



Étude bioinformatique des protéines PfK13 et PfCRT impliquées dans la résistance de Plasmodium falciparum aux antipaludiques

Romain Coppée

► To cite this version:

Romain Coppée. Étude bioinformatique des protéines PfK13 et PfCRT impliquées dans la résistance de Plasmodium falciparum aux antipaludiques. Bio-informatique [q-bio.QM]. Université Paris Cité, 2019. Français. NNT : 2019UNIP5056 . tel-03860217

HAL Id: tel-03860217

<https://theses.hal.science/tel-03860217>

Submitted on 18 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

École doctorale Médicament, Toxicologie, Chimie, Imageries (MTCI ED 563)

Faculté des Sciences Pharmaceutiques et Biologiques

UMR 261 – Mère et enfant en milieu tropical : pathogènes, système de santé et transition épidémiologique (MERIT)

Étude bioinformatique des protéines PfK13 et PfCRT impliquées dans la résistance de *Plasmodium falciparum* aux antipaludiques

Par

Romain Coppée

Thèse de doctorat de bioinformatique

Dirigée par Jérôme Clain

Et co-encadrée par Audrey Sabbagh

Présentée et soutenue publiquement le 6 novembre 2019

Devant un jury composé de :

Dr. Bernard Offmann, Professeur des universités	Rapporteur
Dr. Françoise Benoît-Vical, Directrice de recherche	Rapporteuse
Dr. Antoine Claessens, Chargé de recherche	Examineur
Dr. Martin Picard, Directeur de recherche	Examineur
Dr. Jérôme Clain, Maître de conférences	Directeur de thèse
Dr. Audrey Sabbagh, Maître de conférences	Co-encadrante



Except where otherwise noted, this work is licensed under <https://creativecommons.org/licenses/by-nd/3.0/fr/>

Résumé

Le paludisme est une maladie infectieuse parasitaire causée par diverses espèces de *Plasmodium*, *P. falciparum* étant l'espèce la plus répandue et responsable des cas mortels de la maladie. Les traitements actuels reposent sur des combinaisons thérapeutiques à base d'artémisinine (CTA), couplant un dérivé d'artémisinine (ARTD) à une autre molécule antipaludique.

Des parasites résistants aux ARTDs ont émergé en Asie du sud-est ; ceux-ci sont encore non détectés en Afrique. La résistance se traduit par une durée d'élimination des parasites allongée, et est conférée par des mutations non-synonymes localisées sur le domaine Kelch-repeat propeller (KREP) de la protéine *P. falciparum* K13 (PfK13). Similairement, de multiples mutations non-synonymes sur le transporteur *P. falciparum* Chloroquine Resistance Transporter (PfCRT) confèrent la résistance à la chloroquine (CQ, ancien traitement) et à la pipéraquline (PPQ, une des molécules partenaires dans les CTAs actuels). Le rôle physiologique de ces protéines, essentielles durant le développement intra-érythrocytaire du parasite, demeure mal connu. Ce travail de thèse a donc pour objectif de mieux caractériser PfK13 et PfCRT : *i)* en prédisant les positions qui seraient impliquées dans des interactions protéine-substrat ; et *ii)* en étudiant les altérations structurales et physico-chimiques induites par les mutations de résistance.

Selon les concepts liés à la théorie de l'évolution moléculaire, les mutations touchant des sites exerçant une fonction critique au sein d'une protéine essentielle sont éliminées par la sélection purificatrice. Ces sites sont donc plus conservés que le reste de la protéine. Par des approches bioinformatiques couplant évolution et structure tertiaire, nous avons mis en évidence des régions extrêmement conservées au sein de PfK13 et PfCRT. En comparant ces résultats avec les données expérimentales de protéines / domaines appartenant aux mêmes familles structurales, nous avons identifié plusieurs sites de PfK13 et PfCRT que nous proposons comme candidats pour des interactions protéine-substrat.

À notre surprise, les mutations de résistance aux ARTDs ne sont pas localisées à la surface d'interaction que nous avons prédite sur le domaine KREP de PfK13. Les dynamiques moléculaires que nous avons réalisées sur deux mutations de résistance (C580Y et R539T) ont révélé des déstabilisations structurales locales du domaine KREP. Nous supposons que ces mutations de résistance pourraient perturber la stabilité du domaine KREP et ainsi diminuer l'abondance cellulaire de PfK13. Concernant PfCRT, les deux modèles de structure tertiaire, prédits par homologie structurale, montrent que la majorité des mutations de résistance à la CQ et à la PPQ sont localisées au niveau d'une poche probable de liaison que nous avons identifiée au cœur du transporteur. Ces mutations altèrent fortement le potentiel électrostatique à la surface de cette poche, passant d'un potentiel neutre à un potentiel électronégatif. Ce changement de propriété physico-chimique de la poche du transporteur est probablement un

déterminant majeur de l'acquisition de la propriété de transport de la CQ di-protonée de la vacuole digestive vers le cytoplasme du parasite.

Les sites fonctionnels candidats de PfCRT et PfK13 identifiés doivent maintenant être validés par des approches expérimentales. Nous avons initié des approches biochimiques dites de *pull-down* différentiels pour le domaine KREP de PfK13. Dans un premier temps, certains domaines PfK13, fusionnés à la glutathion S-transférase (GST), ont été exprimés dans la bactérie *Escherichia coli*, purifiés, puis incubés avec un lysat parasitaire total afin de caractériser des protéines interagissant avec PfK13. Les mises au point de ces expériences se poursuivent. Dans un second temps, des approches génétiques directement chez le parasite, par des techniques de transfection et d'édition de gènes, seront mises en place pour tester l'importance des sites fonctionnels candidats.

Mots-clés : Paludisme, *Plasmodium falciparum*, *Apicomplexa*, artémisinine, chloroquine, pipéraquline, *pfk13*, *pfCRT*, bioinformatique, évolution, structure, prédiction fonctionnelle, résistance, sélection naturelle, sélection purificatrice, prédiction structurale, simulation de dynamique moléculaire, phylogénie, mutation non-synonyme

Abstract

Malaria is an infectious parasitic disease caused by *Plasmodium* species. *P. falciparum* is the most prevalent species and is responsible for the majority of the disease burden. Currently, malaria is treated using artemisinin-based combination therapies (ACTs), coupling an artemisinin derivative (ARTD) to another antimalarial drug.

After a decade of use, ARTD-resistant parasites have emerged in Southeast Asia; they are currently not found in Africa. Resistance to ARTD is defined as parasites exhibiting *in vivo* a delayed clearance time following an artemisinin-based treatment, and is conferred by non-synonymous mutations localized in the Kelch-repeat propeller (KREP) domain of the *P. falciparum* K13 protein (PfK13). Similarly, multiple non-synonymous mutations conferring resistance to chloroquine (CQ) or piperazine (PPQ, used as partner drug in one ACT) have been reported on the *P. falciparum* chloroquine resistance transporter (PfCRT). The physiological role of these proteins, essential at least for the intra-erythrocytic development of the parasite, remains poorly understood. This thesis aims to better characterize the PfK13 and PfCRT proteins by: *i*) predicting the protein regions and positions that would be involved in protein-substrate interactions; and *ii*) studying structural and/or physicochemical alterations induced by resistance mutations.

According to concepts related to the theory of molecular evolution, mutations affecting sites exerting a critical function within an essential protein are eliminated by purifying selection. These sites are therefore more conserved than the rest of the protein. Using bioinformatics approaches combining sets of PfK13 and PfCRT orthologous sequences – respectively at *Apicomplexa* and *Plasmodium* scales – and information from protein tertiary structures, we highlighted extremely conserved regions within these two proteins. By comparing these results with those obtained for some proteins / domains belonging to the same structural families and validated at the experimental level, we identified several PfK13 and PfCRT amino acid sites that we propose as candidates for protein-substrate interactions.

Remarkably, ARTD resistance mutations are not localized at the interaction surface that we predicted in the KREP domain of PfK13. Molecular dynamics simulations carried out on KREP domains carrying resistance mutations (C580Y and R539T) indicated that these mutations are associated with local structural destabilizations of the KREP domain. We assume that these resistance mutations may disrupt the stability of the KREP domain and, by extension, the whole PfK13 protein whose cellular abundance would be decreased. Regarding PfCRT, using the two tertiary structure models we predicted by structural homology, the majority of CQ and PPQ resistance mutations were found located in a binding pocket at the core of the transporter. Furthermore, these mutations strongly altered the electrostatic potential at the surface of this pocket, from a neutral to an electronegative potential. This change in physicochemical property

of the pocket is probably a major determinant in the acquisition of the transport property of di-protonated CQ from the digestive vacuole to the cytoplasm of the parasite.

Candidate functional sites of PfCRT and PfK13 identified in our work should be validated by experimental approaches. Here, we initiated differential pull-down biochemical experiments for the KREP domain of PfK13. First, the wild-type PfK13 domains, fused to glutathione S-transferase (GST), were expressed in *Escherichia coli*, purified, and then incubated with parasite lysate to characterize PfK13-interacting proteins. These experiments are still in progress. In a second step, the importance of candidate functional sites will be tested through genetic approaches directly in *P. falciparum* parasites using transfection and gene editing techniques.

Keywords : Malaria, *Plasmodium falciparum*, *Apicomplexa*, artemisinin, chloroquine, piperazine, *pfk13*, *pfcr*, bioinformatics, evolution, structure, functional prediction, antimalarial resistance, natural selection, purifying selection, structure prediction, molecular dynamics simulation, phylogeny, non-synonymous mutation

À mon père

6 novembre 1956 – 27 décembre 2018

Aussi réfléchis soient-ils, aucun de mes mots ne saurait exprimer mon amour et ma reconnaissance envers toi, papa. Tu as su m'inculquer les valeurs du travail, la confiance et la persévérance dans ce que j'ai toujours entrepris. Aujourd'hui, tu n'es plus que dans mon cœur et mes souvenirs et j'espère que là où tu es, tu puisses être fier de ton fils.

Avec tout mon amour...

Remerciements

Je souhaite adresser mes plus profonds remerciements à Françoise Benoît-Vical (Laboratoire de Chimie de Coordination) et Bernard Offmann (UMR 6286, Unité Fonctionnalité et Ingénierie des Protéines) pour avoir pris le temps de juger et évaluer ce travail, ainsi qu'à Martin Picard (UMR 8015, Laboratoire de Cristallographie et RMN Biologiques) et Antoine Claessens (UMR 5235, Dynamique des Interactions Membranaires Normales et Pathologiques) pour avoir fait l'honneur de siéger dans ce jury de thèse.

J'adresse ma gratitude au docteur Philippe Deloron, qui m'a accueilli au sein de l'UMR 261 durant plus de trois ans – de mon stage de Master 2 à mon doctorat – avant de céder ses responsabilités aux docteurs André Garcia et Florence Migot-Nabias, que je remercie chaleureusement pour leur gentillesse et leur confiance.

Je tiens à remercier les docteurs Jérôme Clain et Audrey Sabbagh. Respectivement directeur et co-encadrante de ma thèse, ils sont pour moi ce que j'appelle de belles personnes. Outre le fait de m'avoir fait travailler sur un sujet très passionnant et plein d'avenir mêlant à la fois la bioinformatique et la biologie, ils n'ont eu de cesse que de me soutenir dans les moments les plus difficiles. Je pense sans trop me tromper avoir énormément appris grâce à eux, notamment en termes de rigueur scientifique. Il me semble difficile de trouver une personne aussi disponible et impliquée que Jérôme : malgré la multitude de tâches qui l'incombe, il a toujours su se montrer présent (même durant les congés) et continuellement m'encourager dans mes différents travaux. Quant à Audrey, elle n'a eu de cesse que de me féliciter pour mon investissement et mes résultats : quelle plus grande satisfaction peut-il y avoir que quelqu'un qui soit fier de vous ? En outre, mon seul souhait est de pouvoir continuer à collaborer avec eux dans mes futurs projets. Merci à tous les deux.

Je souhaite remercier les docteurs Anaïs Merckx (UMR 261) et Magali Blaud (CiTCoM - UMR 8038 CNRS) qui m'ont, avec Jérôme, beaucoup appris sur les bonnes pratiques de la biochimie, qui demeuraient pour moi totalement inconnues. Aussi je vous dois ces nouvelles connaissances et vont me permettre à l'avenir de développer plus encore cette double compétence bioinformatique / biochimie. À cela s'ajoute votre gentillesse, vos multiples conseils et vos sourires qui permettent de se sentir rassurer dans son travail.

Merci à Maria Miteva (INSERM U1268 MCTR, CiTCom UMR 8038 CNRS) et Isabelle Callebaut (UMR 7590, CNRS) qui ont apporté une aide précieuse et un regard critique sur une partie de mes résultats bioinformatiques. J'ai apprécié les différents échanges – qu'ils soient physiques ou par messagerie – qui ont permis de rendre les travaux bien plus aboutis. Je souhaite également remercier Frédéric Arieu (Institut Cochin, Biologie Comparative des *Apicomplexa*) pour m'avoir fourni les ressources informatiques nécessaires aux différentes analyses.

Il m'est impossible de faire une liste exhaustive des membres de l'unité dans laquelle j'ai passé près de quatre ans, même si vous avez toutes et tous participé de près ou de loin à rendre mon séjour plaisant. Malgré le fait que le laboratoire soit scindé en deux (une partie dédiée à la parasitologie, immunologie, *etc.* ; la seconde étant la génétique, la bioinformatique, *etc.*), j'ai eu l'honneur de côtoyer chacun de vous. Je souhaiterais tout de même adresser des mots particuliers à Laure Gineau, qui représente pour moi plus que je ne saurai le dire. Elle a toujours eu ce sourire, cette gentillesse et cette générosité que peu de personnes peuvent se vanter d'avoir dans le monde. Ta personne, nos petits goûters et nos discussions sur la vie de tous les jours vont beaucoup me manquer. Merci également à Jean-Gérard Gobert, qui malgré les 80 ans passés, semble garder une jeunesse éternelle, tel Louis Martinet dans le film des années 70' *Le Viager*. Je te considère un peu comme mon grand-père adoptif, avec qui j'ai eu beaucoup de plaisir à échanger. Je souhaite remercier Jacqueline Milet, Brigitte Techer et Jean-Yves Le Hesran, qui furent également pour moi des amis et qui faisaient toujours en sorte de débiter la journée dans les meilleures conditions avec un thé/café, souvent accompagné de quelques biscuits bienvenus. Je remercie Murielle Lohezic, qui m'a souvent rassuré dans les moments les plus délicats, notamment lors du départ de mon père. Tu as toujours su trouver les mots justes pour me redonner le sourire. Enfin, j'adresse mes plus sincères remerciements à Romain Duval, David Courtin, Rachida Tahar, Guillaume Escriou, Gwladys Bertin, Michel Cot, Isabelle Gaugué, Magalie Dambrun, Célia Dechavanne, Carmen Mora, Sayeh Guemouri, Laura Oliveira Martins, Isabelle Guérard, Nicaise Tuikue Ndam, Sandrine Houzé, Justine Bailly et Nicolas Argy : tous m'ont apporté leur sympathie et adressé plus de sourires que je ne saurai les compter.

Au cours de la thèse, on tisse souvent des liens indescriptibles avec les autres doctorants, et il semble que je ne déroge pas à cette règle. Une première équipe s'est formée autour de Margaux Chauvet, Abdoulaye « Gaby » Sissoko et Claire Kamaliddin. Vous avez été pour moi des amis, des soutiens, des confidents, soit tout ce qu'une personne peut avoir besoin dans la vie. Claire, tu fus la première à partir mais tu as été une très grande amie, et pour sûr qu'on le restera malgré la distance qui nous sépare. J'ai pris grand plaisir à travailler avec toi, même si nos objectifs n'ont pas tous été atteints. Vient ensuite Sissoko – car c'est bien comme cela qu'on t'appelait – que je considère aujourd'hui comme l'un de mes meilleurs amis. Je n'oublierai jamais tout ce qu'on a traversé ensemble, ainsi que nos multiples échanges sur la religion, le football – les déboires du Réal Madrid et de Newcastle... – et sur la famille. Enfin, Margaux fut une grande amie pour moi, elle a toujours su trouver les mots et me consoler quand ça n'allait pas : ton rire va me manquer ! L'arrivée de nouveaux doctorants – et stagiaires – ont permis de rendre cette *team* toujours plus vivante et amicale, où chacun pouvait compter les uns sur les autres. Je pense notamment à Diana Roman, Émilie Guillochon et Laurine Rondepierre. Diana, je souhaite te remercier pour ta gentillesse, ton extrême patience et ta précieuse aide en biochimie. Émilie, j'ai apprécié nos échanges en bioinformatique ; on peut vraiment dire que tu étais la seule qui me comprenait vraiment ! Je pense aussi que ton rire est très contagieux et m'a souvent permis d'esquisser un sourire dans les moments de peine. Laurine, tu as été pendant un temps ma partenaire de travail

dans les manipulations. On s'est longtemps serré les coudes et j'avoue avoir pris un grand plaisir à travailler à tes côtés. Qu'importe ce que tu en diras, tu m'as aussi beaucoup appris et rassuré par moment, lorsque je n'étais pas toujours sûr de ce que je faisais. Mon projet est maintenant le tien, et je suis sûr que tu découvriras ce que je n'ai pu montrer durant ma thèse. Gino Agbota, Edwige Pulchérie, Bernard Tornyigah et Cornélia Hounkonnou – également doctorants – ont participé à de plaisantes discussions que je n'oublierai pas. Une fois encore, on s'est tous soutenu pour que chacun puisse accomplir ses objectifs. Je ne vais certainement pas oublier Jérémy Fraering, qui fut également un très bon ami à qui je souhaite le meilleur pour les années à venir. Tu as eu cette capacité à toujours mettre les gens à l'aise – un comble quand tu dis que tu manques de confiance en toi ! – et ta gentillesse et générosité – notamment le partage de réactifs – a, chaque jour, permis de me sortir de certaines situations. Enfin, j'ai rencontré une multitude de stagiaires au cours de ces différentes années : Valentin Joste, Jérémy Filleul, Aurélie Chabry, Puthy Phlek, Xyomara Chavez Pacheco, Marianne Lecat, Margaux Lecanu et Shona Gray--Switzman. Je tiens à remercier spécialement Puthy pour sa gentillesse et sa sincère amitié. J'ai beaucoup apprécié ces quelques séances de marche jusqu'au RER à travers la descende du boulevard Saint-Michel. De même, Shona, tes pâtisseries vont me manquer... mais je suis certain que je te reverrai dans *Le Meilleur Pâtissier* sur M6. Merci à vous tous.

Mes pensées vont également à Jean-Christophe Gelly (INSERM, UMR S1134, DSIMB) et Catherine Etchebest (INSERM, UMR S1134, DSIMB). Vous m'avez donné le goût pour l'enseignement, et vous remercie chaleureusement de m'avoir fait confiance pour forger les esprits de demain.

Mon cœur se bouleverse lorsque je songe à ma famille. Nous avons traversé une épreuve indescriptible, qu'aucun ne mérite de traverser de cette façon. Il me paraît juste de remercier mon père, Michel Coppée. Tu n'es plus parmi nous, et pourtant dans nos cœurs tu demeures. Saches que je suis fier d'être ton fils et, même si je ne te l'ai que trop rarement dit, je t'ai aimé et t'aime de tout mon cœur. Par le passé, tu m'as inculqué les valeurs du travail, tu nous as toujours soutenu, avec mon frère jumeau, à faire de notre mieux. Et aucun ne pouvait douter de l'amour que tu nous portais. Par ces mots, je souhaite donc te dédier ce présent de travail. Je remercie ma mère, Véronique Coppée, qui est certainement la plus belle personne sur cette terre. Toujours présente, elle s'est battue et a toujours été convaincue de ma réussite. Aussi t'aiderai-je en juste retour de mon mieux pour les années à venir. Je remercie mes frères, Florian et Alexis Coppée. Je pense qu'il est rare de trouver des frères aussi aimants sur qui on peut toujours compter. Enfin, bien que n'ayant pas l'usage de la parole, je souhaite adresser ma tendresse à mon chien Pilou, une source de bonheur à lui tout seul et qui me permettait de m'évader lors des moments de peine. Merci à vous.

Pour clore ces remerciements, je souhaite adresser mon plus grand amour à Dieu, et son fils notre seigneur Jésus-Christ, qui m'ont apporté la force spirituelle nécessaire pour avoir tenu durant les épreuves difficiles, et ainsi terminer mon travail de thèse.

Table des matières

Résumé	3
Abstract	5
Remerciements	9
Liste des figures	17
Liste des tableaux	19
Liste des encadrés	19
Liste des annexes	20
Avant-propos	23

1. Introduction.....	25
1.1. Généralités sur le paludisme.....	27
1.1.1. L'histoire du paludisme	28
1.1.2. Épidémiologie de la maladie	29
1.1.3. Biologie du parasite <i>Plasmodium</i>	32
1.1.4. Signes cliniques et méthodes de diagnostic	35
1.1.5. De la prévention aux traitements curatifs	36
1.1.6. Historique d'utilisation des antipaludiques et d'apparition des résistances	39
Les points essentiels	41
1.2. L'ART, de sa découverte à la chimiorésistance des parasites	43
1.2.1. Découverte et historique de l'ART.....	44
1.2.2. Mode d'action des ARTDs	45
1.2.3. De la monothérapie d'ART aux CTAs	46
1.2.4. Émergence et caractérisation de la résistance aux ARTDs	48
1.2.5. La découverte d'un déterminant moléculaire de la résistance aux ARTDs : le gène <i>pfk13</i>	50
1.2.6. Le rôle de PfK13 dans la biologie du parasite : un comparatif sauvage <i>versus</i> mutant	55
1.2.7. D'autres déterminants moléculaires de résistance aux ARTDs à venir ?	60
Les points essentiels	62
1.3. PfCRT, une protéine de transport impliquée dans la résistance à de multiples antipaludiques à base de quinoléine	63
1.3.1. Découverte et développement des médicaments à base de quinoléine	64
1.3.2. Mode d'action des quinoléines.....	66
1.3.3. Rôle physiologique de PfCRT, transporteur impliqué dans la résistance aux médicaments à base de quinoléine.....	67

1.3.4. Caractérisation de la résistance à la CQ, conférée par les mutations du gène <i>pfcr</i> ...	70
1.3.5. Évolution des mutants <i>pfcr</i> et émergence de résistance à d'autres antipaludiques à base de quinoléine.....	72
Les points essentiels	75
1.4. Justification et objectifs du projet de thèse	77
1.4.1. Problématique générale	78
1.4.2. Quelles sont les régions et sites potentiellement impliqués dans la fonction physiologique de PfK13 et PfCRT ?	79
1.4.3. Quel est l'impact des mutations de résistance sur la structure tertiaire des protéines PfK13 et PfCRT ?.....	82
1.4.4. Congrès et publications	83
Les points essentiels	84

2. Méthodologie	85
2.1. La bioinformatique, de la théorie à la pratique	87
2.1.1. Mesure des pressions de sélection agissant sur les gènes <i>pfk13</i> et <i>pfcr</i>	88
2.1.2. Mesure de la vitesse d'évolution dans un contexte structural de PfK13 et PfCRT	90
2.1.3. Approche comparative évolutive et structurale pour la recherche de sites fonctionnels candidats	92
2.1.4. Prédiction de la structure tertiaire des protéines	93
2.1.5. Mesures biophysiques à partir des structures tertiaires de PfK13 et PfCRT	96
2.1.6. Simulations de dynamique moléculaire du domaine KREP de PfK13	98
2.1.7. <i>Docking</i> moléculaire sur le domaine KREP de PfK13	102
Les points essentiels	104
2.2. De la bioinformatique à la biochimie	105
2.2.1. Introduction au <i>pull-down</i>	106
2.2.2. Construction des plasmides.....	108
2.2.3. Transformation par choc thermique dans les souches d'expression	109
2.2.4. Optimisation des conditions d'expression des domaines recombinants	109
2.2.5. Solubilisation des domaines recombinants	110
2.2.6. Purification des domaines recombinants	111
2.2.7. Préparation du lysat parasitaire	111
2.2.8. Réalisation du <i>pull-down</i>	112
Les points essentiels	114

3. Résultats	115
3.1. Études bioinformatiques et expérimentales sur la protéine PfK13	117

3.1.1. Recherche des zones conservées de PfK13	118
3.1.1.1. La haute conservation de <i>k13</i> au cours de l'évolution des <i>Apicomplexa</i>	118
3.1.1.2. Un domaine coiled-coil fortement conservé	118
3.1.1.3. Les similitudes du domaine BTB de PfK13 avec celui des protéines KCTD	119
3.1.1.4. La poche du domaine KREP de PfK13, une possible surface d'interaction protéine-protéine	120
3.1.2. Les mutations de résistance aux ARTDs n'altèrent pas les propriétés structurales de la poche du domaine KREP de PfK13	120
3.1.3. Article 1	121
3.1.4. Analyses expérimentales de la poche du domaine KREP de PfK13	139
3.1.4.1. Analyses de <i>pull-down</i>	139
3.1.4.1.1. Expression des domaines recombinants PfK13	139
3.1.4.1.2. Solubilisation des domaines recombinants PfK13	140
3.1.4.1.3. Purification des domaines recombinants PfK13	143
3.1.4.1.4. La première analyse de <i>pull-down</i>	144
3.1.4.1.5. Le problème des contaminants bactériens après purification	146
3.1.4.2. Recherche d'inhibiteurs de la poche du domaine KREP de PfK13 ?	148
Les points essentiels	152
3.2. Analyses évolutives et structurales de PfCRT	153
3.2.1. À la recherche d'une région fonctionnelle de PfCRT	154
3.2.1.1. Forte conservation de la moitié vacuolaire de CRT	154
3.2.1.2. Prédiction de différents états conformationnels de la structure tertiaire de PfCRT	154
3.2.1.3. Prédiction de sites fonctionnels de PfCRT	155
3.2.1.4. Un potentiel électrostatique altéré chez les structures PfCRT portant les mutations de résistance aux médicaments à base de quinoléine	155
3.2.2. Article 2	156
Les points essentiels	177

4. Discussion	179
4.1. Explications et perspectives des méthodes bioinformatiques	181
4.1.1. Sélection purificatrice : contrainte structurale ou fonctionnelle ?	182
4.1.2. Une nouvelle option pour la détection de <i>patches</i> de sites conservés	183
4.1.3. Étude comparative des niveaux de sélection purificatrice aux échelles inter- et intra- espèces	185
Les points essentiels	186
4.2. Quel avenir pour les protéines PfK13 et PfCRT	187

