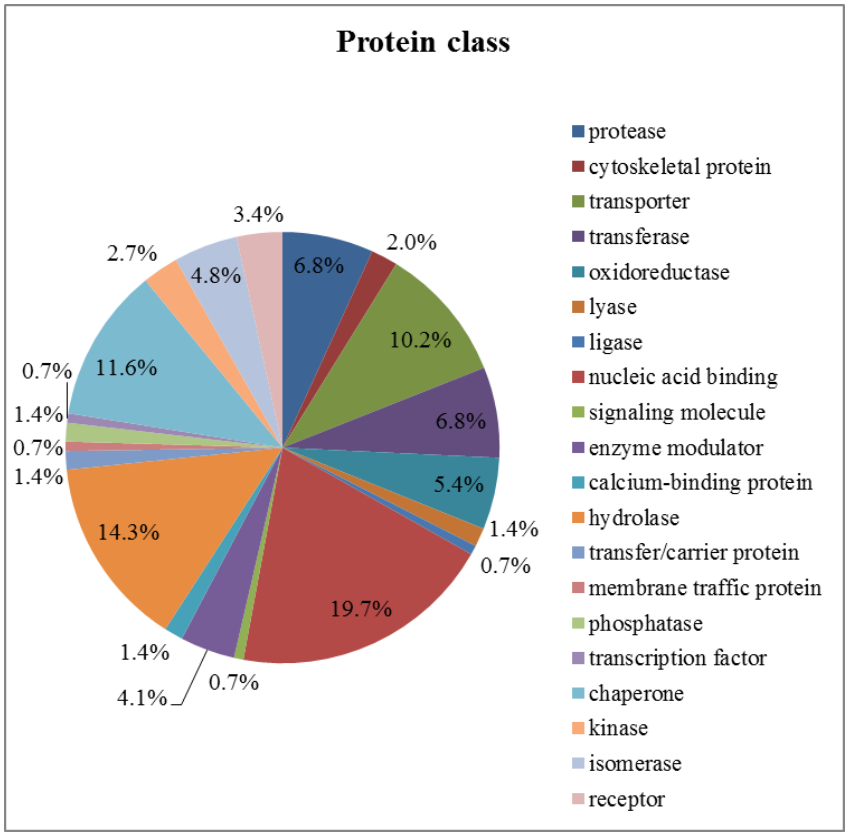
Mass spectrometry results were only obtained for parasite extracts so far. The extracts from RBC are still being processed in the *Centre of Excellence in Mass Spectrometry*, York, UK.

2.3.1. Qualitative analysis

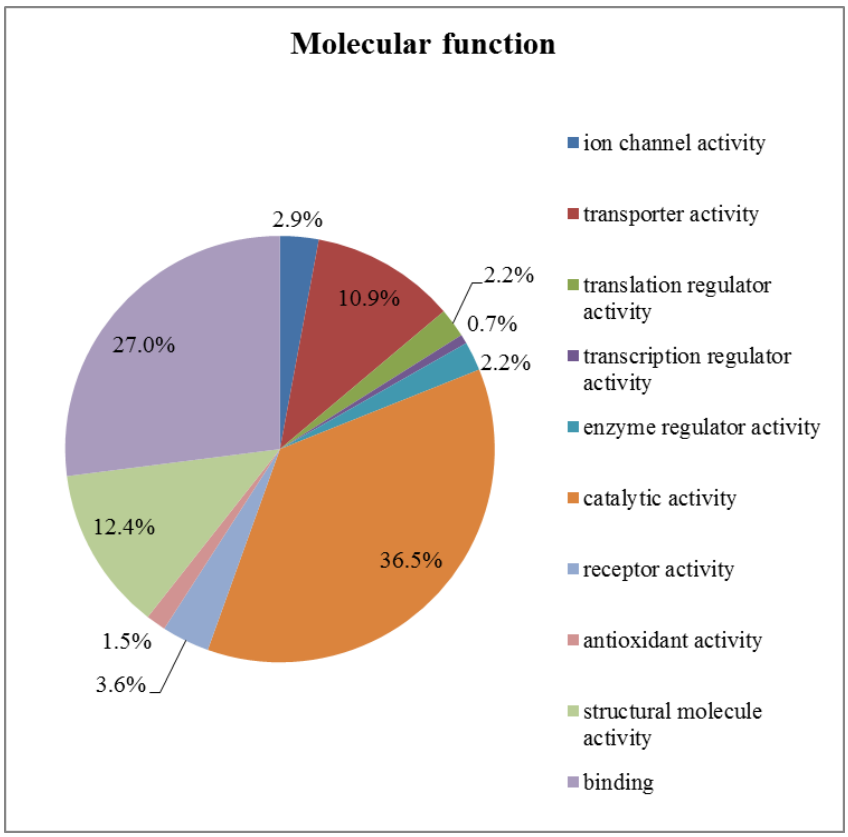
In the present study, 233 different proteins were identified from *Plasmodium* in its trophozoite state: 161 in PK assay (**Table S9**) and 197 in G6PD assay (**Table S10**). The number of proteins identified in both PK and G6PD assays was 125; 36 were identified in PK assay and 72 in G6PD assay, only. When proteins with a single peptide detected were excluded (more confident identification), the numbers dropped to 11 in PK assay and 27 in G6PD assay, resulting in 163 proteins confidently identified, corresponding to 37% of the plasmodial trophozoite proteome (comprising 443 proteins as described in Smit, et al., 2010).

These 163 proteins were classified according to their functional profiles with PANTHER software (**Table S11***). **Figure 8** shows their distribution per class, molecular function and biological process [a), b) and c) respectively]. As expected, a considerable portion of proteins (55, corresponding to 34%) were unable to map (unknown function), so the results relate to the remaining 108. Several classes, biological processes and molecular functions were assigned per protein, since most have diverse biological roles. So, in total, 108 proteins had 197 process hits, 137 function hits and 147 protein class hits and the percentages presented at **Fig. 8** are relative to these numbers.

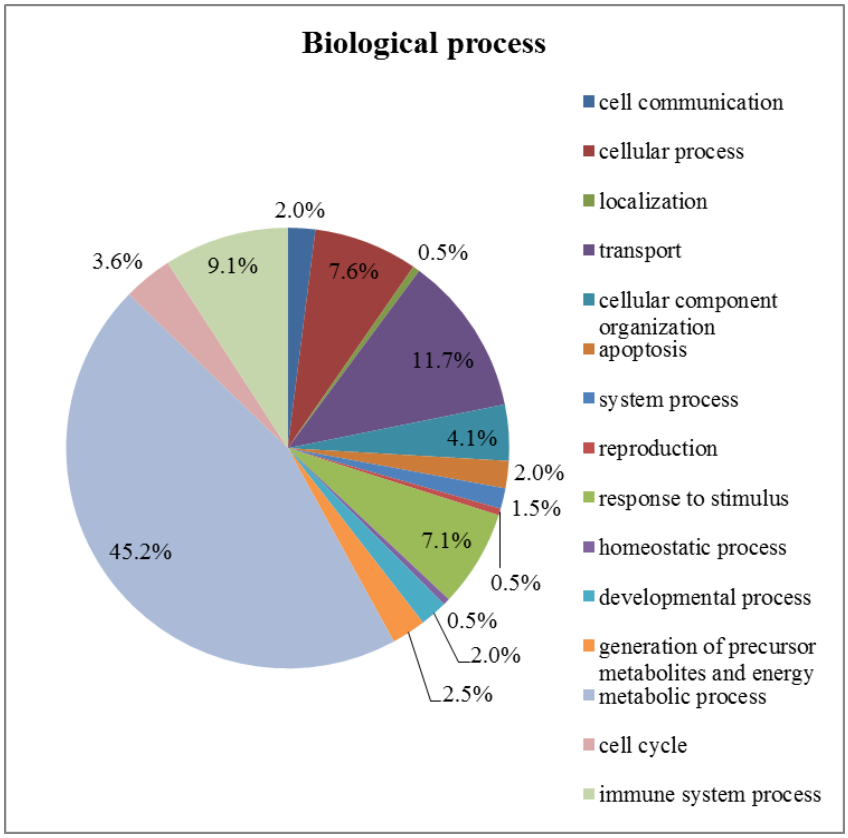
*in digital version only


Protein class
protease
cytoskeletal protein
transporter
transferase
oxidoreductase
lyase
ligase
nucleic acid binding
signaling molecule
enzyme modulator
calcium-binding protein
hydrolase
transfer/carrier protein
membrane traffic protein
phosphatase
transcription factor
chaperone
kinase
isomerase
receptor
6.8%
2.0%
10.2%
6.8%
5.4%
1.4%
0.7%
19.7%
0.7%
4.1%
1.4%
14.3%
1.4%
0.7%
1.4%
0.7%
11.6%
2.7%
4.8%
3.4%

a)


Molecular function
ion channel activity
transporter activity
translation regulator activity
transcription regulator activity
enzyme regulator activity
catalytic activity
receptor activity
antioxidant activity
structural molecule activity
binding
2.9%
10.9%
2.2%
0.7%
2.2%
36.5%
3.6%
1.5%
12.4%
27.0%

b)



c)

Fig. 8. Functional profile of *Plasmodium* expressed proteins defined as a) Protein class; b) Molecular function and c) Biological process; according to PANTHER software (www.pantherdb.org).

It was possible to allocate 108 proteins to 20 different classes, with ten molecular functions and involved in 15 biological processes. Nucleic-acid binding proteins (Panther PC00171) were the most prevalent (19.7%), followed by hydrolases (Panther PC00121) (14.3%) and chaperones (Panther PC00072) (11.6%). Catalytic activity (GO:0003824) and binding (GO:0005488) comprised 63.5% of all molecular functions, in accordance with the most prevalent proteins classes. Metabolism (GO:0008152) was, by far, the most represented process (45.2%). These data are absolutely consistent with previous transcriptome records, reporting that during ring and early trophozoite stage there is an induction of genes associated with transcriptional and translational machinery, glycolysis and ribonbucleotide biosynthesis and that during the trophozoite stage, metabolism is at its peak (Bozdech, et al., 2003).

When we looked to the remaining 55 proteins (**Table 2**), we confirmed that 17 of these were actually conserved or unclassified proteins with unknown function but most of the remaining were exclusive to Apicomplexa protozoans or *Plasmodium* (then not categorized by GO, that standardize the representation of genes and gene products attributes across species). Therefore, manual annotation was performed. Interestingly, 27 of these proteins were annotaded as being involved in parasite-host interactions, putatively localized at cell surface and some of them are specifically expressed in rhoptries (specialized secretory organelles at the apical pole of the parasite with the cellular function of releasing enzymes during the invasion process, consequently important for host-parasite interaction); a few corresponded to proteins exported by the parasite to the RBC to accomplish the host cell remodelling (Goldberg and Cowman, 2010); and other are widely known surface antigens causing immune response in humans and even vaccine candidates, as is the case of merozoite surface proteins (MSPs) 1 and 2 (Aubouy, Migot-Nabias and Deloron, 2003). Two proteins seem to be involved in parasite sexual stage development and the remaining five have probably chaperone functions and are implicated in gene regulation, cell redox homeostase and transport (PLasmoDB database, www.plasmodb.org).

Table 2. Functional profiles of proteins with unknown function according to PANTHER (www.pantherdb.org).

#	Accession	Protein	Function
1	PFI1445w	High molecular weight rhoptry protein-2	Host-parasite interaction (invasion)
2	PFI0265c	RhopH3	Host-parasite interaction (invasion)
3	PF14_0102	rhoptry-associated protein 1, RAP1	Host-parasite interaction (invasion)
4	PFE0080c	rhoptry-associated protein 2, RAP2	Host-parasite interaction (invasion)
5	PFE0075c	rhoptry-associated protein 3, RAP3	Host-parasite interaction (invasion)
6	PF14_0201	surface protein, Pf113	Host-parasite interaction (RBC remodelling)
7	PFE0060w	PIESP2 RBC surface protein	Host-parasite interaction (RBC remodelling)
8	PFE0065w	skeleton-binding protein 1	Host-parasite interaction (RBC remodelling)
9	PFI1735c	ring-exported protein 1	Host-parasite interaction (RBC remodelling)
10	PF14_0678	exported protein 2	Host-parasite interaction (RBC remodelling)
11	PFE1600w	Plasmodium exported protein (PHISTb), unknown function	Host-parasite interaction (RBC remodelling)
12	PFD0080c	Plasmodium exported protein (PHISTb), unknown function	Host-parasite interaction (RBC remodelling)
13	PF14_0744	Plasmodium exported protein, unknown function	Host-parasite interaction (RBC remodelling)
14	MAL13P1.61	Plasmodium exported protein (hyp8), unknown function	Host-parasite interaction (RBC remodelling)
15	PFB0106c	Plasmodium exported protein, unknown function	Host-parasite interaction (RBC remodelling)
16	PF14_0016	early transcribed membrane protein 14.1, etramp14.1	Host-parasite interaction (RBC remodelling)
17	PF10_0019	early transcribed membrane protein 10.1, etramp 10.1	Host-parasite interaction (RBC remodelling)
18	PF10_0323	early transcribed membrane protein 10.2, etramp 10.2	Host-parasite interaction (RBC remodelling)
19	PF10_0159	glycophorin-binding protein 130 precursor	Host-parasite intercation (surface antigen)
20	PF10_0372	Antigen UB05	Host-parasite intercation (surface antigen)
21	PFI1475w	merozoite surface protein 1 precursor	Host-parasite intercation (surface antigen)

22	PF13_0011	plasmodium falciparum gamete antigen 27/25	Host-parasite intercation (surface antigen)
23	PF10_0025	PF70 protein	Host-parasite intercation (surface antigen)
24	PF11_0224	circumsporozoite-related antigen	Host-parasite intercation (surface antigen)
25	PF13_0197	Merozoite Surface Protein 7 precursor, MSP7	Host-parasite intercation (surface antigen)
26	PFL1385c	Merozoite Surface Protein 9, MSP-9	Host-parasite intercation (surface antigen)
27	PFB0915w	liver stage antigen 3	Host-parasite intercation (surface antigen)
28	PFA0110w	DNAJ protein, putative	Protein folding
29	MAL7P1.228	Heat Shock 70 KDa Protein, (HSP70)	Protein folding
30	MAL13P1.221	aspartate carbamoyltransferase	Metabolism
31	PF11_0281	protein phosphatase, putative	Metabolism
32	MAL8P1.72	high mobility group protein	Gene regulation
33	PF10_0063	DNA/RNA-binding protein, putative	Gene regulation
34	PF13_0272	thioredoxin-related protein, putative	Cell redox homeostase
35	MAL8P1.17	protein disulfide isomerase	Cell redox homeostase
36	PFL0795c	male development gene 1	Sexual stage
37	PFD0310w	sexual stage-specific protein precursor	Sexual stage
38	MAL13P1.231	Sec61 alpha subunit, PfSec61	Transport
39	PFI1740c-a	location=Pf3D7_09:1427463-1428011(-) | length=94	Unclassified
40	PF14_0344	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
41	PF11_0302	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
42	PF14_0567	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
43	PFI0605c	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
44	PFL1825w	conserved Plasmodium membrane protein, unknown function	Conserved protein, unknown function
45	MAL7P1.67	conserved Plasmodium protein, unknown function	Conserved protein, unknown function

46	PF14_0301	conserved protein, unknown function	Conserved protein, unknown function
47	MAL8P1.62	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
48	PF11_0179	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
48	PF11_0069	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
50	PFI1270w	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
51	MAL8P1.95	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
52	PF14_0046	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
53	PFC0715c	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
54	PF14_0105	conserved Plasmodium protein, unknown function	Conserved protein, unknown function
55	MAL13P1.237	conserved Plasmodium protein, unknown function	Conserved protein, unknown function

This proteomic profile is in line with previous trophozoite proteome reports (Florens, et al., 2002), describing that the principal modifications during the initial trophozite phase allow the parasite to transfer molecules in and out of the cell, to prepare the surface of the RBC to mediate cytoadherence (where skeleton binding-proteins, RBC surface proteins and exported proteins seem to be involved), and to digest the cytoplasmic contents, particularly hemoglobin, in its food vacuole. Digestion of hemoglobin is a major parasite catabolic process (Klemba and Goldberg, 2002), with proteases (namely plasmepsins and falcilysin) being the fifth (in 20) more prevalent class of proteins identified in this study.

When we looked to the SwissProt database MS search results, the main proteins identified were from human origin (data not shown). The proteins with highest scores were, as expected: Hb (beta, alfa and gamma subunits), band 3 anion transport protein, spectrin, ankyrin, serum albumin precursor, CA and RBC membrane protein 4.2.

Some parasite proteins had considerably differences in the number of peptides (and consequently in sequence coverage and scores) identified in normal and deficient environments. In PK assay (**Table S9**), six proteins [MSP 1 precursor (PFI1475w); rhoptry-associated protein 2, RAP2 (PFE0080c); multidrug resistance protein (PFE1150w); ATP synthase beta chain, mitochondrial precursor (PFL1725w); adenylate kinase (PF10_0086) and heat shock protein 70 (MAL13P1.540)] showed a difference of 15 or more peptides identified in both conditions (considering both replicates). In G6PD assay (**Table S10**), this difference was smaller: the protein with the highest disparity (8 peptides) was DNAJ protein (PFA0110w).

Curiously, in the PK assay, the majority of proteins had more peptides identified in extracts of parasites growing in normal RBC. ATP synthase beta chain was one of the few exceptions, with 25 peptides (sum of peptides from both replicates) identified in parasites from PKD RBC and none in controls, suggesting that this protein may be over-expressed in *Plasmodium* in a PKD environment. No additional ATP synthase peptides were identified in control samples from the G6PD assay and only one was identified in G6PDD RBC. However, no other subunit was identified from mitochondrial ATP synthase (canonical F_1F_0-ATP synthase includes a F1 domain with five subunits and a F_0 domain with six subunits), but all sequenced apicomplexan parasites, including *P.*

falciparum, seem to lack critical subunits of the enzyme (which can be due to the detection incapacity of bioinformatic tools because of a high degree of divergence) (Balabaskaran Nina, et al., 2011).

Mass spectrometry is not inherently quantitative, because proteolytic peptides show great variability in physiochemical properties resulting in variability in mass spectrometric response between runs. Additionally, mass spectrometers only analyze a small percentage of the total peptides because of the overwhelming number of proteotypic peptides in a sample (Bantscheff, et al., 2007). Therefore, the number of peptides is merely suggestive about the abundance of a protein. However, such a big difference in the number of peptides from ATP synthase in both conditions is noteworthy, especially because it is much higher in deficient conditions, counteracting the trend of most identified molecules.

The role of the mitochondrial ATP synthase in *P. falciparum* has remained unclear for decades. Biochemical data indicate that the *Plasmodium* mitochondrion does not seem to be a source of ATP (Fry, Webb and Pudney, 1990) as the major supply of ATP in the parasite is the anaerobic glycolysis pathway (Lang-Unnasch and Murphy, 1998). Yet, the mitochondrial electron transport chain is critical for parasite survival and a target for antimalarial drugs (Mather, Henry and Vaidya, 2007). *Plasmodium falciparum* seems to maintain an active mitochondrial electron transport chain to serve one main metabolic function: regeneration of ubiquinone required as the electron acceptor for dihydroorotate dehydrogenase, an essential enzyme for pyrimidine biosynthesis (Painter, et al., 2007). Therefore, the functions of ATP synthase may then be: providing an adjunct mechanism for the maintenance of electropotential across the mitochondrial inner membrane through ATP hydrolysis and proton pumping; production of ATP for local consumption without making significant contributions to the cytosol (then, not detected on biochemical measurements); and participate in mitochondrial morphogenesis (Balabaskaran Nina, et al., 2011).

2.3.2. Quantitative analysis

It was possible to get quantitative data for 50 *Plasmodium* proteins from PK assay (**Table 3**) and for 40 proteins from G6PD assay (**Table 4**). In order to express relative abundance of each protein, a median ratio was calculated (G6PDD/G6PDN or PKD/PKN). A median ratio of 1 means that the abundance of the protein is exactly the same in deficient and control conditions, a ratio <1 means that the protein is under-expressed and a ratio >1 means that the protein is over-expressed in the deficient condition. Curiously, in PK assay, only three showed a ratio >1; conversely, in G6PD assay, only six showed a ratio <1. Twenty-one were common to both lists and from these, only three showed an expression alteration in the same direction: the MSPs, MSP1 (PFI1475w) and MSP9 (PFL1385c) were under-expressed in both deficient conditions, whereas the ring-exported protein 1 (PFI1735c) was over-expressed.

A cut-off for the median ratio was applied as follows: ≤ 0.55 for under-expressed and ≥ 1.45 for over-expressed, resulting in a total of 45 proteins with alteration in their expression in PK assay and nine in G6PD assay (**Tables 3** and **4**, respectively). As expected, most (4/6) of the proteins with higher difference in the number of detected peptides in parasites growing in normal and deficient conditions (qualitative analysis) were among the proteins with higher difference in quantitative measurements: MSP1 precursor (PFI1475w); multidrug resistance protein (PFE1150w); adenylate kinase (PF10_0086) and heat shock protein 70 (MAL13P1.540). They all presented a notably low PKD/PKN ratio between 0.3 and 0.36. ATP synthase subunits were not identified in quantitative analysis, however, considering the overlap between qualitative and quantitative data, there's a high probability of this enzyme be truly over-expressed in deficient conditions. A possible explanation for no quantitative data may be the relative quantitation method itself, that in absence of signal in one of the two conditions (normal or deficient), gives no output.

Table 3. MS quantitative results: relative abundance of proteins from *P. falciparum* 3D7 in PKD relative to PKN (determined as the median ratio PKD: PKN1+PKN2).

Accession	Protein	MW [kDa]	pI	Scores	Peptides	SC	Median (PKD:PKN1+N2)	# (PKD:PKN1+N2)	CV[%] (PKD:PKN1+N2)
PF14_0377	vesicle-associated membrane protein, putative	27.7	8.8	67.7	2	12	**0.24**	1	0
PF10_0019	early transcribed membrane protein 10.1, etramp 10.1	11.3	10.4	66.5	1	11.2	**0.25**	1	0
PF13_0141	L-lactate dehydrogenase	34.1	7.8	647.5	10	59.2	**0.26**	1	0
PF11_0069	conserved Plasmodium protein, unknown function	30.6	4.8	334.6	6	30.5	**0.27**	1	0
PFI1270w	conserved Plasmodium protein, unknown function	24.7	5.4	458.6	9	47	**0.28**	3	11.77
PF14_0075	plasmepsin IV	51	5.2	610.8	9	31.2	**0.29**	1	0
PF13_0272	thioredoxin-related protein, putative	24	10.1	425.0	9	35.6	**0.29**	3	10.33
PF14_0102	rhoptry-associated protein 1, RAP1	90	6.7	1395.3	26	57.3	**0.29**	4	12.15
PFE1150w	multidrug resistance protein	162.1	9.5	1221.2	22	22.9	**0.3**	3	7.74
PF14_0076	plasmepsin I	51.4	6.9	858.7	14	41.8	**0.3**	4	47.19
PF11_0055	conserved protein, unknown function	49.2	9.8	298.8	9	29	**0.31**	2	1.3
PFI1475w	merozoite surface protein 1 precursor	195.6	6.1	1249.6	23	20.5	**0.32**	1	0
PF11_0062	histone H2B	13.1	10.8	147.9	2	31.6	**0.32**	1	0
PF11_0302	conserved Plasmodium protein, unknown function	51.9	4.8	169.5	3	8.8	**0.33**	1	0
PF14_0301	conserved protein, unknown function	33.2	9.6	131.4	3	17.3	**0.33**	1	0
MAL13P1.540	heat shock protein 70 (hsp70), putative	108.1	5.4	448.8	9	18.8	**0.34**	1	0
PF11_0301	spermidine synthase	36.6	8.8	267.2	5	24.6	**0.34**	3	17.85
MAL8P1.69	14-3-3 protein, putative	30.2	4.7	214.3	4	24.4	**0.35**	1	0
PF10_0086	adenylate kinase	27.6	9.6	422.3	8	45.5	**0.36**	1	0
PFB0210c	hexose transporter, PfHT1	56.4	9.5	143.4	2	6	**0.36**	1	0
MAL8P1.17	protein disulfide isomerase	55.5	5.5	1086.5	18	59.4	**0.36**	4	11.59

PFI0875w	Heat shock protein 70 (HSP70) homologue	72.3	5	1797.3	26	53.1	**0.36**	13	15.2
PF08_0074	DNA/RNA-binding protein Alba, putative	27.2	11.1	121.8	2	17.7	**0.37**	1	0
PFE1590w	early transcribed membrane protein 5, ETRAMP5	19	5.1	189.0	2	20.4	**0.38**	1	0
PF10_0100	conserved Plasmodium protein, unknown function	13.7	10.7	29.1	1	9.3	**0.38**	1	0
PF11_0313	60S ribosomal protein P0	34.9	6.3	442.6	9	53.8	**0.38**	2	5.38
PF13_0304	elongation factor-1 alpha	48.9	9.7	656.7	14	43.6	**0.39**	3	16.16
PFL0740c	10 kd chaperonin	11.1	5.3	53.2	2	23.3	**0.4**	1	0
PF11_0179	conserved Plasmodium protein, unknown function	15.3	10.1	213.5	4	27.3	**0.4**	1	0
PF14_0541	V-type H(+)-translocating pyrophosphatase, putative	76.4	6.1	483.8	8	15.9	**0.4**	2	40.8
PF14_0678	exported protein 2	33.4	4.9	231.8	4	28.6	**0.41**	1	0
PF11_0338	aquaglyceroporin	28.3	7.8	155.3	3	14.3	**0.41**	1	0
MAL13P1.56	m1-family aminopeptidase	126	7.7	639.3	15	26.3	**0.42**	2	1.02
PFI0880c	glideosome-associated protein 50	44.6	9.3	218.9	5	27.5	**0.43**	1	0
PF08_0054	heat shock 70 kDa protein	73.9	5.4	814.4	19	41.5	**0.45**	2	5.51
PFC0725c	formate-nitrate transporter, putative	34.4	9.4	70.0	2	6.5	**0.46**	1	0
PF11_0351	heat shock protein hsp70 homologue	73.3	6.6	706.0	15	37.3	**0.46**	2	14.93
PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	36.6	8.7	827.5	13	61.7	**0.46**	4	9.21
PFE0065w	skeleton-binding protein 1	36.3	4.2	263.0	6	38	**0.47**	1	0
PFC0400w	60S Acidic ribosomal protein P2, putative	11.9	4.3	378.8	5	69.6	**0.48**	2	8.26
PFL1070c	endoplasmin homolog precursor, putative	95	5.1	1609.1	25	41.9	**0.5**	5	15.31
PFL1385c	Merozoite Surface Protein 9, MSP-9	86.6	4.6	64.9	2	3.9	**0.51**	1	0
PF11_0061	histone H4	11.4	11.7	111.5	4	40.8	**0.52**	1	0
PF14_0078	HAP protein	51.7	8.8	644.6	12	36.6	**0.56**	4	51.93
PF10_0153	heat shock protein 60	62.5	6.8	748.7	15	37.1	**0.57**	1	0
PFD1035w	steroid dehydrogenase, putative	37.2	10	81.2	2	8.7	**0.7**	1	0

PF14_0077	plasmepsin II	51.4	5.3	329.9	6	27.6	**0.79**	3	95.9
PFE0625w	Rab1b, GTPase	22.9	6.2	71.0	2	11	**1.27**	1	0
PF14_0630	protein serine/threonine phosphatase	100.7	7	20.7	1	0.8	**1.5**	1	0
PFI1735c	ring-exported protein 1	83	5.3	63.8	3	4.2	**1.78**	1	0

Accession: gene accession number; Protein: protein name; Mw [kDa]: molecular weight; pI: isoelectric point; Scores: protein Mascot scores (reflecting the combined scores of all observed mass spectra that can be matched to amino acid sequences within that protein; a higher score indicates a more confident match); Peptides: number of peptides identified; SC: sequence coverage; Median (PKD/PKN1+N2): median of all peptides ratio based on three technical replicates of each sample (3xPKD1; 3xPKD2; 3xPKN1+PKN2), is indicative of the abundance of protein in PKD relative to control; # (PKD/PKN1+N2): number of peptides present in both PKD and control samples in which the median is based; CV[%](PKD/PKN1+N2): coefficient of variation; PKD: parasites grown in PK-deficient RBC; PKN1+N2: pooled sample of PKN1 and PKN2. 1: replicate 1; 2: replicate 2. In gray, the proteins excluded considering the cut-off ratio ($0.55 \geq$ median (PKD:PKN1+N2) ≥ 1.45).

Table 4. MS quantitative results: relative abundance of proteins from *P. falciparum* 3D7 in G6PDD relative to G6PDN (determined as the median ratio G6PDD: N1+N2).