4.2.1. Quelle(s) relation(s) entre la fonction de PfK13, la poche du domaine KREP, et les mutations de résistance aux ARTDs ?	188
4.2.1.1. PfK13 et sa possible participation dans un complexe d'ubiquitination	188
4.2.1.2. Vers une validation expérimentale du rôle de la poche du domaine KREP de PfK13	189
4.2.1.2.1. Approches biochimiques : <i>pull-down</i> et <i>pull-down</i> différentiels	189
4.2.1.2.2. Approche génétique : étude de parasites <i>P. falciparum</i> génétiquement modifiés	193
4.2.1.2.3. Approche pharmacologique : recherche d'inhibiteurs de la poche	193
4.2.1.3. Les mutations de résistance aux ARTDs : l'hypothèse d'une déstabilisation structurale du domaine KREP de PfK13	194
4.2.2. Quel lien entre la poche de PfCRT et sa fonction de transport ?	197
4.2.2.1. La poche de PfCRT, cœur de son activité ?	197
4.2.2.2. Validation des sites fonctionnels candidats	198
4.2.2.3. La relation entre la structure tertiaire de PfCRT et les mutations de résistance aux médicaments	200
4.2.3. PfK13 et PfCRT, de possibles cibles thérapeutiques	203
Les points essentiels	204

Références bibliographiques	205
Annexes	221

Liste des figures

Figure 1 : Principales figures historiques impliquées dans la caractérisation et la compréhension du paludisme	28
Figure 2 : Phylogénie des <i>Apicomplexa</i>	30
Figure 3 : Endémicité du paludisme dans le monde (2000-2017)	32
Figure 4 : Cycle de vie de <i>Plasmodium</i> chez l'anophèle et l'Homme	33
Figure 5 : Évolution du nombre de décès lié au paludisme entre 2000 et 2015	37
Figure 6 : Émergence de résistance aux principaux antipaludiques utilisés en zone d'endémie palustre	40
Figure 7 : L'artémisinine (ART), extraite d'une armoise annuelle, <i>Artemisia annua</i> , découverte par Tu Youyou	45
Figure 8 : Durée de clairance parasitaire pour des patients traités par monothérapie ou bithérapie à base d'ARTD	49
Figure 9 : Mise en évidence de mutations non-synonymes sur le gène <i>pfk13</i> comme marqueurs moléculaires de la résistance aux ARTDs	50
Figure 10 : Présentation de la protéine PfK13	51
Figure 11 : Localisation des mutations de résistance aux ARTDs sur les structures secondaire et tertiaire du domaine KREP de PfK13	52
Figure 12 : Impact de différentes mutations <i>pfk13</i> sur le taux de survie RSA _{0-3h} à partir de la souche asiatique Dd2	54
Figure 13 : <i>Fitness</i> des parasites portant différentes mutations du gène <i>pfk13</i>	55
Figure 14 : Conséquences de la mutation <i>pfk13</i> C580Y sur diverses voies métaboliques du parasite	57
Figure 15 : Modèle prédictif d'un complexe d'ubiquitination impliquant PfK13 et ses différents domaines	58
Figure 16 : Deux hypothèses permettant d'expliquer la relation entre les mutations <i>pfk13</i> et le plus faible taux de protéines ubiquitinées	59
Figure 17 : Formules chimiques des différents antipaludiques obtenus à partir de la fonction quinoléine	65
Figure 18 : Principales figures historiques ayant participé à la recherche sur la chimie et le mécanisme d'action de la quinoléine	66
Figure 19 : Site d'action de divers antipaludiques à base de quinoléine	67
Figure 20 : Apparition et expansion des parasites <i>P. falciparum</i> résistants à la CQ depuis la fin des années 1950	68
Figure 21 : Représentation schématique de la structure secondaire de PfCRT	68
Figure 22 : Phylogénie des DMTs	69
Figure 23 : Évaluation du rôle des mutations <i>pfCRT</i> dans la résistance à la CQ par mesure de la CI ₅₀ chez des parasites <i>P. falciparum</i> cultivés en culture	71
Figure 24 : Association entre des nouvelles mutations <i>pfCRT</i> (sur un fond génétique <i>pfCRT</i> mutant Dd2) et la résistance à la PPQ	73

Figure 25 : Influences des contraintes structurales et/ou fonctionnelles sur les taux d'évolution spécifiques aux sites	81
Figure 26 : Localisation dans la structure tertiaire des sites les plus conservés (au seuil de 10%) dans la protéine humaine B7-1 en considérant ou non l'environnement spatial des sites pour l'estimation de leur taux de substitution	91
Figure 27 : Recherche de sites fonctionnels candidats par une approche comparative évolutive et structurale	92
Figure 28 : Différentes étapes de la prédiction de la structure tertiaire d'une protéine par modélisation par homologie	94
Figure 29 : Diagramme de Ramachandran	96
Figure 30 : Étapes de préparation, de production et d'analyses relatives aux simulations de dynamique moléculaire via le logiciel GROMACS	102
Figure 31 : Présentation des différentes étapes du <i>docking</i> moléculaire	103
Figure 32 : Approche de <i>pull-down</i> dans la recherche de protéines partenaires à partir d'une protéine d'intérêt (ou appât) étiquetée GST	107
Figure 33 : Schématisation des différentes constructions PfK13 étiquetées GST	108
Figure 34 : Carte du plasmide pGEX-6P-1	108
Figure 35 : Structure tertiaire du domaine coiled-coil de PfK13 prédite par une approche <i>ab initio</i>	119
Figure 36 : Analyse sur gel SDS-PAGE 10% de l'expression des domaines recombinants GST-K, GST-BK et GST-CBK	139
Figure 37 : Analyse sur gel SDS-PAGE 10% de la solubilité du domaine recombinant GST-K	140
Figure 38 : Analyse sur gel SDS-PAGE 10% de la solubilité du domaine recombinant GST-K avec différents détergents	141
Figure 39 : Structure chimique du sodium lauroyl sarcosinate (sarkosyl)	142
Figure 40 : Analyse sur gel SDS-PAGE 10% de la solubilité du domaine recombinant GST-K à l'aide du sarkosyl	142
Figure 41 : Analyse sur gel SDS-PAGE 10% de la purification et de l'élution du domaine recombinant GST-K	143
Figure 42 : Solubilisation des protéines de <i>P. falciparum</i>	144
Figure 43 : <i>Pull-down</i> réalisé avec les constructions GST et GST-K	145
Figure 44 : <i>Pull-down</i> réalisé avec des constructions GST et GST-K/BK/CBK	147
Figure 45 : Tests de suppression des contaminants pour la construction GST-CBK par différents tampons de lavage	147
Figure 46 : Recherche d'inhibiteurs de la poche de KREP ^{KEAP1} et <i>docking</i> moléculaires sur les poches KREP ^{PfK13} et KREP ^{KEAP1}	149
Figure 47 : Résultats des différents <i>docking</i> pour 15 molécules avec les poches KREP ^{PfK13} et KREP ^{KEAP1}	150
Figure 48 : Pourcentage de croissance des parasites <i>P. falciparum</i> après exposition à différentes concentrations en inhibiteurs KEAP1	151

Figure 49 : Mise en évidence des sites les plus conservés du domaine KREP de PfK13 avec l'outil codeml de PAML	182
Figure 50 : Proposition d'une possible amélioration de l'outil FuncPatch dans la recherche de patches de sites conservés	184
Figure 51 : Résultat hypothétique attendu à l'issue d'une expérience de <i>pull-down</i> différentiel révélé par SDS-PAGE	190
Figure 52 : Localisation des positions R529/N530 sur la structure tertiaire du domaine KREP de PfK13	191
Figure 53 : Effet de PfCRT sur la croissance de levures en présence de CQ	200
Figure 54 : Affinités de liaison de la CQ (non- et di-protonée) avec les structures tertiaires de PfCRT sauvage et mutante Dd2 (<i>inward-facing</i>)	202

Liste des tableaux

Tableau 1 : Comparaison de différences espèces de <i>Plasmodium</i> infectant l'Homme	34
Tableau 2 : Liste des différentes molécules antipaludiques	38
Tableau 3 : Différentes activités de transport physiologique pour la protéine PfCRT	70
Tableau 4 : Principaux haplotypes <i>pfcr</i> t présents dans les populations naturelles de <i>P. falciparum</i> en zone d'endémie palustre	71
Tableau 5 : Les différents modèles de substitution des sites proposés par l'outil codeml de PAML	89
Tableau 6 : Modèles de substitution à partitionnement proposés par l'outil codeml de PAML	89
Tableau 7 : Comparaison des méthodes PAML et FuncPatch	91
Tableau 8 : Conditions testées pour l'expression de la protéine de fusion GST-K	109
Tableau 9 : Différentes expériences mesurant la quantité de protéine PfK13 mutante par rapport à leur souche de référence sauvage	195

Liste des encadrés

Encadré 1 : Descriptif des espèces de <i>Plasmodium</i> retrouvées en pathologie humaine	31
Encadré 2 : Les différentes techniques de diagnostic biologique du paludisme	36
Encadré 3 : Présentation des principales CTAs recommandées par l'OMS et mode d'action des molécules partenaires	47
Encadré 4 : Catégorisation des mutations <i>pfk13</i> associées à la résistance aux ARTDs selon divers critères érigés par l'OMS	53
Encadré 5 : Définition des termes relatifs aux principes de l'évolution	80
Encadré 6 : Types d'énergies et d'interactions mises en jeu dans un champ de force	98

Encadré 7 : Composition des tampons de lyse (TLb) testés pour lyser les bactéries et solubiliser les domaines recombinants 110

Encadré 8 : Composition des tampons de lyse (TLp) testés pour lyser et solubiliser les protéines parasitaires 112

Liste des annexes

Publications annexes	221
Poster présenté en congrès	222
Matériel supplémentaire – Article 1	224
Matériel supplémentaire – Article 2	252

Liste des abréviations

Å : Ångström	CRISPR/Cas9 : Clustered regularly interspaced short palindromic repeats / CRISPR associated protein 9
ADN : Acide désoxyribonucléique	CRT : Chloroquine resistance transporter
AL : Artéméther-luméfantrine	CTA : Combinaison thérapeutique à base d'artémisinine
ampR : Résistance à l'ampicilline	DCCM : Dynamical cross correlation map
APBS : Adaptive Poisson-Boltzmann solver	DHA : Dihydroartémisinine
ART : Artémisinine	DHPS : Dihydropteroate synthase
ARTD : Dérivé d'artémisinine	DHFR : Dihydrofolate reductase
ASAQ : Artésunate-amodiaquine	DMT : Drug/metabolite transporter
ASMQ : Artésunate-méfloquine	DOPE : Discrete optimized protein energy
ASPY : Artésunate-pyronaridine	DP : Dihydroartémisinine-pipéraquline
ASSP : Artésunate-sulfadoxine-pyriméthamine	DTT : Dichlorodiphényltrichloroéthane
ATV : Atovaquone	eIF2α : Eukaryotic initiation factor-2α
BLAST : Basic local alignment search tool	Fe²⁺ : Ion ferreux
bp : Base pairs	Fe³⁺ : Ion ferrique
BSA : Bovine serum albumin	GO : Gene ontology
BTB : Broad-complex, tramtrack and bric-à-brac	GR : Globule rouge
Cα : Atome de carbone α	GSH : Glutathion
CI₅₀ : Concentration inhibitrice 50%	GST : Glutathion S-transférase
CQ : Chloroquine	HRP2 : Histidine-rich protein 2

IPTG : Isopropyl β -D-1-thiogalactopyranoside	PCR : Polymerase chain reaction
IVR : Intervening region	PDB : Protein data bank
K13 : Kelch13	PI3K : Phosphatidylinositol 3-kinase
K⁺ : Potassium	PK4 : Protein kinase 4
KCTD : Potassium (K ⁺) channel tetramerization domain	PPQ : Pipéraquine
kDa : Kilodalton	PSA : Piperazine survival assay
KEAP1 : Kelch-like ECH-associated protein 1	PYR : Pyriméthamine
KLHL : Kelch-like	QTL : Quantitative trait loci
KO : Knock-out	RE : Réticulum endoplasmique
KREP : Kelch-repeat propeller	REL : Random-effects likelihood
LDH : Lactate dehydrogenase	RMSF : Root-mean-square fluctuation
LRT : Likelihood-ratio test	RPM : Rotation par minute
MAFFT : Multiple alignment using fast Fourier transform	RSA : Relative solvent accessibility
MDR1 : Multidrug resistance protein 1 and 2	RSA_{0-3h} : Ring-stage survival assay
MRP1, MRP2 : Multidrug resistance associated protein 1 and 2	SDS-PAGE : Sodium dodecyl sulfate polyacrylamide gel electrophoresis
NCBI : National center for biotechnology information	SMS : Smart model selection
Nrf2 : Nuclear factor (erythroid-derived 2)-like 2	TDR : Test de diagnostic rapide
OMS : Organisation mondiale de la santé	TLb : Tampon de lyse bactérien
PAML : Phylogenetic analysis by maximum likelihood	TLp : Tampon de lyse parasitaire
	UPR : Unfolded protein response
	UMR : Unité mixte de recherche
	VP : Vérapamil
	WCN : Weighted contact number

Avant-propos

« Mais que faire après mon stage de master 2 ? Thèse ? Pas Thèse ? ». Ces questions, nombre d'étudiants se les sont posées avant moi. Pour ma part, je ne pensais pas être capable de me lancer dans une thèse de doctorat, nécessitant rigueur et confiance en soi. Cependant, si Jérôme et Audrey m'ont proposé de prolonger l'aventure au sein de l'UMR 261 à plusieurs reprises, peut-être ont-ils remarqué en moi un potentiel dont je n'avais initialement pas conscience moi-même. Aujourd'hui, après trois années de labeur, aucun regret ne me vient à l'esprit : j'ai apprécié mon travail et la thématique de recherche ; j'ai été pris de passion pour l'enseignement ; et j'ai noué des contacts formidables dans le monde scientifique. Ces années sont, je l'espère, les prémices d'une carrière en tant qu'enseignant-chercheur.

Travailler sur une thématique aussi importante et essentielle qu'est le paludisme fut pour moi un honneur. Mon entourage m'a souvent posé la question suivante : « Il n'y a pas un vaccin pour soigner le paludisme ? ». Non, à ce jour, aucun vaccin n'est pleinement efficace... Plus encore, les traitements curatifs actuels perdent en efficacité, alors que le paludisme continue de faire des centaines de milliers de morts chaque année, principalement en Afrique subsaharienne où les populations locales n'ont pas toujours les moyens de suivre les recommandations sanitaires.

La bioinformatique peut apporter de nouvelles pistes / perspectives d'analyse biologique. C'est en tout cas ce que révèle principalement cette thèse, laquelle est articulée autour de l'étude de deux protéines du parasite *Plasmodium falciparum* (responsable des infections palustres) impliquées dans la résistance aux médicaments de première intention actuels : PfK13 et PfCRT. La lecture de ce travail requiert quelques connaissances en biologie et en bioinformatique. Cependant, nous essayerons de clarifier chaque terme, chaque concept et chaque méthode afin que cet ouvrage puisse être accessible au plus grand nombre.

Bien que cette thèse soit majoritairement orientée bioinformatique, l'introduction sera centrée sur de la biologie (présentation de l'agent infectieux *P. falciparum*, connaissances actuelles sur les protéines PfK13 et PfCRT, *etc.*). La bioinformatique apparaît en fin d'introduction et sera pleinement détaillée dans la partie *Méthodologie*.

Partie 1

Introduction



*La vie est un défi à relever, un bonheur
à mériter, une aventure à tenter.*

— **Mère Teresa** (1910 – 1997)

Cette première partie a pour objectif d'introduire des généralités sur la maladie qui est au cœur de ce mémoire : le paludisme. Après un bref historique sur la découverte de la maladie, une description de son épidémiologie et de la biologie de l'agent pathogène impliqué, les parasites du genre *Plasmodium*, nous décrirons ses manifestations cliniques et les principales méthodes de diagnostic. Une dernière sous-section sera consacrée aux médicaments antipaludiques utilisés dans le traitement et la prévention des accès palustres, avec une attention particulière sur l'histoire de l'émergence des résistances à ces drogues, puisque l'étude de protéines impliquées dans certaines résistances est le principal objet de ce travail de thèse.

1.1.1. L'histoire du paludisme

Le terme paludisme provient du latin *paludis* signifiant « marais », référant aux zones de reproduction du vecteur de la maladie, le moustique anophèle. La dénomination « malaria », de l'italien *mal'aria* est également répandue – principalement dans les pays anglophones –, introduite par le médecin du pape Clément XI, Giovanni Maria Lancisi (**Figure 1**), qui se traduit par « mauvais air »^{1,2}. Diverses estimations démontrent que le paludisme affecte l'Homme depuis le pléistocène, il y a plus de 50 000 ans³. Autrefois peu répandues, les populations plasmodiales n'ont cessé de croître en fonction des avancées agricoles et de la migration de l'Homme³. Le paludisme chez l'Homme est rapporté dans les écrits depuis la plus haute antiquité. Dans ses registres datant du IV^{ème} siècle avant Jésus-Christ, le médecin grec Hippocrate fait référence à des fièvres périodiques tandis que l'agronome romain Columelle associe la maladie aux insectes provenant des marais au I^{er} siècle après Jésus-Christ⁴.

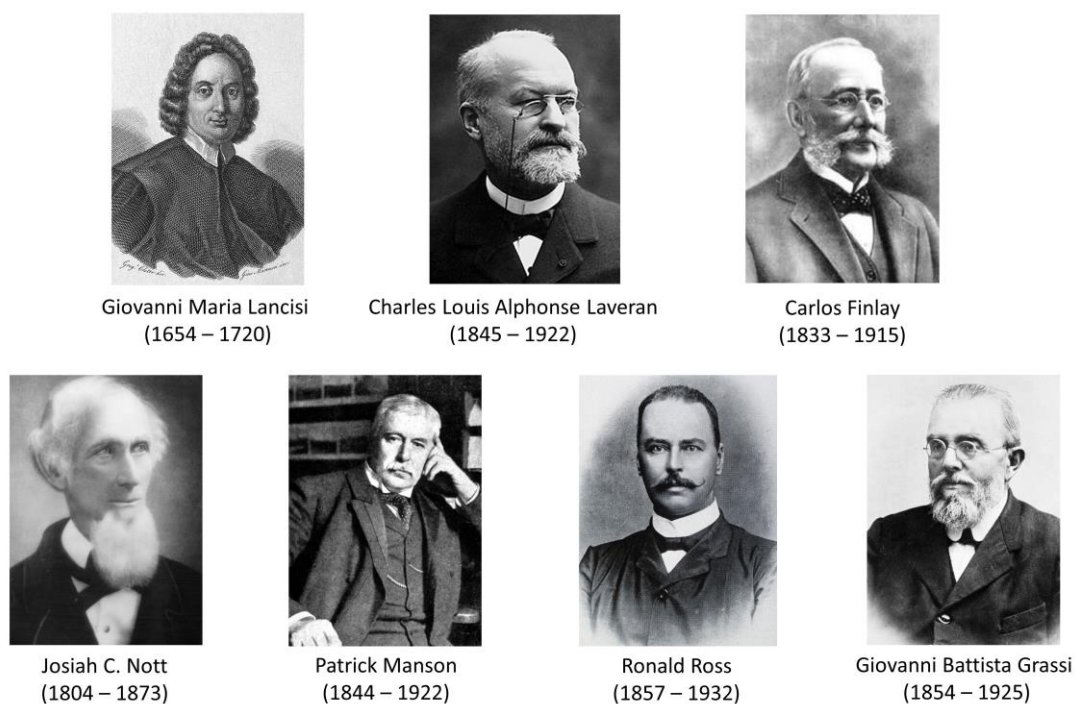


Figure 1 : Principales figures historiques impliquées dans la caractérisation et la compréhension du paludisme. La majorité des grandes découvertes initiales s'est déroulée à la fin du XIX^{ème} siècle et au début du XX^{ème} siècle.

Les parasites humains et leurs vecteurs ont co-évolué avec l'Homme, se dispersant en Afrique, en Eurasie et en Amérique⁵. Le paludisme est d'ailleurs considéré comme contributeur du déclin de l'Empire Romain. En effet, de nombreuses régions de la Rome Antique offraient des conditions favorables pour le développement du vecteur du paludisme, et plus particulièrement le sud de l'Italie et l'Île de la Sardaigne^{6,7}. Les jardins irrigués et les problèmes de drainage dus à la construction de routes ont entraîné une augmentation significative des eaux stagnantes,

permettant aux moustiques de s'y développer. Au XIV^{ème} siècle, il est probable que ce soient les colons européens et leurs esclaves qui ont amené le paludisme en Amérique⁸.

Bien que faisant l'objet de nombreuses études du fait de son fort impact en termes de morbidité et de mortalité, les premières avancées scientifiques majeures ne sont réalisées qu'à partir de 1880 par le médecin de l'armée française Charles Louis Alphonse Laveran (**Figure 1**)⁹. Celui-ci observe que les parasites se développent au sein des globules rouges (GRs, aussi appelés érythrocytes ou hématies), et confirme l'hypothèse selon laquelle un protisteⁱ soit la cause de la maladie. Pour cette découverte et l'ensemble de ses recherches ultérieures, il reçut le prix Nobel de physiologie et de médecine en 1907. Parallèlement, le docteur cubain Carlos Finlay (**Figure 1**) fournit des preuves solides suggérant que les moustiques transmettent le paludisme à l'Homme, hypothèse déjà proposée antérieurement par le chirurgien américain Josiah C. Nott et le médecin britannique Sir Patrick Manson (**Figure 1**)¹⁰. Suite à la collaboration de ce dernier durant quatre ans avec le médecin écossais Sir Ronald Ross (**Figure 1**) – également récompensé par le prix Nobel de physiologie et de médecine en 1902 –, ils explicitent en 1897 le cycle de vie complet du parasite chez le moustique grâce à l'isolement de parasites provenant des glandes salivaires de moustiques qui s'étaient nourris sur des oiseaux préalablement infectés¹¹. Ce cycle est complété et confirmé en 1899 par une étude similaire réalisée chez des humains cette fois par le zoologiste italien Giovanni Battista Grassi (**Figure 1**), lequel montre que le paludisme ne peut être transmis que par les moustiques du genre *Anopheles* via leur système digestif¹². En 1900, les conclusions de Ross et Finlay sont finalement approuvées par une commission médicale présidée par le médecin militaire américain Walter Reed¹³. Ross poursuit ses travaux au sein de la *Liverpool School of Tropical Medicine* pour diriger les efforts de lutte contre le paludisme en Égypte, en Grèce et au Panama. Il propose de nombreuses mesures sanitaires pour limiter le développement de la maladie, mises en place par exemple par le médecin américain William C. Gorgas lors de la construction du canal de Panama. Outre la survie de milliers de travailleurs, ses conseils auront une influence majeure sur les campagnes de santé publique à venir contre la maladie¹⁴.

1.1.2. Épidémiologie de la maladie

Le paludisme est causé par des parasites eucaryotes unicellulaires appartenant au genre *Plasmodium*, du règne des *Chromalveolata*, division *Alveolata*, rattaché à l'embranchement des *Apicomplexa*^{15,16} (**Figure 2**). Les *Apicomplexa* comptent deux classes, nommés *Conoidasida* et *Aconoidasida*. La première se subdivise en *Coccidia* et *Gregarinia*¹⁷, tandis que la seconde regroupe les *Piroplasmida* et les *Haemosporida*¹⁸. Les *Apicomplexa* comprennent uniquement des parasites intracellulaires, parmi lesquels se trouvent *Cryptosporidium parvum* et *Toxoplasma gondii* (*Coccidia*), les espèces de *Theileria* et *Babesia* (*Piroplasmida*), et les

ⁱ Règne obsolète qui regroupait la plupart des eucaryotes unicellulaires.

Plasmodium (*Haemosporida*). Tous les membres des *Apicomplexa* sont des parasites présentant des caractéristiques spécifiques liées à la locomotion et à l'invasion cellulaire : ils se différencient notamment selon l'hôte, le tropisme pour différents tissus, et l'obligation pour certaines espèces – telles que les *Plasmodium* – de se développer chez plus d'un hôte pour compléter leur cycle biologique.

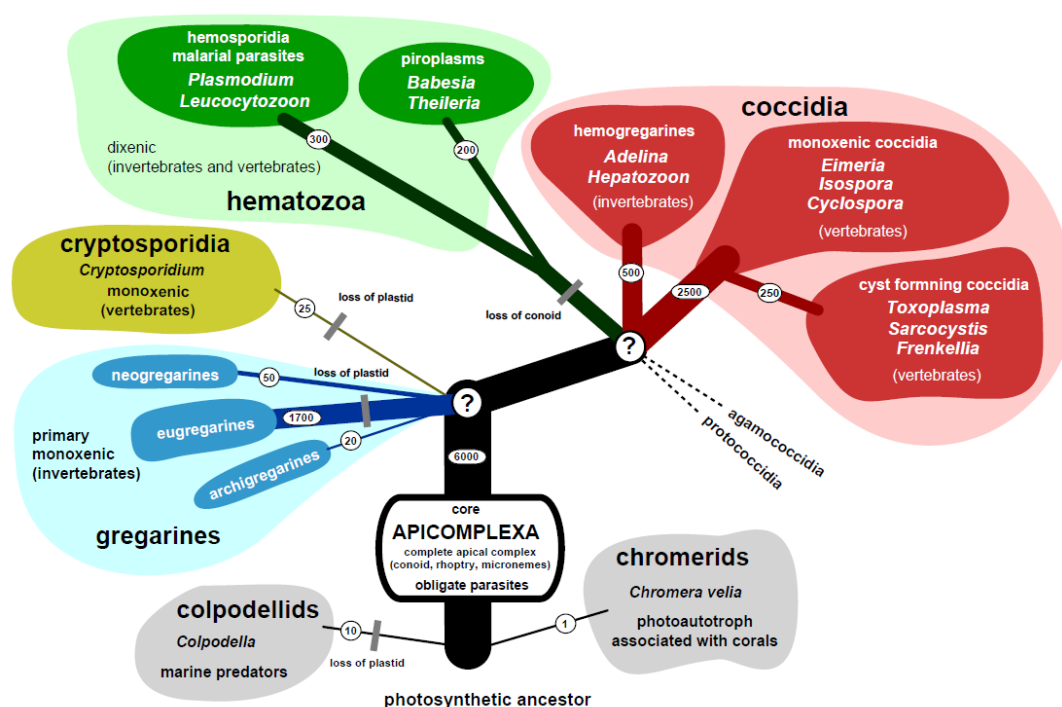


Figure 2 : Phylogénie des *Apicomplexa*. Le genre *Plasmodium* fait partie du groupe *Haemosporidia*, phylogénétiquement proche des *Piroplasmida* comprenant les espèces de *Theileria* et *Babesia* (représentées en vert). La phylogénie est tirée d'un récent ouvrage sur les *Apicomplexa*¹⁵.