Accession	**Protein**	**MW [kDa]**	**pI**	**Scores**	**Peptides**	**SC**	**Median (G6PDD:N1+N2)**	**# (G6PDD:N1+N2)**	**CV [%] (G6PDD:N1+N2)**
PFI1475w	merozoite surface protein 1 precursor	195.6	6.1	1249.6	23	20.5	**0.55**	4	12.83
PF10_0268	merozoite capping protein 1	43.9	10.2	278.4	4	20.9	**0.61**	1	0
PFL1385c	Merozoite Surface Protein 9, MSP-9	86.6	4.6	64.9	2	3.9	**0.63**	1	0
PF14_0016	early transcribed membrane protein 14.1, etramp14.1	11.4	10	53.9	1	12.1	**0.63**	1	0
PFE0660c	purine nucleotide phosphorylase, putative	26.8	6.1	125.5	3	24.1	**0.69**	1	0
PF10_0155	enolase	48.6	6.2	453.3	9	38.1	**0.72**	1	0
PFE0080c	rhoptry-associated protein 2, RAP2	46.7	9.4	1009.6	14	44.7	**1.09**	1	0
PF14_0548	ATPase, putative	48.2	9.2	40.8	1	2.9	**1.11**	1	0
PF11_0179	conserved Plasmodium protein, unknown function	15.3	10.1	213.5	4	27.3	**1.11**	2	47.88
PF14_0231	60S ribosomal protein L7-3, putative	32.7	10.8	52.9	1	6.7	**1.15**	1	0

PFE0065w	skeleton-binding protein 1	36.3	4.2	263.0	6	38	**1.16**	1	0
PF08_0074	DNA/RNA-binding protein Alba, putative	27.2	11.1	121.8	2	17.7	**1.18**	1	0
PFE0850c	60S ribosomal protein L12, putative	18.1	10.2	167.5	4	27.9	**1.18**	2	6.45
MAL13P1.56	m1-family aminopeptidase	126	7.7	639.3	15	26.3	**1.2**	1	0
PFE1150w	multidrug resistance protein	162.1	9.5	1221.2	22	22.9	**1.2**	1	0
PF14_0076	plasmepsin I	51.4	6.9	858.7	14	41.8	**1.2**	1	0
PFC0900w	T-complex protein 1 epsilon subunit, putative	59.1	5.6	31.1	1	2.6	**1.21**	1	0
PF14_0678	exported protein 2	33.4	4.9	231.8	4	28.6	**1.22**	1	0
PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	36.6	8.7	827.5	13	61.7	**1.23**	2	9.04
MAL8P1.17	protein disulfide isomerase	55.5	5.5	1086.5	18	59.4	**1.26**	2	9.39
PF13_0272	thioredoxin-related protein, putative	24	10.1	425.0	9	35.6	**1.28**	1	0
PFL2405c	PFG377 protein	377.2	5.7	21.7	1	0.3	**1.29**	1	0
PFI0875w	Heat shock protein 70 (HSP70) homologue	72.3	5	1797.3	26	53.1	**1.3**	4	3.03
PF14_0517	peptidase, putative	88.4	6.4	411.3	8	15.7	**1.32**	1	0
PF11_0313	60S ribosomal protein P0	34.9	6.3	442.6	9	53.8	**1.34**	1	0
PF07_0029	heat shock protein 86	86.1	4.8	763.5	14	34.1	**1.34**	3	7.41
PF13_0304	elongation factor-1 alpha	48.9	9.7	656.7	14	43.6	**1.34**	3	5.13
PF14_0630	protein serine/threonine phosphatase	100.7	7	20.7	1	0.8	**1.36**	1	0
PF14_0201	surface protein, Pf113	112.5	4.3	365.4	9	12.9	**1.38**	1	0
PFL0740c	10 kd chaperonin	11.1	5.3	53.2	2	23.3	**1.38**	1	0
PFD0310w	sexual stage-specific protein precursor	16.6	5.8	344.8	4	38.9	**1.39**	2	10.87
PFD0305c	vacuolar ATP synthase subunit b	55.8	5.4	265.7	7	23.9	**1.42**	1	0
PF11_0331	TCP-1/cpn60 chaperonin family	60.2	6.8	47.4	1	3.1	**1.48**	1	0
PFL1070c	endoplasmin homolog precursor, putative	95	5.1	1609.1	25	41.9	**1.49**	3	23.56
MAL13P1.221	aspartate carbamoyltransferase	43.2	9.1	182.1	4	14.7	**1.52**	1	0

PF08_0054	heat shock 70 kDa protein	73.9	5.4	814.4	19	41.5	**1.53**	3	8.7
PF14_0391	60S ribosomal protein L1, putative	24.8	10.4	30.4	1	7.8	**1.54**	1	0
PFI1735c	ring-exported protein 1	83	5.3	63.8	3	4.2	**1.67**	1	0
PF11_0351	heat shock protein hsp70 homologue	73.3	6.6	706.0	15	37.3	**1.68**	2	24.81
PF13_0346	60S ribosomal protein L40/UBI, putative	14.6	10.8	142.5	2	31.2	**1.74**	1	0

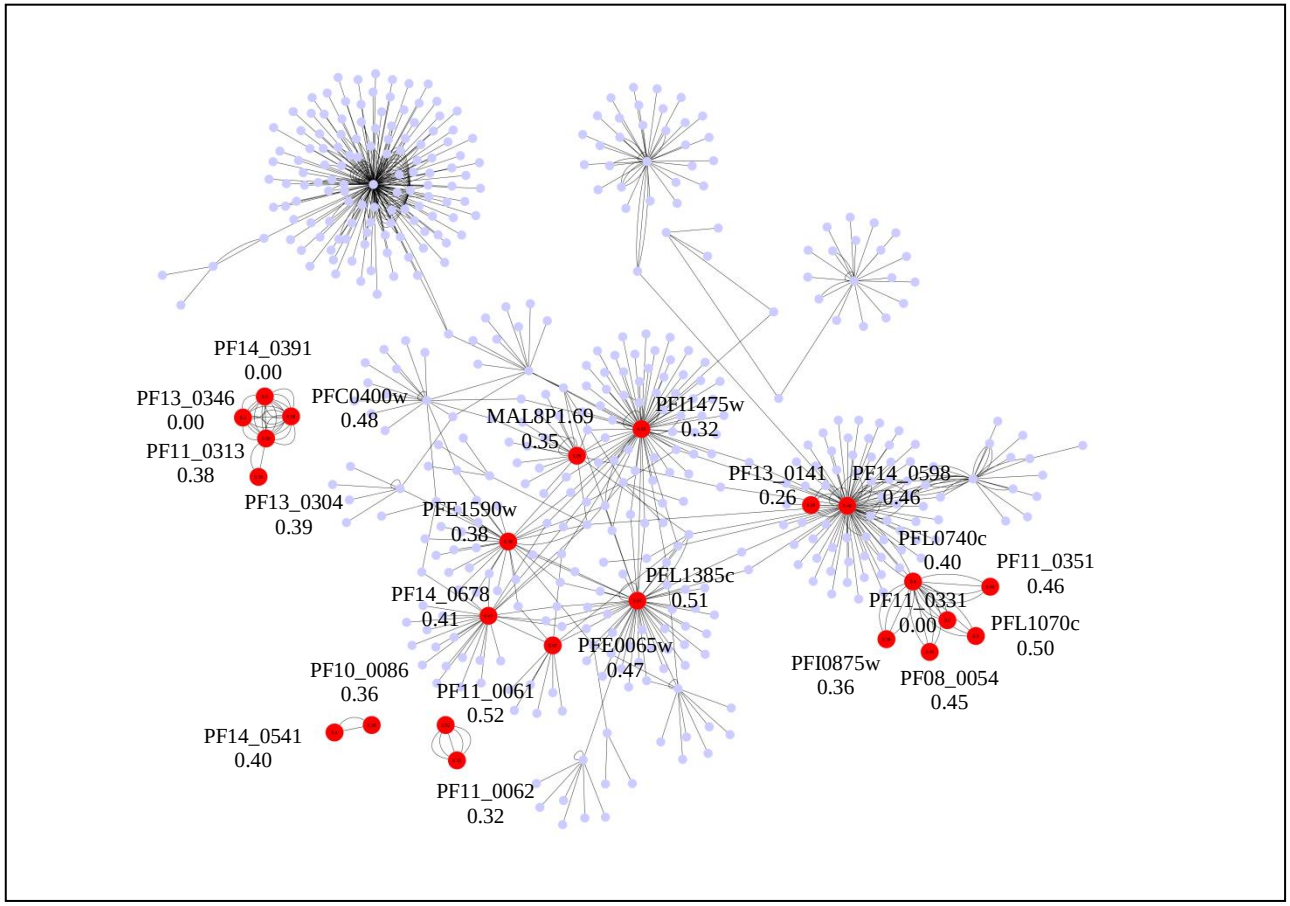
Accession: gene accession number; Protein: protein name; Mw [kDa]: molecular weight; pI: isoelectric point; Scores: protein Mascot scores (reflecting the combined scores of all observed mass spectra that can be matched to amino acid sequences within that protein; a higher score indicates a more confident match); Peptides: number of peptides identified; SC: sequence coverage; Median (G6PDD/N1+N2): median of all peptides ratio based on three technical replicates of each sample (3xG6PDD1; 3xG6PDD2; 3xG6PDN1+G6PDN2), is indicative of the abundance of protein in G6PDD relative to control; # (G6PDD/N1+N2): number of peptides present in both G6PDD and control samples in which the median is based; CV[%](PKD/PKN1+N2): coefficient of variation; G6PDD: parasites grown in G6PD-deficient RBC; N1+N2: pooled sample of G6PDN1 and G6PDN2. 1: replicate 1; 2: replicate 2. In gray, the proteins excluded considering the cut-off ratio ($0.55 \geq$ median (G6PDD:N1+N2) ≥ 1.45).

2.4. Protein-Protein interaction analysis

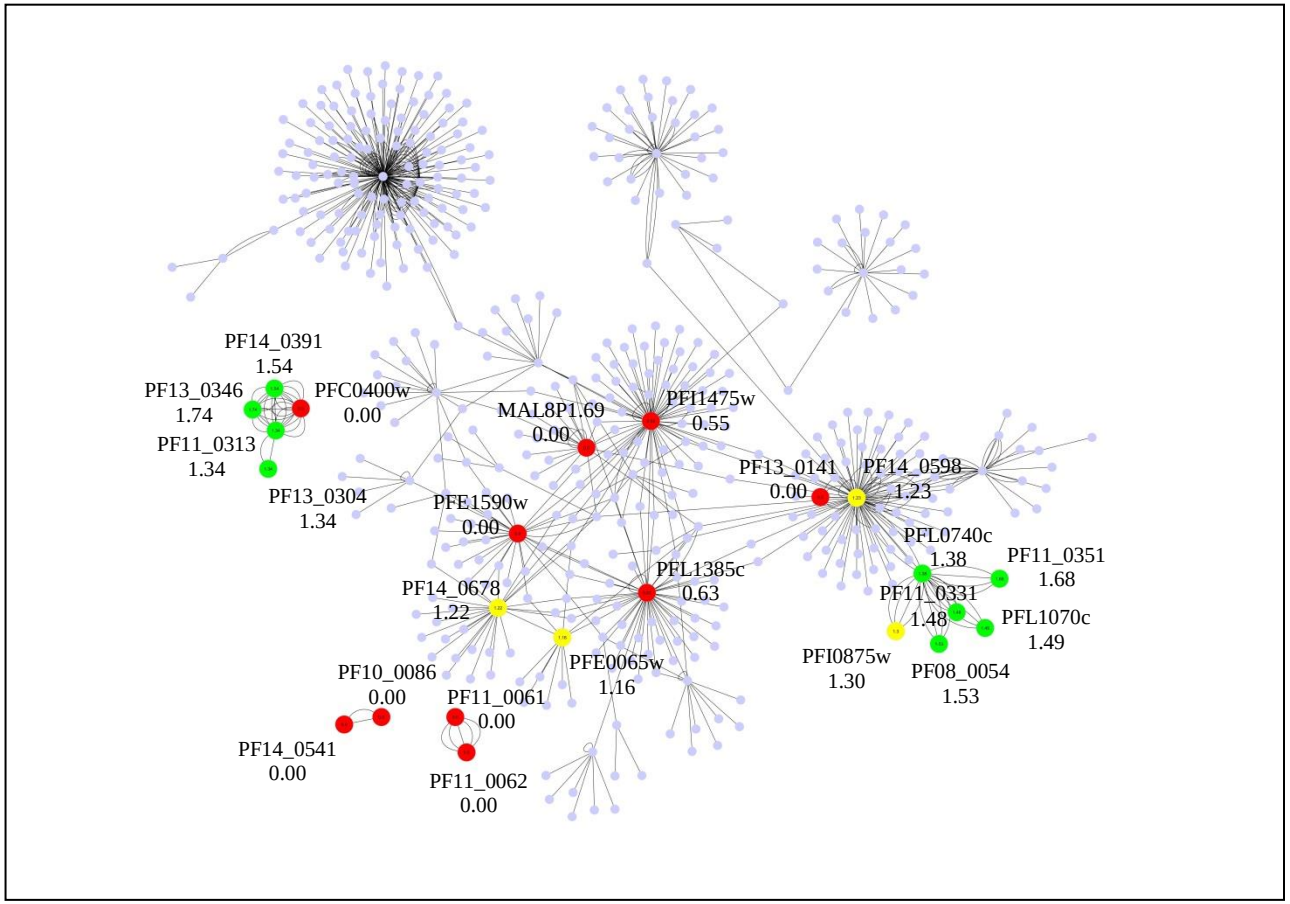
In order to understand the biological relevance of these alterations in protein abundance in parasites growing in PKD and G6PDD RBC relative to normal conditions, interactions among proteins were investigated to identify pathway(s) in which they were involved, in order to try to unveil the mechanism(s) used by the parasite to respond to these stress conditions. The proteins selected to this protein-protein interaction analysis (**Table S12**) were those showing median ratio ≤ 0.55 and ≥ 1.45 and those common to PK and G6PD quantitative lists.

The protein-protein interaction networks presented in **Fig. 9** were imported from Intact (http://www.ebi.ac.uk/intact/) and STRING 9.0 (http://string-db.org) and analyzed with Cytoscape software v2.8.3. Contain 522 proteins and 740 protein-protein interactions. These proteins are distributed in 41 biological pathway terms being the top four: catabolic process (GO:0009056), response to abiotic stimulus (GO:0009628), response to temperature stimulus (GO:0009266) and response to heat (GO:0009408). All the proteins involved in carbon catabolism (glycolysis) and Hb catabolism were included in the catabolic process category whereas the response to abiotic stimulus, to temperature stimulus and to heat included all the chaperones, heat-shock proteins and all the molecules that contribute to cellular redox homeostasis. Altough no heat stress occurred in our cultures (to our knowledge), the high abundance of heat-shock proteins, whose best-known role is the response to temperature, brought this category out.

Four networks were identified: one big network, including 14 proteins, a second network including five proteins and two other networks, with only two proteins each. The largest network included proteins involved in three major biological processes: protein folding/response to stress, glycolysis and host-parasite interaction/protein binding. The second largest network included proteins localized in ribosomes (involved in translation). In PKD condition [**Fig. 9a)**], all except three proteins (these three with no quantitative data available) were under-expressed; conversely, in G6PD condition, proteins involved in protein folding/stress response and from ribosomes showed over-expression [**Fig. 9b)**].


PF14_0391
0.00
PF13_0346
0.00
PFC0400w
0.48
PF11_0313
0.38
PF13_0304
0.39
MAL8P1.69
0.35
PFI1475w
0.32
PF13_0141
0.26
PF14_0598
0.46
PFE1590w
0.38
PFL0740c
0.40
PF11_0351
0.46
PFL1385c
0.51
PF14_0678
0.41
PF11_0331
0.00
PFL1070c
0.50
PFE0065w
0.47
PFI0875w
0.36
PF08_0054
0.45
PF10_0086
0.36
PF11_0061
0.52
PF14_0541
0.40
PF11_0062
0.32

a) PK assay



b) G6PD assay

Fig. 9. Protein-protein interaction networks obtained with Cytoscape in parasites growing in PKD RBC [**a)**] and G6PDD RBC [**b)**] (coloured nodes). Red node: median (PKD/PKN1+N2) or (G6PDD/N1+N2) < 0.65 or protein without quantitative data (0.00); yellow node: 0.65 ≤ median (PKD/PKN1+N2) or (G6PDD/N1+N2) ≤ 1.30; green node: median (PKD/PKN1+N2) or (G6PDD/N1+N2) > 1.30.

Based on this analysis, the parasite response to both enzyme deficiencies seems to be distinct: in PKD condition, the parasite seems to respond with a general under-expression of chaperones, catabolic proteins and host-parasite interaction proteins; in G6PDD, the parasite seems to respond to oxidative stress, enhancing the abundance of stress response proteins. However, protein-protein interaction analysis seems incomplete, because the parasite specific proteins (not categorized by GO) are not included. So, we decided to go further and, again, do a manual search of these parasite proteins.

Tables 5 and **6** show, respectively, the list of proteins whose abundance was considered altered ($1.45 \leq$ median ratio ≤ 0.55) from parasites growing in G6PDD and in PKD RBC and the respective putative function and cellular localization (might not be exaustive) identified by manual search. Proteins with unknown function are not shown; proteins with recently known function (initially classified as unknown) were included.

Table 5. Putative function and cellular localization of parasite proteins with altered expression (1.45 ≤ median ratio ≤ 0.55) in G6PDD conditions (may not include all the organelles where the protein is expressed).

Accession	Protein	Median (G6PDD:N1+N2)	Function	Probable localization
PFI1475w	merozoite surface protein 1 precursor	**0.55**	host-parasite interaction	cell surface
PF11_0331	TCP-1/cpn60 chaperonin family	**1.48**	protein folding/stress response	cytosol and organelles
PFL1070c	endoplasmin homolog precursor, putative	**1.49**	protein folding/stress response	endoplasmic reticulum
MAL13P1.221	aspartate carbamoyltransferase	**1.52**	pyrimidine byosynthetic pathway	cytosol
PF08_0054	heat shock 70 kDa protein	**1.53**	protein folding/stress response	cytosol and organelles
PF14_0391	60S ribosomal protein L1, putative	**1.54**	Translation	ribosome
PFI1735c	ring-exported protein 1	**1.67**	host-parasite interaction	cell surface
PF11_0351	heat shock protein hsp70 homologue	**1.68**	protein folding/stress response	cytosol and organelles
PF13_0346	60S ribosomal protein L40/UBI, putative	**1.74**	Translation	ribosome

Table 6. Putative function and cellular localization of parasite proteins with altered expression (1.45 ≤ median ratio ≤ 0.55) in PKD conditions (may not include all the organelles where the protein is expressed). (In bold and gray background, the proteins putatively associated to Maurer's clefts).

Accession	Protein	Median (PKD: PKN1+N2)	Function	Probable localization
PF14_0377	vesicle-associated membrane protein, putative	**0.24**	transport	vesicle membrane
PF10_0019	**early transcribed membrane protein 10.1, etramp 10.1**	**0.25**	**host-parasite interaction**	**parasitophorous vacuole membrane**
PF13_0141	L-lactate dehydrogenase	**0.26**	glycolysis	cytoplasm
PF14_0075	plasmepsin IV	**0.29**	proteolysis/haemoglobin catabolic process	digestive vacuole
PF13_0272	thioredoxin-related protein, putative	**0.29**	stress response/redox homeostasis	endoplasmic reticulum
PF14_0102	rhoptry-associated protein 1, RAP1	**0.29**	host-parasite interaction	rhoptries
PFE1150w	multidrug resistance protein	**0.3**	transport, response to drug, ATPase activity	endoplasmic reticulum, vacuole membrane
PF14_0076	plasmepsin I	**0.3**	proteolysis, haemoglobin catabolic process	digestive vacuole
PFI1475w	merozoite surface protein 1 precursor	**0.32**	host-parasite interaction	cell surface
PF11_0062	histone H2B	**0.32**	DNA binding	nucleous
PF11_0302	conserved Plasmodium protein, unknown function parasitophorus vacuolar protein 1 (PV1)	**0.33**	protein binding/signal transduction	parasitophorous vacuole
MAL13P1.540	**heat shock protein 70 (hsp70), putative**	**0.34**	**protein folding/response to stress**	**cytosol and organelles**
PF11_0301	spermidine synthase	**0.34**	spermidine biosynthetic process/catalytic activity	cytosol
MAL8P1.69	**14-3-3 protein, putative**	**0.35**	**protein domain binding/host-parasite interaction**	**cytosol and plasma membrane**
PF10_0086	**adenylate kinase**	**0.36**	**nucleotide kinase activity/ATP binding**	**mithocondrion**
PFB0210c	hexose transporter, PfHT1	**0.36**	transport	parasitophorus vacuole, plasma membrane
MAL8P1.17	**protein disulfide isomerase**	**0.36**	**protein folding/stress response**	**endoplasmic reticulum**
PFI0875w	**Heat shock protein 70 (HSP70) homologue**	**0.36**	**protein folding/stress response**	**cytosol and organelles**
PF08_0074	DNA/RNA-binding protein Alba, putative	**0.37**	nucleic acid binding	nucleous

PFE1590w	**early transcribed membrane protein 5, ETRAMP5**	**0.38**	**host-parasite interaction**	**parasitophorous vacuole membrane**
PF10_0100	conserved Plasmodium protein, unknown function	**0.38**	electron flow	mithocondrion
	succinate dehydrogenase subunit 4, putative			
PF11_0313	60S ribosomal protein P0	**0.38**	translation	ribosome
PF13_0304	elongation factor-1 alpha	**0.39**	translation	ribosome
PFL0740c	10 kd chaperonin	**0.4**	protein folding/stress response	cytosol and organelles
PF14_0541	V-type H(+)-translocating pyrophosphatase, putative	**0.4**	vacuolar-type H+ pumping	parasitophorous and digestive vacuoles
PF14_0678	**exported protein 2**	**0.41**	**host-parasite interaction**	**cell surface**
PF11_0338	Aquaglyceroporin	**0.41**	transport/host-parasite interaction	plasma membrane
MAL13P1.56	m1-family aminopeptidase	**0.42**	proteolysis	digestive vacuole
PFI0880c	glideosome-associated protein 50	**0.43**	hydrolase activity	digestive vacuole
PF08_0054	**heat shock 70 kDa protein**	**0.45**	**protein folding/stress response**	**cytosol and organelles**
PFC0725c	formate-nitrate transporter, putative	**0.46**	transport	plasma membrane
PF11_0351	**heat shock protein hsp70 homologue**	**0.46**	**protein folding/stress response**	**cytosol and organelles**
PF14_0598	**glyceraldehyde-3-phosphate dehydrogenase**	**0.46**	**glycolysis**	**cytosol**
PFE0065w	**skeleton-binding protein 1**	**0.47**	**protein transport/binding**	**vacuole membrane**
PFC0400w	60S Acidic ribosomal protein P2, putative	**0.48**	translation	ribosome
PFL1070c	**endoplasmin homolog precursor, putative**	**0.5**	**protein folding/stress response**	**endoplasmic reticulum**
PFL1385c	Merozoite Surface Protein 9, MSP-9	**0.51**	host-parasite interaction	cell surface
PF11_0061	histone H4	**0.52**	DNA binding	nucleous
PF14_0630	**protein serine/threonine phosphatase**	**1.5**	**hydrolase activity/mitosis, meiosis, cell development**	**cytosol and nucleus**
PFI1735c	**ring-exported protein 1**	**1.78**	**host-parasite interaction**	**cell surface**

In G6PDD, the diferentialy altered parasite proteins were identified as being involved in protein folding and stress response (chaperones and heat shock proteins), translation (ribosome subunits), host-parasite interaction, and in the pyrimidine biosynthetic pathway (**Table 5**). They were all over-expressed, except the MSP1.

Glucose-6-phosphate dehydrogenase catalyses the first reaction in the pentose phosphate pathway, providing reducing power to all cells in the form of NADPH. NADPH enables cells to counterbalance oxidative stress that can be triggered by several oxidant agents, and to preserve the reduced form of glutathione (GSH) that is used to mop up free radicals that cause oxidative damage. Since RBC do not contain mitochondria, the pentose phosphate pathway is their only source of NADPH; therefore, defence against oxidative damage is dependent on G6PD (Cappellini and Fiorelli, 2008). As a result, in G6PDD cells, NADPH production is severely restricted and parasites are subjected to constantly increase of endogenous oxidative stress.

Compared to parasites growing in normal RBC, parasites growing in G6PDD cells displayed an increased expression of heat shock proteins and chaperones, showing that parasite was subjected to oxidative stress and responded with increased expression of defence molecules. These highly conserved proteins protect cell structures against thermal, chemical and redox stress. Moreover, play crucial roles in folding, unfolding, assembly and transport of proteins, cell-cycle control and signalling (Li and Srivastava, 2004). This result is according to transcriptomic data that showed an enhanced correspondent mRNA expression of antioxidant enzymes and heat shock proteins in parasites growing in blunted G6PD RBC (Akide-Ndunge, et al., 2009).

Another known cellular stress response is the global down-regulation of protein translation, preventing continued protein synthesis during potentially error-prone conditions (Liu, Han and Qian, 2013; Shalgi, et al., 2013). However, it is becoming increasingly recognized that not all translation is inhibited and that translational control of specific mRNAs is required for survival during growth under stress conditions (Shenton, et al., 2006), as we have previously seen for heat shock proteins. The up representation of 60S ribosomal proteins L1 and L40/UBI in parasites under oxidative stress was not expected. Studies on yeast *Saccharomyces cerevisiae* revealed that translation response depends on stress conditions, namely on hydrogen peroxide (H_2O_2)

concentrations (Shenton, et al., 2006) and this may also happens in *Plasmodium*. A different role for ribosome proteins besides translation could also explain this result. A novel role for *P. falciparum* 60S stalk ribosomal acidic proteins P0 and P2 was indeed identified: these proteins are exported to the RBC surface and P0 seems to have endonuclease activity, participating in cell cycle regulation and RBC invasion (Singh, et al., 2002) and P2 in the formation of a tubovesicular network used for nutrient import (Das, et al., 2012). However, no additional tasks were found for L1 and L40/UBI subunits, but very little has been published on *Plasmodium* ribosomal proteins (Pubmed database retrieve no results on the query "Plasmodium 60S ribosomal protein L1", "Plasmodium 60S ribosomal protein L40/UBI" and even on "60S ribosomal protein L40/UBI"), evidencing that this is an area to be explored. Yet, the great complexity of the translation process may explain this lack of knowledge: Apicomplexans contain a mixture of translation machinery localized in three active compartments: the cytosol, mitochondrion and apicoplast (Jackson, et al., 2011) and manufacture of a 60S ribosomal subunit is extraordinarily complex, involving nearly 200 auxiliary protein and RNA molecules and many serial steps of processing the rRNA together with assembly and disassembly of ribosomal proteins and rRNAs (Zhao, Sohn and Warner, 2003).

In parasites grown in PKD RBC (**Table 6**), a total of 45 proteins displayed a differential expression, the majority being under-expressed. Concerning functional profiles, we were able to attribute to 40 proteins one of the following: protein folding and response to stress, host-parasite interactions (cell surface proteins), transport, proteolysis and hemoglobin catabolism, translation, nucleic-acid binding, cellular energy homeostasis (mithocondria and glycolysis proteins), parasitophorous vacuole proteins and others. Two of the 40 were initially classified as "unknown function" but after a search in PlasmoDB (www.plasmodb.org), the function of genes PF10_0100 and PF11_0302 were identified: succinate dehydrogenase subunit 4 and parasitophorus vacuolar protein 1, PV1, respectively.

Interestingly, despite this diversity of functions, two main cellular processes comprised most proteins: Hb digestion and protein trafficking/RBC remodeling. Moreover, almost 40% of all proteins seem to be related to Maurer's clefts. Maurer's clefts are disc-shaped flattened lamellar organelles in the RBC that occur only in RBC infected with *P. falciparum*. Their function and composition is not fully understood but

are thought to play a vital role in sorting of proteins and assembly of complexes destined for the RBC membrane playing crucial roles in the pathology of malaria infections (Lanzer, et al., 2006).

Hemoglobin hydrolysis by the parasite occurs via the coordinated action of a set of proteases resident within the digestive vacuole (Goldberg, 2005), namely plasmepsins and the m1-family aminopeptidases (Azimzadeh, et al., 2010), which were identified in our analysis (PF14_0075, PF14_0076, MAL13P1.56), to yield either free amino acids or short oligopeptides that may be exported to the cytosol for further degradation. A byproduct of Hb catabolism is the toxic heme, which is sequestered in the digestive vacuole as hemozoin. Detoxification of free heme is a critical process that is exploited by the class of 4-aminoquinoline antimalarials (including chloroquine), which accumulate in the digestive vacuole and are thought to disrupt hemozoin formation (Sanchez, et al., 2010). The importance of digestive vacuole as a site of antimalarial action is reflected in the presence on its membrane of two key drug resistance determinants, the multidrug resistance protein PfMDR (also found in this study, PFE1150w) and the chloroquine resistance transporter PfCRT (Cowman, et al., 1991; Fidock, et al., 2000).