Plus de 200 espèces de *Plasmodium* ont été recensées à ce jour, infectant de nombreuses espèces animales, que ce soit les singes, les lézards, les poulets ou les chauves-souris¹⁹. Six espèces sont habituellement retrouvées en pathologie humaine : *P. falciparum*, *P. vivax*, *P. knowlesi*, *P. malariae* et *P. ovale*²⁰, cette dernière étant maintenant subdivisée en deux nouvelles espèces : *P. ovale curtisi* et *P. ovale wallikeri*. Ces espèces se différencient par diverses particularités biologiques et cliniques ainsi que par leur répartition géographique (**Encadré 1**).

Encadré 1 : Descriptif des espèces de *Plasmodium* retrouvées en pathologie humaine.

P. falciparum : *P. falciparum* est l'espèce la plus largement répandue avec *P. vivax*, présente dans l'ensemble des zones d'endémie palustre et responsable de la majorité des cas cliniques de paludisme²¹. *P. falciparum* présente une période d'incubation minimale comprise entre 7 et 12 jours. Sa transmission est plus élevée lors de la saison des pluies et est interrompue sous 18°C²⁰.

P. vivax : *P. vivax* est principalement répandue en Amérique du sud, en Asie et en Océanie. Elle demeure rare en Afrique, détectée essentiellement en Mauritanie et dans les régions de l'est et des îles environnantes (Madagascar)²². Cette espèce a la particularité d'occasionner des rechutes (accès de reviviscence) dues au réveil d'hypnozoïtes intrahépatiques restés quiescents durant plusieurs semaines ou mois (cf. section 1.1.3). Les cas cliniques associés à ce parasite ont longtemps été considérés comme bénins mais des cas graves ont été décrits. *P. vivax* présente une période d'incubation minimale comprise entre 11 et 15 jours, avec un arrêt de sa transmission sous 15°C.

P. knowlesi : L'espèce *P. knowlesi* est présente uniquement en Asie du sud-est, où sont retrouvés les macaques – son hôte habituel – et son vecteur qui pique à la fois l'Homme et le singe²³. Des cas mortels ont été associés à ce parasite.

P. malariae : L'espèce *P. malariae* sévit en Asie, Océanie (Papouasie-Nouvelle Guinée), Afrique et Amérique du sud, quoique de manière plus épisodique. Elle nécessite une incubation nettement plus longue, comprise entre 15 et 21 jours, et peut occasionner des recrudescences tardives de plusieurs semaines à plusieurs mois, similairement à *P. vivax*. Les infections à *P. malariae* restent toutefois souvent bénignes.

***P. ovale* spp.** : Les deux espèces *P. ovale curtisi* et *P. ovale wallikeri* sévissent en Afrique centrale et Afrique de l'ouest, ainsi que dans quelques régions du Pacifique. Tout comme *P. vivax*, on lui associe des accès de reviviscence. Son incubation est estimée à 15 jours, bien qu'elle puisse être nettement plus longue.

On estime que 3,2 milliards de personnes dans le monde sont exposées au risque de paludisme chaque année²⁴. Bien que la maladie ait été éradiquée en Amérique du nord, en Europe ainsi que dans les États de l'Europe de l'est incluant la Turquie, une vigilance constante existe car *P. vivax* est capable de se réimplanter de façon temporaire dans le sud de l'Europe tel que l'Italie ou la Grèce. Des cas autochtones à *P. vivax* ont d'ailleurs été répertoriés en Grèce en 2011 en raison de travailleurs migrants²⁵. À l'inverse, le paludisme s'avère extrêmement répandu dans toute l'Afrique subsaharienne, représentant un réservoir pour *P. falciparum*, *P. ovale* spp. et *P. malariae*. La région concentre d'ailleurs près de 90% des 435 000 décès recensés en 2017 selon le rapport 2018 de l'Organisation Mondiale de la Santé (OMS), principalement des enfants de moins de 5 ans (**Figure 3**)²⁴. Le nombre de cas en Amérique centrale demeure modéré, majoritairement occasionnés par *P. vivax*, tandis que l'Amérique du sud concentre un nombre d'accès palustres plus élevé incluant la Guyane française (**Figure 3**). L'Asie du sud-est comprend

des infections relatives aux espèces *P. vivax* et *P. falciparum* (**Figure 3**), et est généralement une zone d'émergence de souches résistantes à de multiples antipaludiques²⁴ (cf. section 1.1.6.).

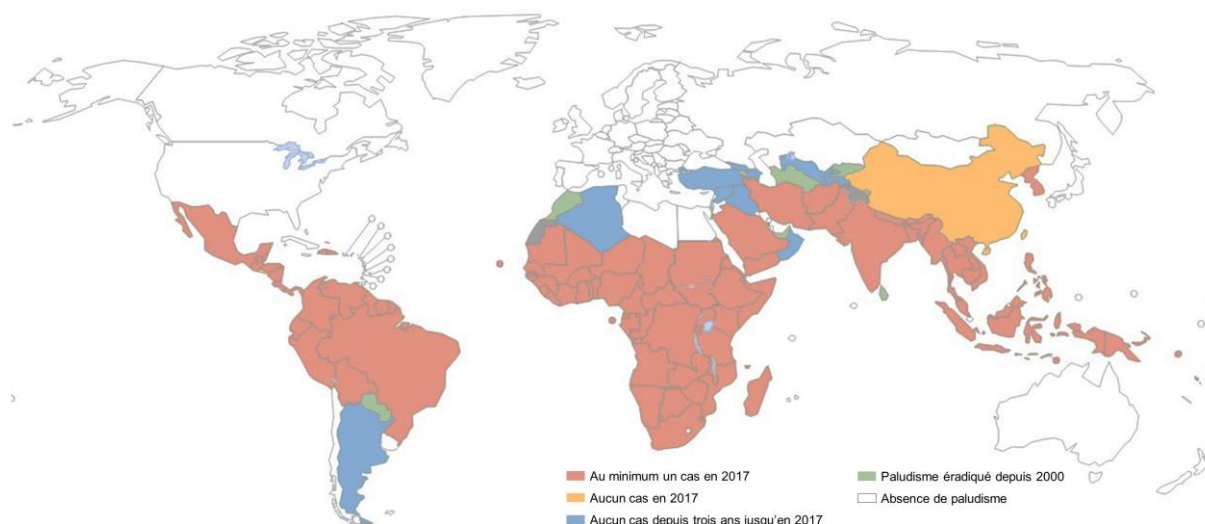


Figure 3 : Endémicité du paludisme dans le monde (2000-2017). En 2000, le paludisme était endémique dans 108 pays. Les moyens mis en œuvre pour lutter contre cette maladie ont drastiquement diminué son taux de mortalité et de morbidité, avec une éradication totale dans une vingtaine de pays. Toutefois, en 2017, une hausse du nombre d'accès palustres a été observée dans les zones les plus sévèrement touchées d'Afrique²⁴.

Le parasite *Plasmodium* est transmis à l'Homme par la piqûre de moustiques femelles appartenant au genre *Anopheles*²⁴. En effet, seules les femelles anophèles sont hématophages. Elles sont principalement retrouvées dans des eaux stagnantes non polluées²⁶. Le développement et la longévité de ces espèces vectrices dépendent de la nature des sols, de la température ou encore de l'altitude, et ont une durée de vie moyenne de 30 jours. Toutes les espèces d'anophèles ne sont toutefois pas capables de transmettre le paludisme, les principaux vecteurs étant les espèces anthropophiles – qui ont une préférence pour les repas sanguins sur l'Homme – et les espèces endophiles – que l'on retrouve communément à l'intérieur des habitations²⁷.

1.1.3. Biologie du parasite *Plasmodium*

Les parasites du genre *Plasmodium* présentent un cycle hétéroxène, c'est-à-dire qu'ils nécessitent plusieurs hôtes successifs au cours de leur vie pour compléter leur cycle évolutif. Une phase sexuée a lieu chez l'anophèle (hôte définitif ; **Figure 4, panel de gauche**), tandis qu'une phase asexuée se déroule chez l'Homme (hôte intermédiaire) chez qui le parasite se développe exclusivement sous sa forme haploïde (**Figure 4, panel de droite**)²⁸. La phase asexuée est elle-même divisée en deux phases : une phase hépatique (aussi connue sous le nom de pré- ou exo-érythrocytaire) qui correspond à la phase d'incubation asymptomatique ; et une phase sanguine

(aussi appelée érythrocytaire) au cours de laquelle se manifestent les symptômes de la maladie (fièvre, fatigue généralisée, vertiges, troubles digestifs)²⁰.

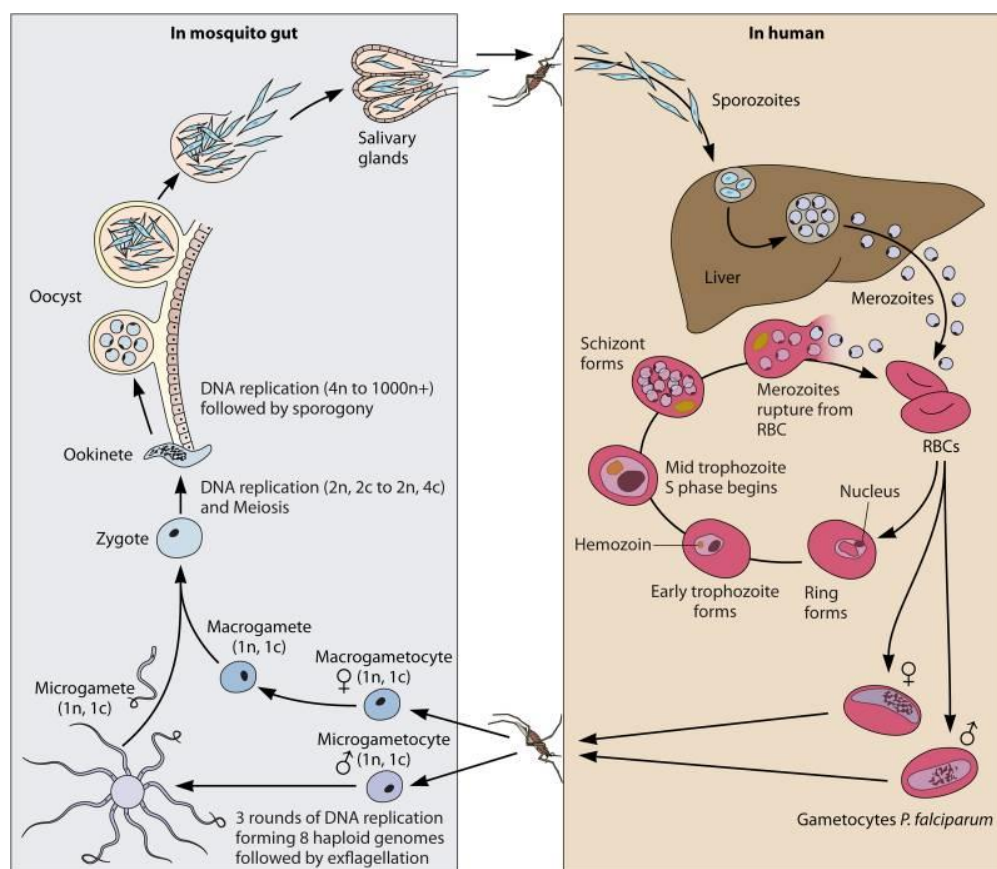


Figure 4 : Cycle de vie de *Plasmodium* chez l'anophèle et l'Homme. Lors de son repas sanguin, l'anophèle transmet à l'Homme le parasite (sous la forme appelée sporozoïte) qui va ensuite rejoindre le foie et s'y multiplier au sein des hépatocytes (phase pré-érythrocytaire, non symptomatique). Les mérozoïtes issus de cette multiplication hépatique sont libérés dans le sang où ils infectent les GRs. Après plusieurs cycles de multiplication érythrocytaire, le nombre de parasites par individu infecté peut atteindre 10^8 à 10^{12} et déclencher les symptômes de la maladie. Certaines formes sanguines vont se transformer en gamétocytes, formes infectantes pour le moustique lors d'un nouveau repas sanguin. Au sein de l'anophèle, les gamétocytes se transforment en gamètes mâles et femelles, puis leur fusion donne naissance à un œuf mobile (ookinète). Cet œuf se transformera ultérieurement en oocyste dans lequel les sporozoïtes se forment et se multiplient. Ces derniers rejoignent ensuite les glandes salivaires, prêts à être inoculés à l'Homme lors d'un repas sanguin. La figure est tirée du travail de Lee et ses collaborateurs (2014)²⁹.

Chez l'Homme, la phase hépatique débute après que le moustique ait effectué son repas sanguin et déposé une certaine quantité de parasites (moins d'une centaine) sous la forme de sporozoïtes. Ceux-ci infectent les hépatocytes dans lesquels ils vont se multiplier : il s'agit de la schizogonie hépatique, durant laquelle plusieurs endo-mitoses se succèdent pour former des parasites multinucléés appelés schizontes hépatiques ou pré-érythrocytaires. Après 7 à 15 jours de maturation hépatique (la durée varie selon l'espèce), ces schizontes éclatent et libèrent quelques milliers de mérozoïtes uninucléés dans le sang. Chez les infections à *P. vivax* et *P. ovale* spp., certaines formes hépatiques initiales peuvent demeurer à un état quiescent qu'on appelle

les hypnozoïtes³⁰. Ces derniers se réactivent quelques semaines voire quelques mois plus tard, pour *in fine* libérer une nouvelle vague de mérozoïtes dans le sang : ces cas particuliers de rechute sont appelés des reviviscences. Dans le sang, les mérozoïtes pénètrent dans les GRs et se différencient d'abord en forme « *ring* », puis s'épaississent et mûrissent pendant environ 12 à 16 heures en jeunes trophozoïtes. Ces formes se développent progressivement en trophozoïtes matures après 16 heures supplémentaires. La croissance du parasite dans le GR s'accompagne d'une digestion massive de l'hémoglobine qui a lieu dans une organite semblable à un lysosome, la vacuole digestive³¹. 10 à 15% des acides aminés libérés par la dégradation de l'hémoglobine dans cette vacuole sont utilisés par le parasite pour sa propre synthèse protéique³². La dégradation de l'hémoglobine libère cependant des molécules d'hème Fe^{2+} toxiques pour le parasite, rapidement oxydées en molécules d'hème Fe^{3+} non solubles qui peuvent altérer la membrane vacuolaire. Pour empêcher cette toxicité, les parasites disposent d'un procédé permettant la bio-minéralisation dans la vacuole digestive des hèmes Fe^{3+} pour former des cristaux inertes et non toxiques, qu'on appelle hémotoïne³³. Les trophozoïtes âgés se transforment en schizontes matures. Après la rupture de la membrane du GR, quelques mérozoïtes (10 à 30) capables d'infecter de nouveaux GRs sont libérés dans le sang. La durée de ce cycle est estimée à 48 heures chez *P. falciparum*, bien qu'elle soit variable selon les espèces (**Tableau 1**). Ainsi, la charge parasitaire au sein de l'hôte augmente graduellement, débouchant sur l'apparition des premiers symptômes de la maladie²⁰.

Tableau 1 : Comparaison de différentes espèces de *Plasmodium* infectant l'Homme.

Espèce	Durée du cycle érythrocytaire	Spécificités	Génome de référence	Taille du génome nucléaire (Mb)	Contenu en GC (%)
<i>P. falciparum</i>	~ 48h	—	3D7 ³⁴	23,3	19,4
<i>P. vivax</i>	~ 48h	Phase de dormance (hypnozoïtes)	Sal1 ³⁵	26,8	42,3
<i>P. knowlesi</i>	~ 24h	Infecte aussi bien le singe que l'Homme	H strain ³⁶	23,5	37,5
<i>P. ovale curtisi</i>	~ 48h	Phase de dormance (hypnozoïtes)	Poc1 ³⁷	34,5	28,5
<i>P. ovale wallikeri</i>	~ 48h	Phase de dormance (hypnozoïtes)	Pow1 ³⁷	35,2	28,9
<i>P. malariae</i>	~ 72h	—	Pm ³⁷	31,9	24,7

Certains trophozoïtes vont subir une maturation plus longue – une dizaine de jours – sans division nucléaire et accompagnée d'une différenciation sexuée. Ces trophozoïtes vont ainsi former des gamétocytes immatures à finalité mâle ou femelle. Ils poursuivront leur maturation dans la moelle osseuse, puis ressortir dans le sang une fois mûres avant d'être potentiellement ingérés par un moustique lors d'un nouveau repas sanguin. Une fois absorbés par le moustique, les gamétocytes se transforment en gamètes mâles et femelles qui, après leur fécondation, forment des œufs mobiles (ou ookinètes) situés dans le tube digestif de l'anophèle. Ces derniers vont ensuite migrer vers la paroi externe de l'estomac pour se transformer en oocystes. Dès lors, la méiose se déroule, produisant des sporozoïtes haploïdes qui vont se multiplier par mitose

toujours au sein des oocystes. Les sporozoïtes se dirigent alors vers les glandes salivaires du moustique, prêts à être inoculés à un nouvel hôte lors d'un repas sanguin²⁰.

Les espèces de *Plasmodium* comprennent trois génomes distincts : un génome nucléaire, un génome mitochondrial et un génome plastidique. À ce jour, de multiples génomes de référence ont été publiés (*P. falciparum* 3D7 en 2002³⁴ ; *P. vivax* Sal1 en 2008³⁵ ; ou encore *P. malariae* et *P. ovale curtisi* et *P. ovale wallikeri* en 2016³⁷) et révèlent une forte variabilité génomique inter-espèce (**Tableau 1**). Ces différents génomes se différencient en termes de longueurs et de contenu en guanine et cytosine (appelé *GC content* en anglais) parmi les *Plasmodium*, contribuant à la spécificité de chaque espèce. Concernant *P. falciparum*, le génome nucléaire est composé de 14 chromosomes comprenant près de 5 300 gènes, bien que 3 200 demeurent présumés³⁴.

1.1.4. Signes cliniques et méthodes de diagnostic

Les signes cliniques apparaissent lors de la phase érythrocytaire, à partir d'un seuil de parasitémie qui peut varier d'un hôte à l'autre selon différents paramètres. On distingue deux principales formes cliniques : les accès palustres simples et les accès palustres graves²⁰. Enfin, une troisième catégorie existe, courante dans la plupart des pays à forte endémicité : les porteurs asymptomatiques. Ces individus ont développé une immunité protectrice au cours des différentes expositions au paludisme, et ne manifestent plus forcément les symptômes de la maladie malgré le portage du parasite³⁸.

L'accès palustre simple est la forme la plus courante. Il se traduit par des frissons, une sensation de chaleur accompagnée de sueurs, survenant périodiquement (tous les 2 à 3 jours selon l'espèce infectieuse)²⁰. Les femmes enceintes et les individus non immuns tels que les jeunes enfants et les voyageurs peuvent subir des conséquences plus importantes sur leur organisme, débouchant sur des accès palustres graves. Il peut être observé des insuffisances rénales, des anémies sévères ou encore des hypoglycémies. Ces symptômes doivent être soignés dans les plus brefs délais afin d'éviter toute aggravation, pouvant aller jusqu'au décès de l'individu. Lorsqu'il y a atteinte cérébrale, on parle de neuropaludisme, souvent diagnostiqué par un coma, auquel peuvent s'ajouter des détresses respiratoires et des convulsions²⁰. Les causes liées à cet accès grave ne sont pas encore clairement définies, bien qu'une obstruction des capillaires sanguins provoquant une anoxie du cerveau soit la principale hypothèse avancée. Un des « dangers » liés au paludisme est la possible rapidité du passage de la forme simple à la forme sévère. Un diagnostic et un traitement rapidement mis en place sont donc essentiels pour éviter des complications.

En cas de suspicion clinique d'infection par *Plasmodium*, divers tests biologiques ont été développés afin d'établir un diagnostic et ainsi proposer un traitement adapté. Ces tests peuvent être classés selon trois catégories : la microscopie, la réaction en chaîne par polymérase (PCR,

pour *Polymerase Chain Reaction*) et le test de diagnostic rapide (TDR)^{24,39}. Une description des différents tests est fournie dans l'**encadré 2**.

Encadré 2 : Les différentes techniques de diagnostic biologique du paludisme.

Méthodes microscopiques^{39,40} : Ce sont toujours les méthodes officielles de référence. Le frottis sanguin est une technique rapide consistant à étaler en couche monocellulaire une goutte de sang sur une lame afin de détecter directement la présence de parasites après une coloration (Giemsa) avec un microscope à objectif x100. La méthode présente toutefois ses limites dans le cas de faibles parasitémiés, en plus de nécessiter un équipement relativement simple mais qui peut faire défaut dans certaines régions endémiques (microscope, électricité, personnel formé). Il peut également être réalisé une goutte épaisse (la goutte de sang n'est pas étalée), concentrant les parasites. Le seuil de détection est plus bas (~20 parasites/μL), mais l'identification de l'espèce s'avère beaucoup plus difficile.

Réaction en chaîne par polymérase (PCR)³⁹ : L'amplification de gènes parasitaires est une technique sensible (1 à 5 parasites/μL, voire moins selon les protocoles utilisés) permettant d'identifier l'espèce infectante. Elle est très utile pour confirmer les résultats obtenus par microscopie et aussi pour identifier le portage asymptomatique du parasite. Cependant, la détection d'ADN parasitaire dans le sang ne signifie pas obligatoirement une infection active. Cette technique demeure assez rare dans de nombreux pays endémiques en raison de son coût, des besoins matériels importants et la nécessité d'avoir un personnel formé.

Tests de diagnostic rapide (TDR)³⁹ : Afin de prendre en charge le patient rapidement, des tests de diagnostic rapide ont vu le jour, pouvant être réalisés à partir d'une goutte de sang (~10 μL sont nécessaires) par une technique d'immunochromatographie avec un anticorps dirigé contre un antigène parasitaire. Les principaux antigènes parasitaires utilisés sont l'Histidine-Rich Protein 2 (HRP2), la Lactate DeHydrogenase (LDH) et la Fructose-biphosphate aldolase (aussi appelée aldolase)⁴¹. La positivité du test doit toutefois être confirmée quand cela est possible par observation microscopique. Le résultat du test, qu'il soit positif ou négatif, est obtenu dans les 15 à 30 minutes suivant le dépôt de l'échantillon sanguin. Des variations du test existent afin de détecter des infections monospécifiques (soit *P. falciparum*, soit *P. vivax*) ou mixtes (*P. falciparum*, *P. vivax*, *P. ovale* spp. et *P. malariae*)⁴².

1.1.5. De la prévention aux traitements curatifs

Les efforts de lutte contre le paludisme n'ont cessé d'évoluer ces dernières années. La prévention a notamment contribué à une réduction significative du nombre de cas chaque année (**Figure 5**). Elle cible principalement le moustique et vise à l'empêcher de piquer l'Homme. Un des aspects majeurs de la prévention s'articule autour d'un contrôle strict des zones d'eaux stagnantes et marécageuses. Dans un second temps, la pulvérisation contrôlée d'insecticides comme le DDT (pour dichlorodiphényltrichloroéthane) a permis d'avoir une

meilleure maîtrise des populations d'anophèles⁴³. La distribution à très grande échelle de moustiquaires imprégnées d'insecticide représente une avancée majeure dans le contrôle du paludisme, et serait la mesure principalement responsable de la diminution des cas observée sur la période 2000-2015⁴⁴. En effet, les moustiques effectuant leur repas sanguin préférentiellement du crépuscule à l'aurore, leur installation permet aux habitants de limiter la possibilité d'une piqûre durant le sommeil. En revanche, l'utilisation massive des insecticides favorise l'émergence de moustiques résistants, requérant une vigilance et un suivi constants de la part des organismes de santé publique. Enfin, des traitements préventifs intermittents pour les nourrissons (uniquement dans les zones à transmission saisonnière) et les femmes enceintes ont été mis en place, consistant en la prise espacée dans le temps de doses curatives d'antipaludiques afin de prévenir une infection pendant la période à risque^{45,46}.

Les modalités de traitements prophylactiques utilisés par les voyageurs se rendant en zone d'endémie ne sont actuellement pas recommandées pour les populations vivant en zone d'endémie³⁹.

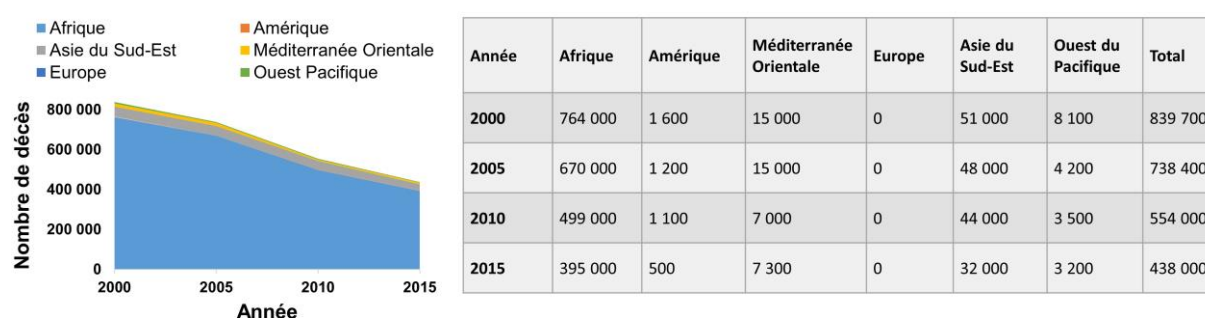


Figure 5 : Évolution du nombre de décès lié au paludisme entre 2000 et 2015. Une diminution drastique du nombre de décès a été observée lors de la dernière décennie, principalement grâce à l'élaboration et l'application des consignes de prévention prodiguées par l'OMS. La majeure partie des décès se concentre en Afrique. La figure a été adaptée à partir du travail de Roser et Ritchie⁴⁷.