To ingest the surrounding material (which mainly is Hb) blood stage malaria parasites perform endocytosis. They digest 70-80% of the RBC's Hb (Francis, Sullivan and Goldberg, 1997) but utilize only about 15% in *de novo* protein synthesis (Krugliak, Zhang and Ginsburg, 2002.). The excess amino acids are exported from the infected RBC by transport pathways created by the parasite (Ginsburg, et al., 1983). Hemoglobin digestion is then dependent on the secretory pathway, the other major biological process that seems to be down-expressed in parasites growing in PKD conditions.

Human RBC lack a secretory system and are rapidly cleared from circulation by the spleen when damaged or infected. To develop within human RBC and to avoid passage through the spleen, *P. falciparum* extensively modifies its host cell (Maier, et al., 2009). So, we predicted that a reduction in protein exporting and RBC remodeling would a) difficult the settlement of young parasites inside the RBC since the exchanges with the extracellular medium will be affected and b) influence the immunological

response by the host since the RBC surface will be differently composed (e.g. cytoadherence).

The *P. falciparum* exportome is 5–10 times larger than that of other malaria parasites, which may reflect the unique pathogenicity of *P. falciparum*, namely its ability to become sequestered in host capillaries (Bonnefoy and Ménard, 2008). In *P. falciparum*, up to 8% of all proteins encoded in its genome is predicted to be exported into the host cell (Marti, et al., 2004; Hiller, et al., 2004). Asexual blood stage parasites are characterized by extensive remodeling of the RBC but it occurs also in gametocytes (Silvestrini, et al., 2010) and was also reported for liver stages (Singh, et al., 2007).

During invasion of RBC (as well as hepatocytes), the parasites become enclosed within an additional membrane layer, the parasitophorous vacuole membrane (PVM), which acts as a semipermeable barrier between parasite and host, allowing for nutrient acquisition and secretion of parasite-derived factors. In early intraerythocytic stages, the parasite initiates the development of membrane structures in the RBC which participate in exported protein trafficking. These include the Maurer's clefts, the tubulovesicular network (TVN), and vesicle-like structures (Wickert, et al., 2003). Several proteins have been established as associated to Maurer's clefts, namely, the ring-exported protein 1, REX1 (PFI1735c) and the skeleton-binding protein 1, SBP1 (PFE0065w) (Lanzer, et al., 2006), which both showed expression alteration in parasites in PKD environment. Others, still under study, have been described as putative Maurer's cleft proteins (Lazer, et al., 2006): early transcribed membrane proteins (PF10_0019; PFE1590w), 14-3-3 protein (MAL8P1.69), adenylate kinase (PF10_0086), disulfide isomerase (MAL8P1.17), exported protein 2 (PF14_0678), heat shock 70 kDa proteins (MAL13P1.540; PFI0875w; PF08_0054; PF11_0351) and glyceraldehyde-3-phosphate dehydrogenase (PF14_0598) are some of these proteins, and were also low-expressed in parasites growing in PKD RBC.

The Maurer's cleft proteins SBP1 and REX1 play a pivotal role in the pathogenesis of *P. falciparum* malaria: SBP1 gene disruption prevented RBC adhesion because of the loss of PfEMP1 (*P. falciparum* erythrocyte membrane protein 1) expression on the surface. In normal conditions, the parasite ligand PfEMP1 is expressed on the surface of infected RBC and adheres to the vascular endothelium

causing the sequestration of the RBC in the microvasculature, being responsible for the high mortality of *P. falciparum* malaria (Cooke, et al., 2006). Similarly, REX1 is also associated to PfEMP1 expression on the RBC surface: removal of the C-terminal domain of REX1 compromises Maurer's cleft architecture and PfEMP1-mediated cytoadherance but permits some trafficking of PfEMP1 to the RBC surface. Deletion of the coiled-coil region of REX1 ablates PfEMP1 surface display, trapping PfEMP1 at the Maurer's clefts (Dixon, et al., 2011). In a previous study (Hanssen, et al., 2008), deletion or truncation of REX1 caused stacking of the Maurer's cleft lamellae which leads to an apparent decrease in Maurer's cleft numbers when examined by immunofluorescence microscopy. So, the loss of functional SBP1 or REX1 directly or indirectly ablates the assembly of the *P. falciparum* virulence complex at the surface of host RBC. However, in our study there was a contrary effect in the expression profiles of both proteins: SPB1 was down-expressed in deficient conditions, whereas REX1 was over-expressed. Some regulatory mechanism, operating in the expression of both proteins, may be counterbalancing the expression of these proteins.

Several other proteins displaying differential expression also seem to be related to RBC remodeling processes, as is the case of the parasite-encoded heat shock proteins (Hiller, et al., 2004) because of their function in folding and unfolding of other proteins. Moreover, they can significantly affect the efficiency of antigen expression by acting at the site of host-targeting exit or the Maurer's clefts (Haldar and Mohandas, 2007). The *P. falciparum* 60S ribosomal acidic protein P2 (PfP2) (PFC0400w) is exported to the infected RBC surface during early schizogony and treatment with anti-P2-antibodies causes disintegration of the TVN, resulting in impaired lipid import, which may be the eventual cause of cell-cycle arrest. The biology of the P0 protein (PF11_0313) is also complex and intriguing, being also transported to the cell surface. It has endonuclease activity, participates in cell cycle regulation and invasion (Singh, et al., 2002).

The parasitic plasma membrane transporters hexose transporter PfHT (PFB0210c) and aquaglyceroporin (PF11_0338) were also down-expressed, meaning that there is a reduced input of glucose, and water and solutes, respectively, to glycolysis. So, it is expected that glycolysis itself will be repressed. Glyceraldehyde is permeant of aquaglyceroporins and is metabolized via glycolysis after phosphorylation to glyceraldehyde 3-phosphate (Pavlovic-Djuranovic, et al., 2006). Malaria parasites

lack energy stores such as glycogen and are therefore extremely sensitive to decreased delivery of glucose. Inhibiting glucose transport in infected RBC or removal of glucose from the medium produces an immediate fall in intraparasitic ATP concentrations (Fry, et al., 1990; Kirk, Horner and Kirk, 1996). Some glycolytic enzymes were indeed under-represented [glyceraldehyde-3-phosphate dehydrogenase (PF14_0598) and L-lactate dehydrogenase (PF13_0141)]. Since ATP is absolutely necessary for parasite survival, the parasite must produce ATP someway and, although biochemical data indicate that the *Plasmodium* mitochondrion does not seem to be a source of ATP (Fry, et al., 1990), the higher peptide number of ATPase subunit beta in parasites from PKD RBC (qualitative data) suggests that ATP synthesis may occur in mitochondria to combat the shortage of energy.

Only two parasitic proteins were regulated in the same direction in PK- and G6PD-deficient conditions: MSP1 (down-expressed) and REX1 (over-expressed). The understanding of their function could provide a clue about a common feature in parasites growing in both enzyme deficiencies. The MSP1 is expressed on the surface of the parasite and mediates the first interaction between the malaria merozoite and the RBC that it will invade. It is essential for RBC invasion and is also targeted by the human immune response (Kadekoppala and Holder, 2010). As above mentioned, REX1 is an important component of the Maurer's clefts associated to PfEMP1 expression on the RBC surface, which mediates citoadherence (Dixon, et al., 2011). An alteration in abundance of two proteins involved in invasion, host-parasite interaction and human immune response may be relevant for the reduced invasion rate observed in parasites growing in both deficient conditions (*in vitro* results - see section 1.1).

Interestingly, when we looked for the functional mechanisms underlying other malaria protective polymorphisms, such as hemoglobinopathies, two studies were found associating Maurer's cleft improper formation with malaria resistance (Cryklaff, et al., 2011; Wellems and Fairhurst, 2012). A significantly reduced actin remodeling and aberrant Maurer's clefts seem to occur in HbCC and HbSC RBC, suggesting that these mutant Hb states may interfere with the installation of actin scaffolds that help to tether Maurer's clefts and support vesicle and protein trafficking to the RBC membrane. A similar protecting mechanism involving Maurer's cleft and protein secretion may also be present in PK deficiency. The determination of the RBC membrane proteomic

profile, with erythrocytic proteins and exported parasitic molecules at the RBC surface, will shed new light on this hypothesis.

CONCLUSION

In this study, invasion and maturation differences in *P. falciparum* 3D7 growing in normal and PKD and G6PDD RBC were analyzed by *in vitro* experiments, and the expression profile of young trophozoites parasites developing in these RBC were determined using a label-free quantitative proteomics approach. The parasite morphology was similar in both normal and deficient conditions but invasion ratios were lower in parasites from deficient RBC. Contrarily, maturation was higher in three growth cycles in deficient conditions. However, none of these differences were statistical significant. These results suggested on one hand, that the parasites have difficulty in invade the deficient RBC, and only a small number of parasites can actually do that, and on the other hand, an adaptation process by parasites living in deficient RBC (more parasites died in control RBC than in deficient during the second half of their cycle). So, we looked to the proteomes of these parasites in order to get some answers about the previous results and some interesting data was obtained: the response from parasites growing in PKD and G6PDD RBC is distinct and proportional to phenotype severity, i.e. a more severe phenotype triggered a more aggressive and wide parasite response. In G6PDD (from an asymptomatic individual), the mainly alteration in proteins abundance was the increase of heat shock proteins and chaperones, showing that parasite was subjected to oxidative stress and responded with increased expression of protective molecules. In PKD (transfusion-dependent individual with regular hemolytic crisis), a more wide and acute response was triggered by the parasite with a high number of proteins involved in diverse pathways displaying significant alterations in their abundance, the majority being down-expressed. The most represented biological processes in this response were Hb digestion and protein trafficking/RBC remodeling, being both connected, since Hb enters in the cell by endocytosis and the excess amino acids are exported from the infected RBC. Moreover, almost 40% of all proteins with abundance alterations were related to Maurer's clefts, that play a vital role in sorting of proteins and assembly of complexes destined for the RBC membrane, playing crucial

roles in the pathology of malaria infections. The loss of functional Maurer's cleft proteins dramatically changes the PfEMP1 RBC surface disposal preventing the assembly of the *P. falciparum* virulence complex at the surface of host RBC (the parasite ligand PfEMP1 is expressed on the surface of infected RBC and adheres to the vascular endothelium causing the sequestration of the RBC in the microvasculature, being responsible for the high mortality of *P. falciparum* malaria). So, from these results we hypothesized that the protection against malaria that seems to be conferred by PK deficiency is associated with the RBC remodeling process by the parasite that reduces invasion and malaria virulence itself.

The fact that almost all proteins were over-expressed in one condition (G6PDD) and down-expressed in the other (PKD) is peculiar and we naturally questioned about the reliability of these results. Each analyzed parasite fraction is, realistically, a mixture of human and parasitic proteins and, as a consequence, different samples may have different proportions of proteins of parasite and human origin. So, we hypothesized that, for instance, more parasite proteins in deficient conditions may reflect a superior percentage of parasitic proteins in the mixture relatively to the percentage of parasitic proteins in normal conditions, instead of a real up-expression. However, several facts play against this hypothesis: 1) technical procedures were the same for control and deficient cultures and performed simultaneously (with the same instruments and equipments) meaning that, to the extent that we can control, errors were performed equitatively; 2) results are based on two biological replicates and three technical replicates for each experiment; 3) there are some exceptional proteins whose expression is in opposite direction of the majority; 4) some of the obtained results are in accordance with previous knowledge: the up-expression of chaperones in G6PDD conditions were totally expected considering previous reports on oxidative stress response; if this was not observed, the results could be jeopardized; 5) a more severe host phenotype (PKD, transfusion-dependent individual) corresponded to a more aggressive and wide parasite response; 6) the protection mechanism suggested for PK deficiency has been reported for other malaria protective polymorphisms.

Nevertheless, our analysis would obviously be more robust and consistent if a new independent assay was carried out. Even because, with more data, significant statistical differences would probably be reached in invasion *in vitro* experiments. It

would also be interesting to test the parasite growth in RBC with other G6PD and PK phenotypes (and genotypes) because the clinical phenotype of both G6PD and PK deficiencies is heterogeneous, ranging from a mild chronic hemolytic anemia to a severe anemia, and the parasite will surely respond differently.

It would also be relevant to explore the ATP-synthase up expression result (this enzyme seems to be more abundant in PKDD, counteracting the trend of down expression in almost all proteins) and, obviously, analyze the proteomic results of the remaining extracts (RBC membranes and cytoplasm), that are still being under MS analysis. In this respect, the membranes profile results will be especially useful to confirm the RBC remodeling as the key process of PK deficiency protection against malaria.

References

Acharya, P., Pallavi, R., Chandran, S., Chakravarti, H., Middha, S., Acharya, J., Kochar, S., Kochar, D., Subudhi, A., Boopathi, A.P., Garg, S., Das, A. and Tatu, U., 2009. A glimpse into the clinical proteome of human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. *Proteomics. Clinical applications*, 3(11), pp.1314-25.

Akide-Ndunge, O.B., Tambini, E., Giribaldi, G., McMillan, PJ., Müller, S., Arese, P. and Turrini, F., 2009. Co-ordinated stage-dependent enhancement of *Plasmodium falciparum* antioxidant enzymes and heat shock protein expression in parasites growing in oxidatively stressed or G6PD-deficient red blood cells, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/8/1/113> [Accessed 18 October].

Alvarez-Llamas, G., de La Cuesta, F., Barderas, M.G., Darde, V.M., Zubiri, I., Caramelo, C. and Vivanco, F., 2009. A novel methodology for the analysis of membrane and cytosolic sub-proteomes of RBC by 2-DE. *Electrophoresis*, 30(23), pp.4095-108.

Aubouy, A., Migot-Nabias, F. and Deloron, P., 2003. Polymorphism in two merozoite surface proteins of *Plasmodium falciparum* isolates from Gabon, *Malaria Journal*, [online] Available at: < http://www.malariajournal.com/content/2/1/12> [Accessed 4 January 2013].

Ayi, K., Liles, W.C., Gros, P. and Kain, K.C., 2009. Adenosine triphosphate depletion of RBC simulates the phenotype associated with pyruvate kinase deficiency and confers protection against *Plasmodium falciparum in vitro*. *The Journal of infectious diseases*, 200(8), pp.1289-99.

Ayi, K., Min-Oo, G., Serghides, L., Crockett, M., Kirby-Allen, M., Quirt, I., Gros, P. and Kain, K.C., 2008. Pyruvate kinase deficiency and malaria. *The New England journal of medicine*, 358(17), pp.1805-10.

Ayi, K., Turrini, F., Piga, A. and Arese, P., 2004. Enhanced phagocytosis of ring-parasitized mutant RBC: a common mechanism that may explain protection against *falciparum* malaria in sickle trait and beta-thalassemia trait. *Blood*, 104(10), pp.3364-71.

Azimzadeh, O., Sow, C., Gèze, M., Nyalwidhe, J., Florent, I., 2010. *Plasmodium falciparum* PfA-M1 aminopeptidase is trafficked via the parasitophorous vacuole and marginally delivered to the food vacuole, *Malaria Journal*, [online] Available at: < http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2914058/> [Accessed 4 January 2013].

Balabaskaran Nina, P., Morrisey, J.M., Ganesan, S.M., Ke, H., Pershing, A.M., Mather, M.W. and Vaidya, A.B., 2011. ATP synthase complex of *Plasmodium falciparum*: dimeric assembly in mitochondrial membranes and resistance to genetic disruption. *The Journal of biological chemistry*, 286(48), pp. 41312-22.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B., 2007. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry*, 389(4), pp. 1017-31.

Bantscheff, M., Lemeer, S., Savitski, M.M. and Kuster, B., 2012. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Analytical and bioanalytical chemistry*, 404(4), pp.939-65.

Beutler, E., 1984. *Red Cell Metabolism: A Manual of Biochemical Methods*. 3rd sub ed. Grune & Stratton, Philadelphia, PA.

Bonnefoy, S. and Ménard, R., 2008. Deconstructing export of malaria proteins. *Cell*, 134(1), pp.20-22.

Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, ED., Zhu. J. and DeRisi, J.L., 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS biology*, 1(1), pp.85-100.

Briolant, S., Almeras, L., Belghazi, M., Boucomont-Chapeaublanc, E., Wurtz, N., Fontaine, A., Granjeaud, S., Fusaï, T., Rogier, C. and Pradines, B., 2010. *Plasmodium falciparum* proteome changes in response to doxycycline treatment, *Malaria Journal*, [online] Available at: <http://www.malariajournal.com/content/9/1/141> [Accessed 4 January 2013].

Cappadoro, M., Giribaldi, G., O'Brien, E., Turrini, F., Mannu, F., Ulliers, D., Simula, G., Luzzatto, L. and Arese, P., 1988. Early phagocytosis of glucose-6-phosphate dehydrogenase (G6PD)-deficient RBC parasitized by *Plasmodium falciparum* may explain malaria protection in G6PD deficiency. *Blood*, 92(7), pp.2527-34.

Cappellini, M.D. and Fiorelli, G., 2008. Glucose-6-phosphate dehydrogenase deficiency. *Lancet*, 371(9606), pp.64-74.

Cytoscape v2.8.3, 2012: http://www.cytoscape.org/ [Accessed September 2012]

Cooke, B.M., Buckingham, D.W., Glenister, F.K., Fernandez, K.M., Bannister, L.H., Marti, M., Mohandas, N. and Coppel, R.L., 2006. A Maurer's cleft-associated protein is essential for expression of the major malaria virulence antigen on the surface of infected red blood cells. *The Journal of cell biology*, 172(6), pp.899-908.

Cowman, A.F., Karcz, S., Galatis, D. and Culvenor, J.G., 1991. A P-glycoprotein homologue of *Plasmodium falciparum* is localized on the digestive vacuole. *The Journal of cell biology*, 113(5), pp.1033-42.

Cyrklaff, M., Sanchez, C.P., Kilian, N., Bisseye, C., Simpore, J., Frischknecht, F. and Lanzer, M., 2011. Hemoglobins S and C interfere with actin remodeling in *Plasmodium falciparum*-infected RBC. *Science*, 334(6060), pp.1283-86.

D'Alessandro, A., Righetti, P.G. and Zolla, L., 2010. The red blood cell proteome and interactome: an update. *Journal of proteome research*, 9(1), pp.144-63.

Das, S., Basu, H., Korde, R., Tewari, R. and Sharma, S., 2012. Arrest of nuclear division in Plasmodium through blockage of RBC surface exposed ribosomal protein P2, *PLoS pathogens*, [online] Available at: <http://www.plospathogens.org/article/info%3Adoi%2F10.1371%2Fjournal.ppat.1002858> [Accessed 16 January 2013].

Delobel, J., Prudent, M., Rubin, O., Crettaz, D., Tissot, J.D. and Lion, N., 2012. Subcellular fractionation of stored red blood cells reveals a compartment-based protein carbonylation evolution. *Journal of Proteomics, 76* Spec No., pp.181-93.

Dixon, M.W., Kenny, S., McMillan, P.J., Hanssen, E., Trenholme, K.R., Gardiner, DL. and Tilley, L., 2011. Genetic ablation of a Maurer's cleft protein prevents assembly of the *Plasmodium falciparum* virulence complex. *Molecular microbiology*, 81(4), pp.982-93.

Durand P.M. and Coetzer, T.L., 2008. Pyruvate kinase deficiency protects against malaria in humans. *Haematologica*, 93(6), pp.939-40.

Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M., Sidhu, A.B., Naudé, B., Deitsch, K.W., Su, X.Z., Wootton, J.C., Roepe, P.D. and Wellems, T.E., 2000. Mutations in the *P. falciparum* digestive vacuole transmembrane protein *PfCRT* and evidence for their role in chloroquine resistance. *Molecular cell*, 6(4), pp.861-71.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., Moch, J.K., Muster, N., Sacci, J.B., Tabb, D.L., Witney, A.A., Wolters, D., Wu, Y., Gardner, M.J., Holder, A.A., Sinden, R.E., Yates, J.R.,´and Carucci, D.J., 2002. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*, 419(6906), pp.520-26.

Fontaine, A., Bourdon, S., Belghaz,i M., Pophillat, M., Fourquet, P., Granjeaud, S., Torrentino-Madamet, M., Rogier, C., Fusai, T. and Almeras, L., 2012. *Plasmodium falciparum* infection-induced changes in RBC membrane proteins. *Parasitology research*, 110(2), pp.545-56.

Francis, S.E., Sullivan, D.J. Jr. and Goldberg, D.E., 1997. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual review of microbiology*, 51(1), pp. 97–123.

Fry, M., Webb, E. and Pudney, M., 1990. Effect of mitochondrial inhibitors on adenosinetriphosphate levels in *Plasmodium falciparum*. *Comparative biochemistry and physiology. B, Comparative biochemistry*, 96(4), pp.775-82.

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. and Barrell, B., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, 419(6906), pp.498-511.

Gelhaus, C., Fritsch, J., Krause, E. and Leippe, M., 2005. Fractionation and identification of proteins by 2-DE and MS: towards a proteomic analysis of *Plasmodium falciparum*. *Proteomics*, 5(16), pp.4213-22.

Ginsburg, H., Krugliak, M., Eidelman, O. and Cabantchik, Z.I., 1983. "New permeability pathways induced in membranes of *Plasmodium falciparum*. *Molecular and biochemical parasitology*, 8(2), pp.177-90.

Goldberg D.E. and Cowman, A.F., 2010. Moving in and renovating: exporting proteins from *Plasmodium* into host RBC. *Nature reviews. Microbiology*, 8(9), pp.617-21.

Goldberg, D.E., 2005. Hemoglobin degradation. In: D.J. Sullivan and S. Krishna, eds. 2005. *Malaria: Drugs, Disease and Post-genomic Biology*. Berlin: Springer Berlin Heidelberg, pp. 275-91.

Graphpad, 2012: http://www.graphpad.com/ [Accessed September 2012)

Haldar, K. and Mohandas, N., 2007. RBC remodeling by malaria parasites. *Current opinion in hematology*, 14(3), pp.203-09.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., Florens, L., Janssen, C.S., Pain, A., Christophides, G.K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M.A., Ormond, D., Doggett, J., Trueman, H.E., Mendoza, J., Bidwell, S.L., Rajandream, M.A., Carucci, D.J., Yates, J.R. 3rd, Kafatos, F.C., Janse, C.J., Barrell, B., Turner, C.M., Waters, A.P. and Sinden, R.E., 2005. A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science*, 307(5706), pp.82-86.

Hanssen, E., Hawthorne, P., Dixon, M.W.A., Trenholme, K.R., McMillan, P.J., Spielmann, T., Gardiner, D.L. and Tilley, L., 2008. Targeted mutagenesis of the ring-exported protein-1 of *Plasmodium falciparum* disrupts the architecture of Maurer's cleft organelles. *Molecular Microbiology*, 69(4), pp.938–53.

Hill, R.J., Konigsbeg, W., Guidotti, G. and Craig, L.C., 1962. The structure of human hemoglobin. I. The separation of the alfa and beta chains and their amino acid composition. *The journal of biological chemistry*, 237(5), pp. 1549-54.

Hiller, N.L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estraño, C. and Haldar, K., 2004. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science*, 306(5703), pp.1934–37.

Intact, 2011: http://www.ebi.ac.uk/intact/ [Accessed March 2012]

Jackson, K.E., Habib, S., Frugier, M., Hoen, R., Khan, S., Pham, J.S., Ribas de Pouplana, L., Royo, M., Santos, M.A., Sharma, A. and Ralph, S.A., 2011. Protein translation in *Plasmodium* parasites. *Trends Parasitology*, 27(10), pp.467-76.

Jacobasch, G. and Rapoport, S.M., 1996. Hemolytic anemias due to RBC enzyme deficiencies. *Molecular aspects of medicine*, 17(2), pp.143-70.

Kirk, K., Horner, H.A. and Kirk, J., 1996. Glucose uptake in *Plasmodium falciparum*-infected RBC is an equilibrative not an active process. *Molecular and biochemical parasitology*, 82(2), pp.195-205.

Kadekoppala, M. and Holder, A.A., 2010. Merozoite surface proteins of the malaria parasite: the MSP1 complex and the MSP7 family. *International journal for parasitology*, 40(10), pp.1155-61.

Kirkman, H.N. and Gaetani, G.F., 1984. Catalase: a tetrameric enzyme with four tightly bound molecules of NADPH. *Proceedings of the National Academy of Sciences of the United States of America*, 81(14), pp.4343-47.

Klemba, M. and Goldberg, D.E., 2002. Biological roles of proteases in parasitic protozoa. *Annual review of biochemistry*, 71, pp.275-305.

Krugliak, M., Zhang, J. and Ginsburg, H., 2002. Intraerythrocytic *Plasmodium falciparum* utilizes only a fraction of the amino acids derived from the digestion of host cell cytosol for the biosynthesis of its proteins. *Molecular and Biochemical Parasitology*, 119(2), pp.249–56.

Kuss, C., Gan, C.S., Gunalan, K., Bozdech, Z., Sze, S.K. and Preiser, P.R., 2012. Quantitative proteomics reveals new insights into erythrocyte invasion by *Plasmodium falciparum*, *Molecular & cellular proteomics : MCP*, [online] Available at <http://www.mcponline.org/content/11/2/M111.010645.long> [Accessed 7 January 2013]

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259), pp.680-85.

Lambros, C. and Vanderberg, J.P., 1979. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *The Journal of parasitology*, 65(3), pp.418-20.

Lang-Unnasch, N. and Murphy, A.D., 1998. Metabolic changes of the malaria parasite during the transition from the human to the mosquito host. *Annual review of microbiology*, 52, pp.561-90.

Lanzer, M., Wickert, H., Krohne, G., Vincensini, L. and Braun Breton, C., 2006. Maurer's clefts: a novel multi-functional organelle in the cytoplasm of *Plasmodium falciparum*-infected RBC. *International journal for parasitology*, 36(1), pp.23-36.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G. and Mann, M., 2002. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature*, 419(6906), pp.537-42.

Li, Z. and Srivastava, P., 2004. Heat-Shock Proteins. *Current Protocols in Immunology*, 58, A.1T.1- A.1T.6.

Liu, B., Han, Y, and Qian, S.B., 2013. Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. *Molecular cell*, 49(3), pp.453-63.

Luzzatto, L., Usanga, F.A. and Reddy, S., 1969. Glucose-6-phosphate dehydrogenase deficient red cells: resistance to infection by malarial parasites. *Science*, 164(3881), pp.839-42.

Maier, A.G., Cooke, B.M., Cowman, A.F. and Tilley L., 2009. Malaria parasite proteins that remodel the host RBC. *Nature reviews. Microbiology*, 7(5), pp.341-54.

Malaria Parasite metabolic Pathways, 2011: http://priweb.cc.huji.ac.il/malaria/ [Accessed April 2012]

Manco, L., Bento, C., Ribeiro, M.L. and Tamagnini, G., 2002. Consequences at mRNA level of the PKLR gene splicing mutations IVS10(+1)G-->C and IVS8(+2)T-->G causing pyruvate kinase deficiency. *British journal of haematology*, 2002, 118(3), pp.927-28.

Manco, L., Ribeiro, M.L., Almeida, H., Freitas, O., Abade, A. and Tamagnini, G., 1999. PK-LR gene mutations in pyruvate kinase deficient Portuguese patients. *British journal of haematology*, 105(3), pp.591-95.

Marti, M., Good, R.T., Rug, M., Knuepfer, E. and Cowman, A.F., 2004. Targeting malaria virulence and remodeling proteins to the host RBC. *Science*, 306(5703), pp.1930-33.

Mather, M.W., Henry, K.W. and Vaidya, A.B., 2007. Mitochondrial drug targets in apicomplexan parasites. *Current drug targets*, 8(1), pp.49-60.

Matzke, M.M., Brown, J.N., Gritsenko, M.A., Metz, T.O., Pounds, J.G., Rodland, K.D., Shukla, A.K., Smith, R.D., Waters, K.M., McDermott, J.E. and Webb-Robertson, B.J., 2012. A comparative analysis of computational approaches to relative protein quantification using peptide peak intensities in label-free LC-MS proteomics experiments. *Proteomics*, 13(3-4), pp.493-503.

Méndez, D., Linares, M., Diez, A., Puyet, A. and Bautista, J.M., 2011. Stress response and cytoskeletal proteins involved in RBC membrane remodeling upon *Plasmodium falciparum* invasion are differentially carbonylated in G6PD A- deficiency. *Free radical biology & medicine*, 50(10), pp.1305-13.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., Kitano, H. and Thomas, P.D., 2005. The PANTHER database of protein families, subfamilies, functions and pathways, *Nucleic acids research*, [online] Available at <http://nar.oxfordjournals.org/content/33/suppl_1/D284.long> [Accessed 6 December 2012].

Nirmalan, N., Sims, P.F. and Hyde, J.E., 2004. Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. *Molecular microbiology*, 52(4), pp.1187-99.

Oehring, S.C., Woodcroft, B.J., Moes, S., Wetzel, J., Dietz, O., Pulfer, A., Dekiwadia, C., Maeser, P., Flueck, C., Witmer, K., Brancucci, N.M., Niederwieser, I., Jenoe, P., Ralph, S.A. and Voss, T.S., 2012. Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum* , *Genome biology*, [online] Available at <http://genomebiology.com/content/13/11/R108> [Accessed 6 December 2012].

Oelshlegel, F.J.Jr., Sander, B.J. and Brewer, G.J., 1975. Pyruvate kinase in malaria host-parasite interaction. *Nature*, 255(5506), pp.345-47.

Painter, H.J., Morrise,y J.M., Mather, M.W. and Vaidya, A.B., 2007. Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. *Nature*, 446(7131), pp.88-91.

Panther, 2012: http://www.pantherdb.org/ [Accessed October 2012].

Pasini, E.M., Kirkegaard, M., Mortensen, P., Lutz, H.U., Thomas, A.W. and Mann, M., 2006. In-depth analysis of the membrane and cytosolic proteome of red blood cells. *Blood*, 108(3), pp.791-801.

Pasini, E.M., Lutz, H.U., Mann, M. and Thomas, A.W., 2010. Red blood cell (RBC) membrane proteomics - Part I: Proteomics and RBC physiology. *Journal of proteomics*, 73(3), pp.403-20.

Pavlovic-Djuranovic, S., Kun, J.F., Schultz, J.E. and Beitz, E., 2006. Dihydroxyacetone and methylglyoxal as permeants of the *Plasmodium* aquaglyceroporin inhibit parasite proliferation. *Biochimica et biophysica acta*, 1758(8), pp.1012-17.

PlasmoDB, 2011: http://plasmodb.org/plasmo/ [Accessed 2012]

Prabakaran, S., Wengenroth, M., Lockstone, H.E., Lilley, K., Leweke, F.M. and Bahn S., 2007. 2-D DIGE analysis of liver and red blood cells provides further evidence for oxidative stress in schizophrenia. *Journal of proteome research*, 6(1), pp.141-49.

Prieto, J.H., Koncarevic, S., Park, S.K., Yates, J3rd. and Becker, K. 2008. Large-scale differential proteome analysis in *Plasmodium falciparum* under drug treatment, *PLoS One*, [online] Available at: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0004098> [Accessed 4 January 2013].

Ray, S., Renu, D., Srivastava, R., Gollapalli, K., Taur, S., Jhaveri, T., Dhali, S., Chennareddy, S., Potla, A., Dikshit, J.B., Srikanth, R., Gogtay, N., Thatte, U., Patankar, S. and Srivastava, S., 2012. Proteomic investigation of *falciparum* and *vivax* malaria for identification of surrogate protein markers, *PLoS One* [online] Available at: <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0041751> [Accessed 4 January 2013].

Rickli, E.E., Ghazanfar, S.A., Gibbons B.H. and Edsall, J.T., 1964., Carbonic Anhydrases from Human RBC Preparation and properties of two enzymes. *The Journal of biological chemistry*, 239(4), pp.1065-78.

Ringrose, J.H., van Solinge, W.W., Mohammed, S., O'Flaherty, M.C., van Wijk, R., Heck, A.J. and Slijper, M., 2008. Highly efficient depletion strategy for the two most abundant RBC soluble proteins improves proteome coverage dramatically. *Journal of proteome research*, 7(7), pp.3060-63.

Roth, E.Jr. and Schulman, S., 1988. The adaptation of *Plasmodium falciparum* to oxidative stress in G6PD deficient human RBC. *British journal of haematology*, 70(3), pp.363-67.

Roth, E.F.Jr., Raventos-Suarez, C., Rinaldi, A. and Nagel, R.L., 1983. Glucose-6-phosphate dehydrogenase deficiency inhibits in vitro growth of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of the United States of America*, 80(1), pp.298-99.

Roux-Dalvai, F., Gonzalez de Peredo, A., Simó, C., Guerrier, L., Bouyssié, D., Zanella, A., Citterio, A., Burlet-Schiltz, O., Boschetti, E., Righetti, P.G. and Monsarrat, B., 2008. Extensive analysis of the cytoplasmic proteome of human RBC using the peptide ligand library technology and advanced mass spectrometry. *Molecular & cellular proteomics : MCP*, 7(11), pp.2254-69.

Sanchez, C.P., Dave, A., Stein, W.D. and Lanzer, M., 2010. Transporters as mediators of drug resistance in *Plasmodium falciparum*. *International journal for parasitology*, 40(10), pp.1109-18.

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S. and Burge, C.B., 2013. Widespread regulation of translation by elongation pausing in heat shock. *Molecular cell*, 49(3), pp.439-52.

Shenton, D., Smirnova, J.B., Selley, J.N., Carroll, K., Hubbard, S.J., Pavitt, G.D., Ashe, M.P. and Grant, C.M., 2006. Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *The Journal of biological chemistry*, 281(39), pp.29011-21.

Sicard, A., Semblat, J.P., Doerig, C., Hamelin, R., Moniatte, M., Dorin-Semblat, D., Spicer, J.A., Srivastava, A., Retzlaff, S., Heussler, V., Waters, A.P. and Doerig C., 2011. Activation of a PAK-MEK signalling pathway in malaria parasite-infected RBC. *Cellular microbiology*, 13(6), pp.836-45.

Silvestrini, F., Lasonder, E., Olivieri, A., Camarda, G., van Schaijk, B., Sanchez, M., Younis Younis, S., Sauerwein, R. and Alano, P., 2010. Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Molecular & cellular proteomics : MCP*, 9(7), pp.1437-48.

Singh, A.P., Buscaglia, C.A., Wang, Q., Levay, A., Nussenzweig, D.R., Walker, J.R., Winzeler, E.A., Fujii, H., Fontoura, B.M. and Nussenzweig, V., 2007. *Plasmodium* circumsporozoite protein promotes the development of the liver stages of the parasite. *Cell*, 131(3), pp.492-504.

Singh, S., Sehgal, A., Waghmare, S., Chakraborty, T., Goswami, A. and Sharma, S., 2002. Surface expression of the conserved ribosomal protein P0 on parasite and other cells. *Molecular and biochemical parasitology*, 119(1), pp.121-24.

Smit, S., Stoychev, S., Louw, A.I. and Birkholtz, L.M., 2010. Proteomic profiling of *Plasmodium falciparum* through improved, semiquantitative two-dimensional gel electrophoresis. *Journal of proteome research*, 9(5), pp.2170-81.

Southworth, P.M., Hyde, J.E., Sims, P.F., 2011. A mass spectrometric strategy for absolute quantification of *Plasmodium falciparum* proteins of low abundance, *Malaria Journal*, [online] Available at: < http://www.malariajournal.com/content/9/1/189> [Accessed 4 January 2013].

Steen, H. and Mann, M., 2004. The ABC's (and XYZ's) of peptide sequencing. *Nature reviews. Molecular cell biology*, 5(9), pp.699-711.

STRING 9.0, 2011: http://string-db.org [Accessed July 2012]

SwissProt, 2012: http://www.uniprot.org/ [Accessed September 2012]

The Gene Ontology (GO) project, 2012: http://www.geneontology.org [Accessed March 2012).

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. and Narechania, A., 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome research*, 13(9), pp.2129-41.

Trager, W. and Jensen, J.B., 1976. Human malaria parasites in continuous culture. *Science*, 193(4254), pp.673-675.

UniProt, 2011: www.uniprot.org [Accessed November 2011].

Usanga, E.A. and Luzzatto, L., 1985. Adaptation of *Plasmodium falciparum* to glucose 6-phosphate dehydrogenase-deficient host red cells by production of parasite-encoded enzyme. *Nature*, 313(6005), pp.793-5.

Vogel, C. and Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics*, 13(4), pp.227-32.

Wellems, T.E. and Fairhurst, R.M., 2012. An evolving picture ofthe interactions between malaria parasites and their host RBC. *Cell research*, 22(3), pp.453-6.

Wickert, H., Wissing, F., Andrews, K.T., Stich, A., Krohne, G. and Lanzer, M., 2003. Evidence for trafficking of PfEMP1 to the surface of *P. falciparum*-infected RBC via a complex membrane network. *European journal of cell biology*, 82(6), pp.271-84.

Zhao, Y., Sohn, J.H. and Warner JR., 2003. Autoregulation in the Biosynthesis of Ribosomes *Molecular and cellular biology,* 23(2), pp.699–707.

Chapter 6 –

General Discussion

GENERAL DISCUSSION

6.1. Results overview and discussion

The major objective of this thesis was the investigation of the association of PK deficiency and malaria in humans. Considering previous results in murine models and *in vitro Plasmodium* cultures growing in PK-deficient RBC, supporting the hypothesis of a protective effect of this enzyme deficiency against malaria severity, data from human source (epidemiological and population genetics data) was clearly missing to complete the body of evidence. So, a focused and tight strategy was defined in order to clarify this main question and, also, contribute to the general understanding of the human genetic factors associated to malaria susceptibility, as well as to the knowledge of the RBC enzyme disorders in the basis of human hemolytic anemias.

Therefore, in a first instance, a study was performed in Cape Verde archipelago (where malaria has an epidemic character), to check if the malaria low morbidity in Santiago island could be a consequence of particular characteristics of the host population genetics. The genes *PKLR* (encoding pyruvate kinase), *HBB* (encoding β-globin) and *G6PD* (encoding glucose-6-phosphate dehydrogenase) were analyzed. The alleles HbS, G6PDB, G6PDA, G6PDA- and G6PDMed, described as protective against malaria for a long time, were searched. In the case of *PKLR*, since no specific allele has been pointed as protective so far, the samples were genotyped for two mutations and two polymorphic loci previously described in the gene. Additionally, new polymorphic loci were investigated, identified and analyzed. The searched mutations were: the substitution 269A>T, previously associated to malaria protection in mice and identified in a human case of pyruvate kinase deficiency; and 1456C>T, the most common mutation associated to PK deficiency in Portuguese (the islands were colonized by European settlers, namely Portuguese) and already identified in Afro-American. The polymorphisms were the binary 1705 A/C (exon 12) and T10/19 (intron 10), highly polymorphic in Sao Tome and Principe. Four new polymorphic loci (STRs) were identified inside (referred as IVS3 and IVS11) and downstream the gene (referred as PKA and PKV). No significant association was detected between any of the *HBB*, *G6PD* and *PKLR* alleles and infection; and the mutations were not identified in any individual. However, the LD test (considering the newly described polymorphic loci)

revealed a more conserved *PKLR* genome region in non-infected individuals (LD significant for all pairs of loci only in this group), justifying further investigation on *PKLR* gene.

The second study, focused on *PKLR* gene only, aimed at searching for selection signatures in the genome region surrounding this gene. Compared to the previous study, a larger number and different type of polymorphic loci were analyzed (SNPs were considered besides STRs), samples were available from two malaria endemic countries, Angola and Mozambique (instead of an epidemic region), and samples were not only characterized in terms of infection, as also in terms of malaria outcome (NI, AI, UM and SM samples available). Overall, two mutations, four STR loci and 13 SNP loci were analyzed in a region of 95 kb long. The two mutations were the only previously identified in individuals with an African ancestry: 1456C>T (detected in Afro-American), and 1614A>T (identified in Sao Tome and Principe). Moreover, the estimated population structure for all African and Portuguese groups (Portuguese were used as control) was determined, through the genotyping of 32 Ancestry Informative INDELs, to make sure that substructure was not skewing the results. In this study several selection signatures were identified: a) data from STR and SNP loci spread along the *PKLR* gene region showed a considerably higher F_{ST} differentiation between African and Portuguese populations (0.10 using STRs and 0.24 using SNPs) than that usually found for neutral markers (0.05 for STRs and 0.10 for SNPs); b) similarly, in AMOVA using STR data, it was determined a significant 10.92% variation between African and Portuguese whereas a percentage of 3.6-5.2% has been reported for variation between major regions of the world using neutral polymorphisms; c) still in AMOVA analysis, variation among populations within Africa was stated to be 3.1% using neutral markers and in the present study a percentage of 0.12% was obtained; d) a wider region showing significant LD was found in the uncomplicated malaria group; and e) the haplotype 9/11/13/34 (PKV/PKA/IVS11/IVS3) was associated with this clinical group (although borderline). Altogether, these data suggested that malaria selective pressure is actually shaping the *PKLR* genomic region in Africa, then increasing the differentiation between endemic and non-endemic malaria regions when *PKLR* markers are considered, and reducing the *PKLR* gene diversity in Africa, where malaria is present. The *PKLR* gene region seems to be highly conserved in Africa and

even more in uncomplicated malaria group, where LD was significant between all loci pairs considered and to which a haplotype was significantly associated.

Latter, to find out if the haplotype associated to uncomplicated malaria included a mutation with a particular phenotype that could somehow be underlying protection against malaria severity, the samples presenting the haplotype were further explored. Each exon from each sample was amplified by PCR and analyzed by SSCP to detect alterations in amplicons mobility, which could indicate the presence of an alteration in the nucleotide sequence. No alterations were detected but subtle differences in migration pattern may go unnoticed with this technique.

The third study focused on PK deficiency prevalence in Africa since there were no previous studies available. A hospital-based study was performed to determine the occurrence of PK deficiency in Mozambique and eventually find a highly prevalent allele that could be under selection by malaria, as it happens for HbS and G6PD A- alleles. In the previous study, samples from Angola, Mozambique and Portugal, already available from other researches, were used and strong evidences were collected supporting the hypothesis of selection by malaria. So, we confidently moved forward to this new approach. The detection of a high frequent mutation in malaria regions would be a great achievement in the context of this thesis. After all, besides all the controversy regarding genetic polymorphisms and malaria protection, their co-distribution is the basis of "malaria hypothesis". So, a stay at Maputo, Mozambique, was planned, with the following objectives: a) to determine the occurrence of PK deficiency in that region, in individuals with distinct infection/malaria outcome status; b) to detect mutation(s) underlying deficiency (low activity); c) last but not the least, to contact with malaria reality away from the lab benches but close to the people who get sick and work with it every day.

After a long period of preparation, submission, and acceptance of a work plan, questionnaires and informed consents, the local Ethical Committee gave its approval to the collection of human isolates. Blood samples were then collected in both Blood Bank (healthy adults, NI and AI) and Pediatric Department (NI, AI, UM and SM children) of Central Hospital of Maputo and the enzyme activity measured in 296 fresh RBC. Overall, 4.1% of samples (12) had an activity of 39-75% of the control, and were

considered PK-deficient (intermediate phenotype). In 41.7% of these, the missense mutation 829G>A, in *PKLR* exon 7, causing the amino acid substitution 277Glu>Lys, was identified. A significant association was found between the allele 829A and PK-deficient activity and the prediction of the substitution effect on the structure and function of the enzyme was "possibly damaging" suggesting that the mutation is likely to be non-functional.

Subsequently, in the same study, the mutation was searched in a second sample set from Mozambique and in other African malaria endemic areas (Angola, Sao Tome and Principe and Equatorial Guinea) and in a non-malaria country (Portugal). The mutation was not found in Portugal. In the African countries, allele 829A frequencies were 3.0%, 3.4%, 1.3%, 1.5% in Mozambique, Angola, Sao Tome and Principe and Equatorial Guinea, respectively. The 829GA heterozygous prevalence was between 2.6 and 6.7%, which is, to our knowledge, the highest estimated so far worldwide, as well as the PK deficiency percentage found in Mozambique (4.1%). However, it must be noted that these values were obtained from hospital samples and not from samples randomly collected in general population. Nevertheless, the overall values should not be significantly different from these, since most deficient and mutant samples were from healthy voluntary blood donors. From all mutant individuals, only one homozygous 829AA was found: an adult blood donor showing no symptoms. This shows that the mutation in homozygosis is not lethal and, in a first approach, seems to counteract the non-functional nature of 277Lys variant predicted *in silico*. However, it is difficult to conclude since we do not know the clinical history of the individual and it is recognized that clinical manifestations of a genetic disease reflect the interactions of physiological and environmental factors.

Samples from Angola and Mozambique were characterized in terms of infection/malaria outcome, so an association analysis was performed trying to associate the infection/malaria disease with the allele 829A presence in children, but no significant association was found. In the same way, no association was found between the 829A allele and infection and no association was detected between PK deficient activity and both infection and malaria outcome. However, only 12 samples (11 NI and 1 SM) were available for testing a possible effect of low enzyme activity on infection/malaria severity and 20 (2 AI, 11 UM and 7 SM) for testing a possible effect

of allele 829A on malaria severity, meaning that this analysis is greatly limited by the small sample number.

The mutation 829G>A has recently been identified in three individuals (in heterozygosis): one with a dubious ancestry (University Medical Center, 2007), one from West Africa and other from Pakistan (Berghout, et al., 2012). Since the haplotypes that include 829G>A mutation in these last two individuals are different, it was suggested that it has arisen separately. In Pakistan, as in sub-Saharan Africa, malaria continues to be a major public health problem, however, contrarily to African region (where *P. falciparum* is the most prevalent *Plasmodium* species), *P. vivax* prevails (WHO, 2012). Berghout and collaborators sequenced the *PKLR* gene in 387 individuals from malaria-endemic (Africa and Middle East) and other regions (Europe) in order to assess genetic variability in different geographical regions and ethnic groups. Coincidentally, neutral testing only suggested positive selection of the gene in sub-Saharan African and Pakistani populations. The only mutation that was found in common in both regions was exactly the substitution 829G>A, suggesting that this locus may be under positive selection.

The highest PK deficiency prevalence (based in activity measurements, since allele frequencies have been determined by different methods in different studies) reported up to the moment seem to occur in sub-Saharan Africa (about 4.1%, as this study shows) and Middle East, namely Saudi Arabia (3.12%, as described in Abu-Melha, et al., 1991) and South Iran (1.9%, described in Yavarian, et al., 2008). These are regions where the burden of malaria has been enormous in the last centuries. In Africa, *P. falciparum* prevails, whereas in the Middle East, *P. vivax* presents higher frequencies (WHO, 2012). In the general white population a prevalence of 0.005% has been estimated (Beutler, et al., 2000). These data shows that PK deficiency geographical distribution presented by López and colleagues (2010) is not correct, simply reflecting the lack of knowledge regarding PK deficiency prevalence in other world regions besides Europe. These frequencies, however, fall far short from those from HbS and G6PD polymorphisms. These polymorphisms can be associated to a more advantageous condition than PK 277Lys. Another possibility is that this variant may have a more recent origin so its frequency is still not very high.

Altogether, these three population studies much contributed to the knowledge of PK deficiency in general, in particular in the African continent, from where there was no data at all. Moreover, they allowed us to gather several evidences supporting the malaria protective effect by PK deficiency. The high frequent variant 277Glu>Lys seems to contribute to this protection, being positively selected by malaria. Four additional studies around this variant would major contribute to clarify the remaining doubts: 1) determination of its date of origin (is ongoing); 2) an association study with a larger sampling effort and longitudinal malaria clinical history characterization of individuals to analyze its association with malaria severity; 3) a large epidemiological study on its worldwide distribution; and 4) an *in vitro* study growing *Plasmodium* parasites in RBC presenting the mutation (homo and heterozygous) and comparison with growth in normal RBC.

Our fourth study had a totally different nature, since it was focused on the biological mechanism underlying malaria protection and on the global infection dynamics. We tried to look to an old problem (malaria) with innovative approaches, with the following characteristics: explore the problem under a dynamic perspective (the perspective of the host, as in the previous studies, and the perspective of the parasite); analyze a different biological material (proteins); and use of cutting edge technology (quantitative label free MS). In this study, we had four main objectives: 1) to assess parasite invasion and maturation of *P. falciparum* 3D7 growing *in vitro* in PK and G6PD-deficient and normal RBC; 2) to analyze the proteomic profile of non-infected and infected PK and G6PD-deficient and normal RBC (membrane and cytoplasmic fractions); 3) to analyze the proteomic profile of *P. falciparum* 3D7 parasites that grew in both deficient and normal RBC and 4) to correlate all these data from both host and parasite and understand their interactions in terms of protein exchanges and metabolism as well as the process in the basis of protection against malaria in the human host.

We thought that this would be a pertinent study since it would bring new important data on the total proteome from: normal RBC infected with *Plasmodium*, infected and non-infected PK-deficient RBC, infected and non-infected G6PD-deficient RBC and *Plasmodium* growing in different conditions (normal, PK and G6PD-deficient RBC). G6PD deficiency was considered to be included in this study because it would be

a control to PK deficiency experiments (as it is the most studied enzymopathy in association with malaria) and, additionally, because it has not been studied under this perspective.

No significant differences were observed in invasion and maturation of parasites growing *in vitro* in normal and deficient RBC (both PK and G6PD) considering three growth cycles. However, the reduced number of replicates may have contributed to this result. Invasion ratios were lower (although not significantly) in parasites from deficient RBC, indicating that the invasion step should be further analyzed.

Up to now, only proteomic data from parasites were obtained. The response from parasites growing in PK-deficient and G6PD-deficient RBC was distinct and proportional to phenotype severity. In parasites growing in G6PD-deficient RBC (from an asymptomatic individual), the main alteration in protein abundance was the increase of parasitic heat shock proteins and chaperones, showing that parasites are responding to oxidative stress conditions increasing the expression of defensive molecules. In PK-deficient (transfusion-dependent individual with regular hemolytic crisis), a more wide and acute response was triggered by the parasite with a high number of proteins involved in diverse pathways displaying significant alterations in their abundance, the majority being down-expressed. The most represented biological processes in this response were hemoglobin digestion and protein trafficking/RBC remodeling. Moreover, almost 40% of all proteins with abundance alterations seemed to be related to Maurer's clefts, which have functions in sorting of proteins and assembly of complexes destined for the RBC membrane, playing crucial roles in the pathology of malaria infections (Lanzner, et al., 2006) The loss of functional Maurer's cleft proteins dramatically changes the PfEMP1 RBC surface disposal, preventing the assembly of the *P. falciparum* virulence complex at the surface of host RBC (Cooke, et al., 2006; Dixon, et al., 2011; Hanssen, et al., 2008). So, from these results, we hypothesized that the protection against malaria that seems to be conferred by PK deficiency is associated with the RBC membrane remodeling process by the parasite, which may lead to a reduction in invasion by new parasites and malaria virulence itself.

These results are in agreement with PK deficiency pathophysiology data, which indicate the membrane as one of the most affected cellular component. The key

abnormalities in PK deficiency are ATP depletion and increased content of 2,3-DPG. ATP-depleted cells lose large amounts of potassium and water, becoming dehydrated and rigid and cell destruction appears to be brought about mostly by the phagocytosis of metabolic unabled cells, the surface of which is recognized by the phagocytic cells (Zanella and Bianchi, 2000). Several abnormalities of PK-deficient RBC membranes have actually been reported (Zanella, et al., 1979; Allen, et al., 1983). Additionally, and unexpectedly, two studies were recently found associating Maurer's cleft improper formation in hemoglobinopathies with malaria resistance (Cryklaff, et al., 2011; Wellems, et al., 2012). On the other hand, our results do not support previous studies suggesting that reduced RBC ATP levels provide a model system to define the molecular basis of protection in PK deficiency (Ayi, et al., 2009).

To complete this proteomic analysis, it will be essential to get the results from the RBC proteome, in particular the membrane fraction. Soon, it will be possible to look "inside out and outside in" both the RBC and parasite through their proteomic profile and confirm the RBC remodeling as the key process of PK deficiency protection against malaria. The complete proteome profile from both RBC host and parasite will surely open new avenues of exciting research.

As previously mentioned, although abnormalities in *PKLR* gene may result in alterations of both RBC and liver enzyme, clinical symptoms are confined to RBC, since the hepatic deficiency is usually compensated by the persistent enzyme synthesis in hepatocytes (Nakashima, et al., 1977). However, since the malaria parasite has an initial hepatic phase, we considered that it would also be important to look to PKL. We found a study from Prudêncio, et al. (2008) describing a kinome-wide RNAi screening in hepatocytes, to identify kinases that could be implicated in *Plasmodium* sporozoite infection. The results suggested that *PKLR* was not implicated on liver infection. However, as the authors stressed, the obtained data did not rule out the possible involvement of other genes among those tested, since negative results in RNAi screens are generally inconclusive. So, we didn't immediately exclude the hypothesis of *PKLR* be implicated on malaria hepatic infection and contacted this group to share the results of our investigation and discuss this possibility. Then, a second screening of 20 genes was performed: 19 in which they were specifically interested and *PKLR* in which we were interested. Yet, the results were similar: *PKLR* knock-down did not led to any

alteration of parasite load in hepatocytes, indicating that this gene is not important to hepatic infection; nevertheless, the possibility of an inefficient *PKLR* knock-down was not totally excluded (Prudêncio, et al., unpublished results). Still, since this was the second study including *PKLR* silencing without relevant results, there were no further experiments. We still think that it would be pertinent to study PKL, especially in murine models (instead of *in vitro* hepatocytes). It would be interesting to infect normal and PK-deficient mice with luminescent sporozoites and compare the liver parasitemias. Gene knock-down is limited on time (after 2 cycles the gene is no longer silenced); if murine models were used, the silencing would be constant. Additionally, it would be possible to do the follow up of the parasites in the erythrocytic phase and see if the parasites arising from a PKL-deficient liver would have the same fitness as parasites originated from normal hepatocytes.

6.2. Major constraints of the study

During the development of the work described in this thesis we came across several constraints and difficulties. The major constraint in the population studies was the reduced sampling. For instance, with such a limited number of samples presenting both the allele 829A and characterization for malaria outcome it was not possible to definitely conclude about an association between the presence of allele 829A and malaria severity.

In invasion/maturation and proteomics studies the obstacles were greater. The main difficulties were the low volume of PK-deficient blood available and the absence of previous reports and protocols that we could use as reference (e.g. describing the quantity of protein extracts that was possible to get in these specific conditions, describing the preparation of protein extracts from both *Plasmodium* and RBC from the same culture). Concerning the volume of blood, only 10 ml were provided for both invasion/maturation and proteomics experiments, corresponding to a final volume of about 4 ml of RBC, approximately, considering a percentage of 45% of RBC in whole blood sample plus the volume of RBC that is wasted with washings. It was not possible to collect a higher volume of blood from the PK-deficient individual, since this is an

anemic person requiring frequent blood transfusions; however it clearly limited our experiments and conclusions.

Due to its innovative character, our priority was the proteomic experiments rather than the invasion/maturation assays. So, we were conservative in the volume of RBC used in invasion/maturation experiments, to be sure that we had sufficient RBC to get enough parasite extracts for MS analysis (in the end, we had 32 flasks with 15 ml cultures each). So, we only worked with two replicates in invasion/maturation assays, which proved to be insufficient. We should have worked with a higher number of replicates with a lower volume (1 ml instead of 3 ml cultures, for instance). This was not done because we fear that such a small initial volume of deficient RBC did not stand all the three parasite growth cycles and daily blade smears (remember that new RBC were never added to cultures during these assays). Additional difficulties included the contamination of parasite fractions with host proteins, particularly hemoglobin. Much time was spent trying to identify a good method of hemoglobin depletion and it was not totally efficient.

Besides technical constraints, some other factors may have also influenced the results, namely: the concentration of normal RBC in cultures with PK-deficient RBC (from where we obtained the deficient extracts); the proportion of non-infected RBC mixed with infected RBC (from where we obtained the infected extracts); and the percentage of reticulocytes in PK-deficient cultures. All these issues may lead to some noise in the MS analysis. However, we intended to get close to human infection physiological conditions (where we do have reticulocytes in these conditions and non-infected RBC mixed with infected ones) and we had several different controls (normal RBC, non-infected RBC, etc.) to ensure the accuracy of the analysis.

References

Abu-Melha, A.M., Ahmed, M.A., Knox-Macaulay, H., Al-Sowayan, S.A. and el-Yahia A., 1991. Erythrocyte pyruvate kinase deficiency in newborns of eastern Saudi Arabia. *Acta haematologica*, 85(4), pp.192-94.

Allen, D.W., Groat, J.D., Finkel, B., Rank, B.H., Wood, P.A. and Eaton, J.W., 1983. Increased adsorption of cytoplasmic proteins to the erythrocyte membrane in ATP-depleted normal and pyruvate kinase-deficient mature cells and reticulocytes. *American journal of hematology*, 14(1), pp.11-25.

Ayi, K., Liles, W.C., Gros, P. and Kain, K.C., 2009. Adenosine triphosphate depletion of erythrocytes simulates the phenotype associated with pyruvate kinase deficiency and confers protection against Plasmodium falciparum in vitro. *The Journal of infectious diseases*, 200(8), pp.1289-99.

Berghout, J., Higgins, S., Loucoubar, C., Sakuntabhai, A., Kain, K.C. and Gros, P., 2012. Genetic diversity in human erythrocyte pyruvate kinase. *Genes and immunity*, 13(1), pp.98-102.

Beutler, E. and Gelbart, T., 2000. Estimating the prevalence of pyruvate kinase deficiency from the gene frequency in the general white population. *Blood*, 95(11), pp.3585-88.