Une fois le diagnostic de paludisme posé, diverses molécules – qu'elles soient naturelles ou synthétiques, et appartenant à diverses classes chimiques – peuvent être utilisées pour éliminer les parasites (**Tableau 2**). L'antipaludique prescrit dépend de l'espèce infectieuse, de la sévérité de la maladie (accès simple ou forme sévère) et de l'origine géographique du parasite³⁹. Par exemple, les infections à *P. vivax*, *P. ovale* spp. et *P. malariae* sont traitées avec la chloroquine (CQ), cette dernière n'étant plus indiquée pour *P. falciparum* en raison de sa résistance à ce composé. Dans la mesure où notre sujet porte sur *P. falciparum*, nous présenterons dans cette section clinique uniquement les molécules et les traitements relatifs à cette espèce.

Lors d'un accès palustre simple et en absence de vomissements persistants qui imposeraient un traitement intraveineux (par artésunate ou quinine), on privilégie une option par voie orale. Aujourd'hui, plusieurs traitements sont couramment prescrits et recommandés par l'OMS. En

zone d'endémie, le traitement par voie orale de première intention repose sur des bithérapies à base d'artémisinine (ART) qu'on appelle communément combinaisons thérapeutiques à base d'ART (CTAs), que nous détaillerons dans la seconde grande section de cette partie introductive. En revanche, chez la femme enceinte, la quinine ou l'artéméther-luméfamtrine (uniquement aux deuxième et troisième trimestres) doivent être privilégiés d'après l'OMS afin d'éviter de potentiels effets indésirables sur le fœtus³⁹.

Tableau 2 : Liste des différentes molécules antipaludiques.

Molécule antipaludique	Classe chimique et utilisation
4-aminoquinolés	
Chloroquine (CQ)	Autrefois utilisée en monothérapie comme traitement de première intention pour les infections <i>P. falciparum</i> ; toujours utilisée comme traitement de première intention contre <i>P. vivax</i> et les autres espèces.
Amodiaquine	Utilisée comme molécule partenaire avec l'artésunate en tant que CTA contre les infections à <i>P. falciparum</i> .
Pipéraquine (PPQ)	Utilisée comme molécule partenaire avec la dihydroartémisinine (DHA) en tant que CTA contre les infections à <i>P. falciparum</i> .
Pyronaridine	Utilisée avec l'artésunate (CTA) contre les infections à <i>P. falciparum</i> .
8-aminoquinolés	
Primaquine	Capables d'éliminer les gamétocytes de <i>P. falciparum</i> et les formes dormantes pour les espèces <i>P. vivax</i> et <i>P. ovale</i> spp. ; également utilisées en tant que traitements prophylactiques.
Tafénoquine	
Arylamino alcools	
Quinine	Utilisée en 2 ^{de} intention contre les formes simples et sévères de paludisme à <i>P. falciparum</i> .
Méfloquine	Utilisée avec l'artésunate (CTA) contre les infections à <i>P. falciparum</i> .
Luméfamtrine	Utilisée avec l'artéméther (CTA) contre les infections à <i>P. falciparum</i> .
Antifolates	
Pyriméthamine	Utilisée en combinaison avec la sulfadoxine contre les infections à <i>P. falciparum</i> .
Proguanil / cycloguanil	Utilisée en combinaison avec l'atovaquone en tant que traitement prophylactique ou curatif.
Sulfadoxine	Utilisée en combinaison avec la pyriméthamine contre les infections à <i>P. falciparum</i> .
Dérivés de l'artémisinines (ARTDs)	
Dihydroartémisinine (DHA)	Utilisée avec la pipéraquine (CTA) contre les infections à <i>P. falciparum</i> .
Artéméther	Utilisée avec la luméfamtrine (CTA) contre les infections à <i>P. falciparum</i> .
Artésunate	Utilisée avec l'amodiaquine, la méfloquine ou la pyronaridine (CTA) contre les infections à <i>P. falciparum</i> ; également utilisée contre les formes sévères de la maladie par voie intraveineuse.
Hydroxynaphtoquinone (analogue de l'ubiquinone)	
Atovaquone	Utilisée en combinaison avec le proguanil en tant que traitement prophylactique ou curatif.
Tétracycline	
Doxycycline	Utilisée en tant que traitement prophylactique ; également utilisée en curatif en combinaison avec la quinine.

Dans le cas d'accès palustres graves, une hospitalisation en urgence en soins intensifs ou en réanimation est nécessaire, avec comme traitement de première intention l'artésunate par voie intraveineuse avec relais par voie orale dès que possible avec une CTA^{39,48}. Cette thérapie est

préférée à la quinine en raison de son action très rapide sur la baisse de la parasitémie et de son mécanisme d'action inhibant la cytoadhérence¹. En cas de contre-indication – selon les antécédents du patient – une administration de quinine par voie intraveineuse est privilégiée³⁹.

Enfin, pour les voyageurs désirant se rendre dans des zones d'endémie palustre, trois traitements prophylactiques sont recommandés : l'atovaquone-proguanil, la doxycycline et la méfloquine³⁹.

1.1.6. Historique d'utilisation des antipaludiques et d'apparition des résistances

L'une des barrières majeures à l'efficacité des traitements antipaludiques réside dans l'apparition de parasites résistants aux molécules thérapeutiques^{49,50}. Suite à leur utilisation massive, les parasites développent des stratégies adaptatives afin de survivre aux thérapies. Cette notion de résistance fut pleinement investiguée chez *P. falciparum* depuis l'émergence et la propagation de parasites résistants à la CQ, autrefois utilisée comme traitement de première intention. En effet, après une décennie d'utilisation massive (1945 à 1957), la CQ a vu son efficacité diminuer en raison de la mise en place par le parasite d'un mécanisme d'efflux, dont le déterminant génétique a été identifié en 2000^{51,52}. Nous n'entrerons toutefois pas dans le détail des mécanismes de résistance pour chaque molécule thérapeutique, mis à part ceux de l'ART et de la CQ (et des molécules chimiquement voisines : quinoléines), respectivement dans les sections 2 et 3 de la partie introductive. L'historique de la politique d'utilisation des antipaludiques et les émergences successives de résistance sont présentés en **figure 6**. L'émergence de parasites résistants aux antipaludiques apparaît suivant une durée d'utilisation plus ou moins longue et le plus souvent d'abord en Asie du sud-est, avant de s'étendre – par migration ou émergence indépendante – dans les autres zones d'endémie incluant l'Afrique⁵³.

Ainsi, la CQ a progressivement laissé place à l'amodiaquine à la fin des années 1950 en Asie du sud-est, qui a été utilisée environ deux décennies avant de voir son efficacité totalement remise en question. Durant l'utilisation massive d'amodiaquine, la bithérapie sulfadoxine-pyriméthamine fut brièvement introduite avant d'être retirée en raison d'une faible efficacité liée à des résistances. En 1977 est officiellement conseillée la méfloquine qui, après seulement cinq années d'utilisation, ne fut plus recommandée en raison de l'émergence de parasites résistants à cette molécule⁵⁴. À la fin des années 1990 en Asie et au début des années 2000 en Afrique sont recommandées par l'OMS les CTAs. Tout comme les autres antipaludiques, des parasites résistants aux dérivés de l'ART (ARTDs) ont émergé, malgré leur utilisation en bithérapies, et se propagent depuis quelques années maintenant en Asie du sud-est⁵⁵. Les différentes émergences de résistance aux antipaludiques ont historiquement contribué à une

¹ Adhérence des GRs infectés par des parasites matures sur les cellules endothéliales (tapissant la face interne des vaisseaux sanguins).

augmentation significative du nombre de décès liés au paludisme, principalement en Afrique. Pour l'heure, les CTAs semblent rester efficaces en Afrique.

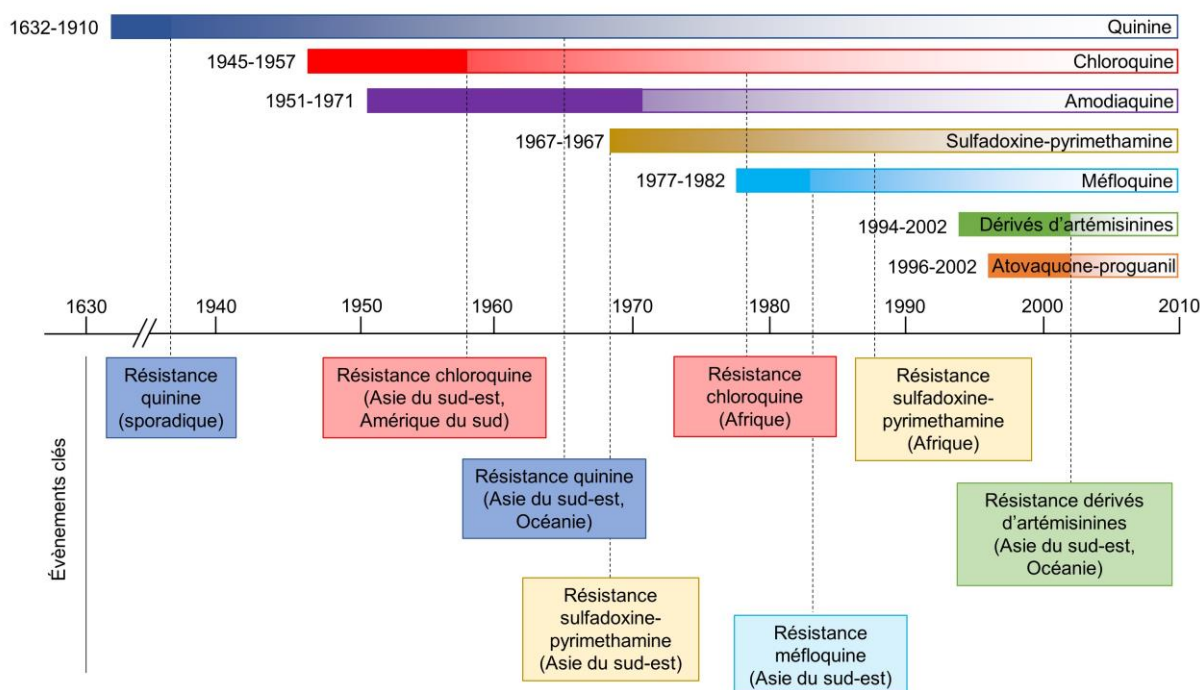


Figure 6 : Émergence de résistance aux principaux antipaludiques utilisés en zone d'endémie palustre. Chaque barre horizontale colorée représente un antipaludique (prescrit en monothérapie ou bithérapie). Les dates à gauche des barres indiquent l'année d'introduction du médicament et l'année du premier cas de résistance rapporté en zone d'endémie palustre. Aujourd'hui, les CTAs représentent les traitements de première intention, malgré l'émergence et l'expansion de la résistance aux ARTDs, uniquement en Asie du sud-est et en Océanie pour le moment⁵⁶. La figure est adaptée du travail de Eklund et Fidock (2008)⁵⁴.

Le présent travail de thèse vise à approfondir nos connaissances sur la fonction de protéines impliquées dans la résistance aux ARTDs et à diverses quinoléines (CQ et pipéraquine ou PPQ) par l'utilisation d'approches bioinformatiques. Nous consacrerons les deux prochaines grandes sections à décrire l'état des connaissances actuelles sur le mécanisme d'action des ARTDs et des quinoléines, et des protéines impliquées dans ces phénomènes de résistance.

- Le paludisme est causé par des parasites eucaryotes unicellulaires appartenant au genre *Plasmodium*, dont six espèces infectent l'Homme : *P. falciparum*, *P. vivax*, *P. knowlesi*, *P. malariae*, *P. ovale curtisi* et *P. ovale wallikeri*.
- *P. falciparum* est l'espèce la plus largement répandue, présente dans l'ensemble des zones d'endémie palustre et responsable de la majorité des cas cliniques et des formes mortelles.
- Le paludisme cause encore aujourd'hui 435 000 décès recensés en 2017 selon l'OMS.
- La transmission du paludisme se fait par un vecteur, un moustique femelle du genre *Anopheles*, lequel inocule les parasites à l'hôte humain lors d'un repas sanguin.
- Le cycle du parasite implique une phase sexuée chez l'anophèle et une phase asexuée chez l'Homme.
- La phase asexuée comprend une phase hépatique asymptomatique puis une phase sanguine pendant laquelle peuvent se manifester les symptômes de la maladie.
- La phase sanguine comprend un cycle érythrocytaire au cours de laquelle les parasites se multiplient ; une faible proportion d'entre eux se différencie sans se multiplier en formes érythrocytaires transmissibles au moustique, les gamétocytes.
- Diverses méthodes préventives comme les moustiquaires imprégnées d'insecticide ont permis de réduire considérablement le nombre de décès annuel au cours des deux dernières décennies. Des médicaments à titre préventif sont également disponibles pour certaines populations à risque plus élevé de paludisme grave.
- Divers médicaments à titre curatif ont été introduits au cours du temps pour soigner le paludisme...
- ... mais le parasite a systématiquement développé des résistances à ces médicaments.
- Le traitement de première intention actuel pour les formes non compliquées de la maladie repose sur une bithérapie composée d'une molécule ARTD couplée à un autre antipaludique appelé *molécule partenaire*.
- Des parasites résistants aux ARTDs et à certaines molécules partenaires ont émergé en Asie du sud-est ces 15 dernières années et menacent les progrès réalisés au cours de la dernière décennie dans la lutte contre le paludisme.

L'ART, de sa découverte à la chimiorésistance des parasites

Section 1.2



Même sans espoir, la lutte est encore un espoir.

— **Romain Rolland** (1866 – 1944)

Les ARTDs constituent, combinés avec un autre antipaludique, les traitements de première intention contre les formes simples du paludisme. Ces bithérapies sont appelées CTAs. Dans cette partie, nous rappellerons l'historique de la découverte de l'ART et de ses dérivés, puis de son mode d'action. Des parasites résistants à cette molécule ont émergé en Asie du sud-est au cours des années 2000, mettant potentiellement en péril les nombreux progrès réalisés jusqu'alors. Cette résistance, caractérisée cliniquement par un retard de clairance parasitaire chez le patient traité, est due à des mutations non-synonymesⁱ au sein du gène *P. falciparum k13* (*pfk13*)ⁱⁱ. Les connaissances actuelles concernant la fonction physiologique de la protéine PfK13 seront explicitées.

ⁱ Mutation nucléotidique qui entraîne un changement de l'acide aminé sur la séquence protéique.

ⁱⁱ Nous faisons référence au gène *k13* lorsqu'on parle du gène indépendamment de l'espèce qui le porte, tandis que *pfk13* est spécifiquement axé sur l'espèce *P. falciparum*.

1.2.1. Découverte et historique de l'ART

Lors du début de la guerre du Vietnam en 1955, l'armée nord-vietnamienne recensa un nombre inquiétant de soldats infectés par le paludisme. En effet, l'élaboration de réseaux souterrains favorisa le cumul de l'eau de pluie, offrant aux anophèles un lieu propice à la reproduction. De cela fut né, le 23 mai 1967, le projet secret « 523 » avec pour objectif premier de découvrir une solution efficace contre le paludisme à partir de la médecine traditionnelle chinoise⁵⁷. Près de 500 chercheurs répartis à travers 60 instituts scientifiques ou militaires participèrent à son accomplissement suite aux ordres promulgués par leur chef militaire, Mao Zedong. Diverses équipes ont été constituées : une partie s'est concentrée sur le développement de combinaisons thérapeutiques supplémentaires ; d'autres scientifiques passèrent au crible quelques dizaines de milliers de molécules connues ; tandis qu'un dernier groupe s'attela à la littérature chinoise et entreprit plusieurs expéditions pour recueillir diverses herbes ou plantes, certaines utilisées dans des villages reculés d'Asie pour traiter la fièvre. Une équipe dirigée par Tu Youyou (**Figure 7A**) – récompensée en 2015 par le prix Nobel de physiologie et de médecine pour l'ensemble de ses travaux liés au paludisme – découvrit, peu de temps après le lancement du projet le « qinghao », une décoction utilisée dès les premiers signes de la maladie dans un village pourtant peu touché par ce trouble. Rapidement, le groupe détermina la plante associée à ce traitement, identifiée comme une armoise annuelle : *Artemisia annua*⁵⁸ (**Figure 7B**).

Les principes actifs furent isolés entre 1972 et 1973, parmi lesquels figure l'ART (**Figure 7C**), dont l'activité antipaludique a été investiguée par le pharmacologue et médecin Li Shizhen⁵⁸. Les premiers essais en laboratoire furent spectaculaires et démontrèrent des résultats encourageants dans l'élimination rapide des parasites liés au paludisme⁵⁷.

Malgré la fin de la guerre du Vietnam en 1976 et l'avortement définitif du projet « 563 » en 1981, les travaux cliniques sur l'ART ont continué, la molécule représentant une aubaine pour les régions d'endémie palustre^{58,59}. La structure de l'ART, obtenue par diffraction aux rayons X, révèle la présence d'un pont endopéroxyde, lequel lie de façon covalente deux atomes d'oxygène sur l'un des trois cycles de la molécule. Ces connaissances permirent de concevoir les ARTDs semi-synthétiques tels que l'artésunate et l'artéméther en 1987, et la DHA en 1992 (**Figure 7C**)⁶⁰. Pourtant, bien que les propriétés indéniables de l'ART dans la lutte contre le paludisme soient ouvertement reconnues depuis le début des années 1970, l'OMS n'inclut les ARTDs et les CTAs qu'au début des années 2000 comme traitements essentiels. Fin 2014, 81 pays ont adopté les CTAs comme traitements de première intention contre les accès palustres simples^{24,50}.

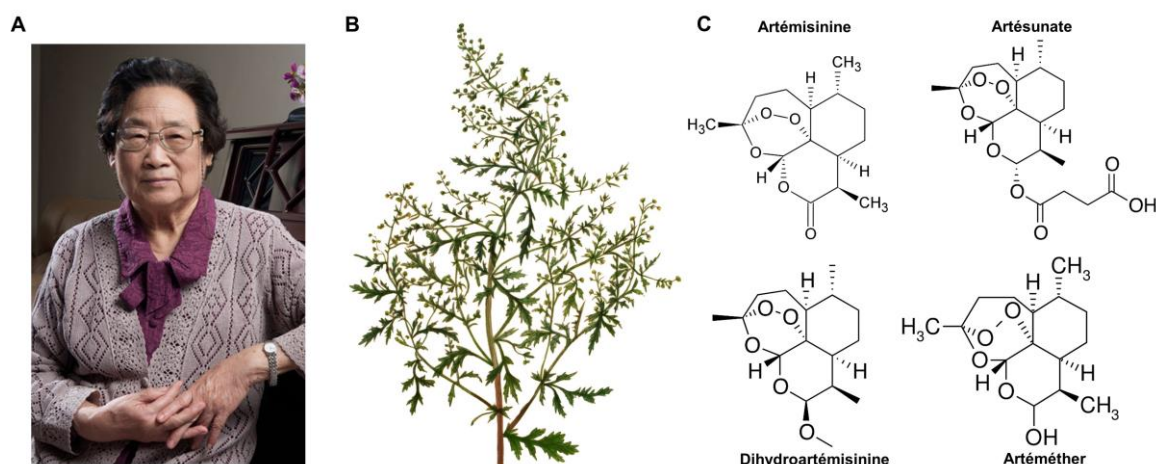


Figure 7 : L'ART, extraite d'une armoise annuelle, *Artemisia annua*, découverte par Tu Youyou. (A) Portrait de Tu Youyou, principale investigatrice dans la recherche sur *Artemisia annua*. Elle est récompensée en 2015 par le prix Nobel de physiologie et de médecine pour l'ensemble de ses travaux liés au paludisme. **(B)** Gravure de *Artemisia annua* (Kops *et al.* 1906). **(C)** Structures moléculaires de l'ART et de ses principaux dérivés semi-synthétiques (ARTDs).

1.2.2. Mode d'action des ARTDs

L'ART et l'ensemble de ses dérivés sont des lactones sesquiterpéniques⁶¹ avec une activité puissante contre la majorité des stades sanguins des parasites du genre *Plasmodium*^{62,63}, mais de courte durée du fait de leur demi-vie d'élimination plasmatique chez le patient comprise entre 1 et 2 heures. Notamment, les ARTDs ont un impact sur tous les stades érythrocytaires asexués (*rings*, trophozoïtes, schizontes) – lesquels provoquent les manifestations cliniques de la maladie – et les gamétocytes immatures. Cependant, les parasites au stade *ring* sont rapportés comme moins sensibles à l'ART, par comparaison avec les autres stades asexués⁶⁴. En outre, les différents stades sont sensibles aux ARTDs en raison de la digestion active de l'hémoglobine à mesure de leur développement dans les GRs. Il est aujourd'hui admis que le clivage du pont endopéroxyde de l'antipaludique par l'ion Fe^{2+} associé à l'hème est nécessaire à l'activation des ARTDs⁶¹. Ce clivage engendre la formation de dérivés réactifs qui ciblent les groupements nucléophilesⁱ des protéines et des lipides du parasite. Les récents travaux de Wang *et al.* (2015) ont démontré qu'environ 120 protéines du parasite se lient de façon covalente avec l'ART, la majorité d'entre elles étant impliquées dans des processus biologiques essentiels à la survie du parasite^{65,66}. Notamment, ces protéines auraient des rôles dans les processus de dégradation de l'hémoglobine, de la défense anti-oxydante, de la synthèse des protéines, et de la dégradation de protéines par le protéasome^{67,68}.

ⁱ Composé chimique possédant un excès de charges négatives et attiré par les espèces chargées positivement.

Chez les patients infectés, l'élimination des parasites par les ARTDs ne peut être étudiée que dans le sang périphérique, où les *rings* se développent et circulent au sein des GRs pour une durée comprise entre 12 et 16 heures avant de former les jeunes trophozoïtes séquestrés dans les microvaisseaux^{62,69}. Lorsque les *rings* sont exposés aux ARTDs, ils se condensent en formes pycnotiques (dégradation du noyau des cellules du parasiteⁱ), puis sont éliminées efficacement du flux sanguin sans destruction des GRs de l'hôte⁷⁰. Quant aux formes séquestrées (trophozoïtes, schizontes et gamétocytes immatures), elles sont éliminées *in situ* dans les microvaisseaux après exposition aux ARTDs. Il est estimé que les ARTDs réduisent par un facteur 10^4 la densité parasitaire au cours des premières 48 heures suivant l'initiation du traitement, détruisant rapidement les *rings* avant qu'ils ne puissent se transformer en trophozoïtes matures et schizontes. Cette propriété pharmacologique est une spécificité remarquable des ARTDs car trophozoïtes matures et schizontes sont des formes intra-érythrocytaires particulièrement virulentes du fait qu'elles séquestrent dans la microvasculature et induisent des symptômes cliniques⁶⁹.

1.2.3. De la monothérapie d'ART aux CTAs

Au début des années 1990, nombreuses sont les études à démontrer une aggravation des phénomènes de résistance des parasites à différents antipaludiques utilisés en monothérapie (comme la CQ et l'association sulfadoxine-pyriméthamine qui est considérée comme une monothérapie par l'OMS)⁵⁴. Dans la mesure où la clairance des ARTDs dans le sang est extrêmement rapide et nécessite un nombre élevé de prises pouvant favoriser l'émergence de parasites résistants à ces molécules, l'OMS a recommandé en 2006 l'arrêt de leur utilisation en monothérapie au profit des bithérapies de type CTAs²⁴. Cinq CTAs sont actuellement utilisées : artésunate-sulfadoxine-pyriméthamine (ASSP), artéméter-luméfantrine (AL), dihydroartémisinine-pipéraquine (DP), artésunate-amodiaquine (ASAQ) et artésunate-méfloquine (ASMQ)^{24,39}. Une sixième bithérapie est en train d'être déployée, artésunate-pyronaridine (ASPY)⁷¹, alors que ASSP est presque abandonnée. Le mode d'action des différentes molécules partenaires aux ARTDs et une brève description de leur utilisation en bithérapie sont présentés dans l'**encadré 3**. Le principe sous-jacent des CTAs est que l'ARTD élimine la majorité des parasites en quelques heures après chaque dose quotidienne du traitement de 3 jours (contre 7 en monothérapie), tandis que la molécule antipaludique partenaire tue les parasites résiduels sur plusieurs jours – en raison d'une activité nettement plus longue que les ARTDs – via un autre mécanisme d'action^{72,73}. Cette multiplicité et complémentarité des modes d'action visent à éliminer complètement les parasites chez le patient et aussi limiter l'apparition de résistance. Par exemple, l'action rapide des ARTDs, par un simple effet de diminution rapide de la biomasse parasitaire, réduit considérablement le risque que les parasites restants développent une mutation conférant la résistance à la molécule partenaire. Inversement, la

ⁱ État d'altération proche de la mort cellulaire.

molécule partenaire protège l'apparition de résistance aux ARTDs. Pourtant, malgré les précautions prises par l'OMS, des phénomènes de résistance aux ARTDs ont été observés à la fin des 2000 en Asie du sud-est, plus récemment en Papouasie-Nouvelle Guinée⁵⁶ et possiblement en Inde⁷⁴⁻⁷⁶, relançant la crainte d'une nouvelle propagation de résistance aux CTAs à l'ensemble des zones d'endémie palustre⁵⁵.

Encadré 3 : Présentation des principales CTAs recommandées par l'OMS et mode d'action des molécules partenaires.

Artéméter-luméfantrine (AL) : Cette bithérapie est actuellement la plus largement utilisée dans les zones d'endémie palustre, représentant près de 75% des CTAs achetées (hors marché noir et trafic)²⁴. Le mode d'action de la luméfantrine est encore mal compris, mais elle interférerait avec la détoxification de l'hème⁷⁷ et/ou inhiberait directement la protéine *P. falciparum* multidrug resistance 1 (PfMDR1 ; transporteur impliquée dans de multiples résistances aux antipaludiques)⁷⁸. Si certaines études montrent une légère diminution de l'efficacité de cette bithérapie, notamment en Angola ou encore en Gambie⁷⁹, son utilisation n'est pour l'heure pas remise en cause. La diminution de la sensibilité à la luméfantrine serait associée à une variation du nombre de copies du gène *pfmdr1* (hypothèse non confirmée par des méta-analyses⁸⁰) et à la présence d'allèles sauvages des gènes *pfmdr1* et *pfcr1*⁸¹.