Cooke, B.M., Buckingham, D.W., Glenister, F.K., Fernandez, K.M., Bannister, L.H., Marti, M., Mohandas, N. and Coppel, R.L., 2006. A Maurer's cleft-associated protein is essential for expression of the major malaria virulence antigen on the surface of infected red blood cells. *The Journal of cell biology*, 172(6), pp.899-908.

Cyrklaff, M., Sanchez, C.P., Kilian, N., Bisseye, C., Simpore, J., Frischknecht, F. and Lanzer, M., 2011. Hemoglobins S and C interfere with actin remodeling in Plasmodium falciparum-infected erythrocytes. *Science*, 334(6060):1283-6.

Dixon, M.W., Kenny, S., McMillan, P.J., Hanssen, E., Trenholme, K.R., Gardiner, DL. and Tilley, L., 2011. Genetic ablation of a Maurer's cleft protein prevents assembly of the *Plasmodium falciparum* virulence complex. *Molecular microbiology*, 81(4), pp.982-93.

Hanssen, E., Hawthorne, P., Dixon, M.W.A., Trenholme, K.R., McMillan, P.J., Spielmann, T., Gardiner, D.L. and Tilley, L., 2008. Targeted mutagenesis of the ring-exported protein-1 of *Plasmodium falciparum* disrupts the architecture of Maurer's cleft organelles. *Molecular Microbiology*, 69(4), pp.938–53.

Lanzer, M., Wickert, H., Krohne, G., Vincensini, L. and Braun Breton, C., 2006. Maurer's clefts: a novel multi-functional organelle in the cytoplasm of Plasmodium falciparum-infected erythrocytes. International journal for parasitology, 36(1), pp.23-36.

López, C., Saravia, C., Gomez, A., Hoebeke, J. and Patarroyo, M.A., 2010. Mechanisms of genetically-based resistance to malaria. Gene, 467(1-2), pp.1-12.

Nakashima, K., Miwa, S., Fujii, H., Shinohara, K., Yamauchi, K., Tsuji, Y. and Yanai, M., 1977. Characterization of pyruvate kinase from the liver of a patient with aberrant erythrocyte pyruvate kinase, PK Nagasaki. The Journal of laboratory and clinical medicine, 90(6), pp.1012-20.

Prudêncio, M., Rodrigues, C.D., Hannus, M., Martin, C., Real, E., Gonçalves, L.A., Carret, C., Dorkin, R., Röhl, I., Jahn-Hoffmann, K., Luty, A.J., Sauerwein, R., Echeverri, C.J. and Mota, M.M., 2008. Kinome-wide RNAi screen implicates at least 5 host hepatocyte kinases in Plasmodium sporozoite infection, PLoS pathogens, [online] Available at: < http://www.plospathogens.org/article/info%3Adoi%2F10.1371%2Fjournal.ppat.1000201> [Acessed 4 January 2013].

University Medical Center, Laboratory for Red Blood Cell Research., 2007. PKLR Mutation Database. [online] Available at: http://www.pklrmutationdatabase.com/ [Accessed 15 December 2012].

Wellems, T.E. and Fairhurst, R.M., 2012. An evolving picture ofthe interactions between malaria parasites and their host erythrocytes. Cell research, 22(3), pp.453-6.

World Health Organization (WHO), 2012. World Malaria Report. [pdf] Geneva: WHO Press. Available at: http://www.who.int/malaria/publications/world_malaria_report_2012/en/> [Accessed 8 December 2012].

Yavarian, M., Karimi, M., Shahriary, M. and Afrasiabi, A.R., 2008. Prevalence of pyruvate kinase deficiency among the south Iranian population: quantitative assay and molecular analysis. Blood cells, molecules & diseases, 40(3), pp.308-11.

Zanella, A. and Bianchi, P., 2000. Red cell pyruvate kinase deficiency: from genetics to clinical manifestations. Baillière's best practice and research. Clinical haematology. 13(1):57-81.

Zanella, A., Brovelli, A., Mantovani, A., Izzo, C., Rebulla, P. and Balduini, C., 1979. Membrane abnormalities of pyruvate kinase deficient red cells. British journal of haematology, 42(1), pp.101-08.

Chapter 7 –

Conclusions

CONCLUSIONS

One of the challenges in the fight against malaria is to describe the host determinants of disease susceptibility and decipher the involved mechanisms to eventually use them as new targets for antimalarial drugs or vaccines.

This thesis has given an important input to the knowledge of human genetic factors associated to malaria protection and malaria infection dynamics between *Plasmodium* parasite and RBC host. The strategy followed included several distinct approaches (molecular human genetics and proteomic analyses, enzymatic assays, field work in Africa) and was developed in laboratories with very different characteristics (IHMT, CIAS, IPATIMUP, Faculty of Medicine at Maputo, Centre of Excellence in Mass Spectrometry, the last two at opposite ends in terms of technology) that enriched the present study with data of diverse nature. This proved to be efficient since we could get a global picture of the association between PK deficiency and malaria, as we initially aimed. Further research as that described in the previous chapter, based on the results obtained in the present work, will be important to complete our *understanding* of the complex interactions between host and parasite.

"*I do not understand. That is so vast that it surpasses all understanding. Understanding is always limited. But do not understand can not have borders. I feel I am much more complete when I do not understand. Not understanding (…) is a strange blessing like experiencing madness without being mad. It is a gentle disinterestedness, is a sweetness of stupidity. Only once in a while comes the concern: I want to understand a little. Not too much, but at least understand that I do not understand.*"

Clarice Lispector

Supplementary Information

CHAPTER 1 – General Introduction

Table S1. Classification of countries by stage of elimination (data from December 2012) (WHO, 2012).

Region	**Pre-elimination**	**Elimination**	**Prevention of re-introduction**	**Recently certified as malaria free**
African	Cape Verde	Algeria		
Region of the Americas	Argentina Costa Rica Ecuador El Salvador Mexico Paraguay			
Eastern Mediterranean		Islamic Replublic of Iran Saudi Arabia	Egypt Iraq Oman Syrian Arab Republic	Morocco- 2010 United Arab Emirates- 2007
European		Azerbaijan Kyrgyzstan Tajikistan Turkey Uzbekistan	Georgia	Armenia- 2011 Turkmenistan- 2010
South-east Asia	Bhutan Democratic People's Republic of Korea	Sri Lanka		
Western Pacific	Malaysia	Republic of Korea		

Table S2. Epidemiological profile, intervention strategies and antimalarial policy from the five studied African countries (WHO, 2012)

		Cape Verde	**Mozambique**	**Angola**	**Equatorial Guinea**	**Sao Tome and Principe**
EPIDEMIOLOGICAL PROFILE	Phase	Pre-elimination*	Control	Control	Control	Control*
	High transmission area (≥1 case/1000)	0%	100%	100%	100%	100%
	Low transmission area (0-1 case/1000)	26%	0%	0%	0%	0%
	Malaria-free area	74%	0%	0%	0%	0%
	Plasmodium species	*P. falciparum*: 100%	**P. falciparum: 95% P. malariae and P. ovale: 5% P. vivax rare**	**P. falciparum: 90% P. ovale: 5% P. vivax: 5%**	**P. falciparum: 85% P. malariae, P. ovale and P. vivax: 15%**	**P. falciparum: 85% P. malariae, P. ovale: 15% P. vivax rare**
	Major *Anopheles* species	*An. gambiae* *An. arabiensis*	*An. gambiae* *An. arabiensis* *An. funestus*	*An. gambiae* *An. funestus* *An. nili*	*An. gambiae* *An. cinctus* *An. melas*	*An. gambiae*
INTERVENTION STRATEGY	ITNs/LLINs	Not distributed	Distributed free of charge to all age groups	Distributed free of charge to all age groups	Distributed free of charge not to all age groups	Distributed free of charge not to all age groups
	IRS	Recommended; DDT not used	Recommended; DDT is used	Recommended; DDT not used	Recommended; DDT not used	Recommended; DDT not used
	IPT	Not used	Used during pregnancy	Used during pregnancy	Used during pregnancy	Used during pregnancy
ANTIMALARIAL POLICY	First line treatment	Arthemether/ lumefantrine	Arthemether/ lumefantrine	Arthemether/ lumefantrine	Artesunate/ amodiaquine	Artesunate/ Amodiaquine
	For treatment failure of *P. falciparum*	Quinine	-	Quinine	Quinine	Quinine
	Treatment of severe malaria	Quinine	Quinine	Quinine	Quinine	Quinine
	Drug resistance	Chloroquine	Chloroquine	Chloroquine	Chloroquine	Chloroquine
	Prophylaxis	Not applicable	Atovaquone/ proguanil, doxycycline or mefloquine	Atovaquone/ proguanil, doxycycline or mefloquine	Atovaquone/ proguanil, doxycycline or mefloquine	Atovaquone/ proguanil, doxycycline or mefloquine

*>75% decrease in case incidence 2000-2011.

Table S3. Intervention coverage estimation and reported malaria cases and deaths in the countries studied in the present thesis (data from 2011, WHO 2012).

	Cape Verde	**Mozambique**	**Angola**	**Equatorial Guinea**	**Sao Tome and Principe**
United Nations population	500 585	23 929 708	19 618 432	720 213	168 526
Nr of probable and confirmed malaria cases	36	1 756 374	2 534 549	33 830	6 504
Nr of deaths	4	3 086	6 909	52	19
% IRS coverage	100	36	4	-	69
% of population potentially protected by ITNs delivered	-	46	40	1	87
% any antimalarial coverage/ % ACTs coverage	100/100	64/64	73/73	8/8	100/100

CHAPTER 2 - Analysis of malaria associated genetic traits in Cabo Verde, a melting pot of European and sub Saharan settlers

Detection of Hemoglobin S Allele (HbS)

Primers were used in a multiplex reaction mixture and PCR conditions were 35 cycles, each of 94ºC (1') for DNA denaturation, 65ºC (1') for primer annealing and 72ºC (2') for extension, followed by an elongation period of 10' at 72ºC; PCR reaction was performed in a final volume of 25µl with 50mM KCL, 10mM Tris pH 8.3, 7mM of $MgCl_2$, 200µM of dNTP's, 1µM of WT-AS, WT-CP517 and Mut-AS primers and 0.8µM of Mut-CP267, 0.1µg/µl of BSA and 0.02U/µl of GoTaq Flexi DNA Polymerase (Promega).

Homozygous HBSS status was confirmed by a PCR-RFLP technique. A DNA fragment of 390bp containing the 5' end of the HBB gene was amplified using the primers: 5'-ACCTCACCCTGTGGAGCCAC-3' (forward) and 5'-ACCAGCAGCCTAAGGGTGGGAAAATACACC-3' (reverse). The PCR reaction was performed in a volume of 50µL, containing 150ng of genomic DNA, 25pmol of each primer, 16.6mM $(NH_4)_2SO_4$, 67mM Tris-HCl, pH 8.8, 6.7mM $MgCl_2$, 6.7µM Na_2EDTA, 1.4µg/mL BSA, 10mM β-mercaptoethanol, 0.2mM dNTPs and 1U/µl Taq polymerase. Amplification was performed through an initial denaturation at 94ºC (5') followed by 28 cycles of denaturation at 94ºC (1'), annealing at 64ºC (1') and extension at 72ºC (1'), with a final extension at 72ºC (10'). The sickle cell mutation was searched in PCR fragments by restriction with Bsu36I endonuclease, according to the manufacturer's instructions (New England Biolabs).

Detection of Glucose-6-phosphate Dehydrogenase Polymorphisms

The G6PD B, A, and A^- alleles were distinguished by PCR amplification of exons 3 and 4 followed by digestion with NlaIII restriction enzyme (New England Biolabs), and by

amplification of exon 5 followed by digestion with FokI (New England Biolabs). Alleles who lack both restriction sites were classified as B, those who lacked the NlaIII site but contained the FokI site were classified as A, and those who had both NlaIII and FokI restriction sites were classified as A^-. The Med mutation was detected by amplification of exon 6 followed by digestion with MobII (New England Biolabs).

Detection of Pyruvate Kinase Polymorphisms

Analysis of binary polymorphisms

Exon 3 was amplified by PCR with specific primers as follows: [3D:5'-GGTGACATGCAGTCCCTGAG-3' (forward), 3R: 5'-AGATGAAGAAGCACCTCAAG-3' (reverse)], denaturation at 94ºC for 5min followed by 30 cycles of 45sec at 94ºC, 45sec at 58ºC, 1min at 72ºC and final extension for 1min at 72ºC. In all cases (exons 3, 11 and 12, intron 10), 1μl of DNA template was used in the amplification reaction. PCR reactions were carried out in a total volume of 25μl, containing 3mM MgCl2, 50mM KCl, 10mM Tris, pH 8.3 (HCl), 0.2mM of each dNTP, 50ng of each oligonucleotide primers and 0.1 units of Taq DNA Polymerase (Fermentas). PCR products from exon 3, 11 and 12 were first visualized under UV transillumination after electrophoresis on agarose gels, stained with 1.5% ethidium bromide and then further analyzed as follows. 269T>A mutation was screened by specific restriction with SfaNI endonuclease according to the manufacturer's instructions (New England Biolabs). 1456C>T and 1705A/C screening was performed by single-strand conformational polymorphism (SSCP) analysis [modified from Orita et al (Proc.Natl.Acad.Sci. USA 86 (1989) 2766–2770.)]: PCR products were mixed (1:1) with a denaturing solution [0.1% (w/v) each of bromophenol blue and xylene-cyanol; 10mM EDTA; 0.1% SDS and 95% (v/v) deionized formamide]; the mixture was heated at 96°C for 5min, placed immediately on ice and then 5μl were loaded on a vertical non-denaturing polyacrylamide minigel containing 12% (w/v) acrylamide–bisacrylamide (75.9/1), 10% (v/v) glycerol and 50mM TBE buffer, pH 8.3; the electrophoresis was performed in 0.5× TBE buffer at 200V for 4h; the DNA bands were stained with silver nitrate. In case of doubtful mobility patterns, isolates were screened for these two polymorphisms with

BsmAI and BspHI endonucleases, respectively according to the manufacturer's instructions (New England Biolabs). The T10/19 repeat (intron 10) was screened through horizontal polyacrylamide gel electrophoresis of PCR products.

Analysis of STRs

DNA was amplified using a Multiplex PCR with labeled forward primers (Table). PCR reactions were carried out in a total volume of 5µl, containing 2.5µl of Qiaqen Multiplex PCR Master Mix (Qiaqen Multiplex PCR Kit), 0.5µl of Primer Mix (IVS3 0.5µM, PKA 0.5µM, PKV 0.25µM, IVS11 0.75µM) and 0.5µl of genomic DNA, as follows: denaturation at 94ºC for 15min, 30 cycles of 30seg at 94ºC, 1min30seg at 62ºC and 1min at 72ºC followed for a final extension at 72ºC for 1h.

Amplified fragments were analysed on ABI Prism 3100 or 3130 sequencers (Applied Biosystems) and results were analysed with GeneScan 3.1.2 software. In order to determine the sequence and number of repeats of each *locus*, some samples with alleles of different size were selected. After separation by polyacrylamide gel electrophoresis, the band of smaller size was extracted and 0.5µl of this DNA was amplified with specific primers in a total volume of 25µl as above.

Products were purified with Microspin S-300 HR columns (Amersham Pharmacia Biotech) and sequenced using the BigDye Terminator Cycle Sequencing ready reaction kit (Applied Biosystems) as follows: reaction mixture of 2.5µl of DNA, 2µl of labeled dNTPs and 0.5µl of reverse primer with the following conditions: 96ºC for 4min, 35 cycles of 96ºC for 10seg, 58ºC for 5seg, 60ºC for 2min and 60ºC for 10min. Products were purified again with Sephadex (Amersham Biosciences) - 750µl of Sephadex was put in columns which were centrifuged at 1 000 x g for 4min; after transferring the columns for new tubes, product was add to the columns and centrifuged again at 1 000 x g for 4min. Eight µl of formamide was added before sequencing and analysis was performed on ABI Prism 3100 sequencer (Applied Biosystems) and Sequencing Analysis 3.7 software.

Table. STR multiplex amplification primers (labeled forward primers).

Loci	Repetitive region	Primers	Amplicon size (bp)
IVS3 (intron 3)	several (see text for details)	IVS3 F - 5'CCTAGGTGACAGACGAGACC3' IVS3 R - 5'CCGGCCAACTTTCACTCC3'	300
IVS11 (intron 11)	(ATT)n	IVS11 F - 5'GCC TTGATGTGGTGAAAGGT3' IVS11 R - 5'CTGGGGACAGAGCAAGACTC3'	167
PKA (25kb downstream)	(AAAT)n	PKA F2 - 5'ATGCCACTGCACATCAGTCT3' PKA R - 5'TGGCTCCAACTGGGTAAAAC3'	221
PKV (65kb downstream)	(TTTA)n	PKV F - 5'GATGCTGACTCCGAACACAA3' PKV R - 5'GGAGGCTGAAGGAGGAGAAT3'	175

Pyruvate Kinase Polymorphisms

STRs

The IVS3 locus is the most polymorphic with 8 repeat regions and it is interrupted. The consensus sequence determined is

...TC **$(CTTT)_n(CT)_{0-1}(CTTT)_n(CCTT)_n$**
CTTTCTTTTCTTTCTTTCTTTCTTGCCTGCTTGCTTTCTTTCCTTCCTTCCTTCCCT
CCCTCCCTCCCTCCTTCCTTCCTTCCTTCTTT **$(CT)_{2-4}(CTTT)_n(CCTT)_n(CTTT)_n$**
CTC...

and for simplicity, the following one was considered

$(CTTT)_{nA}(CT)_{0-1}(CTTT)_{nB}$ $(CCTT)_{nC}$ [89] $(CT)_{2-4}(CTTT)_{nD}(CCTT)_{nE}(CTTT)_{nF}$

The allele classification was assessed through the sum of the number of repeats, as nA+0or1+nB+nC+nD+nE+nF; when $(CT)_2$ is present, the allele is classified as .1 and $(CT)_4$, classified as .2.

CHAPTER 3 - Malaria: looking for selection signatures in the human *PKLR* gene region

Supplementary Table I. SNP loci selected for analysis (ordered according to localization), allelic frequencies and primers used for multiplex PCR.

SNP (along 40970 bp)	Allelic Frequency		Primer	Product (bp)	Primers Sequence (5'-3')
11055 bp after TGA					
refSNP rs7549276 – pk_276 – gene *HCN3*	G	A	pk_276	442	GCTGTCCCTAGTGCTGAAGG
chr1:153515199..153515199	0.500	0.500			GACTAGAAAAGGCGCACTGG
(5008 bp)					
refSNP rs7520184 – pk_184 – gene *HCN3*	G	A	pk_184	413	CTGCACCCACTAACTCGTCA
chr1:153520207..153520207	O.583	0.417			CAGCCTGGCAAATTCTCTTC
(2254 bp)					
refSNP rs11264352 – pk_352 – gene *HCN3*	T	C	pk_352	127	ATCCTACTTTGGGGGTCAGC
chr1:153522461..153522461	0.542	0.458			GGCTGGAGCTCTGTGATTCT
(1655 bp)					
refSNP rs11264355 – pk_355 – gene *HCN3*	C	G	pk_355	393	TGAGTACCAGTCCCCTGACC
chr1:153524116..153524116	0.569	0.431			GTACCAGTGGCTCCCACAGT
(2604 bp)					
chr1: 153526254 – *pkLR* gene TGA					
refSNP rs932972 - pk_972 - EXON 12	C	T			
chr1:153526720..153526720	0.583	0.417			
(254 bp)					
refSNP rs1052177 – pk_177 - EXON 12	T	C	pk_972_177_176	406	CTGGTGATTGTGGTGACAGG
chr1:153526974..153526974	0.585	0.415			AACCAGCCAAACTGGGATTA
(33 bp)					
refSNP rs1052176 – pk_176 - EXON 12	C	A			
chr1:153527007..153527007	0.583	0.417			
(1168 bp)					

1614A>T – pk_1614 - EXON 11 chr1:153528175..153528175 (158 bp) 1456C>T – pk_1456 - EXON 11 chr1:153528333..153528333 (904 bp)		Mutations associated to PK-deficiency	pk_mut	372	TGACACCTGGAACTGGAACA GACCACAGGAGAGAGGCAAG
refSNP rs4620533 – pk_533 - INTRON 10 chr1:153529237..153529237 (3048 bp)	C 0.517	G 0.483	pk_533	180	TCCTGTTAATCCTGCCAACC GCTCAGAGGCAAGTCCATTC
refSNP rs8177970 – pk_970 - INTRON 3 chr1:153532285..153532285 (9299 bp)	A 0.892	G 0.108	pk_970	151	AGGGAAGGGGAGTCTGTGAT TCACGTTCAGACAACGTTCC
chr1: 153537835 – ATG of *pkLR* gene					
refSNP rs12032720 – pk_720 chr1:153541584..153541584 (4522 bp)	G 0.700	C 0.300	pk_720	321	GGCACCCATAGGAGATGAGA CTCCACTATCTGGGCCTGAA
refSNP rs2297480 – pk_480 - gene *FDPS* chr1:153546106..153546106 (3347 bp)	A 0.783	C 0.217	pk_480	357	GAAGACCCCCACAGATCTCA TCCTTTCAGCCCCTAATCCT
refSNP rs11264359 – pk_359 - gene *FDPS* chr1:153549453..153549453 (6716 bp)	A 0.375	G 0.625	pk_359	206	TCCAAAGGCTATTCAGAAGCA GCAGAAGTTGCATCCACTCA
refSNP rs11264361 – pk_361 - gene *FDPS* chr1:153556169..153556169 (5227 bp)	T 0.817	G 0.183	pk_361	212	CACCAGCTTCACTCCTCCTC GGCACCTTCAGGATCTGGTA
18 334 bp before ATG					

(**refSNP rs...– pk_"nr"– "x"** - SNP reference in HapMap – SNP designation in the study – localization; chr1: ... – localization in chromosome 1; **("nr" bp)** – distance between adjacent SNPs; **Allelic Frequency** – allelic frequencies in Nigerian population, African populations reference in Hapmap; **ATG** – *pkLR* gene start codon; **TGA** – STOP codon of *pkLR* gene.

Supplementary Table II. Single Base Extension (SBE) primers used for SNaPshot reaction.

Target region	SNP	Mutation	Detection	Conc. (μM)	SBE-primer sequence
pk_972_177_176	pk_177		A>G	0.4	**GTAGGCTGGGCCAGAGG**
pk_352	pk_352		T>C	0.4	GTCTGACAA**GCTCTGGGTCCCTGCC**
pk_972_177_176	pk_972		G>A	1.22	TCTGACAA**CTGAGCAGATTGGATGCAG**
pk_184	pk_184		G>A	0.4	**CCTATCTATAAGATGAGAGAAATAAGAAACT**
pk_276	pk_276		G>A	0.4	GTGAAAGTCTGACAA**CCCATTGTTCCTTTCACTCCT**
pk_355	pk_355		C>G	1.22	GCCACGTCGTGAAAGTCTGACAA**CCCACCCCATCCTGATA**
pk_720	pk_720		C>G	0.4	AGGTGCCACGTCGTGAAAGTCTGACAA**GGGCAAGGGTGTTGGTAAA**
pk_mut	pk_1614		T>A	0.2	GCCACGTCGTGAAAGTCTGACAA**GAAGGTCTAGGTAGCTCACCACT**
pk_480	pk_480		A>C	0.4	AAACTAGGTGCCACGTCGTGAAAGTCTGACAA**CCAGATAACTCCCACCCC**
pk_972_177_176	pk_176		G>T	0.4	GACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA**CAGGATATGCTTAGCACCC**
pk_361	pk_361		A>C	1.22	TGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA**CAGCAAAAGAGGAAGGATG**
pk_mut	pk_1456		C>T	1.22	CAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA**CTCAGCCCAGCTTCTGTCT**
pk_970	pk_970		A>G	0.4	CAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA**GGTTGCATCAGGGAATAAAG**

pk_359	pk_359	T>C	0.4	CCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA**AGTGAGCTGCCAGTTTTCAAT**
pk_533	pk_533	G>C	0.12	CCCCCCCCCCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA**AGAAATGTAGCTCTATTAGCCTGCT**

Target region – Multiplex PCR product; **Detection** – alternative alleles detected; **Conc. (µM) –** concentration in the SBE-primer mix; **bold nucleotides in SBE-primer sequence** – target sequence of the SBE-primes; **nucleotides not in bold** – neutral sequence as described in Sanchez *et al,* 2005.

Supplementary Table III. STR loci allele frequencies found in Angola (ANG), Mozambique (MOZ), control Portuguese (PT-C) and PK-deficient Portuguese (PT-PKD).

Loci	Allele	ANG	MOZ	PT-C	PT-PKD
IVS11	**7**	0.007	0.012	0.000	0.000
	9	0.000	0.000	0.006	0.000
	10	0.058	0.067	0.006	0.000
	11	0.000	0.000	0.000	0.000
	12	0.273	0.287	0.156	0.071
	13	0.054	0.146	0.063	0.000
	14	0.115	0.063	0.506	0.452
	15	0.155	0.071	0.188	0.262
	16	0.119	0.118	0.0313	0.167
	17	0.209	0.197	0.038	0.048
	18	0.011	0.039	0.006	0.000
PKV	**8**	0.011	0.008	0.025	0.024
	9	0.162	0.197	0.406	0.452
	10	0.428	0.433	0.481	0.476
	11	0.381	0.354	0.075	0.048
	12	0.018	0.008	0.013	0.000
PKA	**8**	0.004	0.008	0.000	0.000
	9	0.248	0.252	0.688	0.929
	10	0.054	0.047	0.075	0.024
	11	0.216	0.244	0.019	0.000
	12	0.151	0.079	0.013	0.000
	13	0.162	0.193	0.044	0.000
	14	0.104	0.134	0.069	0.024
	15	0.043	0.016	0.075	0.000
	16	0.018	0.028	0.019	0.000
	17	0.000	0.000	0.000	0.024
IVS3	**30**	0.004	0.000	0.006	0.000
	31	0.000	0.004	0.013	0.000
	31.2	0.000	0.004	0.000	0.000
	32	0.025	0.043	0.013	0.000
	32.2	0.000	0.000	0.013	0.000
	33	0.061	0.067	0.031	0.000
	34	0.176	0.220	0.056	0.024
	34.2	0.007	0.000	0.031	0.048
	35	0.198	0.177	0.031	0.024
	35.2	0.036	0.008	0.050	0.024
	36	0.097	0.087	0.056	0.000
	36.2	0.032	0.016	0.044	0.214
	37	0.061	0.047	0.056	0.024
	37.2	0.076	0.083	0.163	0.048
	38	0.050	0.031	0.019	0.024
	38.2	0.054	0.075	0.194	0.381
	39	0.022	0.024	0.025	0.000
	39.2	0.040	0.051	0.100	0.167
	40	0.004	0.008	0.013	0.024
	40.2	0.014	0.031	0.088	0.000

41	0.004	0.000	0.000	0.000
41.2	0.040	0.024	0.000	0.000

Supplementary Table IV. SNP loci allelic frequencies observed in Angola, Mozambique and Portuguese groups.

SNP loci	**Allele**	**Population groups**			
		ANG	**MOZ**	**PT-C**	**PT – PKD**
pk_276	**A**	0.610	0.646	0.222	0.053
	G	0.390	0.354	0.778	0.947
pk_184	**A**	0.566	0.549	0.213	0.079
	G	0.434	0.451	0.788	0.921
pk_352	**C**	0.610	0.612	0.220	0.079
	T	0.390	0.388	0.780	0.921
pk_355	**C**	0.404	0.373	0.768	0.895
	G	0.596	0.627	0.232	0.105
pk_972	**A**	0.588	0.566	0.219	0.105
	G	0.412	0.434	0.781	0.895
pk_177	**A**	0.423	0.393	0.781	0.895
	G	0.577	0.607	0.219	0.105
pk_176	**G**	0.414	0.399	0.769	0.895
	T	0.586	0.601	0.231	0.105
pk_1614	**A**	1.000	1.000	1.000	1.000
	T	0.000	0.000	0.000	0.000
pk_1456	**C**	1.000	1.000	1.000	0.737
	T	0.000	0.000	0.000	0.263
pk_533	**C**	0.726	0.715	0.225	0.105
	G	0.274	0.285	0.775	0.895
pk_970	**A**	0.826	0.818	1.000	1.000
	G	0.174	0.182	0.000	0.000
pk_720	**C**	0.515	0.565	0.800	0.868
	G	0.485	0.435	0.200	0.132
pk_480	**A**	0.632	0.680	0.800	0.868
	C	0.368	0.320	0.200	0.132
pk_359	**C**	0.790	0.794	0.219	0.132
	T	0.210	0.206	0.781	0.868
pk_361	**A**	0.771	0.758	0.805	0.868
	C	0.229	0.242	0.195	0.132

CHAPTER 4 - Pyruvate Kinase Deficiency in Sub-Saharan Africa: Identification of a Highly Frequent Missense Mutation (G829A;Glu277Lys) and Association with Malaria

Supporting Table S1. List of primers and annealing temperatures (a.t.) used in the amplification of PKLR promoter (Prom) and coding.