Artésunate-amodiaquine (ASAQ) : Cette bithérapie représente près de 20% des ventes totales des CTAs et est utilisée à ce jour uniquement en Afrique subsaharienne. L'amodiaquine est également utilisée en association avec la sulfadoxine-pyriméthamine en tant que traitement prophylactique chez les jeunes enfants durant la saison des pluies dans la région du Sahel³⁹. Le mode d'action de l'amodiaquine est similaire à celui de la CQ, c'est-à-dire que la molécule s'accumule dans la vacuole digestive du parasite à de fortes concentrations, empêche la détoxification de l'hème et conduit à la mort du parasite⁸². La résistance à l'amodiaquine est associée à des mutations non-synonymes des gènes *pfcr1* et *pfmdr1*, et plus spécifiquement celles conférant la résistance à la CQ⁸³.

Dihydroartémisinine-PPQ (DP) : Utilisée par certains pays d'Asie du sud-est et depuis peu en Afrique, la combinaison DP est également prescrite comme traitement prophylactique pendant la grossesse⁸⁴. Elle est d'ailleurs massivement utilisée dans les régions africaines proches d'une élimination totale du paludisme. Toutefois, l'utilisation de cette bithérapie est actuellement remise en question dans la sous-région du Grand-Mékong en raison de l'émergence de parasites résistants à la PPQ⁸⁵. L'échec du traitement fut observé pour la première fois dans l'ouest du Cambodge, atteignant 24,1% d'échecs thérapeutiques en 2010⁸⁶. Le mode d'action de la PPQ n'est pas entièrement élucidé, mais pourrait être similaire à celui de la CQ, c'est-à-dire s'accumulant dans la vacuole digestive et inhibant la voie de dégradation de l'hémoglobine⁸⁷. Récemment, la résistance à la PPQ a été associée avec une variation du nombre de copies des gènes *plasmepsin II* et *III*⁸⁸, ainsi que des mutations localisées dans le gène *pfcr1*^{87,89}. La résistance à la PPQ semble différente de celle à la CQ, malgré leur mode d'action similaire⁸⁹.

Artésunate-Méfloquine (ASMQ) : La combinaison ASMQ fut la première CTA à avoir été introduite en Asie du sud-est, limitant ainsi la propagation de la résistance à la méfloquine (jusqu'alors utilisée en monothérapie depuis les années 1990)⁹⁰. La bithérapie est actuellement utilisée en Asie du sud-est. Son utilisation n'a cessé de décroître au fil des années et se limite dorénavant à quelques pays en raison d'une résistance déjà présente à la méfloquine²⁴. Le mode d'action de la méfloquine n'est pas encore pleinement élucidé mais, tout comme la CQ et la PPQ, elle agirait par inhibition de la détoxification de l'hème⁷⁷, de la fonction physiologique de *pfmdr1*⁹¹, et/ou de la synthèse protéique via une interaction avec le ribosome cytoplasmique (*pf80S*)⁹². La résistance à la méfloquine est associée à une variation du nombre de copies du gène *pfmdr1*⁹³, et potentiellement à des mutations non-synonymes des gènes *P. falciparum* multidrug resistance protein 1 (*pfmrp1*) et 2 (*pfmrp2*)^{94,95}.

Artésunate-Pyronaridine (ASPY) : Pas encore officiellement recommandée par l'OMS, cette bithérapie a toutefois reçue une opinion positive de la part de l'Agence européenne des médicaments et devrait par conséquent être introduite sous peu. L'association des deux molécules a montré un taux d'efficacité élevé en Afrique, à un niveau similaire aux combinaisons DP et AL⁹⁶. Toutefois, l'efficacité a montré des limites dans la région ouest du Cambodge, probablement en raison de la résistance aux ARTDs et à une résistance croisée entre la pyronaridine et la PPQ⁹⁷. Le mode d'action de la pyronaridine est peu caractérisé : elle interférerait avec la formation de l'hémozoïne⁹⁸. Pour l'heure, la résistance à la pyronaridine n'a pas été clairement démontrée, mais des variations de la sensibilité à cette drogue pourrait être liées à des mutations non-synonymes du gène *pfcr*⁹⁹.

1.2.4. Émergence et caractérisation de la résistance aux ARTDs

La résistance aux ARTDs fut initialement rapportée en 2008 sur des parcelles de Battambang, province située à l'ouest du Cambodge¹⁰⁰. En mesurant la durée de clairance parasitaire suite à un traitement à base d'artésunate en monothérapie chez 60 patients atteints de formes simples du paludisme, il a été observé que pour deux d'entre eux, le temps nécessaire pour éliminer les parasites était significativement allongé (**Figure 8A**). La présence de parasites résistants dans la province de Pailin au Cambodge – démontré en 2009 via un procédé similaire⁵⁵ (**Figure 8B**) – confirma ces résultats initiaux. Depuis, les pays de la sous-région du Grand Mékong font l'objet d'incessantes investigations pour rapporter les cas de résistance : aujourd'hui, la résistance aux ARTDs s'est également propagée au Myanmar, au Vietnam, à la République démocratique populaire Lao – communément appelé Laos – et à la Chine⁵⁰. Quelques cas de résistance ont été observés en Inde^{74,75}, tandis qu'ils restent extrêmement rares en Afrique¹⁰¹. Cependant, cette émergence de résistance initiale en Asie du sud-est n'est pas sans rappeler les précédents scénarios d'émergence de résistance pour les premières générations d'antipaludiques comme la CQ ou la sulfadoxine-pyriméthamine, lesquels ont vu leurs résistances respectives apparaître en Asie du sud-est avant de se propager ou d'émerger de façon indépendante en Afrique plusieurs années plus tard⁵⁴ (cf. section 1.1.6, **Figure 6**). En conséquence, il convient de mener des actions

urgentes pour prévenir et lutter contre l'apparition de résistance, notamment dans les pays les plus vulnérables comme ceux de l'Afrique subsaharienne.

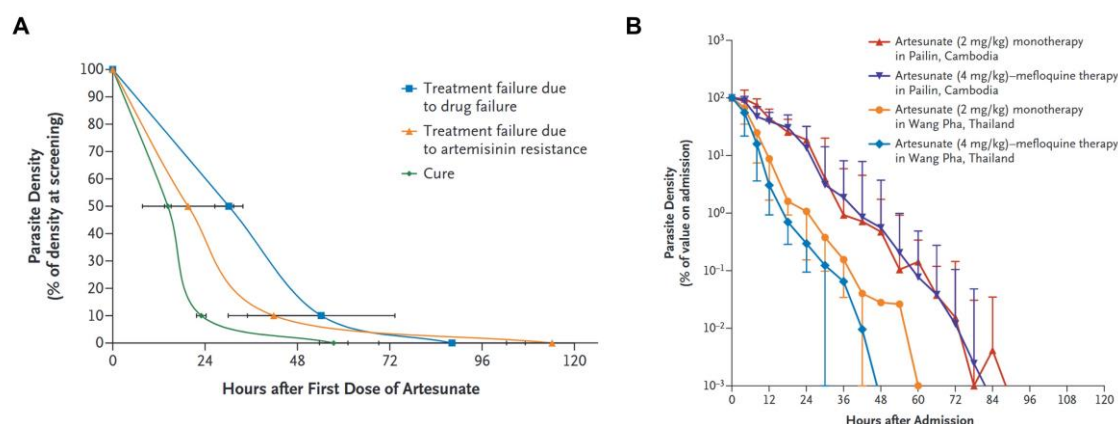


Figure 8 : Durée de clairance parasitaire pour des patients traités par monothérapie ou bithérapie à base d'ARTD. (A) Évolution des densités parasitaires 120 heures après la première prise d'artésunate chez un patient guéri (courbe verte) et deux patients présentant un échec au traitement du fait d'une infection résistante à l'artésunate (courbes jaune et bleu). La figure a été reprise sans modifications du travail de Noedl et ses collaborateurs (2008)¹⁰⁰. **(B)** Évolution des densités parasitaires médianes (normalisées en log₁₀). Diverses modalités de traitement dans différentes zones d'endémie palustre ont été appliquées afin de mesurer la variation de la durée d'élimination des parasites. La clairance parasitaire est significativement allongée pour les patients vivant à Pailin, au Cambodge (courbes rouge et violette). La figure a été tirée sans modifications du travail de Dondorp *et al.* (2009)⁵⁵.

Cliniquement, la résistance aux ARTDs se traduit par une réduction de la vitesse d'élimination des parasites chez le patient traité (mesurée par l'estimation de la demi-vie de clairance parasitaire) que ce soit en monothérapie ou en CTAs. Selon la définition de l'OMS, la résistance clinique aux ARTDs est avérée lorsque la demi-vie de clairance parasitaire est supérieure à 5 heures⁵⁰. Cette méthodologie s'avère à la fois longue et coûteuse car basée sur le suivi clinique des patients avec notamment plusieurs prélèvements de sang dans les 24 premières heures après initiation du traitement. Ce suivi requiert d'importants moyens logistiques et la demi-vie de clairance parasitaire peut être biaisée par d'autres facteurs tels que l'observance et l'immunité partielle du patient¹⁰². Par la suite, une méthodologie *in vitro* / *ex vivo* développée en 2013 et nommée « Ring-stage Survival Assay » (RSA_{0-3h}) a été proposée afin de détecter la résistance aux ARTDs¹⁰³. Cette technique mesure le taux de parasites survivants suite à une courte exposition (6 heures) des stades précoces érythrocytaires (*rings*) à de hautes concentrations d'ARTD (exemple : 700 nM de DHA, classiquement observée dans le plasma des patients traités). On estime qu'un taux de survie des parasites supérieur à 1% correspond à un phénotype de résistance suite à des corrélations *in vitro* / *in vivo*¹⁰³.

1.2.5. La découverte d'un déterminant moléculaire de la résistance aux ARTDs : le gène *pfk13*

Si la résistance aux ARTDs est aujourd'hui bien documentée en Asie du sud-est, elle n'est ni confirmée ni suspectée en Afrique, se limitant à une poignée de cas cliniques recensés^{50,101}. Un marqueur moléculaire fiable pour la surveillance rapide à large échelle de la résistance a été découvert par plusieurs équipes françaises. Des analyses génomiques comparatives entre une lignée résistante aux ARTDs (F32-ART5 ; générée *in vitro* par pression de sélection médicamenteuse en ART) par rapport à la souche sauvage initiale (F32-TEM), ont permis d'identifier huit mutations non-synonymes sur sept gènes différents de la souche F32-ART5 (**Figure 9A**)¹⁰⁴. Parmi ces mutations figure M476I sur le locus PF3D7_1343700, apparue après 30 cycles de pression de sélection médicamenteuse et associée à un RSA_{0-3h} supérieur à 1%, suggérant son implication dans la résistance aux ARTDs. Après génotypage de 49 isolats parasitaires collectés entre 2010 et 2011 au Cambodge et adaptés en culture, plusieurs mutations non-synonymes situées sur le locus PF3D7_1343700 ont été répertoriées et associées à un RSA > 1% alors qu'elles sont absentes des isolats et souches de référence sensibles aux ARTDs (souches 3D7, 89F5 et K1992). Collectivement, ces résultats ont montré l'association de mutations de ce locus, dénommé par la suite *P. falciparum k13* (*pfk13*), avec la résistance aux ARTDs (**Figure 9B**)¹⁰⁴. Par la suite, des expériences de transfection sur *pfk13* ont révélé que la seule présence de ces mutations peut conférer la résistance aux ARTDs^{105,106}, permettant de considérer *pfk13* comme un marqueur, mais aussi comme un déterminant moléculaire de la résistance.

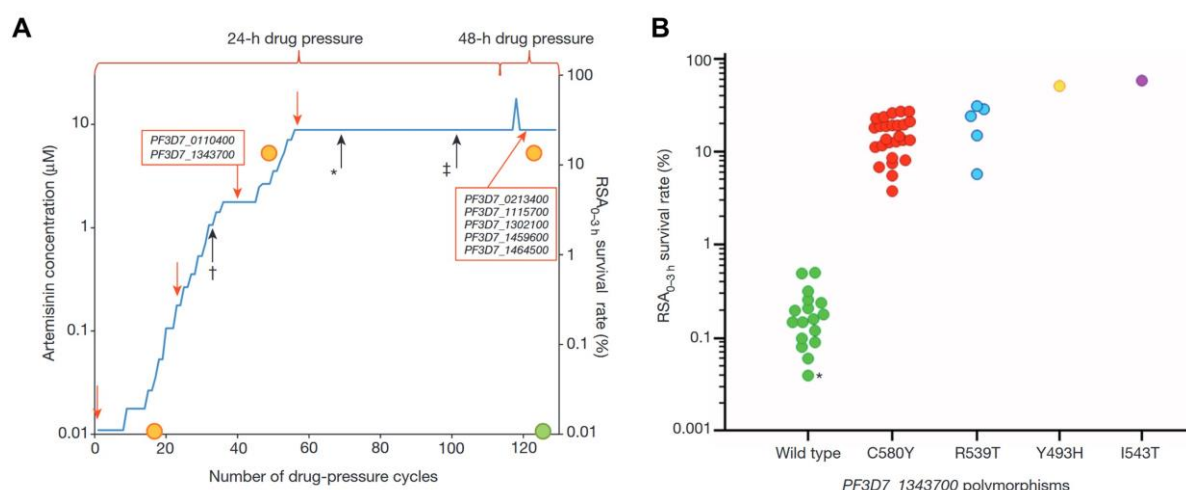


Figure 9 : Mise en évidence de mutations non-synonymes dans le gène *pfk13* comme marqueurs moléculaires de la résistance aux ARTDs. (A) Acquisition temporelle de mutations chez la souche F32-ART5, soumise à une pression de concentration croissante d'ART (échelle des ordonnées de gauche). Les loci mutés sont indiqués dans les encadrés rouges. Les cercles orange et vert indiquent le taux de survie RSA_{0-3h} (échelle des ordonnées de droite) pour les souches soumises (F32-ART5) ou non (F32-TEM) à une pression de sélection médicamenteuse, respectivement. La mutation non-synonyme sur le locus PF3D7_1343700 (gène *pfk13*) est apparue après le 30^{ème} cycle (symbole † sur la figure). (B) Taux de survie d'isolats parasitaires provenant du Cambodge et portant diverses mutations du gène *pfk13*. Un taux de survie RSA_{0-3h} supérieur à 1% est systématiquement observé pour les mutants *pfk13* par rapport aux isolats sauvages. Les figures ont été tirées du travail de Arieu et ses collaborateurs (2014)¹⁰⁴.

Le gène *pfk13* code la protéine PfK13 d'une longueur de 726 acides aminés et appartenant à la superfamille de protéines contenant un domaine en forme de propulseur ou hélice, communément appelé *propeller* (signifiant hélice en anglais) ou Kelch^{104,107,108}. Cette protéine contient trois domaines annotés : un domaine coiled-coil (c'est-à-dire deux hélices α parallèles qui s'enroulent entre elles ; acides aminés 212 à 341), un domaine BTB (pour Broad-Complex, Tramtrack and Bric-à-brac; acides aminés 350 à 437) et un domaine propeller de type Kelch (acides aminés 443 à 726 ; **Figure 10A**). La structure cristallographique a été résolue pour les domaines BTB et propeller dans un état dimérique (**Figure 10B**), accessible avec les identifiants PDB 4yy8 et 4zgc (avec une résolution de 1,81 Å et 2,5 Å, respectivement ; ces deux structures diffèrent également par la présence d'un pont disulfure entre les résidus d'acides aminés C532 et C580)^{109,110}. Le domaine propeller de PfK13 est composé de la répétition de six motifs kelch (aussi appelé *blade*) ; chaque motif kelch est long d'une cinquantaine d'acides aminés et forme un feuillet β impliquant quatre brins β antiparallèles^{107,108}. Ce domaine, que nous appellerons KREP (Kelch-REpeat Propeller) dans la suite de ce manuscrit de thèse, concentre la grande majorité des mutations de résistance aux ARTDs associées à un taux de survie RSA_{0-3h} supérieur à 1% et/ou à une clairance parasitaire retardée *in vivo*. Enfin, des expériences de mutagenèse ont suggéré le rôle essentiel du gène *k13* pour la croissance intra-érythrocytaire de *P. falciparum* et *P. berghei* lors de la phase asexuée^{111,112}.

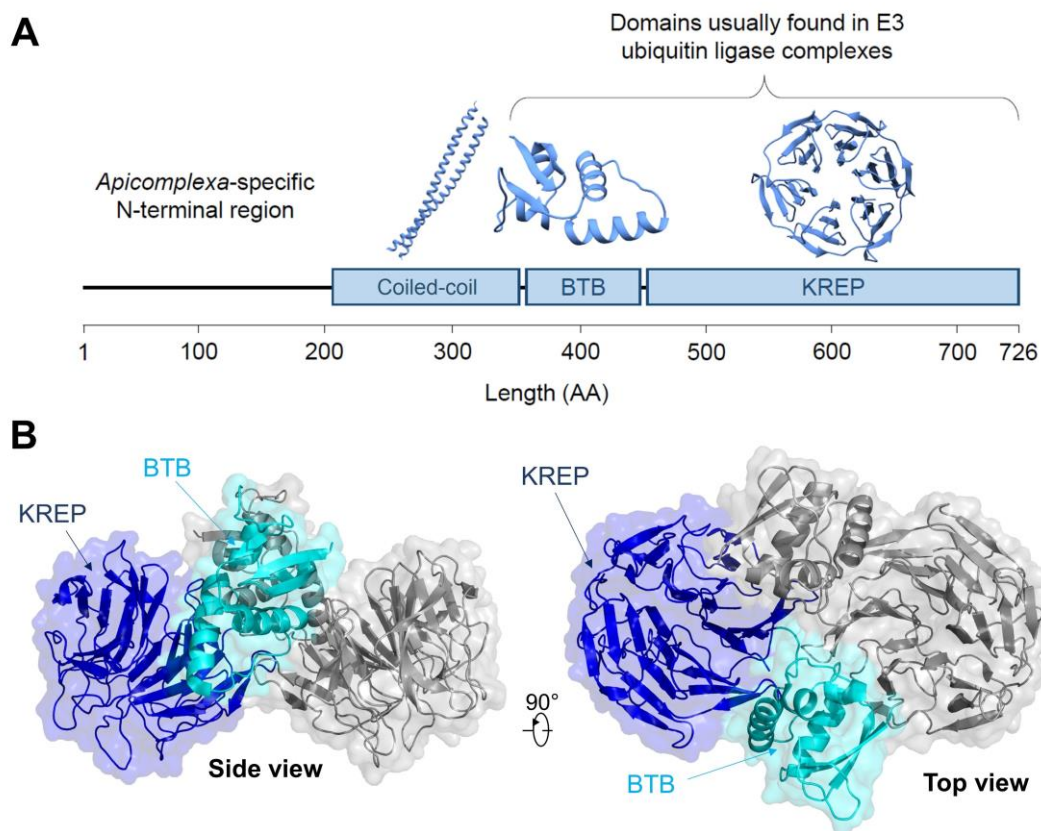


Figure 10 : Présentation de la protéine PfK13. (A) Annotation des domaines composant PfK13. La protéine contient une région N-terminale variable parmi les *Apicomplexa*, et trois domaines annotés : un domaine coiled-coil, un domaine BTB et un domaine KREP. Ces deux derniers domaines sont généralement retrouvés dans des

complexes d'ubiquitination, dans lesquels un substrat protéique est polyubiquitiné puis conduit en voie de dégradation. **(B)** Représentation dimérique des domaines BTB et KREP de PfK13. Les domaines BTB et KREP du premier monomère sont colorés en cyan et en bleu, respectivement ; le second monomère est représenté en gris. La structure est affichée du dessus (panel de gauche) et de côté (panel de droite).

En étudiant les séquences de *pfk13* obtenues à partir de 1 876 patients ayant contracté un paludisme non compliqué au cours de la période 2001-2014 le long de la frontière Myanmar-Thaïlande, Anderson *et al.* (2017) ont montré que la mutation E252Q (localisée sur le domaine coiled-coil de PfK13) prédominait initialement dans cette région et était associée à un retard de clairance légèrement allongé¹¹³. Au cours de ces dernières années, sa fréquence n'a cessé de décroître au profit de mutations localisées sur le domaine KREP. En 2018, une trentaine de mutations non-synonymes ont été associées à un phénotype de résistance dans diverses populations parasitaires d'Asie du sud-est⁵⁰. Catégorisées selon leur niveau de caractérisation par des approches cliniques et des expérimentations *in vitro* (**Encadré 4**), neuf mutations non-synonymes du gène *pfk13* ont été officiellement validées par l'OMS comme conférant la résistance aux ARTDs (F446I, N458Y, M476I, Y493H, R539T, I543T, P553L, R561H, C580Y). Vingt-deux mutations supplémentaires (P441L, G449A, D452E, C469F, C469Y, K479I, A481V, R515K, S522C, P527H, N537D, N537I, G538V, V568G, P574L, R575K, M579I, D584V, P667T, F673I, A675V, H719N), considérées comme mutations de résistance « candidates » par l'OMS (**Encadré 4**), ont également été associées à un retard de clairance dans des études cliniques, retrouvées à des fréquences très variables dans les populations naturelles. Ces listes sont évidemment appelées à être régulièrement actualisées en fonction des nouvelles investigations réalisées dans les régions endémiques. La localisation de l'ensemble de ces mutations sur les structures secondaire et tertiaire du domaine KREP de PfK13 est indiquée en **figure 11**.

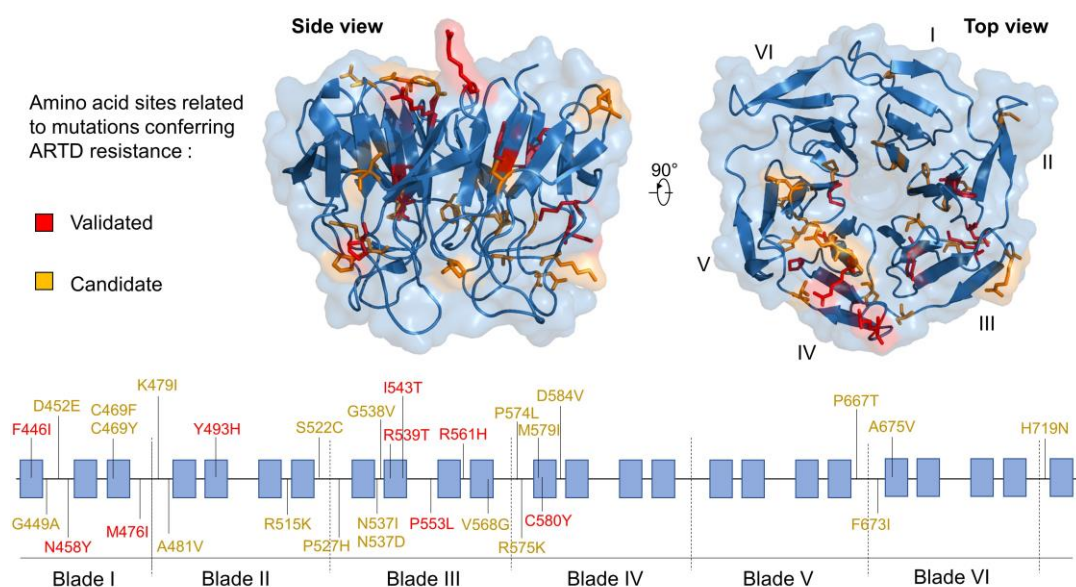


Figure 11 : Localisation des mutations de résistance aux ARTDs sur les structures secondaire et tertiaire du domaine KREP de PfK13. Les sites associés aux mutations de résistance aux ARTDs validées et candidates sont respectivement colorés en rouge et en orange. Les motifs kelch (ou *blades*) sont labellisés en chiffres romains (de I à VI) sur la structure tertiaire.

Encadré 4 : Catégorisation des mutations *pfk13* associées à la résistance aux ARTDs selon divers critères érigés par l'OMS.

Mutation de résistance validée : Une mutation de résistance est considérée comme validée lorsque celle-ci répond aux deux critères suivants :

- i) une association statistique ($P < 0,05$) entre une mutation et une demi-vie de clairance parasitaire supérieure à 5 heures chez des patients ;
- ii) une association statistique ($P < 0,05$) entre une mutation et une sensibilité réduite aux ARTDs par des tests *in vitro* (RSA_{0-3h}) à partir d'isolats parasitaires ; ou une association statistique ($P < 0,05$) entre une sensibilité réduite aux ARTDs et des parasites mutants *pfk13* produits *in vitro* par des techniques de transfection et d'édition du génome.

Mutation de résistance candidate : Une mutation de résistance est considérée comme candidate lorsque celle-ci est associée statistiquement ($P < 0,05$) à une demi-vie de clairance parasitaire supérieure à 5 heures *in vivo*, mais sans validation *in vitro*. Cette liste inclut également les mutations retrouvées à faible fréquence dans les populations naturelles, associées à un retard de clairance mais sans soutien statistique (en raison de faibles échantillonnages).

Trois études ont examiné la diversité moléculaire de *pfk13* de quelques milliers de parasites provenant d'Asie et d'Afrique¹¹⁴⁻¹¹⁶. Les mutations de résistance aux ARTDs décrites en Asie ne sont pas détectées en Afrique^{115,116}. De plus, une élévation significative du taux de substitution non-synonyme par rapport au taux de substitution synonymeⁱ pour le gène *pfk13* est uniquement observée en Asie, suggérant l'absence de sélection positiveⁱⁱ agissant sur ce gène en Afrique¹¹⁵. Si quelques mutations non-synonymes de *pfk13* sont rapportées en Afrique, celles-ci sont très rares et spécifiques à l'Afrique¹¹⁴⁻¹¹⁶. Cependant, de récentes études ont documenté des échecs thérapeutiques pour des patients porteurs de parasites mutants *pfk13* infectés en Afrique, dont la mutation de résistance A675V retrouvée aussi en Asie^{101,117}.