Table A. List of primers and annealing temperatures (a.t.) used in the amplification of *PKLR* promoter (Prom) and coding regions by PCR.

Exon	Product (bp)	Forward Primer (5'-3')	Reverse Primer (5'-3')	PCR a.t. (ºC)
Prom/1	495	AGCTAACTTCAGTAAAGTAC	GATGTGGATCATTTATGC	54
3	286	GGTGACATGCAGTCCCTGA	AGATGAAGAAGCACCTCAAG	56
4	253	CGTTCTGAGAATGGTAATGG	GAGGGTTTCAGGGGAAGGT	60
5	239	CCACCTTCCCCTGAAACC	CTGGGCCCAACCCTACAG	54
6	304	ACTCCGGGGCTCAGAACT	CTGATGGGGGAGCCAAGG	62
7	350	ACCGCAGCTGGCTCTTTC	GTGATGGGGAATAGCGACAG	60
8	252	CACCTTTCTTCTCCTGCCTG	CAGGTGTCCCTAAAACCCAC	60
9-10	500	CAGTGTGAGTCCTACAAC	CTGACCCAAAGCTCCATC	56
11	413	AGTGACACCTGGAACTGG	GATATCTCAGTCTTAGTG	52
12	259	CCTTGGCTTCCCAAAGTG	GCTGGAGAACGTAGACTG	60

CHAPTER 5 - Quantitative proteomics approach for the analysis of the human malaria parasite *Plasmodium falciparum* (trophozoite stage) and its red blood cell host – a preliminary study

Supplemental Figures

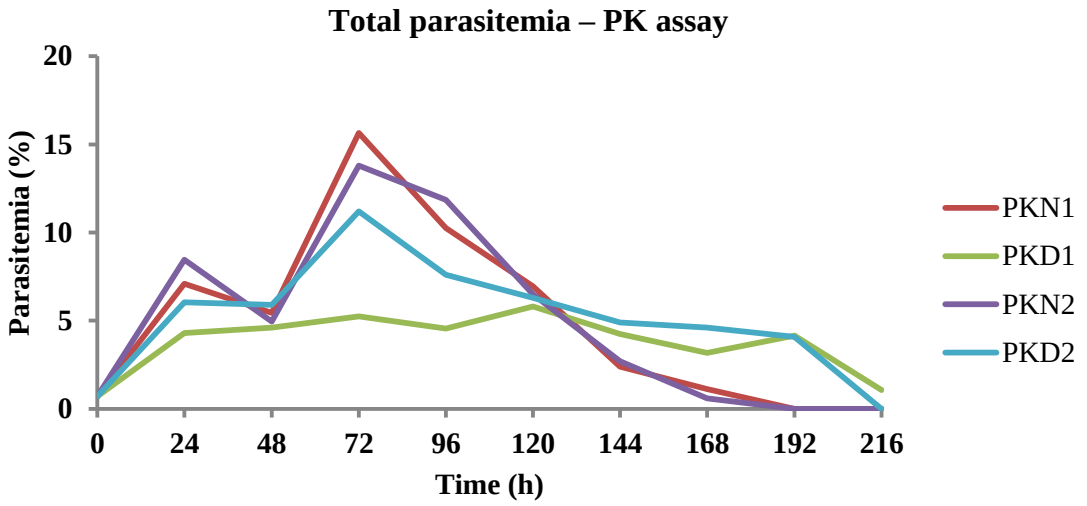


Fig. S1. Total parasitemias along the invasion/maturation PK assay. PKN: normal RBC; PKD: PK-deficient RBC; 1: replicate 1; 2: replicate 2.

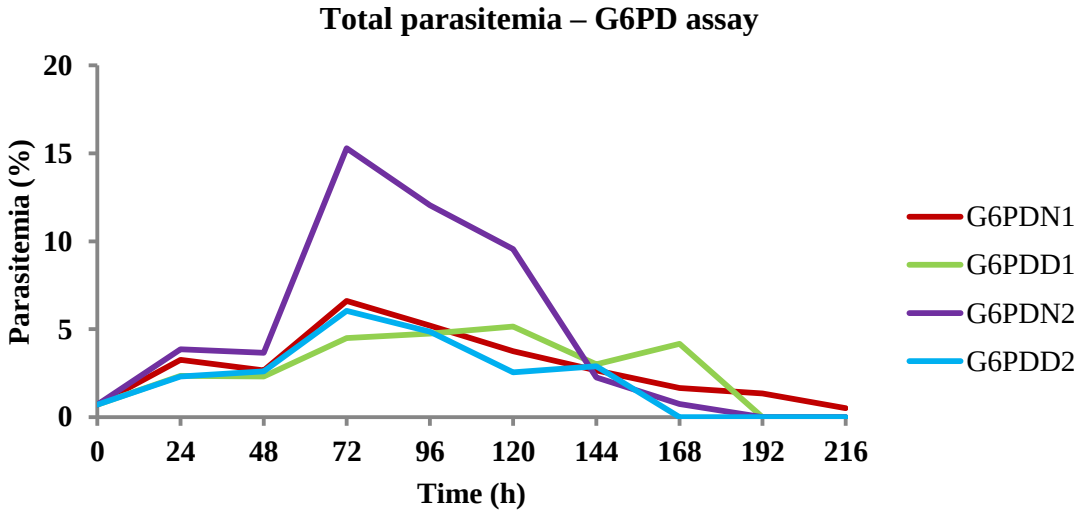


Fig. S2. Total parasitemias along the invasion/maturation G6PD assay. G6PDN: normal RBC; G6PDD: G6PD-deficient RBC; 1: replicate 1; 2: replicate 2.

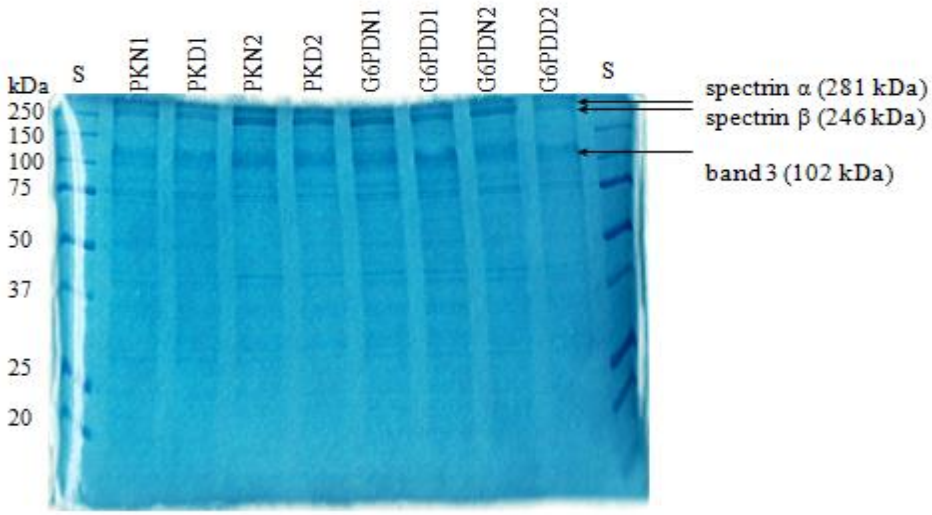


Fig. S3. Parasite extracts (5 µg loaded on each well) run in a 12.5% acrylamide:bisacrylamide 37.5:1 gel and stained with Coomassie blue brilliant reagent. PKN and G6PDN: extracts from parasites grown in normal RBC; PKD: extracts from parasites grown in PK-deficient RBC; G6PDD: extracts from parasites grown in G6PD-deficient RBC; 1: replicate 1; 2: replicate 2; S: protein standard (BioRad). The arrows indicate some proteins predicted to be present at those band levels.

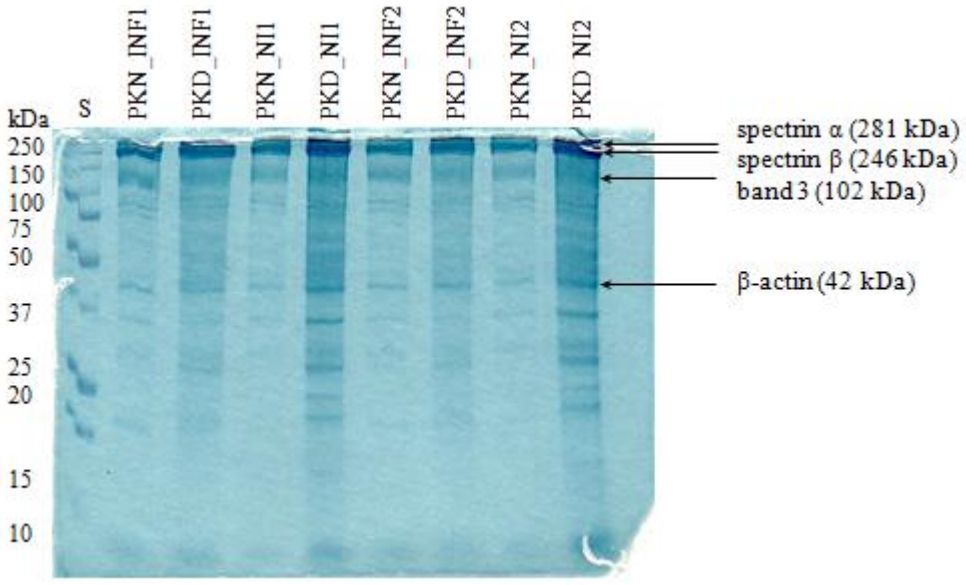


Fig. S4. RBC membrane extracts (5 µg loaded on each well) run in a 12.5% acrylamide:bisacrylamide 37.5:1 gel and stained with Coomassie blue brilliant reagent. PKN: extracts from normal RBC; PKD: extracts from PK-deficient RBC; INF: extracts from infected RBC; NI: extracts from non-infected RBC; 1: replicate 1; 2: replicate 2; S: protein standard (BioRad). The arrows indicate some proteins predicted to be present at those band levels.

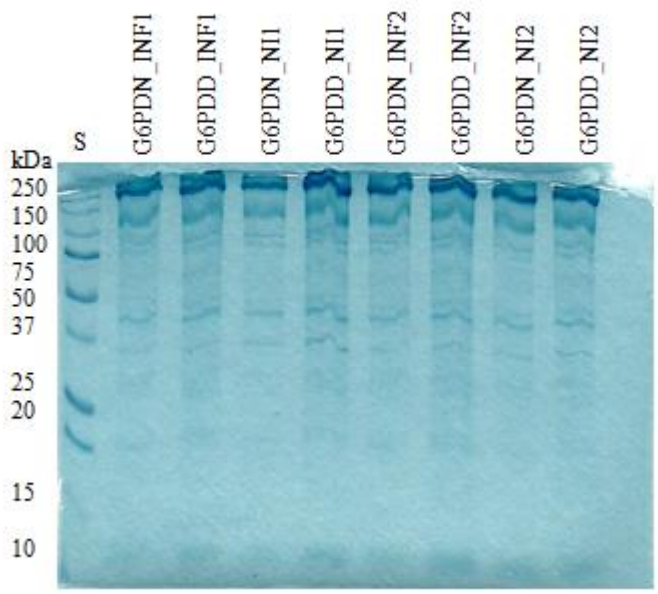


Fig. S5. RBC membrane extracts (5 µg loaded on each well) run in a 12.5% acrylamide:bisacrylamide 37.5:1 gel and stained with Coomassie blue brilliant reagent. G6PDN: extracts from normal RBC; G6PDD: extracts from G6PD-deficient RBC; INF: extracts from infected RBC; NI: extracts from non-infected RBC; 1: replicate 1; 2: replicate 2; S: protein standard (BioRad).

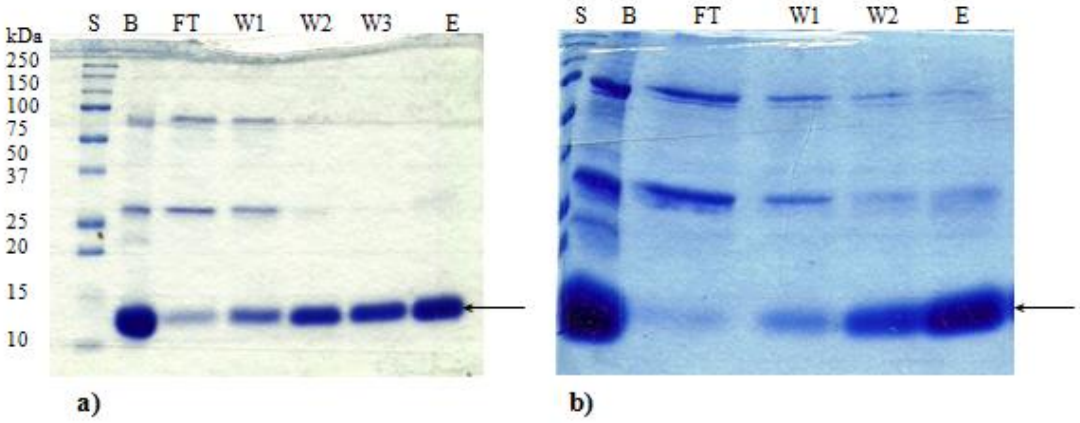


Fig. S6. RBC cytoplasmic extracts prepared with the Ni-NTA resin (Qiagen) for Hb removal. **a)** 5 µg loaded on each well); **b)** 20 µg loaded on each well. B: before the resin use; FT: flow-through fraction (all proteins except Hb expected); W1, W2 and W3: washes 1, 2 and 3 of the column (protein remains expected); E: eluate (only Hb expected); S: protein standard (BioRad); the arrows indicates the Hb monomers band.

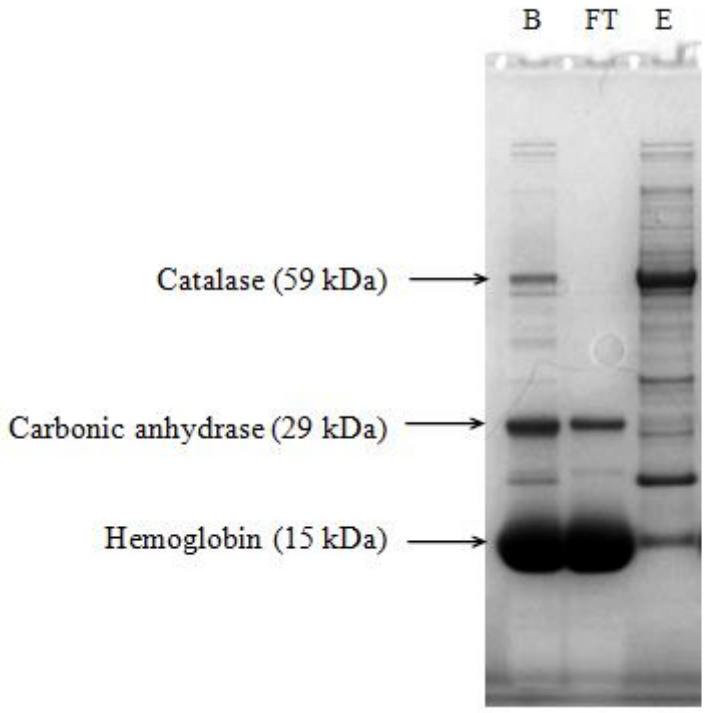


Fig. S7. RBC cytoplasmic extracts prepared with the Hemovoid reagent (Biotech Support Group) for Hb removal (5 µg loaded on each well). B: before the reagent use; FT: flow-through fraction (only Hb expected); E: eluate (all proteins except Hb expected); the arrow indicates the Hb monomers band.

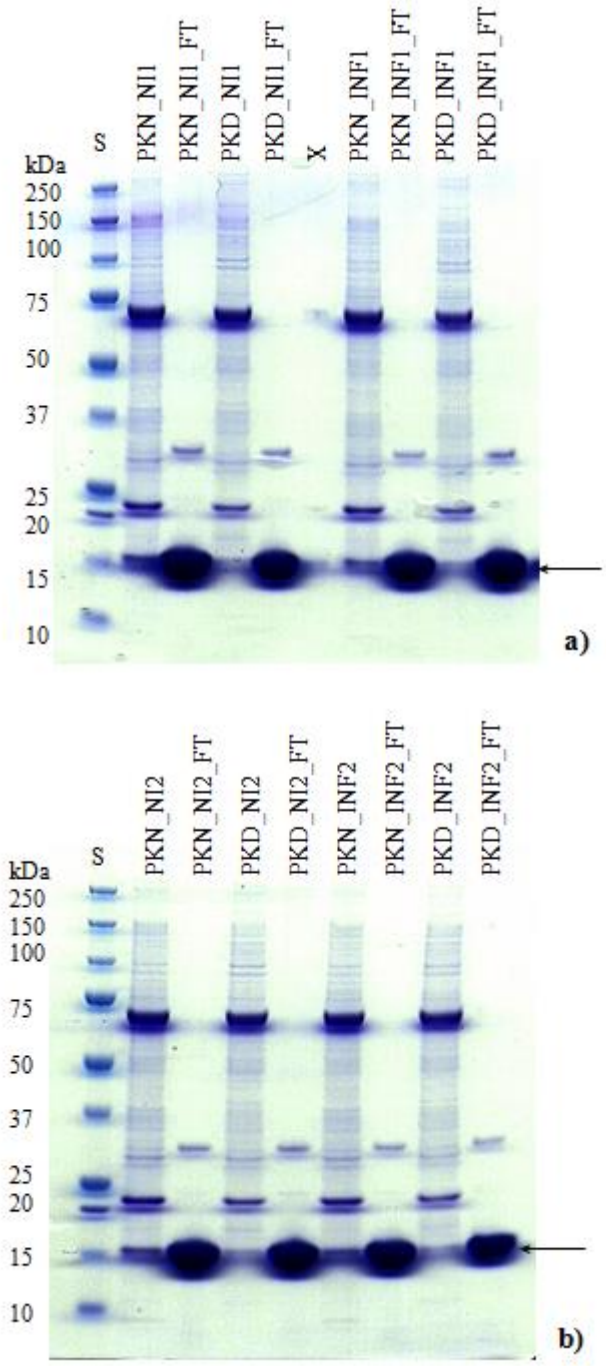


Fig. S8. RBC cytoplasmic extracts [eluates and flow-through (FT)] prepared with the Hemovoid reagent (Biotech Support Group) (5 µg loaded on each well). **a)** Replicate 1; **b)** Replicate 2. PKN: extracts from normal RBC; PKD: extracts from PK-deficient RBC; INF: extracts from infected RBC; NI: extracts from non-infected RBC; S: protein standard (BioRad); X: empty well; the arrow indicates the Hb monomers band.

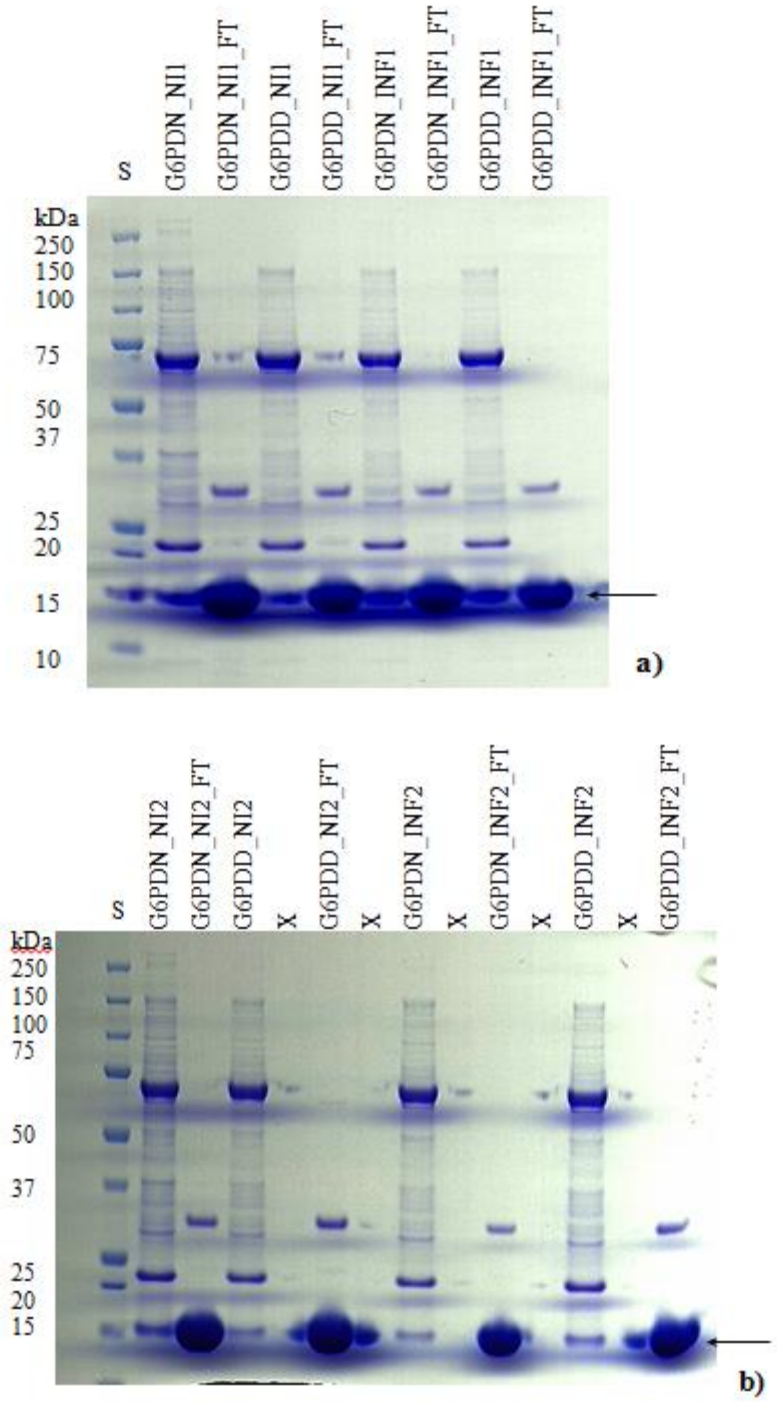


Fig. S9. RBC cytoplasmic extracts [eluates and flow-through (FT)] prepared with the Hemovoid reagent (Biotech Support Group) (5 µg loaded on each well). **a)** Replicate 1; **b)** Replicate 2. G6PDN: extracts from normal RBC; G6PDD: extracts from G6PD-deficient RBC; INF: extracts from infected RBC; NI: extracts from non-infected RBC; S: protein standard (BioRad); X: empty wells; the arrow indicates the Hb monomers band.

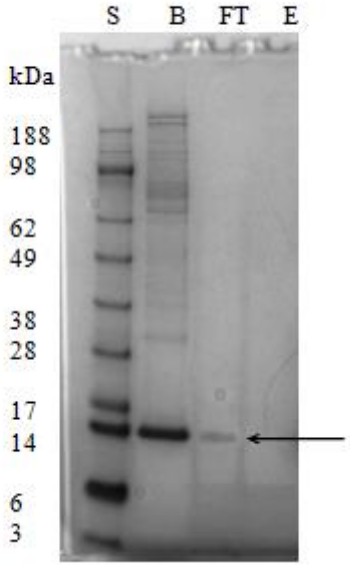


Fig. S10. Parasite extracts prepared with the Hemovoid reagent (Biotech Support Group) for Hb removal [5 µg loaded on well B and < 1 µg loaded on FT and E (the maximum volume was loaded on the wells but samples were low-concentrated]. B- before the reagent use; FT- flow-through fraction (only Hb expected); E- eluate (all proteins except Hb expected); S- protein standard (Invitrogen); the arrow indicates the Hb monomers band.

Supplemental Tables

Table S1. Parasite invasion (ring parasitemia) and maturation (schizont parasitemia) in normal and PK-deficient RBC.

	Time (h)	PKN1	PKD1	PKN2	PKD2	Wilcoxon *p*
Invasion	24	6.90	4.25	8.25	6.00	
(% rings)	72	15.15	4.80	13.65	10.75	0.50
	120	4.40	5.50	3.60	5.85	
Maturation	48	4.85	3.90	4.50	4.35	
(% schizonts)	96	7.70	3.40	7.55	5.85	0.75
	144	1.20	2.75	0.85	3.15	

PKN: normal RBC; PKD: PK-deficient RBC; 1: replicate 1; 2: replicate 2.

Table S2. Parasite invasion (ring parasitemia) and maturation (schizont parasitemia) in normal and G6PD-deficient RBC.

	Time (h)	G6PDN1	G6PDD1	G6PDN2	G6PDD2	Wilcoxon *p*
Invasion	24	3.25	2.30	3.70	2.25	
(% rings)	72	6.00	3.85	15.05	5.70	0.50
	120	2.80	3.70	2.15	1.50	
Maturation	48	2.40	1.80	2.85	1.85	
(% schizonts)	96	2.60	2.95	8.60	2.85	0.50
	144	0.85	1.45	1.45	1.70	

G6PDN: normal RBC; G6PDD: G6PD-deficient RBC; 1: replicate 1; 2: replicate 2.

Table S3. Parasite invasion and maturation ratios in three growth cycles in normal and PK-deficient RBC.

	Cycle	PKN1	PKD1	PKN2	PKD2	Wilcoxon *p*
Invasion	1 (24h/0h)	9.86	6.07	11.79	8.57	
ratios	2 (72h/48h)	3.12	1.23	3.03	2.47	0.50
(R/S)	3 (120h/96h)	0.57	1.62	0.48	1.00	
Maturation	1 (48h/24h)	0.70	0.92	0.55	0.73	
ratios	2 (96h/72h)	0.51	0.71	0.55	0.54	0.25
(S/R)	3 (144h/120h)	0.27	0.50	0.24	0.54	

Invasion ratios: ratios between the percentage of ring-stage parasites (R) at 24h, 72h and 120h and the percentage os schizont-stage parasites (S) at 0h, 48h and 96h, respectively. Maturation ratios: ratios between the percentage of schizont-stage parasite (S) at 48h, 96h and 144h and the percentage of ring-stage parasites (R) at 24h, 72h and 120h, respectively. PKN: normal RBC; PKD: PK-deficient RBC; 1: replicate 1; 2: replicate 2.

Table S4. Parasite invasion and maturation ratios obtained in three cycles in the G6PD assay.

	Cycle	G6PDN1	G6PDD1	G6PDN2	G6PDD2	Wilcoxon *p*
Invasion	1 (24h/0h)	4.64	3.29	5.29	3.21	
Ratios	2 (72h/48h)	2.50	2.14	5.28	3.08	0.50
(R/S)	3 (120h/96h)	1.08	1.25	0.25	0.53	
Maturation	1 (48h/24h)	0.74	0.78	0.77	0.82	
ratios	2 (96h/72h)	0.43	0.77	0.57	0.50	0.25
(S/R)	3 (144h/120h)	0.30	0.39	0.67	1.13	

Invasion ratios: ratios between the percentage of ring-stage parasites (R) at 24h, 72h and 120h and the percentage os schizont-stage parasites (S) at 0h, 48h and 96h, respectively. Maturation ratios: ratios between the percentage of schizont-stage parasite (S) at 48h, 96h and 144h and the percentage of ring-stage parasites (R) at 24h, 72h and 120h, respectively. G6PDN: normal RBC; G6PDD: G6PD-deficient RBC; 1: replicate 1; 2: replicate 2.

Table S5. Parasite extracts quantification.

#	Sample	Av. Conc. (µg/ml)	Volume (ml)	Total (µg)
1	PKN1	467.30	0.50	233.65
2	PKD1	466.60	0.50	233.30
3	PKN2	368.65	0.50	184.33
4	PKD2	458.40	0.50	229.20
5	G6PDN1	504.95	0.50	252.48
6	G6PDD1	408.15	0.50	204.08
7	G6PDN2	436.85	0.50	218.43
8	G6PDD2	393.85	0.50	196.93

PKN and G6PDN: extracts from parasites grown in normal RBC; PKD and G6PDD: extracts from parasites grown in PK or G6PD-deficient RBC, respectively; 1: replicate 1; 2: replicate 2.

Table S6. RBC membrane extracts quantification.

#	Sample	Concentration (µg/ml)	Volume (ml)	Total (µg)
1	PKN_INF1	2188.40	0.50	1094.20
2	PKD_INF1	2778.65	0.50	1389.33
3	PKN_NI1	2151.25	0.50	1075.63
4	PKD_NI1	2743.85	0.50	1371.93
5	PKN_INF2	2528.70	0.50	1264.35
6	PKD_INF2	2526.55	0.50	1263.28
7	PKN_NI2	2415.05	0.50	1207.53
8	PKD_NI2	2244.30	0.50	1122.15
9	G6PDN_INF1	2368.15	0.50	1184.08
10	G6PDD_INF1	2781.55	0.50	1390.78
11	G6PDN_NI1	1639.85	0.50	819.93
12	G6PDD_NI1	2446.50	0.50	1223.25
13	G6PDN_INF2	2209.25	0.50	1104.63
14	G6PDD_INF2	2736.50	0.50	1368.25
15	G6PDN_NI2	2159.55	0.50	1079.78
16	G6PDD_NI2	2044.20	0.50	1022.10

PKN and G6PDN: extracts from normal RBC; PKD and G6PDD: extracts from PK or G6PD-deficient RBC; INF: infected RBC; NI: non-infected RBC; 1: replicate 1; 2: replicate 2.

Table S7. RBC cytoplasm extracts quantification (eluates and flow-through fractions after hemoglobin removal with Hemovoid reagent, Biotech Support Group).

#	Sample	Concentration (µg/ml)	Volume (ml)	Total (µg)
1	PKN_INF1	419.30	0.30	125.79
2	PKD_INF1	341.50	0.30	102.45
3	PKN_NI1	503.00	0.30	150.90
4	PKD_NI1	408.80	0.30	122.64
5	PKN_INF2	377.80	0.30	113.34
6	PKD_INF2	299.40	0.30	89.82
7	PKN_NI2	540.20	0.30	162.06
8	PKD_NI2	369.20	0.30	110.76
9	G6PDN_INF1	329.20	0.30	98.76
10	G6PDD_INF1	369.40	0.30	110.82
11	G6PDN_NI1	554.90	0.30	166.47
12	G6PDD_NI1	352.80	0.30	105.84
13	G6PDN_INF2	423.80	0.30	127.14
14	G6PDD_INF2	287.30	0.30	86.19
15	G6PDN_NI2	554.10	0.30	166.23
16	G6PDD_NI2	318.90	0.30	95.67
1	PKN_INF1_FT	5238.00	0.30	1571.40
2	PKD_INF1_FT	3328.00	0.30	998.40
3	PKN_NI1_FT	5350.00	0.30	1605.00
4	PKD_NI1_FT	4320.00	0.30	1296.00
5	PKN_INF2_FT	5794.00	0.30	1738.20
6	PKD_INF2_FT	5467.00	0.30	1640.10
7	PKN_NI2_FT	6461.00	0.30	1938.30
8	PKD_NI2_FT	4106.00	0.30	1231.80
9	G6PDN_INF1_FT	3446.00	0.30	1033.80
10	G6PDD_INF1_FT	4049.00	0.30	1214.70
11	G6PDN_NI1_FT	4250.00	0.30	1275.00
12	G6PDD_NI1_FT	3795.00	0.30	1138.50
13	G6PDN_INF2_FT	4798.00	0.30	1439.40
14	G6PDD_INF2_FT	4597.00	0.30	1379.10
15	G6PDN_NI2_FT	5467.00	0.30	1640.10
16	G6PDD_NI2_FT	3638.00	0.30	1091.40

PKN and G6PDN: extracts from normal RBC; PKD and G6PDD: extracts from PK or G6PD-deficient RBC, respectively; INF: infected RBC; NI: non-infected RBC; 1: replicate 1; 2: replicate 2; FT: flow-through fraction.

Table S8. Parasite extracts quantification (eluate and flow-through fractions after hemoglobin removal with Hemovoid reagent, Biotech Support Group).

#	Sample	**Concentration (µg/ml)**	**Volume (ml)**	**Total (µg)**
1	G6PDN1	130.90	0.05	6.545
2	G6PDD1	56.40	0.05	2.82
3	G6PDN2	124.30	0.05	6.215
4	G6PDD2	63.70	0.05	3.19
1	G6PDN1_FT	101.50	0.40	40.60
2	G6PDD1_FT	159.60	0.50	79.80
3	G6PDN2_FT	165.20	0.35	57.82
4	G6PDD2_FT	53.70	0.40	21.48

G6PDN: extracts from parasites grown in normal RBC; G6PDD: extracts from parasites grown in G6PD-deficient RBC; 1: replicate 1; 2: replicate 2; FT: flow-through fraction.

Table S9. MS qualitative results: identified proteins from *P. falciparum* 3D7 grown in normal and PK-deficient RBC.

			Score				Peptides				SC [%]			
#	**Accession**	**Protein**	**PN1**	**PN2**	**PD1**	**PD2**	**PN1**	**PN2**	**PD1**	**PD2**	**PN1**	**PN2**	**PD1**	**PD2**
1	PFE0965c	vacuolar ATP synthetase	47.01	27.81	38.46	38.83	1	2	1	1	10.91	10.91	10.91	10.91
2	PF14_0296	60S ribosomal protein L14, putative	29.74	0	0	0	1	0	0	0	7.879	0	0	0
3	PF11_0043	60S ribosomal protein P1, putative	46.84	28.03	30.26	32.78	1	1	1	1	8.475	8.475	8.475	8.475
4	PF10_0372	antigen UB05	37.59	28.55	0	26.42	2	3	0	1	5.882	5.882	0	5.882
5	PFE0625w	Rab1b, GTPase	0	29.56	54.29	69.42	0	1	2	2	0	5.5	11	11
6	PF13_0346	60S ribosomal protein L40/UBI, putative	0	90.46	63.3	0	0	3	1	0	0	26.56	7.031	0
7	PFL0185c	nucleosome assembly protein 1, putative	21.81	0	0	0	1	0	0	0	2.594	0	0	0
8	PF14_0083	40S ribosomal protein S8e, putative	33.27	0	0	0	1	0	0	0	5.046	0	0	0
9	PFF0510w	histone H3	20.76	29.7	0	50.8	1	1	0	1	5.147	5.147	0	5.147
10	PFF1025c	pyridoxine/pyridoxal 5-phosphate biosynthesis enzyme	56.559	0	0	0	2	0	0	0	11.3	0	0	0
11	**PFI1475w**	**merozoite surface protein 1 precursor**	**271.6**	**353.11**	**0**	**0**	**9**	**11**	**0**	**0**	**7.151**	**8.663**	**0**	**0**
12	MAL8P1.72	high mobility group protein	0	57.01	0	0	0	1	0	0	0	14.14	0	0
13	PFF0860c	histone h2a	21.89	0	0	41.72	1	0	0	1	6.818	0	0	6.818
14	PFE0865c	splicing factor, putative	47.15	25.97	0	0	2	1	0	0	7.047	4.362	0	0
15	PF08_0074	DNA/RNA-binding protein Alba, putative	70.1	66.66	0	0	3	4	0	0	14.11	17.74	0	0
16	MAL13P1.288	conserved Plasmodium protein, unknown function	0	39.21	0	0	0	2	0	0	0	10.46	0	0
17	PFI1740c-a	ring-exported protein 2, REX2	27.63	65.01	0	0	1	2	0	0	11.7	11.7	0	0
18	PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	354.03	223.9	145.46	80.68	9	7	5	2	37.39	29.38	18.4	10.68
19	PFI1735c	ring-exported protein 1	0	21.89	0	0	0	1	0	0	0	0.982	0	0
20	MAL13P1.231	Sec61 alpha subunit, PfSec61	0	64.3	0	0	0	2	0	0	0	2.542	0	0
21	PF13_0011	plasmodium falciparum gamete antigen 27/25	0	52.96	0	0	0	1	0	0	0	5.991	0	0
22	PF14_0361	Sec62, putative	0	41.22	0	0	0	2	0	0	0	5.57	0	0
23	PFA0110w	DNAJ protein, putative	0	20.76	0	0	0	1	0	0	0	0.922	0	0

24	PF11_0065	40S ribosomal protein S4, putative	54.81	0	0	0	2	0	0	0	11.88	0	0	0
25	PFL0590c	non-SERCA-type Ca2+ -transporting P-ATPase	107.06	137.47	0	0	2	4	0	0	2.07	4.305	0	0
26	**PFE0080c**	**rhoptry-associated protein 2, RAP2**	**291.89**	**542.79**	**30.94**	**0**	**8**	**12**	**1**	**0**	**22.61**	**35.43**	**4.774**	**0**
27	PFF1300w	pyruvate kinase	28.96	0	0	0	1	0	0	0	2.348	0	0	0
28	PFE0660c	purine nucleotide phosphorylase, putative	89.38	0	0	0	2	0	0	0	14.29	0	0	0
29	PFI0720w	transporter, putative	20.51	167.38	0	0	1	6	0	0	3.295	13.76	0	0
30	PF14_0486	elongation factor 2	161.94	109.54	0	0	2	2	0	0	5.288	5.288	0	0
31	PFF0290w	long chain polyunsaturated fatty acid elongation enzyme, putative	39.57	24.92	0	0	2	1	0	0	6.143	3.413	0	0
32	PF14_0678	exported protein 2	31.83	120.44	0	44.64	1	3	0	3	6.272	13.24	0	6.272
33	PF13_0143	phosphoribosylpyrophosphate synthetase	41.56	77.15	0	0	2	3	0	0	7.094	9.382	0	0
34	PF14_0016	early transcribed membrane protein 14.1, etramp14.1	53.17	31.39	0	23.58	1	2	0	1	12.15	12.15	0	12.15
35	PFE0850c	60S ribosomal protein L12, putative	0	48.45	0	0	0	2	0	0	0	14.55	0	0
36	PF14_0425	fructose-bisphosphate aldolase	123.48	169.87	0	0	4	7	0	0	16.8	24.93	0	0
37	PF14_0344	conserved Plasmodium protein, unknown function	22.96	0	0	0	1	0	0	0	2.216	0	0	0
38	PF14_0548	ATPase, putative	22.44	26.73	0	32.28	1	1	0	1	2.864	2.864	0	2.864
39	PF10_0063	DNA/RNA-binding protein, putative	51.23	0	27.87	0	1	0	1	0	17.76	0	22.43	0
40	PF11_0351	heat shock protein hsp70 homologue	229.06	262.93	69.69	0	11	7	3	0	19.31	15.69	6.335	0
41	PF14_0359	HSP40, subfamily A, putative	0	104.76	0	0	0	1	0	0	0	4.717	0	0
42	PF14_0377	vesicle-associated membrane protein, putative	31.26	47.34	0	0	1	1	0	0	4.979	4.979	0	0
43	PFL1545c	chaperonin, cpn60	67.88	216.33	0	0	3	7	0	0	7.382	11.98	0	0
44	PF10_0025	PF70 protein	0	20.22	0	0	0	1	0	0	0	2.377	0	0
45	PFF1375c	ethanolaminephosphotransferase, putative	20.37	72.55	0	0	1	1	0	0	5.115	5.115	0	0
46	PF10_0019	early transcribed membrane protein 10.1, etramp 10.1	58.16	55.83	25.06	43.84	3	4	1	3	11.21	11.21	11.21	11.21
47	PF11_0302	conserved Plasmodium protein, unknown function	111.86	110.98	0	0	4	4	0	0	7.08	7.08	0	0
48	PFI1445w	high molecular weight rhoptry protein-2	46.179	51.78	0	0	2	2	0	0	2.395	1.742	0	0
49	MAL13P1.221	aspartate carbamoyltransferase	0	39.91	0	0	0	1	0	0	0	3.2	0	0
50	PF11_0224	circumsporozoite-related antigen	94.19	102.45	28.49	0	3	5	1	0	17.9	17.9	11.11	0

51	PF14_0368	thioredoxin peroxidase 1	125.07	0	0	0	3	0	0	0	25.64	0	0	0
52	PF14_0567	conserved Plasmodium protein, unknown function	65.35	0	0	0	2	0	0	0	7.941	0	0	0
53	PFI0935w	DNAJ-like molecular chaperone protein, putative	0	65.64	0	0	0	2	0	0	0	9.189	0	0
54	PFD1035w	steroid dehydrogenase, putative	0	33.8	0	21.75	0	2	0	1	0	3.738	0	3.738
55	**PFE1150w**	**multidrug resistance protein**	**605.58**	**775.85**	**141.52**	**36.9**	**21**	**23**	**3**	**2**	**15.86**	**17.97**	**3.312**	**2.326**
56	PF10_0268	merozoite capping protein 1	182.22	121.16	21.26	0	8	3	1	0	15.52	7.125	3.817	0
57	PFC0725c	formate-nitrate transporter, putative	61.15	46.49	0	0	2	2	0	0	6.472	6.472	0	0
58	PF08_0054	heat shock 70 kDa protein	320.63	249.61	105.25	122.5	12	12	3	5	21.12	16.25	5.908	9.897
59	PF14_0077	plasmepsin II	145.13	215.08	0	0	6	7	0	0	11.48	20.09	0	0
60	PFI0605c	conserved Plasmodium protein, unknown function	0	35.26	0	0	0	1	0	0	0	2.018	0	0
61	PFI0880c	glideosome-associated protein 50	0	57.26	0	0	0	2	0	0	0	5.303	0	0
62	**PFL1725w**	**ATP synthase beta chain, mitochondrial precursor, putative**	**0**	**0**	**261.28**	**231.11**	**0**	**0**	**12**	**13**	**0**	**0**	**9.72**	**9.72**
63	PFL1825w	conserved Plasmodium membrane protein, unknown function	56.1	132.41	0	0	3	3	0	0	11.43	11.43	0	0
64	PFI0755c	6-phosphofructokinase, putative	21.45	0	0	0	1	0	0	0	1.058	0	0	0
65	MAL7P1.67	conserved Plasmodium protein, unknown function	62.03	127.15	0	0	2	4	0	0	17.22	29.67	0	0
66	PFI0875w	Heat shock protein 70 (HSP70) homologue	1088.2	1285.3	601.91	621.73	42	50	20	23	42.02	48.77	31.75	31.75
67	PF07_0033	Cg4 protein	184.55	0	0	0	3	0	0	0	6.415	0	0	0
68	MAL13P1.233	nucleic acid binding protein, putative	65.97	57.29	27.03	0	1	2	1	0	6.161	14.69	6.161	0
69	PF13_0214	elongation factor 1-gamma, putative	20.77	0	0	0	1	0	0	0	2.676	0	0	0
70	PF14_0301	conserved protein, unknown function	0	0	31.59	0	0	0	1	0	0	0	3.806	0
71	MAL8P1.62	conserved Plasmodium protein, unknown function	73.64	0	0	0	3	0	0	0	15.88	0	0	0
72	PF11_0338	aquaglyceroporin	128.21	120.63	49.35	58.46	5	7	2	2	14.34	14.34	3.876	7.364
73	PF11_0164	peptidyl-prolyl cis-trans isomerase	133.5	78.76	0	31.03	5	3	0	1	27.69	13.85	0	8.718
74	PF07_0054	histone H2B	96.53	73.72	0	38.57	2	4	0	2	12.2	12.2	0	11.38
75	MAL13P1.56	m1-family aminopeptidase	235.96	143.11	73.87	0	9	7	4	0	11.89	7.558	4.516	0
76	PFI0265c	RhopH3	23.85	136.37	0	0	1	4	0	0	1.226	5.797	0	0
77	PFE0065w	skeleton-binding protein 1	0	0	0	20.62	0	0	0	1	0	0	0	2.671

78	PF11_0055	conserved protein, unknown function	134.55	160.83	0	0	5	6	0	0	14.15	13.44	0	0
79	**PF10_0086**	**adenylate kinase**	**217.7**	**248.91**	**0**	**0**	**6**	**9**	**0**	**0**	**30.58**	**45.04**	**0**	**0**
80	PF11_0179	conserved Plasmodium protein, unknown function	81.26	188.04	22.37	0	3	5	1	0	25	27.34	7.031	0
81	**MAL13P1.540**	**heat shock protein 70 (hsp70), putative**	**240.76**	**154.68**	**0**	**0**	**10**	**5**	**0**	**0**	**10.3**	**6.009**	**0**	**0**
82	PF14_0517	peptidase, putative	26.94	142.54	0	0	1	3	0	0	1.44	6.021	0	0
83	PF14_0201	surface protein, Pf113	131.22	167.83	0	28.91	7	7	0	1	7.74	7.637	0	1.032
84	PF10_0366	ADP/ATP transporter on adenylate translocase	0	91.75	0	0	0	4	0	0	0	14.29	0	0
85	PFE1590w	early transcribed membrane protein 5, ETRAMP5	49.78	90.06	21.04	29.76	2	4	1	2	7.182	20.44	7.182	7.182
86	PF11_0069	conserved Plasmodium protein, unknown function	134.47	125.91	40.23	53.29	3	4	1	1	10.53	11.28	4.511	4.511
87	PF11_0301	spermidine synthase	113.67	190.14	0	31.81	5	6	0	1	14.02	22.43	0	3.115
88	PF11_0175	heat shock protein 101, putative	56.69	124.26	0	0	3	4	0	0	4.305	4.857	0	0
89	PFB0210c	hexose transporter, PfHT1	61.44	58.73	0	27.8	1	1	0	1	2.976	2.976	0	2.976
90	PF14_0078	HAP protein	348.03	451.74	174.36	223.14	10	13	6	8	19.07	27.94	17.74	21.95
91	PF11_0313	60S ribosomal protein P0	167.58	119.16	50.35	0	4	6	2	0	19.3	16.46	9.81	0
92	PF14_0230	60S ribosomal protein L5, putative	0	28.72	0	0	0	1	0	0	0	2.721	0	0
93	PF11_0352	protein disulfide isomerase	49.21	0	0	0	2	0	0	0	8.511	0	0	0
94	PF13_0141	L-lactate dehydrogenase	236.64	330.25	95.07	0	5	9	3	0	22.15	39.56	16.46	0
95	PF13_0102	DnaJ/SEC63 protein, putative	21.8	0	0	0	1	0	0	0	2.458	0	0	0
96	PFD0310w	sexual stage-specific protein precursor	227.17	199.61	227.25	179.94	10	12	11	8	33.12	33.12	33.12	38.85
97	PF14_0075	plasmepsin IV	419.77	461.56	177.64	180.81	14	12	7	7	30.07	30.96	23.16	16.93
98	PFC0400w	60S Acidic ribosomal protein P2, putative	223.94	292.08	195.86	86.0895	5	5	4	3	59.82	52.68	59.82	42.86
99	PFF0940c	cell division cycle protein 48 homologue, putative	163.17	127.26	0	0	6	6	0	0	9.058	8.454	0	0
100	MAL8P1.95	conserved Plasmodium protein, unknown function	43.06	62.25	0	0	2	2	0	0	9.524	7.619	0	0
101	PF13_0272	thioredoxin-related protein, putative	356.09	273.6	70.53	78.49	12	13	3	2	35.58	31.73	10.1	8.654
102	PFI1270w	conserved Plasmodium protein, unknown function	404.14	248.55	217.39	239.9	23	14	8	11	47	26.73	35.94	43.78
103	PF11_0062	histone H2B	61.33	137.6	55.25	39.88	3	5	1	2	12.82	31.62	12.82	12.82
104	PF14_0076	plasmepsin I	647.24	590.84	454.55	292.739	20	24	14	9	35.84	34.51	35.84	22.79

105	PFL1070c	endoplasmin homolog precursor, putative	637.68	638.74	360.83	390.66	27	20	16	13	21.56	24.12	14.98	15.96
106	PFI0930c	nucleosome assembly protein	102.52	52.61	0	0	5	2	0	0	19.33	12.27	0	0
107	PF11_0208	phosphoglycerate mutase, putative	80.85	30.54	0	23.22	2	1	0	1	17.2	13.2	0	13.2
108	PF14_0102	rhoptry-associated protein 1, RAP1	468.58	472.94	94.23	104.26	13	16	3	4	19.57	21.87	5.371	5.627
109	MAL8P1.17	protein disulfide isomerase	572.63	745.53	396.06	206.27	21	24	10	6	45.55	51.35	30.85	18.22
110	PF11_0099	heat shock protein DnaJ homologue Pfj2	80.31	40.68	0	0	1	1	0	0	3.889	3.889	0	0
111	PF10_0153	heat shock protein 60	96.26	168.59	0	0	2	4	0	0	4.31	10	0	0
112	PF10_0121	hypoxanthine phosphoribosyltransferase	79.03	24.7	0	0	3	1	0	0	23.81	5.628	0	0
113	PF11_0174	cathepsin C, homolog	63.78	47.19	0	0	2	1	0	0	8	1.286	0	0
114	PF14_0046	conserved Plasmodium protein, unknown function	67.27	0	0	0	2	0	0	0	8.754	0	0	0
115	PFE0585c	myo-inositol 1-phosphate synthase, putative	57.01	0	0	0	1	0	0	0	2.318	0	0	0
116	PF11_0061	histone H4	63.66	40.43	36.69	40.43	2	2	2	2	23.3	19.42	19.42	17.48
117	PF14_0164	NADP-specific glutamate dehydrogenase	41.63	0	0	0	2	0	0	0	6.596	0	0	0
118	PF10_0068	RNA binding protein, putative	44.44	0	0	0	1	0	0	0	6.911	0	0	0
119	PF14_0159	root hair defective 3 GTP-binding protein (RHD3) homolog, putative	34.66	0	0	0	1	0	0	0	2.134	0	0	0
120	PF13_0252	nucleoside transporter 1	38.18	22.86	0	0	2	1	0	0	3.791	1.896	0	0
121	PF14_0391	60S ribosomal protein L1, putative	30.45	0	0	0	1	0	0	0	7.834	0	0	0
122	PFL0795c	male development gene 1	33.8	0	26.6	0	1	0	1	0	8.597	0	8.597	0
123	PF11_0161	falcipain-2B	29.78	0	0	0	1	0	0	0	2.905	0	0	0
124	PFL1880w	acyl-CoA synthetase, PfACS11	26.6	0	0	0	1	0	0	0	2.399	0	0	0
125	PFE0810c	40S ribosomal protein S14, putative	27.37	22.65	0	0	1	1	0	0	8.609	8.609	0	0
126	PFF1350c	acetyl-CoA synthetase	24.19	0	0	0	1	0	0	0	2.006	0	0	0
127	PFC0975c	peptidyl-prolyl cis-trans isomerase	25.53	0	0	0	1	0	0	0	7.018	0	0	0
128	PFC0920w	histone H2A variant, putative	0	30.86	0	0	0	1	0	0	0	6.329	0	0
129	PF10_0328	bromodomain protein, putative	23.49	0	0	0	1	0	0	0	4.303	0	0	0
130	PF14_0231	60S ribosomal protein L7-3, putative	22.005	52.865	0	0	1	2	0	0	3.887	6.714	0	0
131	PF14_0541	V-type H(+)-translocating pyrophosphatase, putative	269.32	323.77	0	154.35	11	9	0	4	13.11	15.9	0	7.531

132	PF11_0280	small nuclear ribonucleoprotein F, putative	0	34.47	0	0	0	1	0	0	0	10.47	0	0
133	PF11_0272	40S ribosomal protein S18, putative	0	30.17	0	0	0	2	0	0	0	5.128	0	0
134	PF13_0197	merozoite Surface Protein 7 precursor, MSP7	0	42.45	0	0	0	3	0	0	0	6.268	0	0
135	PF14_0439	M17 leucyl aminopeptidase	0	23.89	0	0	0	1	0	0	0	2.645	0	0
136	PF13_0276	membrane-associated histidine rich protein 2, (MARHP2)	0	22.08	0	0	0	1	0	0	0	12.41	0	0
137	MAL7P1.27	chloroquine resistance transporter	0	29.59	0	0	0	1	0	0	0	1.887	0	0
138	PF11_0096	casein kinase II, alpha subunit	0	26.42	0	0	0	1	0	0	0	3.582	0	0
139	PF14_0494	ribosome biogenesis protein tsr1, putative	0	20.68	0	0	0	1	0	0	0	0.486	0	0
140	PFC0730w	HVA22/TB2/DP1 family protein, putative	0	21.12	0	0	0	1	0	0	0	4.525	0	0
141	PFI0695c	phospholipid or glycerol acyltransferase, putative	0	21.14	0	0	0	1	0	0	0	1.435	0	0
142	MAL7P1.228	Heat Shock 70 KDa Protein, (HSP70)	0	0	0	143.71	0	0	0	5	0	0	0	6.808
143	PF13_0242	isocitrate dehydrogenase (NADP), mitochondrial precursor	0	0	34.11	0	0	0	2	0	0	0	2.35	0
144	MAL8P1.69	14-3-3 protein, putative	117.12	122.65	63.57	22.61	6	2	3	1	17.18	13.36	9.924	6.107
145	PFL0930w	clathrin heavy chain, putative	0	0	37.19	0	0	0	1	0	0	0	0.601	0
146	PF14_0630	protein serine/threonine phosphatase	0	0	20.23	20.73	0	0	2	1	0	0	0.787	0.787
147	MAL13P1.224	conserved Plasmodium protein, unknown function	0	0	0	26.35	0	0	0	1	0	0	0	2.679
148	PFD1070w	eukaryotic initiation factor, putative	0	0	0	22.86	0	0	0	1	0	0	0	4.103
149	PFL1465c	heat shock protein hslv	0	0	0	20.92	0	0	0	1	0	0	0	2.415
150	MAL7P1.29	conserved Plasmodium membrane protein, unknown function	0	0	27.86	0	0	0	1	0	0	0	0.346	0
151	PF07_0029	heat shock protein 86	244.95	137.73	151.05	92.65	8	5	5	5	13.83	10.2	8.054	5.235
152	PFD0860w	conserved Plasmodium protein, unknown function	0	0	22.34	0	0	0	1	0	0	0	1.365	0
153	PFE0290c	conserved Plasmodium protein, unknown function	0	0	20.65	0	0	0	1	0	0	0	8.73	0
154	PFC0715c	conserved Plasmodium protein, unknown function	0	0	20.56	0	0	0	2	0	0	0	0.496	0
155	PF13_0304	elongation factor-1 alpha	279.5	220.33	143.73	91.47	8	9	6	3	27.31	26.41	18.96	5.192
156	PFB0405w	transmission-blocking target antigen s230	36.98	0	0	0	1	0	0	0	0.67	0	0	0
157	PF10_0155	enolase	109.27	125.38	75.98	20.49	3	7	4	1	16.37	17.94	19.51	4.036
158	PF07_0112	proteasome subunit alpha type 5, putative	0	22.94	0	0	0	1	0	0	0	6.25	0	0

159	PF14_0323	calmodulin	0	0	58.35	0	0	0	1	0	0	0	11.41	0
160	PFL1385c	merozoite Surface Protein 9, MSP-9	35.21	0	0	0	1	0	0	0	2.423	0	0	0
161	PF10_0100	conserved Plasmodium protein, unknown function	29.08	0	0	0	1	0	0	0	9.322	0	0	0
	TOTAL NUMBER OF COMPOUNDS		**115**	**113**	**49**	**48**								

Accession: gene accession number; Protein: protein name; Score: Protein Mascot score (reflecting the combined scores of all observed mass spectra that can be matched to amino acid sequences within that protein; a higher score indicates a more confident match); Peptides: number of peptides identified; SC [%]: sequence coverage. PN: parasites grown in normal RBC; PD: parasites grown in PK-deficient RBC; 1: replicate 1; 2: replicate 2. In bold, the proteins with higher differences in the number of detected peptides (> or equal to 15) between parasites growing in normal and PK-deficient RBC.

Table S10. MS qualitative results: identified proteins from *P. falciparum* 3D7 grown in normal and G6PD-deficient RBC.