Jusqu'à récemment, un balayage sélectif modéré (ou *soft selective sweep* en anglais) était décrit comme le modèle de sélection initial agissant sur *pfk13* en Asie du sud-est, avec plusieurs mutations ciblées par la sélection positive coexistant à des fréquences plus ou moins élevées¹¹³. Ce balayage sélectif s'est intensifié dans lequel seules les lignées de parasites portant la mutation C580Y se sont propagées à travers la région est de l'Asie du sud-est¹¹³. La raison pour laquelle cette mutation est plus compétitive que les autres demeure inconnue. En effet, par modification génétique de *pfk13* au sein d'un isolat cambodgien adapté de longue date à la culture (souche Dd2), Straimer *et al.* (2015) ont montré par mesure du taux de survie RSA_{0-3h} que cette mutation n'est pas associée à un niveau de résistance supérieur par rapport à d'autres mutations *pfk13* (**Figure 12**)¹⁰⁵. Les mutations R539T et I543T sont associées à un taux de survie RSA_{0-3h} largement supérieur à 10%, contrairement à la mutation C580Y qui atteint 8%. Sachant que la protéine K13

ⁱ Mutation nucléotidique qui n'altère pas la séquence protéique (mutation silencieuse).

ⁱⁱ Se produit lorsqu'une mutation bénéfique survient dans une population et augmente en fréquence, généralement en raison d'un avantage sélectif.

est démontrée comme essentielle durant le cycle asexué érythrocytaire chez *P. falciparum* et *P. berghei*^{111,112}, un moindre *fitness cost*ⁱ pourrait être à l'origine du succès de C580Y par rapport aux autres mutations^{105,118}. Diverses études ont cependant révélé des résultats contradictoires. Des expériences de compétition *in vitro* avec des lignées isogéniques sauvage et mutantes *pfk13* (R539T, I543T ou C580Y) ont montré des déficits de croissance variables en fonction de la mutation et de la souche parasitaire testées (**Figure 13A**). Il a notamment été observé que C580Y avait un *fitness cost* supérieur à R539T lorsqu'elle est portée par la souche vietnamienne V1/S datant de 1976. En revanche, à partir de souches cambodgiennes récentes (Cam3.II et CamWT), C580Y est associée à un plus faible *fitness cost* par rapport aux autres mutations. Parallèlement à ce travail, Nair *et al.* (2018) ont comparé les *fitness cost* associés à C580Y et R561H¹¹⁹. Les mutations ont été introduites par la méthode CRISPR/Cas9 à partir d'un isolat parasitaire *pfk13* sauvage provenant de la bordure Thaïlande-Myanmar. Dans cette étude *in vitro*, C580Y est associée à un *fitness cost* plus élevé que R561H, et la mutation R561H remplace C580Y lorsque les deux types de parasites mutants sont placés en concurrence directe (**Figure 13B**). Des analyses supplémentaires sont donc nécessaires pour mieux comprendre la relation entre les mutations *pfk13*, le fond génétique et le *fitness* du parasite.

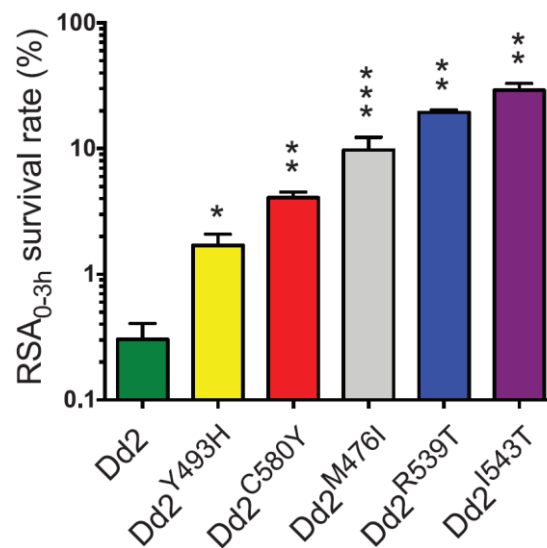


Figure 12 : Impact de différentes mutations *pfk13* sur le taux de survie RSA_{0-3h} à partir de la souche asiatique Dd2. Seuil de significativité des tests de Student par comparaison des mutants au sauvage Dd2 : * $P < 0,05$; ** $P < 0,01$; *** $P < 0,001$. La figure a été tirée du travail de Straimer *et al.* (2015)¹⁰⁵.

ⁱ Désigne la perte de *fitness* d'un organisme associé à une mutation (ou un génotype) donné. Le *fitness* (ou valeur adaptative en français) désignant la capacité d'un organisme à survivre et à se reproduire dans un environnement donné.

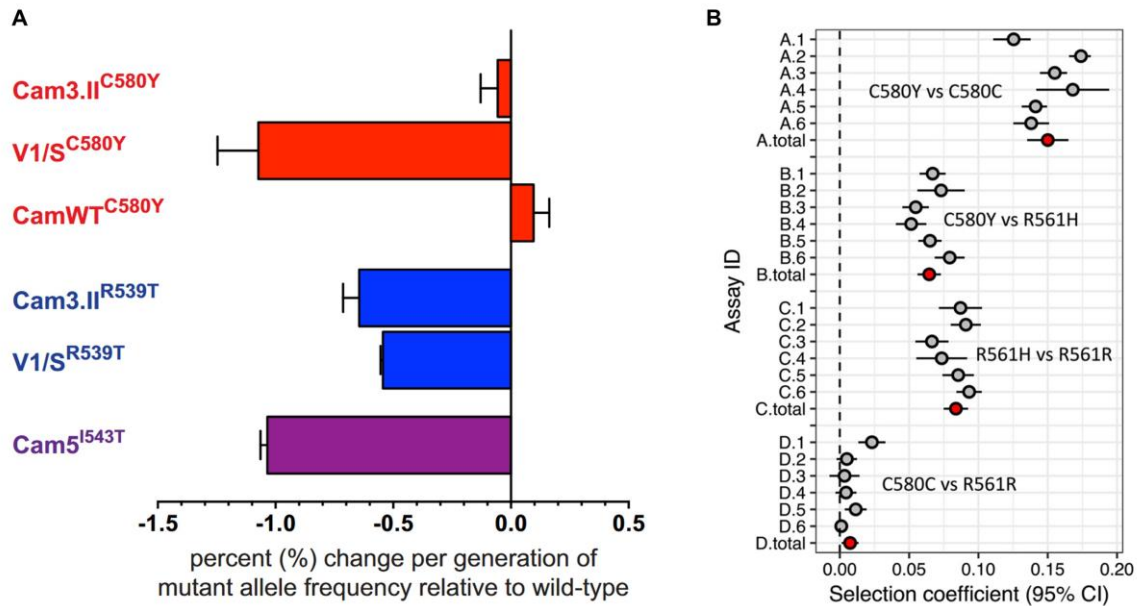


Figure 13 : Fitness des parasites portant différentes mutations du gène *pfk13*. (A) Les différences de taux de croissance ont été calculées en tant que pourcentage de changement de la fréquence de l'allèle mutant *pfk13* par génération asexuée (48 heures) durant 60 jours de co-culture d'un mutant avec un sauvage. D'après Straimer *et al.* (2017)¹²⁰. (B) Résultats de la compétition entre parasites sauvages vs. mutants et mutants vs. mutants. Les valeurs sont exprimées par des coefficients de sélection. Chaque expérience de compétition a été réalisée en six réplicats (points gris), avec les résultats de la méta-analyse (points rouges). Le génotype associé à un meilleur *fitness* est indiqué à droite pour chaque comparaison. D'après Nair et ses collègues (2018)¹¹⁹.

1.2.6. Le rôle de PfK13 dans la biologie du parasite : un comparatif sauvage versus mutant

La fonction physiologique de PfK13, protéine essentielle chez *P. falciparum*¹¹¹, n'est pas encore bien comprise. Toutefois, différents travaux ont démontré plusieurs fonctions de régulation de PfK13. L'inactivation fonctionnelle de cette protéine, par le biais d'une mauvaise localisation dans la cellule ou par un *knock-out* (KO) inductible, conduit à un arrêt précoce du cycle intra-érythrocytaire au stade *ring*, suivi d'une transformation lente des jeunes *rings* en formes pycnotiques¹²¹. Cet arrêt de croissance des parasites induit par l'inactivation de PfK13 est cohérent avec le phénotype de résistance aux ARTDs médié par *pfk13* qui s'exprime principalement au stade jeune *ring*¹²². De même, l'étude de parasites mutants *pfk13* résistants révèle un allongement de la durée du stade *ring*, tandis que celles des stades trophozoïte et schizonte se trouvent elles raccourcies, par rapport aux parasites sauvages¹²³. Par analyse du transcriptome *in vivo* de 1 043 isolats *P. falciparum* collectés chez des patients atteints de paludisme non compliqué, il a été constaté que la résistance aux ARTDs est associée à une expression accrue des gènes impliqués dans la voie cellulaire *Unfolded Protein Response* (réponse UPR), impliquant plusieurs protéines chaperonnes dont la fonction est d'assister d'autres protéines dans leur maturation¹²⁴. En effet, le stimulus causé par les ARTDs semble perturber le bon repliement des protéines, conduisant à l'activation d'une cascade de voies de

signalisation – constituant la réponse UPR – pour diminuer la biosynthèse des protéines afin d'en réduire leur accumulation tout en augmentant la biosynthèse de protéines chaperonnes ou de protéines impliquées dans la machinerie de dégradation des protéines. Ces résultats sont cohérents avec les travaux de Zhang et ses collaborateurs (2017)¹²³, lesquels démontrent une phosphorylation du facteur d'initiation 2A (PfeIF2 α), par la kinase PfPK4 et régulant la réponse UPR chez le parasite mutant C580Y au stade jeune *ring* (à partir d'une souche parasitaire Dd2). Au final, tous ces éléments suggèrent que le parasite résistant aux ARTDs modifie son cycle cellulaire de sorte qu'il passe plus de temps à un stade de son développement qui est moins sensible aux ARTDs (*ring*), tout en préparant une réponse atténuant les dommages causés par la molécule.

Dogovski *et al.* (2015) ont montré une diminution de l'accumulation de protéines parasitaires ubiquitinées chez les parasites mutants *pfk13*, relativement aux sauvages, après une brève exposition (1,5 heures) à une forte concentration d'ART (1 μ M), suggérant que la voie du protéasome a également un lien avec la résistance¹²⁵. PfK13 pourrait donc contrôler – directement ou indirectement – le niveau d'ubiquitination de protéines cibles. L'une d'entre elles pourrait être l'enzyme phosphatidylinositol-3 kinase (PfPI3K), qui a été immunoprécipitée avec PfK13, indiquant une interaction – directe ou non – entre ces deux protéines. L'ubiquitination et la dégradation par le protéasome de la PfPI3K sont par ailleurs diminuées chez les parasites *pfk13* C580Y¹²⁶. L'augmentation du taux cellulaire de PfPI3K entraîne un accroissement de la concentration cellulaire de son produit lipidique, le phosphatidylinositol-3 phosphate (PI3P), qui est corrélée avec le degré de résistance *in vitro*. Cette observation fut par la suite confirmée via des élévations synthétiques de ce lipide en l'absence de mutations *pfk13*¹²⁷. Un récapitulatif du lien entre les différentes voies métaboliques liées aux mutations *pfk13* est présenté en **figure 14**.

Un lien entre le lipide PI3P et la voie UPR – représentant actuellement deux facteurs clés de la résistance aux ARTDs – a récemment été proposé. Par microscopie immuno-électronique à haute résolution, Bhattacharjee *et al.* (2018) ont démontré que PfK13 se concentre au niveau des vésicules PI3P du réticulum endoplasmique (RE)¹²⁷. L'isolement de ces vésicules et leur analyse par protéomique ont permis de déterminer qu'elles sont enrichies en de multiples systèmes protéostasiquesⁱ tels que l'exportation de protéines, le contrôle de la qualité et du repliement des protéines dans le RE et le cytoplasme, et les voies de signalisation relatives à la réponse UPR. De façon remarquable, les vésicules PI3P atteignent, une fois amplifiées, tous les compartiments cellulaires chez les mutants *pfk13* C580Y. L'ensemble de ces données suggère que l'expansion des vésicules PI3P – liées à la protéostasie du RE – associée à la mutation *pfk13* C580Y dissémine une capacité protéostatique étendue incluant la réponse UPR, pouvant neutraliser la protéopathie toxique des ARTDs.

ⁱ Représente le concept de voies biologiques intégrées dans les cellules qui contrôlent la biogenèse, le repliement, le trafic et la dégradation des protéines.

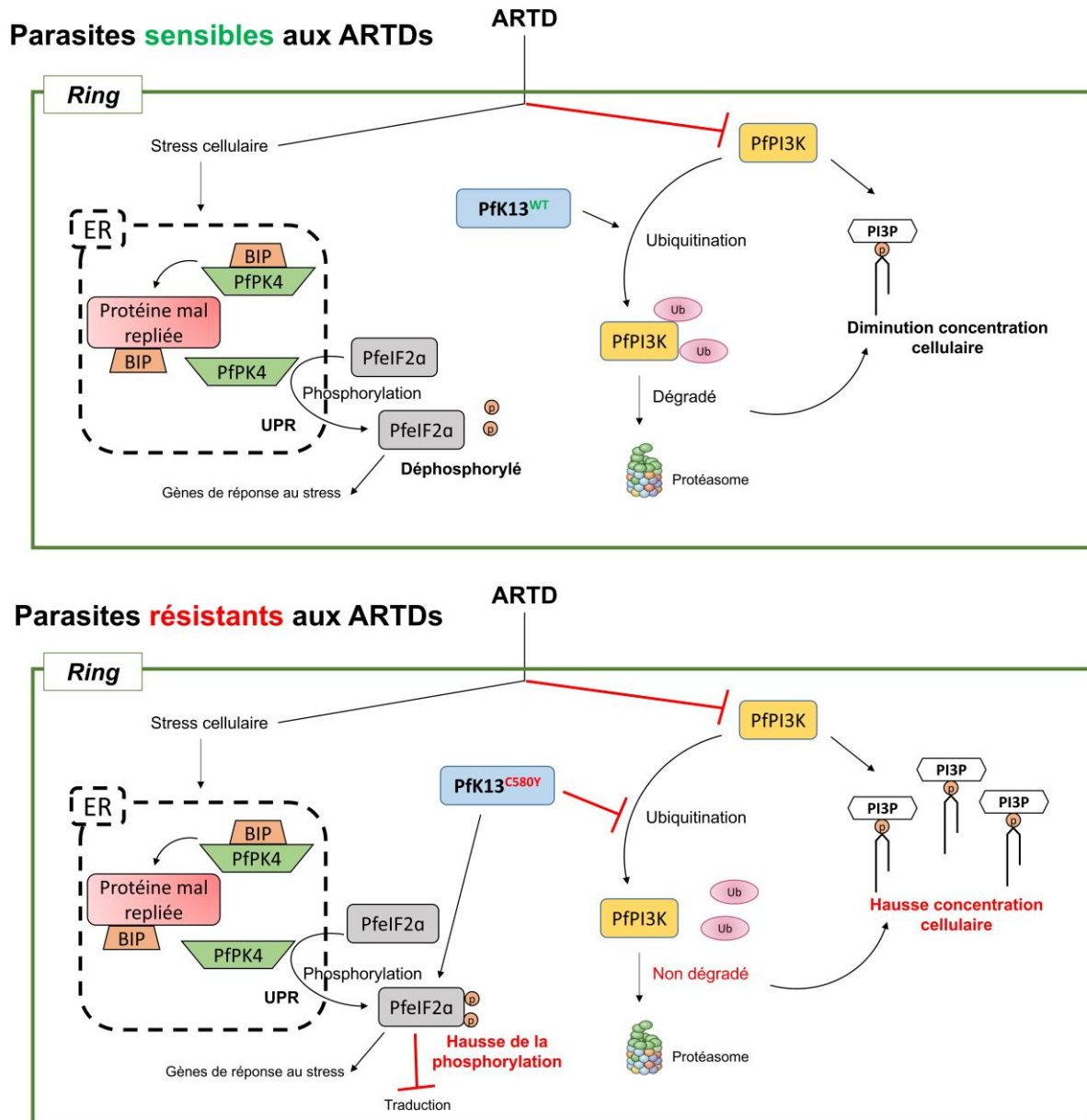
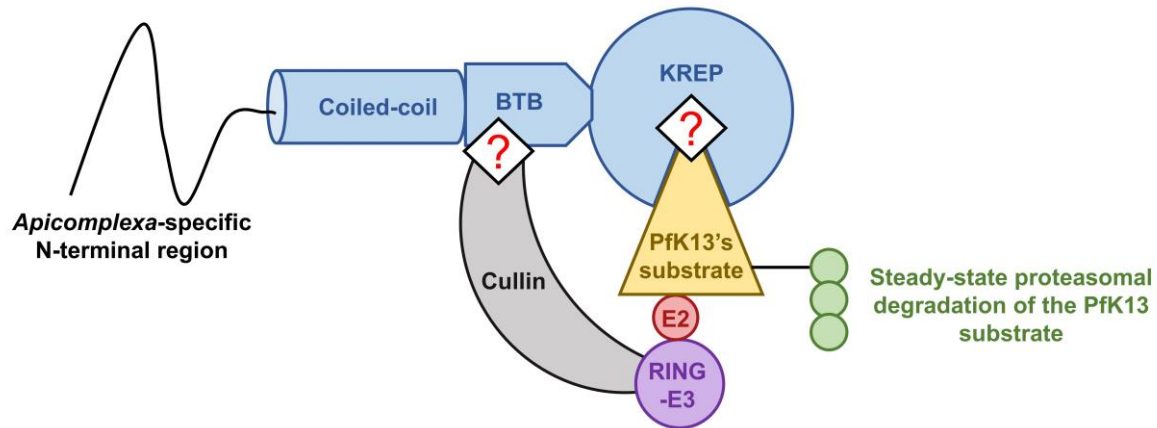


Figure 14 : Conséquences de la mutation *pfk13* C580Y sur diverses voies métaboliques du parasite. La mutation *pfk13* C580Y est associée à une diminution de protéines ubiquitinées incluant PfPI3K, engendrant une hausse de la concentration cellulaire de son produit lipidique, PI3P. En parallèle, cette mutation est associée à une hausse de la phosphorylation au stade jeune *ring* de PfeIF2α, médiée par la kinase PfPK4, qui est un homologue de PERK chez l'Homme et interagit avec des protéines chaperonnes (BIP). Il est supposé que la phosphorylation de PfeIF2α par PfPK4 soit proche du mécanisme de PERK, c'est-à-dire via la dissociation du complexe PfPK4-BIP. La hausse de phosphorylation de PfeIF2α conduit à un arrêt de la traduction générale, associée à une phase de latence (stade jeune *ring*) induite par les ARTDs.

Ces différentes études, couplées aux descriptions automatiques de PfK13 générées par la Gene Ontology (GO)¹²⁸ et à sa composition en domaines (**Figure 10**), suggèrent que cette protéine joue un rôle d'adaptateur dans un complexe d'ubiquitination⁶². Un modèle prédictif a été élaboré, par homologie avec les protéines BTB-Kelch chez les mammifères⁶² (**Figure 15**), bien qu'il reste à confirmer aujourd'hui. En effet, aucune des interactions décrites dans ce modèle n'a été démontrée de façon directe, par exemple avec des protéines recombinantes purifiées.



Identifiant PlasmoDB	Type de protéine	Description	Masse moléculaire	Longueur séquence (AA) ^a	Expression au cours de la phase asexuée érythrocytaire
PF3D7_0629800	Cullin	Cullin-like protein, putative	136 kDa	1 129	Fortement exprimée durant l'intégralité du cycle
PF3D7_0811000	Cullin	Cullin-1, putative	100 kDa	829	Fortement exprimée en début et fin du cycle
PF3D7_0312100	RING E3	E3 ubiquitin-protein ligase	99 kDa	836	Fortement exprimée dans les dernières 24 heures
PF3D7_0316900	RING E3	E3 ubiquitin-protein ligase, putative	71 kDa	600	Fortement exprimée au début du cycle
PF3D7_0305700	E2	Ubiquitin-conjugating enzyme E2, putative	16 kDa	142	Fortement exprimée durant l'intégralité du cycle
PF3D7_1203900	E2	Ubiquitin-conjugating enzyme E2	17 kDa	147	Fortement exprimée durant l'intégralité du cycle

Figure 15 : Modèle prédictif d'un complexe d'ubiquitination impliquant PfK13 et ses différents domaines. Chez les autres protéines à domaine BTB et/ou KREP, le domaine BTB contrôle l'état d'oligomérisation de la protéine et contribue au recrutement d'une protéine Cullin, tandis que le domaine KREP sert de récepteur pour une protéine à ubiquitiner qui sera ensuite dégradée par le protéasome. À noter que sur ce modèle, PfK13 est présentée comme un monomère pour simplifier la compréhension. Les symboles « ? » correspondent à de possibles surfaces d'interaction des domaines BTB et KREP, encore non caractérisées. Le tableau sous la figure liste quelques protéines de *P. falciparum* qui pourraient participer à ce complexe. ^a AA, acides aminés.

La majorité des protéines contenant à la fois un domaine BTB et un domaine KREP sont retrouvées dans des complexes d'ubiquitination Cullin-RING E3 (RING : Really Interesting New Gene) dans lesquels des protéines substrats cibles sont ubiquitinées^{129,130}. Dans ces complexes, le domaine BTB de protéines BTB-Kelch engendre différentes architectures oligomériques de la protéine, et contribue au recrutement de la protéine Cullin et de sa machinerie associée (RING-E3 et E2), tandis que le domaine KREP sert de récepteur pour le substrat protéique. Le domaine KREP présente plusieurs zones d'interaction, bien que la plupart des substrats se lient à travers une poche peu profonde localisée sur la face inférieure de la structure. L'activité de RING catalyse ensuite le transfert de l'ubiquitine de l'enzyme de conjugaison d'ubiquitine (E2) aux substrats protéiques^{107,129}. Chez les mammifères, une protéine BTB-Kelch a été très étudiée, nommé Kelch-like ECH-Associated Protein 1 (KEAP1) et impliquée dans des cancers, et sert souvent de modèle pour mieux comprendre la fonction de PfK13. Chez KEAP1, le domaine BTB permet la dimérisation de la protéine et l'interaction avec une Cullin 3, tandis que la poche peu

profonde de KREP permet la liaison avec le facteur de transcription Nrf2 (Nuclear Factor (erythroid-derived 2)-like 2) qui sera ensuite ubiquitiné et conduit en voie de dégradation. Des mutations au sein de la protéine induisent des changements conformationnels locaux ou globaux de KEAP1 et altèrent l'interaction avec Nrf2 qui est dès lors moins dégradé et peut faciliter la progression de cellules cancéreuses^{131,132}. Concernant Pfk13, le(s) substrat(s) reste(nt) encore à identifier, bien que les kinases PfPI3K et à un moindre degré PfkPK4 représentent de sérieux candidats^{123,126}. De plus, les régions et plus spécifiquement les sites (ou positions) de Pfk13 impliqués dans ces possibles interactions protéine-protéine demeurent jusque-là inconnues. La caractérisation de ces régions et sites représente l'un des objectifs majeurs de ce travail de thèse.

Plusieurs hypothèses peuvent être avancées pour expliquer le plus faible taux de protéines ubiquitinées – incluant PfPI3K – chez les parasites mutants *pfk13* C580Y. Les mutations étant localisées sur le domaine KREP de Pfk13, elles pourraient : i) engendrer une diminution de l'affinité de liaison entre le domaine KREP et le substrat de Pfk13 ; ii) déstabiliser la protéine qui serait moins abondante – bien que toujours fonctionnelle ; iii) déstabiliser l'homo-dimérisation des domaines BTB-KREP. Les deux premières hypothèses, qui nous semblent les plus probables, sont illustrées dans la **Figure 16** et seront discutées dans la section 4.1.2.3. de la thèse en analysant nos résultats et la littérature.

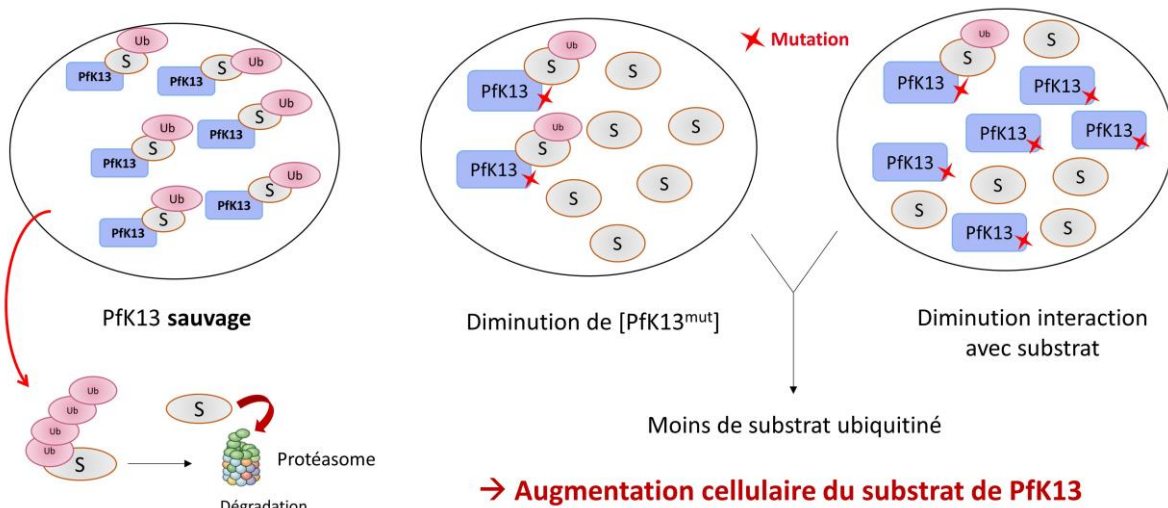


Figure 16 : Deux hypothèses permettant d'expliquer la relation entre les mutations *pfk13* et le plus faible taux de protéines ubiquitinées. La représentation de gauche reflète le cas sauvage, et les deux autres, le cas muté (au milieu, une diminution de l'abondance de la protéine Pfk13 ; à droite, une diminution de l'affinité de liaison entre Pfk13 et son substrat cible).

1.2.7. D'autres déterminants moléculaires de résistance aux ARTDs à venir ?

Les mutations de *pfk13* ont clairement été démontrées comme étant causales dans la résistance aux ARTDs en Asie du sud-est, et seraient également présentes à fréquence élevée en Papouasie-Nouvelle Guinée⁵⁶. Cependant, la recherche d'autres déterminants moléculaires de la résistance est continuellement menée. Parmi les échecs thérapeutiques aux CTAs associés à des infections contractées en Afrique, deux impliquent une mutation de *pfk13* (M579I et A675V)^{101,117,133–135}. Ces observations isolées n'ont pas été confirmées depuis et confortent la possibilité d'autres déterminants moléculaires situés sur d'autres loci.

Des mutations de plusieurs gènes ont été identifiées par des approches *in vivo* ou *in vitro* comme allongeant la durée de clairance parasitaire après traitement à la DHA. En 2018, Demas *et al.* (2018) ont effectué une sélection *in vitro* de deux souches parasitaires provenant de l'ouest de l'Afrique avec une augmentation croissante de la concentration en DHA sur une période de quatre ans – un procédé similaire au travail de Witkowski et ses collaborateurs (2010)¹³⁶ ayant ensuite permis la découverte du gène *pfk13*¹⁰⁴. Parmi les clones associés à une meilleure survie (test RSA_{0-3h} avec de la DHA) ont été identifiées des mutations non-synonymes dans le gène *pfcoronin* qui code pour une protéine de la famille des WD40, arborant un domaine propeller similaire à celui de Pfk13¹³⁷. Le rôle de ces mutations dans la résistance aux ARTDs a ensuite été confirmé par l'approche CRISPR/Cas9 couplée à des tests RSA_{0-3h}. Bien que les mutations de ce gène ne soient pour l'heure pas retrouvées dans les populations naturelles de *P. falciparum* – représentant la limite majeure de cette découverte –, découvrir d'autres marqueurs ou déterminants moléculaires de la résistance aux ARTDs permettrait une surveillance plus large de la résistance, notamment en Afrique.

Des mutations non-synonymes localisées dans la région C-terminale de la protéine codée par le gène *pcubp1* ont également été démontrées comme déterminant la résistance aux ARTDs par des approches *in vivo* dans un modèle murin de paludisme (*P. chabaudi*)^{138,139}. Cette protéine est une ubiquitine-hydrolase et pourrait jouer un rôle dans les processus de polyubiquitination. Des variants de la protéine codée par le gène *pfubp1* ont également été associées à une diminution de la susceptibilité aux ARTDs, bien que ces variants requièrent des validations expérimentales¹⁴⁰. À cela s'ajoutent des mutations dans le gène orthologue *pcap2-mu* qui code pour une protéine vraisemblablement impliquée dans le processus d'endocytose par lequel les cellules absorbent des métabolites ou des protéines¹⁴¹. Ces gènes mutés ont par la suite été partiellement validés chez *P. falciparum* comme conférant la résistance aux ARTDs par des approches *in vivo* (quelques cas d'échecs thérapeutiques suite à des infections en Afrique ; et sélection post-traitement)^{133,142,143} et par la génération de parasites transgéniques *in vitro*¹⁴². La plus large étude de génomique des populations africaines de *P. falciparum* menée à ce jour vient d'identifier un fragment du chromosome 12 soumis à une sélection positive récente, plus

spécifiquement au Ghana et au Malawi : ce fragment contient en particulier le gène *pfap2-mu*¹⁴⁴. Il est fort probable que l'agent sélectif soit les ARTDs contenus dans les traitements CTAs. Il a également été montré que le gène *falcipain 2a* est un contributeur de la résistance aux ARTDs, puisque des versions mutées sont associées à une diminution de la digestion de la partie protéique de l'hémoglobine, ainsi qu'à un taux élevé de parasites survivants après traitement à la DHA¹⁴⁵. Des études d'association pangénomique et des expériences de manipulation de gènes ont identifié d'autres loci possiblement associés à la résistance aux ARTDs comme le gène *pfatg18*^{146,147}, bien que les validations fonctionnelles restent à faire.

Au final, ces différents travaux montrent que la résistance aux ARTDs peut être conférée par différentes protéines mutées. Toutefois, nous nous sommes focalisés uniquement sur PfK13 dans ce travail de thèse, car elle représente pour l'heure le déterminant moléculaire majeur de la résistance aux ARTDs et le seul dont la relevance clinique a été largement démontrée.

- L'ART provient d'une armoise annuelle : *Artemisia annua*, utilisée dans la médecine traditionnelle chinoise.
- Plusieurs dérivés semi-synthétiques de l'ART (les ARTDs) ont été conçus pour utilisation clinique : l'artésunate, la DHA ou encore l'artéméter.
- L'ART contient un pont endopéroxyde qui réagit avec un atome de fer – probablement lié à l'hème – pour former des radicaux libres. La formation de ces derniers entraîne des dommages moléculaires (exemple : alkylation des protéines) conduisant à la mort du parasite.
- Les ARTDs sont couplés au sein de bithérapies, appelées CTAs, avec d'autres antipaludiques de modes d'action différents afin de limiter les phénomènes de résistance. Les ARTDs, présentent très brièvement dans le sang, éliminent la majorité des parasites en quelques jours, tandis que la seconde molécule antipaludique – dite partenaire – reste dans le sang plus longtemps et tue les parasites résiduels.
- En Asie du sud-est, la résistance aux ARTDs se traduit par une réduction de la vitesse d'élimination des parasites chez le patient traité.
- Des parasites résistant aux ARTDs ont été identifiés en 2008, avant de s'étendre dans toute l'Asie du sud-est. La résistance aux ARTDs n'a pour l'heure pas été clairement caractérisée en Afrique.
- Des mutations non-synonymes localisées sur le gène *pfk13* ont été identifiées comme conférant la résistance aux ARTDs.
- Initialement, plusieurs mutations du gène *pfk13* auraient coexisté par sélection positive avec des fréquences différentes. Le balayage sélectif s'est intensifié au cours duquel seules les lignées portant la mutation C580Y se sont propagées.
- La protéine Pfk13 contient trois domaines annotés (coiled-coil, BTB et KREP) et une région N-terminale spécifique aux *Apicomplexa* ; sa séquence est très conservée au cours de l'évolution, reflétant son rôle essentiel dans la biologie du parasite ; elle agirait vraisemblablement dans un complexe d'ubiquitination.
- Les parasites mutants *pfk13* se caractérisent par un ralentissement de leur développement au début du stade intra-érythrocytaire (période où ils sont par ailleurs le moins sensibles aux ARTDs) et sont associés à diverses modifications de voies métaboliques permettant d'atténuer les dommages causés par les ARTDs.
- La protéine Pfk13 a été immunoprécipitée avec la protéine PfPI3K.
- La fonction précise de Pfk13 est inconnue, ainsi que ses sous-régions et les sites impliqués dans de probables interactions protéine-protéine.
- D'autres marqueurs / déterminants moléculaires de résistance aux ARTDs ont été identifiés.

PfCRT, une protéine de transport impliquée dans la résistance à de multiples antipaludiques à base de quinoléine

Section 1.3



Dans la vie, on ne fait pas ce que l'on veut mais on est responsable de ce que l'on est.

— Jean-Paul Sartre (1905 – 1980)

Des mutations localisées sur la protéine *Plasmodium falciparum* chloroquine resistance transporter (PfCRT)ⁱ confèrent la résistance à la CQ ainsi qu'à d'autres médicaments à base de quinoléine, découverte qui a bouleversé la politique d'utilisation des antipaludiques et précipité l'adoption des CTAs. De nouvelles mutations de *pfcr*t ont récemment été identifiées et associées à la résistance à la PPQ, une des molécules partenaires des CTAs. Dans cette partie, nous rappellerons l'histoire de l'utilisation des quinoléines et de la résistance à la CQ, son mode d'action, l'identification de PfCRT et son rôle dans la physiologie du parasite et dans la résistance, et l'évolution des haplotypes *pfcr*t.

ⁱ Nous faisons référence à la protéine CRT lorsqu'on parle de la protéine indépendamment de l'espèce qui la porte, tandis que PfCRT est spécifiquement axée sur l'espèce *P. falciparum*.

1.3.1. Découverte et développement des médicaments à base de quinoléine

Avant l'adoption des CTAs, de multiples antipaludiques ont été utilisés pour combattre le paludisme, incluant divers médicaments à base de quinoléine (quinine, CQ, méfloquine, amodiaquine, PPQ)⁵⁴. Certaines quinoléines sont d'ailleurs toujours utilisées, soit en tant que molécule partenaire avec un ARTD (comme la bithérapie DHA-PPQ utilisée en Asie du sud-est et Afrique), soit en monothérapie (CQ) pour combattre des infections palustres dues à des espèces autres que *P. falciparum*³⁹.

Les dérivés 4-aminoquinoléines (en raison du groupe amine à la position 4 de la quinoléine ; **Figure 17**) ont constitué la première classe de composés de synthèse utilisés pour le traitement du paludisme. La quinoléine a été initialement extraite du goudron de houilleⁱ dans lequel elle est naturellement présente par le chimiste allemand Friedlieb Ferdinand Runge (**Figure 18**) en 1834¹⁴⁸. Quelques années après, le chimiste alsacien Charles Frédéric Gerhardt (**Figure 18**) obtint par distillation l'une des huiles issues de la quinoléine : le Leukol (huile blanche), déjà décrite par Runge¹⁴⁹. Cette huile peut s'obtenir en grandes quantités, favorisant ainsi la chimie de synthèse ayant pour finalité de développer des analogues. Plus tard, il a également été rapporté que cette huile existe chez diverses plantes de la famille des *Rutaceae*¹⁵⁰ et, plus surprenant encore, chez certains insectes néoptères incluant des phasmes (comme *Oreophoetes peruana*, lequel sécrète de la quinoléine en cas de stress). En 1856, le chimiste anglais William Henry Perkin (**Figure 18**) se lance dans la recherche d'une forme synthétique de la quinine, extrait naturel d'un arbuste originaire d'Amérique du sud, le quinquina. Ses efforts l'amènent non pas à une nouvelle forme de quinine, mais à un colorant synthétique appelé mauvéine¹⁴⁹. De son côté, le scientifique allemand Paul Ehrlich (**Figure 18**) a remarqué que le bleu de méthylène était particulièrement efficace pour la coloration des parasites du paludisme, pouvant même être sélectivement toxique pour le parasite. Ainsi, en 1891, Ehrlich – aidé par le pathologiste Paul Guttman (**Figure 18**) – a guéri deux patients atteints du paludisme avec le bleu de méthylène, devenant ainsi le premier médicament de synthèse utilisé en thérapie. Cette molécule n'a plus été utilisée par la suite (elle est à nouveau utilisée depuis quelques années pour son activité contre les formes du parasite transmissibles au moustique), mais elle constitue la base du développement des antipaludiques de synthèse, notamment grâce au remplacement d'un groupe méthyle par un dialkylaminoalkyle par les laboratoires Bayer d'Elberfeld en Allemagne^{149,151}. Cette chaîne latérale a ensuite été connectée à différents systèmes hétérocycliques tels que la quinoléine, conduisant au premier antipaludique de synthèse en 1925, la plasmoquine¹⁵¹. Ce médicament a été peu utilisé en raison d'effets secondaires observés lors de l'évaluation clinique. Son développement a toutefois conduit à la découverte de la primaquine, bien mieux tolérée et introduite en 1952¹⁵¹. En 1934, Johann Andersag (également connu sous le nom de Hans Andersag), travaillant au sein des laboratoires Bayer, introduit une nouvelle chaîne latérale pour mettre au point une substance aux propriétés antipaludiques

ⁱ Produit issu de la distillation de la houille (roche carbonée sédimentaire).

appelée résochine¹⁵². Ce composé est rapporté comme étant toxique et ne fait donc l'objet d'aucun développement par Bayer. Andersag poursuit ses efforts et produit la sontochine, un dérivé moins toxique de la résochine¹⁵². Quelques années plus tard, les chercheurs se décident à réévaluer les propriétés de la résochine et de la sontochine. Des essais cliniques sont menés et fournissent des résultats très encourageants, démontrant une toxicité négligeable chez l'Homme. La résochine est finalement distribuée sous le nom de CQ en février 1946, et deviendra la molécule clé dans le contrôle du paludisme au cours des quatre décennies suivantes¹⁵².

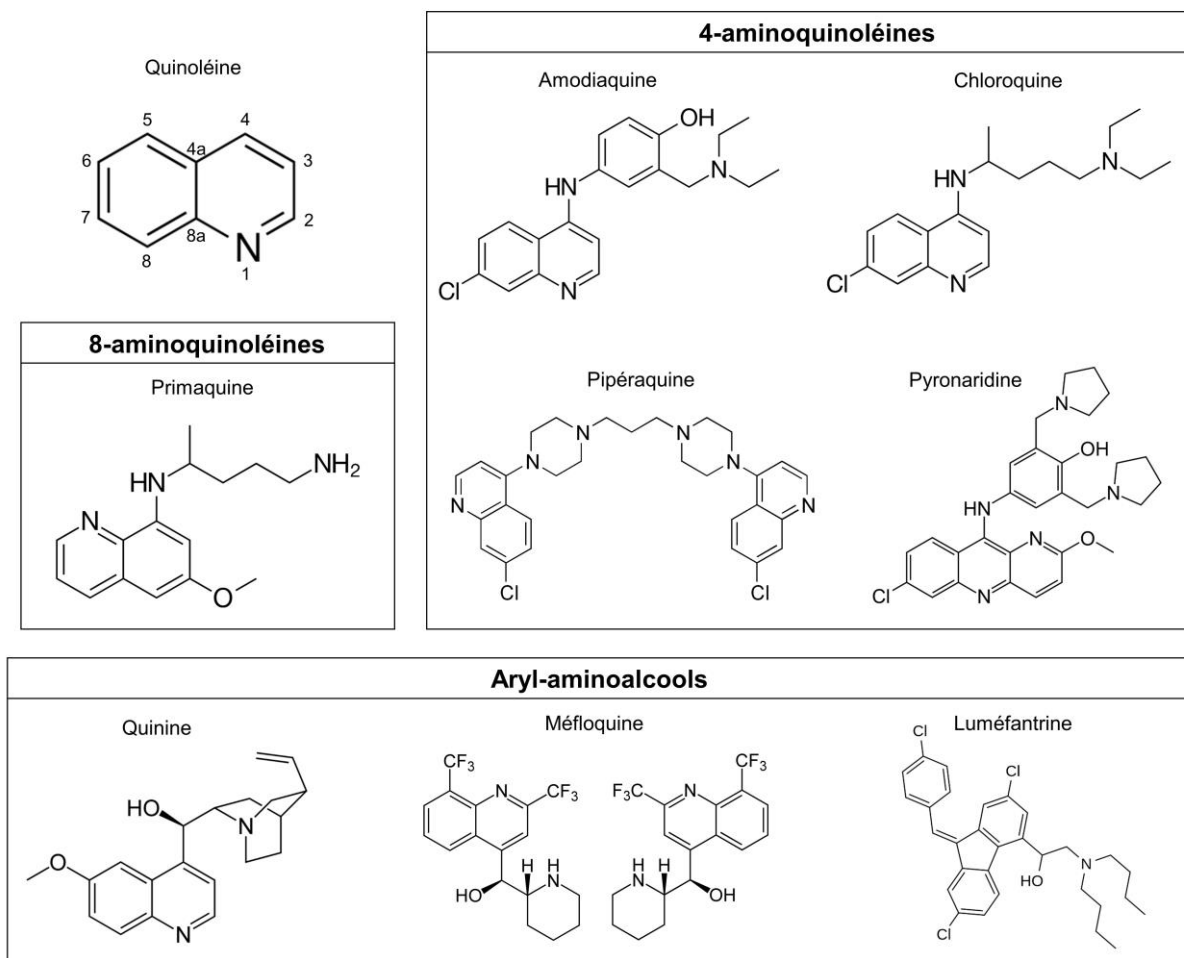


Figure 17: Formules chimiques des différents antipaludiques obtenus à partir de la fonction quinoléine. L'ensemble de ces molécules inclut le composé organique aromatique hétérocyclique de la quinoléine. Toutes sont actuellement utilisées en clinique.

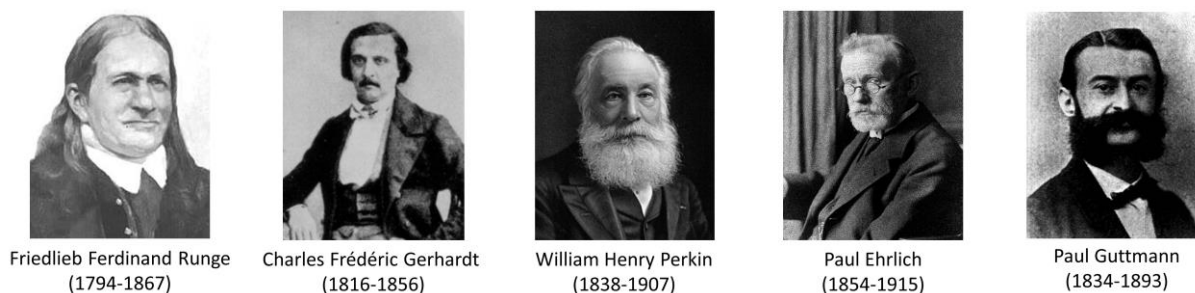


Figure 18 : Principales figures historiques ayant participé à la recherche sur la chimie et le mécanisme d'action de la quinoléine. Johann Andersag (1902-1955), autre figure, n'est pas représenté en raison de l'absence de portrait disponible dans les archives.

De nombreux travaux ont ensuite été dédiés à la modification des 4-aminoquinoléines, générant un grand nombre de dérivés – dont les formules de ceux actuellement utilisés en clinique sont présentées en **figure 17** – qui ont permis de combler pour un temps les émergences successives de résistance. Les modifications portent aussi bien sur l'allongement ou le raccourcissement de la chaîne latérale dialkylaminoalkyle que sur l'introduction de noyaux aromatiques (amodiaquine, pyronaridine) ou la dimérisation des 4-aminoquinoléines (PPQ) à l'aide d'un connecteur de nature et de longueur variables¹⁵¹. D'autres dérivés du noyau quinoléine ont été synthétisés tels que la luméfantrine et la méfloquine. Ces molécules sont actuellement utilisées en clinique pour le traitement et/ou la prévention du paludisme (cf. section 1.1.5.).

1.3.2. Mode d'action des quinoléines

Le mode d'action de toutes les molécules à base de quinoléine n'est pas encore élucidé. La CQ, et probablement les autres molécules 4-aminoquinoléines et la quinine, sont supposées interférer dans le processus de conversion de l'hème toxique, issu de la dégradation de l'hémoglobine, en hémozoïne non toxique (cf. section 1.1.3.). Plus spécifiquement, ces drogues s'accumulent à de très fortes concentrations dans la vacuole digestive des parasites¹⁵³ : il a notamment été montré que près de 85% de la CQ est retrouvée dans la vacuole digestive. Bien que le mode d'action de ces antipaludiques demeure débattu, l'hypothèse majeure suppose qu'ils forment des complexes avec l'hème de fer Fe^{3+} , empêchant ainsi sa polymérisation en hémozoïne et conduisant à la mort du parasite^{33,153,154}. Ce mécanisme n'est toutefois pas universel pour toutes les molécules à base de quinoléine, puisque la luméfantrine, elle, agirait dans d'autres compartiments cellulaires que la vacuole digestive. Une représentation schématique du site d'action des divers antipaludiques à base de quinoléine est donnée en **figure 19**.

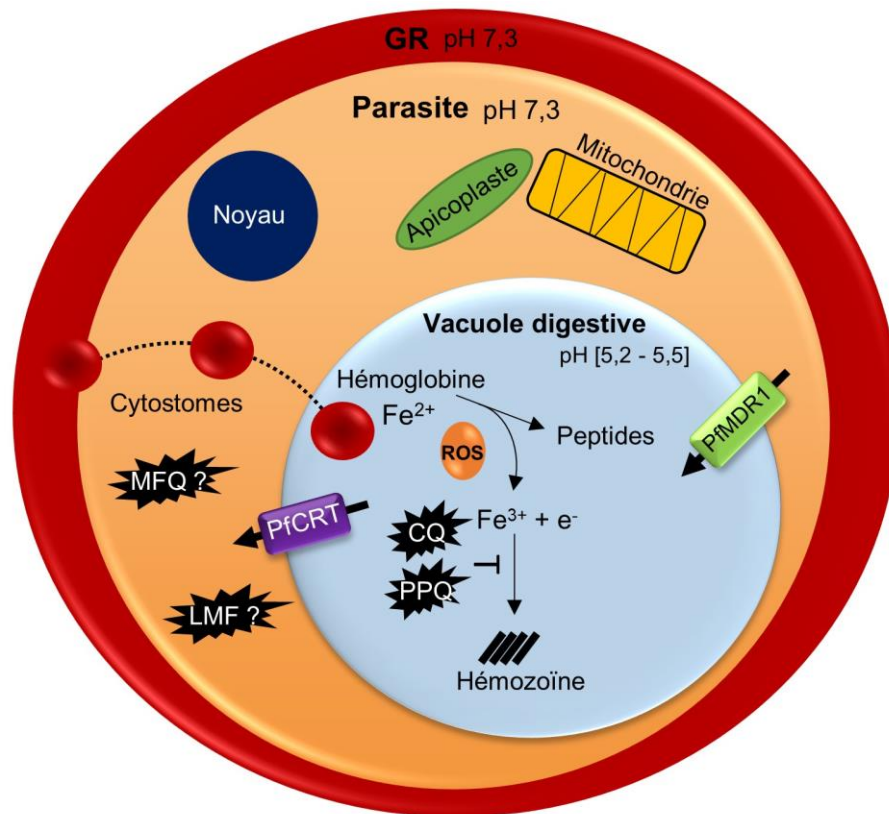


Figure 19 : Sites d'action de divers antipaludiques à base de quinoléine. CQ, chloroquine ; PPQ, pipéraquline ; LMF, luméfantrine ; MFQ, méfloquine ; PfcRT, *P. falciparum* chloroquine resistance transporter; PfMDR1, *P. falciparum* multidrug resistance 1. La CQ et la PPQ sont supposées traverser les membranes, probablement par diffusion passive, pour se concentrer au sein de la vacuole digestive du parasite et empêcher la conversion de l'hème toxique en hémotoïne non toxique pour le parasite. La méfloquine et la luméfantrine, elles, agiraient dans d'autres compartiments cellulaires ; pour ces deux dernières molécules, leur stockage dans la vacuole digestive pourrait être un mécanisme de résistance du parasite.

1.3.3. Rôle physiologique de PfcRT, transporteur impliqué dans la résistance aux médicaments à base de quinoléine

L'utilisation massive de la CQ a permis de sauver des millions de vies¹⁵⁵. Outre son efficacité, cette molécule était aussi reconnue pour sa bonne tolérance, son accessibilité et son faible coût. À la fin des années 1950, des parasites résistants à la CQ ont initialement émergé en Asie du sud-est et en Amérique du sud, avant de se propager une décennie plus tard de l'Asie du sud-est à l'ouest du Pacifique puis à l'Afrique, causant une hausse significative du nombre de décès liés au paludisme (**Figure 20**)¹⁵⁵. La résistance à la CQ est mesurée *in vitro* via la concentration inhibitrice 50% (CI₅₀), c'est-à-dire en évaluant la concentration de la molécule capable d'inhiber la croissance de 50% des parasites durant une incubation de 48 à 72 heures avec la drogue. On estime qu'une CI₅₀ supérieure à 100 nM pour la CQ est associée à une résistance à cette molécule.

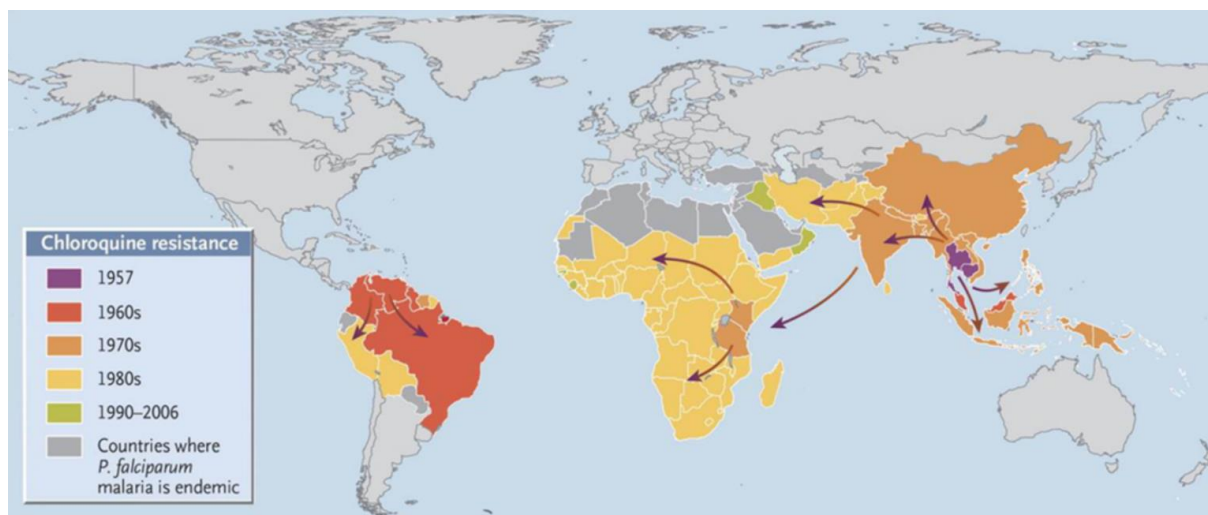


Figure 20 : Apparition et expansion des parasites *P. falciparum* résistants à la CQ depuis la fin des années 1950. La résistance à la CQ a émergé de façon indépendante en Asie du sud-est et en Amérique du sud à la fin des années 1950, puis à nouveau en Océanie. Les parasites résistants à la CQ d'Asie se sont ensuite propagés vers l'Afrique une vingtaine d'années plus tard. La figure est tirée de la revue de Tanner et ses collaborateurs (2015)¹⁵⁶.

Un croisement génétique entre souches parasitaires résistante (souche Dd2) et sensible (souche Hb3) à la CQ a permis de mettre en évidence l'implication de mutations du gène *P. falciparum* chloroquine resistance transporter (*pfcr*t) dans ce phénomène de résistance^{51,52,157}. Localisé sur le chromosome 7 et s'étendant sur 3,1 kb, le gène *pfcr*t code pour la protéine PfCRT comprenant 424 acides aminés, constituée de 10 hélices transmembranaires (**Figure 21**). Un screen génétique à l'échelle génomique a démontré son rôle essentiel dans le développement intra-érythrocytaire du parasite¹¹¹. Sa structure secondaire, sa localisation à la membrane de la vacuole digestive et sa proximité phylogénétique avec la superfamille des *drug/metabolite transporter* (DMT) suggèrent sa participation dans le transport de divers métabolites et drogues entre le cytoplasme et la vacuole digestive du parasite (**Figure 22**)^{158,159}.