			Score				Peptides				SC [%]			
#	**Accession**	**Protein**	**GN1**	**GN2**	**GD1**	**GD2**	**GN1**	**GN2**	**GD1**	**GD2**	**GN1**	**GN2**	**GD1**	**GD2**
1	PFE0965c	vacuolar ATP synthetase	0	0	0	36.48	0	0	0	1	0	0	0	10.91
2	PF14_0543	signal peptide peptidase	0	0	0	35.39	0	0	0	1	0	0	0	2.427
3	PF14_0296	60S ribosomal protein L14, putative	0	0	0	33.68	0	0	0	1	0	0	0	7.879
4	PF10_0187	60S ribosomal protein L30e, putative	0	0	0	31.63	0	0	0	1	0	0	0	12.04
5	PF11_0043	60S ribosomal protein P1, putative	0	21.73	0	31.06	0	1	0	1	0	8.47	0	26.27
6	PFE0050w	Plasmodium exported protein, unknown function	0	0	0	30.68	0	0	0	1	0	0	0	4.231
7	PF14_0448	40S ribosomal protein S2, putative	0	0	0	28.37	0	0	0	1	0	0	0	5.515
8	PF10_0372	Antigen UB05	0	0	0	26.89	0	0	0	3	0	0	0	5.882
9	PFE0625w	Rab1b, GTPase	0	0	0	47.03	0	0	0	2	0	0	0	11
10	PF13_0014	40S ribosomal protein S7, putative	0	0	0	101.5	0	0	0	3	0	0	0	17.01
11	PF13_0133	plasmepsin V	0	0	0	23.55	0	0	0	1	0	0	0	2.881
12	PF10_0203	ADP-ribosylation factor	68.58	0	0	22.5	2	0	0	1	19.34	0	0	7.735
13	PF08_0076	40S ribosomal protein S16, putative	0	0	0	21.74	0	0	0	1	0	0	0	8.333
14	PF11_0258	co-chaperone GrpE, putative	0	0	0	21.45	0	0	0	1	0	0	0	4.651
15	MAL7P1.38	regulator of chromosome condensation, putative	0	0	0	21.28	0	0	0	1	0	0	0	2.038
16	PF08_0091	conserved Plasmodium protein, unknown function	0	0	0	21.1	0	0	0	1	0	0	0	1.322
17	PFL1170w	polyadenylate-binding protein, putative	0	0	0	20.92	0	0	0	1	0	0	0	1.257
18	PFL0185c	nucleosome assembly protein 1, putative	40.88	0	0	0	2	0	0	0	6.052	0	0	0
19	PFE0075c	rhoptry-associated protein 3, RAP3	43.73	91.1	29.29	35.72	2	3	1	2	8.5	11.3	4.25	7
20	PFL1500w	Rab2, GTPase	0	0	0	42.73	0	0	0	1	0	0	0	6.103
21	PF14_0083	40S ribosomal protein S8e, putative	0	0	0	73.28	0	0	0	2	0	0	0	11.93
22	PFE0785c	metabolite/drug transporter, putative	28.38	0	0	0	1	0	0	0	2.412	0	0	0
23	PFI1475w	merozoite surface protein 1 precursor	676	583.5	604.4	273.8	21	18	24	10	12.91	12.8	13.6	7.965

24	PFF1025c	pyridoxine/pyridoxal 5-phosphate biosynthesis enzyme	0	0	0	52.21	0	0	0	2	0	0	0	11.96
25	PFL2405c	PFG377 protein	0	0	0	21.7	0	0	0	1	0	0	0	0.256
26	PFE1600w	Plasmodium exported protein (PHISTb), unknown function	0	0	0	48.94	0	0	0	2	0	0	0	2.947
27	MAL8P1.72	high mobility group protein	0	0	0	78.87	0	0	0	1	0	0	0	14.14
28	PFD0080c	Plasmodium exported protein (PHISTb), unknown function	0	0	0	95.26	0	0	0	2	0	0	0	4.107
29	PFF0160c	dihydroorotate dehydrogenase, mitochondrial precursor	0	0	0	51.85	0	0	0	1	0	0	0	1.582
30	PFL0740c	10 kd chaperonin	29.27	21.6	0	53.24	1	1	0	2	15.53	15.5	0	23.3
31	PF13_0076	Plasmodium exported protein, unknown function	0	0	28.14	0	0	0	1	0	0	0	7.051	0
32	PFE0865c	splicing factor, putative	0	0	25.63	0	0	0	1	0	0	0	4.362	0
33	PF10_0159	glycophorin-binding protein 130 precursor	27.6	0	28.31	30.13	1	0	1	1	2.306	0	2.306	2.306
34	PF08_0074	DNA/RNA-binding protein Alba, putative	51.25	0	28.19	94.03	1	0	1	2	9.677	0	4.435	14.11
35	MAL13P1.413	membrane associated histidine-rich protein, MAHRP-1	0	0	23.64	0	0	0	1	0	0	0	8.835	0
36	MAL13P1.288	conserved Plasmodium protein, unknown function	0	0	23.37	0	0	0	1	0	0	0	5.229	0
37	PFI0820c	RNA binding protein, putative	0	0	24.56	102	0	0	1	1	0	0	4.639	4.639
38	PFI1740c-a	ring-exported protein 2, REX2	55.84	0	23.67	0	2	0	1	0	23.4	0	11.7	0
39	PFE1195w	karyopherin beta	47.38	0	21.13	27.58	1	0	1	1	1.781	0	2.048	1.425
40	PF14_0598	glyceraldehyde-3-phosphate dehydrogenase	599.9	447.7	573.4	555	13	12	17	13	50.45	50.4	49.26	50.74
41	PFI1735c	ring-exported protein 1	21.07	0	20.87	0	1	0	1	0	1.262	0	1.964	0
42	PFE1155c	mitochondrial processing peptidase alpha subunit, putative	0	0	23.33	0	0	0	1	0	0	0	2.06	0
43	PFB0685c	acyl-CoA synthetase, PfACS9	0	0	23.13	0	0	0	1	0	0	0	1.13	0
44	MAL13P1.231	Sec61 alpha subunit, PfSec61	0	0	20.67	0	0	0	1	0	0	0	1.907	0
45	PF13_0065	vacuolar ATP synthase, catalytic subunit a	111.6	0	0	23.9	6	0	0	1	12.27	0	0	2.782
46	PF13_0011	plasmodium falciparum gamete antigen 27/25	49.28	0	20.24	70.24	2	0	1	2	18.89	0	4.147	18.89
47	PF11_0250	high mobility group-like protein NHP2, putative	0	0	0	61.38	0	0	0	1	0	0	0	13.1
48	PF13_0247	transmission blocking target antigen precursor	0	0	34.98	0	0	0	1	0	0	0	2.232	0
49	PF14_0361	Sec62, putative	0	0	35.3	0	0	0	1	0	0	0	4.509	0
50	PFC0290w	40S ribosomal protein S23, putative	0	0	0	29.81	0	0	0	1	0	0	0	18.62

51	PFA0110w	DNAJ protein, putative	0	0	35.85	193.7	0	0	2	6	0	0	2.857	7.281
52	PF11_0331	TCP-1/cpn60 chaperonin family	0	0	0	47.44	0	0	0	1	0	0	0	3.125
53	PF11_0065	40S ribosomal protein S4, putative	0	0	37.61	96.03	0	0	1	3	0	0	4.598	15.33
54	PF13_0322	falcilysin	68.07	0	32.28	0	2	0	1	0	3.185	0	0.754	0
55	PFE1370w	hsp70 interacting protein, putative	0	0	0	26.69	0	0	0	1	0	0	0	4.803
56	PF10_0323	early transcribed membrane protein 10.2, etramp 10.2	0	21.68	33.7	41.62	0	1	1	1	0	7.61	7.606	7.606
57	PFI0155c	PfRab7, GTPase	0	0	0	26.94	0	0	0	1	0	0	0	6.796
58	PFL0590c	non-SERCA-type Ca2+ -transporting P-ATPase	79.35	26.84	34.54	141.9	2	1	1	3	2.897	1.41	1.407	5.05
59	PFF1300w	pyruvate kinase	0	0	34.83	0	0	0	1	0	0	0	4.892	0
60	PF11_0461	PfRab6, GTPase	0	0	29.49	0	0	0	1	0	0	0	5.314	0
61	PFE0395c	6-cysteine protein, putative	40.76	0	0	0	1	0	0	0	7.163	0	0	0
62	PFE0660c	purine nucleotide phosphorylase, putative	36.16	23.78	30.68	24.21	1	1	1	1	9.796	8.16	9.796	9.796
63	PFC0900w	T-complex protein 1 epsilon subunit, putative	0	0	31.11	0	0	0	1	0	0	0	2.617	0
64	PFI0720w	transporter, putative	71.78	24.83	32.23	29.24	2	1	1	1	5.814	2.52	2.519	3.295
65	PF14_0105	conserved Plasmodium protein, unknown function	81.15	0	28.58	0	1	0	1	0	5.689	0	5.689	0
66	PFI1670c	vacuolar ATP synthase subunit E, putative	29.31	0	28.74	0	1	0	1	0	5.532	0	5.532	0
67	PF14_0744	Plasmodium exported protein, unknown function	0	0	28.98	23.61	0	0	1	1	0	0	6.429	6.429
68	PF14_0421	apicoplast 1-acyl-sn-glycerol-3-phosphate acyltransferase, putative	0	0	29.26	0	0	0	1	0	0	0	6.507	0
69	PF14_0486	elongation factor 2	51.46	40.22	51.73	93.75	1	1	1	3	2.163	2.16	2.163	5.409
70	PFF0290w	long chain polyunsaturated fatty acid elongation enzyme, putative	26.1	36.02	49.16	36.89	1	1	2	1	3.413	6.14	9.556	6.143
71	PF14_0678	exported protein 2	0	123	47.8	43.49	0	2	2	1	0	15	6.62	11.85
72	PF13_0143	phosphoribosylpyrophosphate synthetase	116.4	48.36	47.62	121.9	4	2	2	3	13.96	7.09	7.094	9.84
73	PF14_0016	early transcribed membrane protein 14.1, etramp14.1	27.35	24.98	53.86	42.91	1	1	2	1	12.15	12.1	12.15	12.15
74	PFE0850c	60S ribosomal protein L12, putative	79.25	108.1	53.64	143.5	2	3	2	3	16.97	22.4	14.55	22.42
75	PF14_0425	fructose-bisphosphate aldolase	68.88	49.1	52.81	66.34	2	3	2	5	10.57	16.3	7.588	11.65
76	PF14_0344	conserved Plasmodium protein, unknown function	0	0	52.39	69.95	0	0	2	2	0	0	3.927	3.927
77	PF14_0548	ATPase, putative	0	24.16	40.81	34.65	0	1	1	1	0	2.86	2.864	2.864

78	PF10_0063	DNA/RNA-binding protein, putative	47.33	74.67	39.06	166	2	2	1	3	40.19	17.8	17.76	50.47
79	PF14_0359	HSP40, subfamily A, putative	123	0	38.61	40.95	1	0	2	1	4.717	0	7.547	4.717
80	PF14_0377	vesicle-associated membrane protein, putative	20.34	22.25	37.81	40.03	1	1	1	1	7.054	4.98	4.979	4.979
81	PFB0120w	early transcribed membrane protein 2, ETRAMP2	0	0	45.34	0	0	0	1	0	0	0	11.32	0
82	PFL1545c	chaperonin, cpn60	24.5	31.55	44.77	37.44	1	1	1	1	1.95	1.95	3.343	1.95
83	PF10_0025	PF70 protein	34.95	0	44.6	95.1	1	0	1	3	1.743	0	1.743	3.487
84	PF11_0384	cleft lip and palate associated transmembrane protein-related	79.82	0	43.55	0	2	0	1	0	4.79	0	3.193	0
85	PFF1375c	ethanolaminephosphotransferase, putative	0	42.81	68.06	41.59	0	1	1	1	0	5.12	5.115	5.115
86	PFE0060w	PIESP2 erythrocyte surface protein	25.85	0	28.88	20.38	1	0	1	2	3.676	0	3.676	3.676
87	PF11_0188	heat shock protein 90	0	0	71.57	46.76	0	0	3	2	0	0	5.376	1.828
88	PF10_0019	early transcribed membrane protein 10.1, etramp 10.1	51.67	43.24	66.49	32.86	1	2	4	1	11.21	11.2	11.21	11.21
89	PF11_0302	conserved Plasmodium protein, unknown function	25.72	35.83	67.92	0	1	1	2	0	2.876	2.88	4.646	0
90	PFI1445w	high molecular weight rhoptry protein-2	107.4	115.5	75.68	194.6	5	5	3	8	5.443	4.86	3.048	5.951
91	MAL13P1.221	aspartate carbamoyltransferase	104.4	54.6	77.74	0	2	1	2	0	8.8	3.2	5.867	0
92	PF11_0224	circumsporozoite-related antigen	49.89	57.16	72.4	68.97	2	2	1	3	11.11	11.1	11.11	11.73
93	PF14_0368	thioredoxin peroxidase 1	0	62.06	72.77	21.14	0	2	1	1	0	17.4	6.667	10.77
94	PF14_0567	conserved Plasmodium protein, unknown function	69.42	53.32	58.26	0	3	2	3	0	12.65	7.94	13.82	0
95	MAL13P1.237	conserved Plasmodium protein, unknown function	0	0	58.66	32.42	0	0	2	2	0	0	5.645	7.796
96	PFB0915w	liver stage antigen 3	34.78	0	56.31	43.18	1	0	1	2	1.155	0	1.155	2.567
97	PFI0935w	DNAJ-like molecular chaperone protein, putative	51.96	28.05	57.12	57.01	1	1	1	3	4.324	4.32	4.324	6.486
98	PFD1035w	steroid dehydrogenase, putative	31.3	31.2	65.58	81.19	1	1	2	2	3.738	3.74	8.723	8.723
99	PF10_0268	merozoite capping protein 1	83.65	104.5	65.8	0	2	3	2	0	4.071	7.89	6.361	0
100	PFC0725c	formate-nitrate transporter, putative	0	0	61.71	0	0	0	2	0	0	0	6.472	0
101	MAL13P1.61	Plasmodium exported protein (hyp8), unknown function	28.67	0	63.26	0	1	0	1	0	10.2	0	10.2	0
102	PF14_0077	plasmepsin II	93.66	80.74	93.6	105.4	6	5	4	3	11.48	9.05	11.48	9.051
103	PFI0605c	conserved Plasmodium protein, unknown function	0	65.27	88.27	0	0	1	2	0	0	2.47	4.484	0
104	PFI0880c	glideosome-associated protein 50	117.2	55.27	95.12	129.1	4	2	3	3	22.22	9.09	15.91	12.12

105	PFL1725w	ATP synthase beta chain, mitochondrial precursor, putative	0	0	0	23.33	0	0	0	1	0	0	0	3.364
106	PFL1825w	conserved Plasmodium membrane protein, unknown function	115.5	51.88	94.62	0	2	1	2	0	11.43	5.71	11.43	0
107	PFI0755c	6-phosphofructokinase, putative	37.57	0	105.8	27.63	1	0	3	1	1.693	0	3.385	1.269
108	MAL7P1.67	conserved Plasmodium protein, unknown function	56.37	0	101.9	24.69	1	0	4	1	11.96	0	29.67	11.96
109	PF07_0033	Cg4 protein	97.2	0	114.1	0	3	0	3	0	5.155	0	6.3	0
110	MAL13P1.233	nucleic acid binding protein, putative	84.66	49.74	114.1	106.7	2	1	4	4	14.69	6.16	20.38	20.38
111	PFD0305c	vacuolar ATP synthase subunit b	183.9	26.32	81.82	57.48	6	1	3	2	13.36	2.63	10.53	7.49
112	PF13_0214	elongation factor 1-gamma, putative	0	0	79.31	103.3	0	0	2	2	0	0	7.299	7.299
113	PF14_0301	conserved protein, unknown function	0	0	82.26	74.44	0	0	3	2	0	0	7.958	10.38
114	MAL8P1.62	conserved Plasmodium protein, unknown function	78.41	29.25	81.85	26.52	4	1	4	1	15.88	4.33	14.8	4.332
115	PF11_0281	protein phosphatase, putative	0	0	83.78	44.12	0	0	2	1	0	0	11.15	5.575
116	PF11_0338	Aquaglyceroporin	74.5	81.79	82.62	105.3	4	6	6	4	10.85	10.9	10.85	10.85
117	PF11_0164	peptidyl-prolyl cis-trans isomerase	0	50.79	85.15	21.93	0	1	2	1	0	5.13	11.79	5.128
118	PF07_0054	histone H2B	118.5	74.48	84.32	105.6	3	2	2	2	20.33	12.2	12.2	12.2
119	MAL13P1.56	m1-family aminopeptidase	202.6	113.5	140.5	333.5	7	3	6	11	10.69	4.15	9.309	14.38
120	PFI0265c	RhopH3	185.1	86.53	148.5	126.9	7	2	3	4	10.03	4.01	5.574	8.027
121	PFE0065w	skeleton-binding protein 1	174.9	43.28	149.4	209.4	6	2	6	4	28.19	10.7	27.6	28.19
122	PF11_0055	conserved protein, unknown function	45.31	33.62	159.2	24.55	3	1	6	1	5.896	4.01	17.22	2.358
123	PF10_0086	adenylate kinase	258.9	133.9	161.9	189.8	7	5	5	6	24.79	24.4	24.79	25.21
124	PF11_0179	conserved Plasmodium protein, unknown function	92.75	60.69	164.3	23.55	3	1	4	1	17.97	10.2	25	7.031
125	MAL13P1.540	heat shock protein 70 (hsp70), putative	147.9	68.93	171.2	176.7	8	2	4	6	7.189	3.97	6.33	7.296
126	PF14_0517	peptidase, putative	201.6	29.34	183.8	113.3	5	1	4	4	9.817	1.44	8.901	8.639
127	PF14_0201	surface protein, Pf113	158.1	56.69	115.6	152.4	4	3	4	5	5.366	3.2	3.406	5.573
128	PF10_0366	ADP/ATP transporter on adenylate translocase	108.8	125.4	116	141.8	4	4	6	4	13.29	19.6	16.94	18.27
129	PFE1590w	early transcribed membrane protein 5, ETRAMP5	139.2	73.41	116.7	75.23	4	2	3	2	13.26	13.3	20.44	13.26
130	PF11_0069	conserved Plasmodium protein, unknown function	220.8	54.06	118	73.16	4	2	3	2	24.44	8.27	10.53	8.271
131	PF11_0301	spermidine synthase	73.57	109.3	119	194.5	3	3	4	6	7.477	11.8	22.43	22.43

132	PF11_0175	heat shock protein 101, putative	55.14	0	119.1	0	2	0	4	0	2.428	0	6.843	0
133	PFB0210c	hexose transporter, PfHT1	108.1	38.85	132.6	78.87	2	1	2	1	5.952	2.98	5.952	2.976
134	PF14_0078	HAP protein	220.3	169.6	137.7	86.78	8	7	6	3	21.06	20.6	11.31	8.204
135	PF11_0313	60S ribosomal protein P0	55.57	62.48	295.6	231.1	2	2	6	6	10.13	6.96	30.06	36.71
136	PF13_0141	L-lactate dehydrogenase	204.7	175.2	335.1	361.9	6	6	10	10	38.92	35.8	43.99	44.3
137	PFD0310w	sexual stage-specific protein precursor	325.6	260.9	257.9	248.2	15	9	13	13	33.12	33.1	33.12	33.12
138	PF14_0075	plasmepsin IV	352.8	261.7	282.2	205.5	6	7	6	8	27.62	22.3	18.71	19.15
139	PFC0400w	60S Acidic ribosomal protein P2, putative	192.7	175.2	236.5	305.5	3	3	3	5	42.86	42.9	42.86	59.82
140	PFF0940c	cell division cycle protein 48 homologue, putative	192.5	34.31	249.1	149.6	9	2	10	5	13.53	2.05	16.18	8.816
141	PF13_0272	thioredoxin-related protein, putative	95.29	151.5	191.2	155	2	5	5	4	18.27	23.1	23.08	23.08
142	PFI1270w	conserved Plasmodium protein, unknown function	173.5	109	222.6	201.8	7	4	9	9	27.65	27.6	36.41	32.26
143	PFA0310c	calcium-transporting ATPase, putative	29.35	0	0	0	1	0	0	0	1.629	0	0	0
144	PF14_0076	plasmepsin I	547.2	372.2	498.4	430.4	18	12	11	14	34.96	31	30.97	29.87
145	PF14_0102	rhoptry-associated protein 1, RAP1	711.2	497.2	705.6	549.9	20	12	20	14	38.36	22.9	32.61	29.54
146	MAL8P1.17	protein disulfide isomerase	407	478.7	464.1	504.9	13	11	16	17	34.16	31.9	35.82	37.06
147	PF10_0153	heat shock protein 60	330.3	295.5	491.7	460.1	7	4	11	10	18.28	12.1	25.86	25
148	PF07_0029	heat shock protein 86	145.1	165.7	368.5	398.4	7	5	11	11	13.83	9.26	19.19	17.32
149	PF10_0084	tubulin beta chain, putative	0	0	21.43	0	0	0	1	0	0	0	4.045	0
150	PFL2215w	actin I	54.05	0	97.11	54.01	1	0	2	2	7.979	0	7.713	12.77
151	PF13_0346	60S ribosomal protein L40/UBI, putative	24.08	21.72	29.51	0	3	1	2	0	12.5	12.5	12.5	0
152	PFB0106c	Plasmodium exported protein, unknown function	22.79	27.38	62.35	0	1	1	1	0	4.828	4.83	4.828	0
153	PFE0080c	rhoptry-associated protein 2, RAP2	524.8	645.5	613.6	673.3	12	14	13	13	34.17	39.2	31.41	39.7
154	PF11_0351	heat shock protein hsp70 homologue	300.7	116.7	431.9	266.4	12	5	10	9	18.7	11.3	20.66	18.85
155	PFE1150w	multidrug resistance protein	595.4	463.1	432.6	525.6	18	13	13	15	15.72	10.9	11.84	12.83
156	PF08_0054	heat shock 70 kDa protein	314.7	228.7	334.7	490.5	10	8	14	19	15.81	10.6	18.17	27.03
157	PFI0875w	heat shock protein 70 (HSP70)	734.9	725.7	1104	848.1	27	26	45	30	26.38	27.6	39.42	35.12
158	PFB0765w	conserved Plasmodium protein, unknown function	0	0	0	23.89	0	0	0	1	0	0	0	0.651

159	PF14_0230	60S ribosomal protein L5, putative	0	0	0	36.76	0	0	0	2	0	0	0	10.54
160	PF11_0352	protein disulfide isomerase	0	0	0	20.27	0	0	0	1	0	0	0	5.437
161	PF13_0102	DnaJ/SEC63 protein, putative	0	0	0	20.1	0	0	0	1	0	0	0	2.611
162	MAL7P1.61	erythrocyte membrane protein 1 (PfEMP1)	0	0	0	20.01	0	0	0	1	0	0	0	2.591
163	PF13_0338	cysteine-rich surface protein	20.39	0	0	0	1	0	0	0	2.387	0	0	0
164	PFA0210c	conserved Plasmodium protein, unknown function	20.3	0	0	0	1	0	0	0	5.794	0	0	0
165	PF11_0062	histone H2B	75.55	63.04	0	53.44	3	3	0	4	21.37	21.4	0	21.37
166	PFD1055w	40S ribosomal protein S19, putative	46.6	0	0	0	1	0	0	0	10	0	0	0
167	PFL1070c	endoplasmin homolog precursor, putative	975.7	623.9	954.6	802.5	32	20	29	28	31.18	24.8	33.62	29.11
168	PFI0930c	nucleosome assembly protein	54.32	42.65	0	59.68	1	1	0	1	8.55	8.55	0	8.55
169	PF11_0208	phosphoglycerate mutase, putative	50.26	0	0	37.57	1	0	0	1	13.2	0	0	13.2
170	PF11_0099	heat shock protein DnaJ homologue Pfj2	0	0	0	51.95	0	0	0	1	0	0	0	3.889
171	PF10_0121	hypoxanthine phosphoribosyltransferase	59.16	0	0	0	3	0	0	0	35.93	0	0	0
172	PF11_0174	cathepsin C, homolog	21.79	0	0	38.69	1	0	0	1	3.571	0	0	3.571
173	PF11_0061	histone H4	26.89	27.59	0	0	1	1	0	0	11.65	11.7	0	0
174	PF14_0164	NADP-specific glutamate dehydrogenase	40.21	0	0	0	2	0	0	0	6.596	0	0	0
175	PF10_0068	RNA binding protein, putative	28.16	0	0	0	1	0	0	0	4.065	0	0	0
176	PFC0975c	peptidyl-prolyl cis-trans isomerase	20.13	0	0	0	1	0	0	0	7.018	0	0	0
177	PF13_0116	conserved Plasmodium protein, unknown function	21.33	0	0	0	1	0	0	0	1.59	0	0	0
178	PF14_0541	V-type H(+)-translocating pyrophosphatase, putative	266	146.1	240.3	113.1	7	4	9	2	9.902	6.42	12.41	3.347
179	PF11_0272	40S ribosomal protein S18, putative	21.72	21.92	0	20.48	1	1	0	1	14.74	5.13	0	14.74
180	PF13_0197	Merozoite Surface Protein 7 precursor, MSP7	0	88.23	0	0	0	3	0	0	0	9.4	0	0
181	PF14_0439	M17 leucyl aminopeptidase	25.14	0	0	0	1	0	0	0	1.983	0	0	0
182	PFC0730w	HVA22/TB2/DP1 family protein, putative	20.84	0	0	0	1	0	0	0	4.525	0	0	0
183	PFL1550w	lipoamide dehydrogenase	22.74	0	0	0	1	0	0	0	3.32	0	0	0
184	PFL0405w	conserved Plasmodium protein, unknown function	22.98	0	0	0	1	0	0	0	0.179	0	0	0
185	PF08_0113	vacuolar proton translocating ATPase subunit A, putative	21.37	0	0	0	1	0	0	0	1.33	0	0	0

186	PFF0420c	proteasome subunit alpha type 2, putative	0	25.62	0	0	0	1	0	0	0	5.96	0	0
187	PFD1037w	conserved Plasmodium protein, unknown function	22.24	0	0	0	1	0	0	0	7.273	0	0	0
188	MAL7P1.228	heat Shock 70 KDa Protein, (HSP70)	0	255.8	0	0	0	11	0	0	0	6.81	0	0
189	PFE0120c	merozoite Surface Protein 8, MSP8	41.01	0	0	0	1	0	0	0	2.848	0	0	0
190	PFF1155w	hexokinase	66.33	0	0	0	1	0	0	0	3.651	0	0	0
191	MAL8P1.69	14-3-3 protein, putative	55.5	42.54	153.7	96.24	1	2	4	3	6.107	9.92	24.43	18.32
192	PFI0180w	alpha tubulin	23.06	0	0	0	1	0	0	0	3.091	0	0	0
193	PF14_0716	proteosome subunit alpha type 1, putative	23.18	0	0	0	1	0	0	0	7.087	0	0	0
194	PF11_0172	folate/biopterin transporter, putative	34.99	0	0	0	1	0	0	0	3.077	0	0	0
195	PF13_0304	elongation factor-1 alpha	214.7	165.4	315.6	312.4	7	5	9	13	18.51	18.5	25.73	32.28
196	PF10_0155	enolase	365	134.9	181.8	244.3	10	6	6	7	34.75	20	19.96	23.32
197	PFL1385c	merozoite Surface Protein 9, MSP-9	31.23	54.88	64.95	0	1	2	2	0	2.423	3.9	3.903	0
	TOTAL NUMBER OF COMPOUNDS		**125**	**89**	**126**	**136**								

Accession: gene accession number; Protein: protein name; Score: Protein Mascot score (reflecting the combined scores of all observed mass spectra that can be matched to amino acid sequences within that protein; a higher score indicates a more confident match); Peptides: number of peptides identified; SC [%]: sequence coverage. GN: parasites grown in normal RBC; GD: parasites grown in G6PD-deficient RBC; 1: replicate 1; 2: replicate 2.

Table S12. List of proteins for protein-protein interaction analysis with Cytoscape software (accession numbers from human homologous genes were also used since *P. falciparum* has several proteins with not noted function).

Accession ***P. falciparum***	**Accession** ***Homo sapiens***	**Median (PKD:PKN1+N2)**	**Median (G6PDD:N1+N2)**
PF14_0377	EAX01590.1	0.24	-
PF10_0019	low homology	0.25	-
PF13_0141	ADG58108.1	0.26	-
PF11_0069	no significant homology found	0.27	-
PFI1270w	no significant homology found	0.28	-
PF14_0075	AAR03502.1	0.29	-
PF13_0272	NP_001139021.1	0.29	1.28
PF14_0102	low homology	0.29	-
PFE1150w	NP_000918.2	0.3	1.2
PF14_0076	AAA60364.1	0.3	1.2
PF11_0055	NP_005733.1	0.31	-
PFI1475w	low homology	0.32	0.55
PF11_0062	AAH66243.1	0.32	-
PF11_0302	no significant similarity found	0.33	-
PF14_0301	no significant similarity found	0.33	-
MAL13P1.540	BAD96476.1	0.34	-
PF11_0301	AAA36633.1	0.34	-
MAL8P1.69	NP_006752.1	0.35	-
PF10_0086	NP_037543.1	0.36	-
PFB0210c	NP_006507.2	0.36	-
MAL8P1.17	NP_004902.1	0.36	1.26
PFI0875w	AAF13605.1	0.36	1.3
PF08_0074	NP_680544.1	0.37	1.18
PFE1590w	no significant similarity found	0.38	-
PF10_0100	no significant similarity found	0.38	-
PF11_0313	NP_000993.1	0.38	1.34
PF13_0304	BAD96766.1	0.39	1.34
PFL0740c	NP_002148.1	0.4	1.38
PF11_0179	no significant similarity found	0.4	1.11
PF14_0541	no significant similarity found	0.4	-
PF14_0678	low homology	0.41	1.22
PF11_0338	NP_536354.2	0.41	-
MAL13P1.56	CAA10709.1	0.42	1.2
PFI0880c	CAG33359.1	0.43	-
PF08_0054	NP_005337.2	0.45	1.53
PFC0725c	no significant similarity found	0.46	-
PF11_0351	AAH00478.1	0.46	1.68

PF14_0598	NP_002037.2	0.46	1.23
PFE0065w	low homology	0.47	1.16
PFC0400w	NP_000995.1	0.48	-
PFL1070c	CAI64497.1	0.5	1.49
PFL1385c	no significant similarity found	0.51	0.63
PF11_0061	EAW55528.1	0.52	-
PF14_0630	AAA36475.1	1.5	1.36
PFI1735c	no significant similarity found	1.78	1.67
PF11_0331	NP_110379.2	-	1.48
MAL13P1.221	AAA51907.1	-	1.52
PF14_0391	AAA86463.1	-	1.54
PF13_0346	NP_003324.1	-	1.74