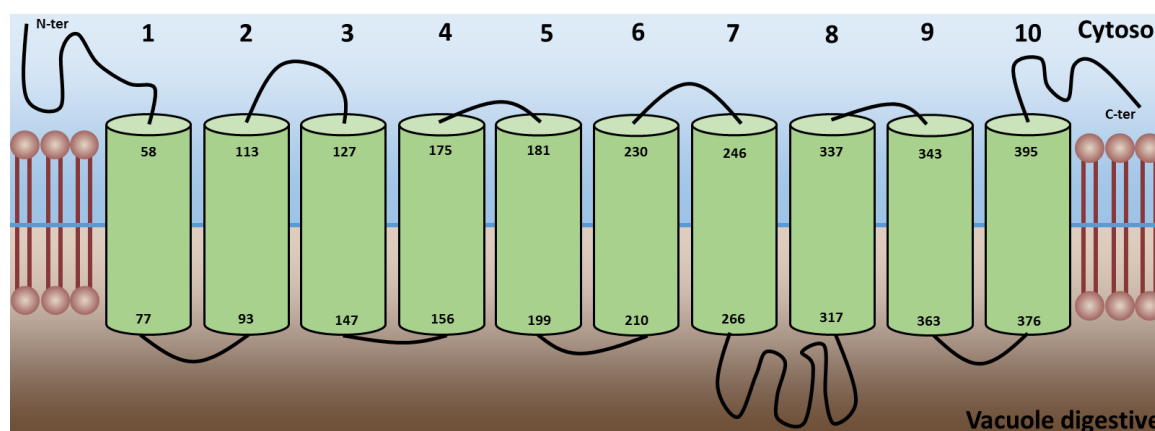


Figure 21 : Représentation schématique de la structure secondaire de PfCRT. La limite des hélices transmembranaires (en acides aminés) a été définie par un consensus de plusieurs méthodes bioinformatiques spécialisées dans la prédiction de ces structures secondaires. La partie cytoplasmique est représentée en bleu, tandis que la vacuole digestive est représentée en marron. Cette représentation sera toutefois mise à jour dans nos travaux de recherche.

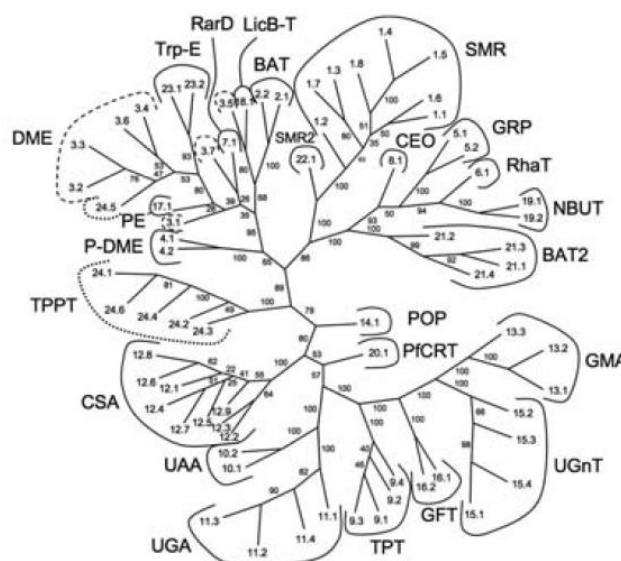


Figure 22 : Phylogénie des DMTs. PfCRT s'avère être proche de protéines *Nucleotide-sugar transporter* (NST), une sous-famille incluant les protéines *CMP-sialate:CMF antiporter* (CSA). La phylogénie est tirée du chapitre axée sur les transporteurs et leur lien avec la résistance aux médicaments, par Yen et ses collaborateurs (2010)¹⁵⁹.

Plusieurs activités de transport ont été rapportées pour cette protéine (**Tableau 3**). Tout d'abord, Juge *et al.* (2015) ont préparé des protéoliposomes contenant la protéine PfCRT purifiée¹⁶⁰. Ils ont observé que PfCRT participait au transport de divers substrats cationiques (comme le tétraéthylammonium). Ce résultat est en accord avec des études antérieures dans d'autres systèmes d'expression hétérologues (levure et ovocytes de Xénope) proposant un rôle de pompe à protons et/ou d'activateur d'échangeurs Na^+/H^+ ^{161–163}. Une autre étude s'est intéressée à trois homologues de PfCRT chez la plante *Arabidopsis thaliana*¹⁶⁴, et suggère un rôle essentiel dans le transport de glutathion. Toutefois, l'expression de PfCRT au sein d'ovocytes de Xénope a révélé le transport de glutathion uniquement via les formes mutantes de PfCRT conférant la résistance à la CQ, également confirmé en utilisant des lignées de parasites *P. falciparum* isogéniques portant le gène *pfcr*t sauvage ou mutant¹⁶⁵. Enfin, des expériences électrophysiologiques et de flux conduites avec des ovocytes de Xénope exprimant PfCRT à leur surface ont démontré son rôle dans le transport de fer libre Fe^{2+} et Fe^{3+} , avec une activité de transport diminuée en présence des mutations conférant la résistance à la CQ¹⁶⁶. Dans un contexte de résistance, deux études indépendantes ont révélé que des mutations du gène *pfcr*t peuvent altérer la morphologie de la vacuole digestive^{89,167}. Enfin, des analyses QTL (pour locus de traits quantitatifs, ou *Quantitative Trait Locus* en anglais) ont montré que les mutations *pfcr*t provoquent une accumulation parasitaire de divers peptides issus de la dégradation de l'hémoglobine, par comparaison avec le sauvage¹⁶⁸.

Malgré ces nombreuses investigations, la fonction précise de PfCRT demeure mal comprise, d'autant plus que la majorité d'entre elles se sont basées sur des systèmes d'expression hétérologues et non directement chez le parasite (**Tableau 3**). Fait notable, aucune structure tertiaire de la protéine n'a encore été déterminée, limitant les investigations structurales.

Tableau 3 : Différentes activités de transport physiologique pour la protéine PfCRT.

Système d'expression	Année	Séquence optimisée	Fonction / substrat	Référence
<i>S. cerevisiae</i> <i>P. pastoris</i>	2002	Oui ; régions N- et C-terminales modifiées	Transport de H ⁺ pour PfCRT sauvage	161
<i>X. laevis</i>	2004	Oui	Activateur d'échangeurs Na ⁺ /H ⁺ pour PfCRT sauvage	163
<i>A. thaliana</i> (étude d'homologues de PfCRT)	2010	—	Transport de glutathion	164
<i>P. falciparum</i> <i>X. laevis</i>	2013	— Oui	Transport de glutathion pour PfCRT mutante	165
<i>E. coli</i>	2015	Oui	Transport de substrats cationiques pour PfCRT sauvage et mutantes (Dd2, 7G8)	160
<i>X. laevis</i>	2017	Oui	Transport d'ions Fe ²⁺ et Fe ³⁺ pour PfCRT sauvage et mutante (Dd2)	166

1.3.4. Caractérisation de la résistance à la CQ, conférée par les mutations du gène *pfCRT*

Les souches résistantes à la CQ arborent différentes combinaisons de mutations (ou haplotypes) selon leur région géographique d'origine. Ces haplotypes *pfCRT* mutants sont composés de quatre à dix changements simultanés d'acides aminés, toutes ayant en commun la substitution d'une lysine (K) en thréonine (T) à la position 76 dans la première hélice transmembranaire de PfCRT (**Figure 23** et **Tableau 4**). Les mutations accompagnant la mutation K76T sont supposées diminuer le *fitness cost* de ce changement et/ou participer à la fonction de résistance. En effet, il a été montré que la mutation K76T seule ne suffisait pas à conférer la résistance à la CQ^{169,170}. Considérant la fonction de transport de PfCRT, il a été suggéré que les mutants *pfCRT* ont acquis la capacité d'évacuer la CQ hors de la vacuole digestive vers le cytoplasme^{51,170}. L'efflux de CQ se faisant à son état di-protoné en raison de l'acidité de la vacuole digestive^{171,172}, la suppression de la charge positive liée à la position 76 pourrait effectivement favoriser ce transport. Cette observation a été initialement rapportée dans des systèmes d'expression hétérologues chez *Pichia pastoris*¹⁶² et *Dictyostelium discoideum*¹⁷³, puis plus tard au sein d'ovocytes de Xénope¹⁷⁰. Dans ce dernier système, il a été montré que la réversion – au sein de l'haplotype mutant Dd2 – de la seule mutation critique K76T diminue l'accumulation de CQ dans l'ovocyte¹⁷⁰.

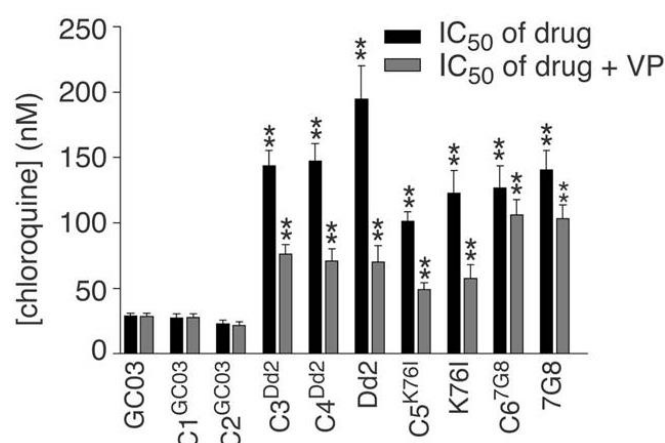


Figure 23 : Évaluation du rôle des mutations *pfcr*t dans la résistance à la CQ par mesure de la IC_{50} chez des parasites *P. falciparum* cultivés *in vitro*. Les souches GC03, C1 et C2 portent un haplotype *pfcr*t sauvage. C3, C4 portent l'haplotype *pfcr*t de la souche asiatique Dd2. C5 modifie l'haplotype Dd2 sur deux positions : K76I (au lieu de K76T) et absence de la mutation I356T. C6 porte l'haplotype *pfcr*t de la souche brésilienne 7G8. Le vérapamil (VP) est un inhibiteur de canaux calciques et réverse partiellement la IC_{50} des parasites *pfcr*t mutants. Les tests statistiques ont été réalisés entre la souche de référence C2^{GC03} et les autres (** : $P < 0.01$). D'après Sidhu *et al.* (2002)⁵².

Tableau 4 : Principaux haplotypes *pfcr*t présents dans les populations naturelles de *P. falciparum* en zone d'endémie palustre.

Région d'origine	Isolats	Phénotype	72	74	75	76	97	144	148	160	194	220	271	326	333	356	371	Ref
Haplotype sauvage																		
Monde	HB3, 3D7	CQS	C	M	N	K	H	A	L	L	I	A	Q	N	T	I	R	51
Haplotypes mutants																		
Afrique	106/1	CQS	C	I	E	K	H	A	L	L	I	S	E	S	T	I	I	51
Asie, Afrique	Dd2	CQR	C	I	E	T	H	A	L	L	I	S	E	S	T	T	I	51
Asie, Afrique	FCB	CQR	C	I	E	T	H	A	L	L	I	S	E	S	T	I	I	51
Asie, Afrique	GB4	CQR	C	I	E	T	H	A	L	L	I	S	E	N	T	I	I	174
Asie, Afrique	Cam783	CQR	C	I	E	T	H	A	L	L	I	S	E	N	T	T	I	174, 175
Asie	Cam738	CQR	C	I	D	T	H	A	I	L	T	S	E	N	S	I	R	174
Asie	Cam734	CQR	C	I	D	T	H	F	I	L	T	S	E	N	S	I	R	174
Pacifique ouest	PH1	CQR	C	M	N	T	H	T	L	Y	I	A	Q	D	T	I	R	176
Amérique, Afrique	7G8	CQR	S	M	N	T	H	A	L	L	I	S	Q	D	T	L	R	51
Amérique	Jav	CQR	C	M	E	T	Q	A	L	L	I	S	Q	N	T	I	T	51

Note – D'autres haplotypes non présentés ici existent, retrouvés à faible fréquence en zone d'endémie. Les acides aminés sur fond gris représentent les positions mutantes. CQS, sensible à la CQ ; CQR, résistant à la CQ ; Ref, référence.

En supposant que les mécanismes par lesquels *P. falciparum* devient résistant aux médicaments soient similaires à ceux des cellules néoplasiquesⁱ, Martin *et al.* (1987) ont émis l'hypothèse que des agents inhibant la chimiorésistance des cellules cancéreuses puissent avoir le même effet sur *P. falciparum*¹⁷⁷. Parmi ces agents se trouve le vérapamil (VP), un médicament agissant comme antagoniste des canaux calciques. L'étude révèle que la molécule a complètement inversé la résistance à la CQ ($IC_{50} < 100$ nM) pour deux souches résistantes (Dd2 et 7G8). Le VP a par la suite été repris dans de multiples travaux comme expérience contrôle pour caractériser le phénotype de résistance à la CQ (Figure 23)⁵².

ⁱ Cellules cancéreuses.

S'il a clairement été démontré que les mutations du gène *pfcr*t sont le déterminant principal de la résistance à la CQ, d'autres gènes pourraient contribuer ou moduler ce phénotype. Par des approches variées incluant des transfections¹⁷⁸ et des analyses QTL, des mutations faux sens du gène *P. falciparum* multidrug resistance 1 (*pfmdr1*) ont également un rôle de modulateur de la résistance à la CQ^{168,179}. Nous n'entrerons pas dans les détails relatifs à cette protéine car elle n'a pas été investiguée dans ce travail de thèse. Toutefois, il est à noter que ce gène code une protéine de transport présente chez de très nombreux eucaryotes. *P. falciparum* en possède deux différentes : PfMDR1 et PfMDR2. La première citée est localisée à la membrane de la vacuole digestive du parasite et module le degré de résistance à plusieurs antipaludiques dont la CQ¹⁸⁰.

1.3.5. Évolution des mutants *pfcr*t et émergence de résistance à d'autres antipaludiques à base de quinoléine

Suite à l'adoption des CTAs recommandées par l'OMS, l'utilisation de la CQ – et de la pression de sélection associée – a diminué. Après quelques années, il a été observé en Afrique une augmentation de la fréquence des parasites *pfcr*t sauvages et sensibles à la CQ^{181,182}, certaines régions telles que la Zambie affichant même l'absence totale de parasites *pfcr*t mutants ($n = 302$ échantillons)¹⁸³. Le retour des parasites *pfcr*t sauvages serait lié principalement à un moindre *fitness* des parasites *pfcr*t mutants qui se reproduiraient moins vite que les sauvages, en absence de la sélection exercée par la CQ. Cet effet de compétition serait particulièrement efficace dans le contexte épidémiologique africain où la polyclonalité des infections est fréquente¹⁸⁴. Les haplotypes de résistance, notamment Dd2 (**Tableau 4**), demeurent très fréquents en Asie du sud-est, et ce malgré l'adoption des CTAs à la place de la CQ¹⁷⁵. Cette différence peut s'expliquer soit par un plus faible *fitness cost* du gène *pfcr*t mutant pour les parasites d'Asie du sud-est, soit par un plus faible taux d'infections polyclonales engendrant moins d'opportunités de compétition directe entre sauvages et mutants au sein d'une même infection.

Une étude récente basée sur les données génomiques de milliers d'isolats africains et asiatiques démontre la présence en Afrique de deux haplotypes *pfcr*t mutants dominants en Afrique¹⁷⁵. De façon surprenante, il ne s'agit pas de l'haplotype mutant asiatique Dd2 qui a initialement envahi l'Afrique et y est maintenant indétectable, mais de deux haplotypes (GB4 et Cam783) proches de Dd2 à une ou deux mutations près. Par rapport à Dd2, l'haplotype GB4 présente les allèles sauvages aux positions 326 et 356 et Cam783 présente l'allèle sauvage à la position 326¹⁷⁵. Les haplotypes GB4 et Cam783 sont associés à de plus faibles niveaux de résistance à la CQ et à l'amodiaquine, mais surtout à un bien meilleur *fitness* que Dd2 (taux de croissance *in vitro* sans drogue). Ces résultats suggèrent que ces mutations, en plus de leur rôle important dans la résistance, surviennent à des positions jouant probablement un rôle clé dans la fonction physiologique de PfCRT.

Récemment, des échecs thérapeutiques relatifs à la combinaison DHA-PPQ ont été observés en Asie du sud-est^{85,88,185,186}. Si la résistance à la DHA est conférée par des mutations localisées dans le gène *pfk13* (cf. section 1.2.5.), de nouvelles mutations situées sur *pfCRT* confèrent la résistance à la PPQ. Ces nouvelles mutations s'ajoutent sur un haplotype mutant de résistance à la CQ. La mutation C101F fut la première associée à une diminution de la sensibilité à la PPQ chez les parasites⁸⁷. D'autres mutations sur un gène *pfCRT* mutant ont par la suite émergé (F145I, H97Y, M343L et G353V ; **Figure 24A**), associées à un haut taux de survie des parasites tel que mesuré par la méthode « Piperaquine Survival Assay » (PSA_{0-3h} ; **Figure 24B**)¹⁸⁷, récemment développée et reprenant le concept du RSA_{0-3h} pour les ARTDs¹⁰³. D'autres isolats sont associés à un taux de survie élevé en absence de mutations (**Figure 24B**), suggérant la contribution de facteurs et/ou loci additionnels incluant le nombre de copies de gènes *P. falciparum* *plasmepsin* 2-3⁸⁸. En éditant le gène *pfCRT* chez des parasites Dd2 résistants à la CQ pour inclure ces différentes mutations, il a été confirmé le rôle de ces mutations dans la résistance à la PPQ (**Figure 24C**), tout en rendant les parasites sensibles à la CQ, l'amodiaquine et la quinine (**Figure 24D**). La mesure du transport de PPQ révèle toutefois un mécanisme de résistance différent de celui de la CQ : là où la résistance à la CQ se manifeste par un efflux de la molécule de la vacuole digestive vers le cytoplasme, mesurée dans des parasites par une diminution de l'accumulation cellulaire de CQ radioactive, aucune différence majeure dans le niveau d'accumulation cellulaire de la PPQ n'a été observée entre les souches sauvage et mutantes⁸⁹.

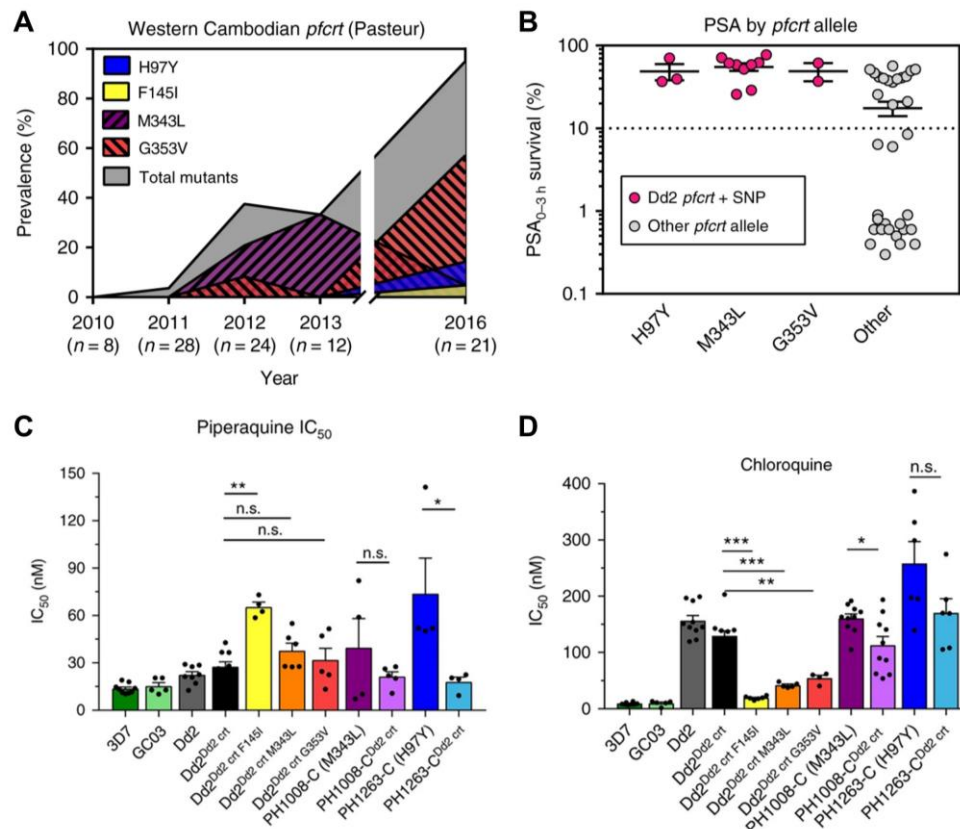


Figure 24 : Association entre des nouvelles mutations *pfCRT* (sur un fond génétique *pfCRT* mutant Dd2) et la résistance à la PPQ. (A) Prévalence de mutations *pfCRT* parmi 93 isolats issus du Cambodge, collectés

entre 2010 et 2016. Une augmentation rapide de mutations *pfcr*t sur fond génétique Dd2 peut être observée. Parmi les 93 isolats, un sous-ensemble de 51 isolats sera analysé par la méthode PSA_{0-3h}. **(B)** Valeurs PSA_{0-3h} pour différentes mutations *pfcr*t apparues sur fond génétique Dd2. Les mutations *pfcr*t H97Y, M343L et G353V sont associées à un taux de survie PSA_{0-3h} supérieur à 10%, suggérant leur rôle dans la résistance à la PPQ. D'autres isolats sont associés à un PSA_{0-3h} supérieur à 10% sans mutation *pfcr*t additionnelle, suggérant l'implication d'autres facteurs et/ou loci. **(C)** et **(D)** CI₅₀ relatives à des parasites génétiquement modifiés au locus *pfcr*t. Les mutations F145I, M343L et G353V ont été introduites dans un fond génétique Dd2. Ces mutations ont également été supprimées par édition de gène (PH1008-C^{Dd2 crt} et PH1263-C^{Dd2 crt}). 3D7, GC03 portent un haplotype sauvage *pfcr*t. Bien que non présentés, les résultats relatifs à la quinine et à l'amodiaquine révèlent des résultats similaires à ceux obtenus avec la CQ. D'après Ross *et al.* (2018)⁸⁹.

Au final, l'ensemble de ces données démontre le rôle de PfCRT dans la résistance à de multiples antipaludiques. Des analyses supplémentaires sont nécessaires pour comprendre le mécanisme de résistance relatif à chaque molécule thérapeutique.

Dans cette thèse, nos travaux sur PfCRT se concentreront sur la prédiction de sites fonctionnels et l'analyse *in silico* des mutations de résistance à la CQ et à la PPQ.

- Plusieurs molécules dérivées de la quinoléine ont été synthétisées pour former une classe d'antipaludiques : la CQ, la méfloquine, l'amodiaquine, l'halofantrine, la primaquine et la PPQ ; cette classe comprend aussi la quinine, produit naturel.
- La CQ s'accumule à de très fortes concentrations dans la vacuole digestive du parasite où elle interfère dans le processus de conversion des molécules d'hème toxique, issues de la dégradation de l'hémoglobine du GR, en hémozoïne non toxique.
- La CQ, utilisée durant quatre décennies, était un médicament remarquable : haute efficacité, très bonne tolérance, très bien acceptée par les populations, et faible coût.
- L'efficacité de la CQ comme antipaludique a diminué suite à l'émergence et l'expansion dans toutes les zones d'endémie de parasites *P. falciparum* résistants à cette molécule.
- La résistance à la CQ est associée à plusieurs combinaisons de changements d'acides aminés (ou haplotypes) sur la protéine transmembranaire PfCRT, avec un rôle clé de la mutation K76T.
- PfCRT est localisée à la membrane de la vacuole digestive, et les mutations permettent l'efflux de la CQ à l'état di-protoné de la vacuole digestive vers le cytoplasme, donc hors de son site d'action principal.
- Divers haplotypes mutants *pfcr*t – différemment répartis selon les zones géographiques – confèrent la résistance à la CQ.
- La fonction physiologique de PfCRT demeure mal caractérisée, mais aurait un rôle dans le transport de divers substrats : cations et fer.
- Aucune structure tertiaire de PfCRT n'a été déterminée expérimentalement à ce jour.
- De nouvelles mutations localisées sur *pfcr*t qui s'ajoutent à celles conférant la résistance à la CQ ont émergé en Asie du sud-est et confèrent la résistance à la PPQ, molécule partenaire utilisée dans les CTAs. Ces nouvelles mutations diminuent le degré de résistance à la CQ.
- Les résistances à la PPQ et à la CQ, *via* les mutations *pfcr*t, sont associées à des mécanismes différents. Celui relatif à la PPQ demeure encore inconnu.

Justification et objectifs du projet de recherche

Section 1.4



La vie n'est facile pour aucun d'entre nous. Mais quoi, il faut avoir de la persévérance, et surtout de la confiance en soi. Il faut croire que l'on est doué pour quelque chose, et que, cette chose, il faut l'atteindre coûte que coûte.

— **Marie Curie** (1867 – 1934)

Cette partie est dédiée à la présentation du projet de recherche faisant l'objet de cette thèse et de ses objectifs. La finalité première de ce travail est de mieux caractériser, par des approches bioinformatiques, les protéines PfK13 et PfCRT. Nous chercherons à : i) identifier des régions et des sites qui pourraient jouer un rôle dans leurs fonctions physiologiques respectives ; et ii) évaluer les altérations structurales et/ou physico-chimiques engendrées par les mutations de résistance aux antipaludiques. Nous expliciterons les principaux défis posés par ce projet de thèse et les solutions méthodologiques que nous avons tenté d'y apporter. Enfin, nous résumerons les différents moyens de valorisation et de diffusion employés pour les travaux réalisés.

La présente thèse a été réalisée au sein de l'UMR 261 MERIT « Mère et enfant en milieu tropical : pathogènes, système de santé et transition épidémiologique », une unité mixte de recherche de l'Université Paris Descartes et l'Institut de Recherche pour le Développement, dirigée par le Dr. André Garcia.

Ce travail de doctorat fait suite aux travaux initiés lors d'un stage de Master 2 effectué en 2016 dans ce même laboratoire, portant sur l'étude des pressions de sélection sur les protéines impliquées dans la résistance de *P. falciparum* aux antipaludiques.

La thèse a été menée sur une période de trois ans sous la direction du docteur Jérôme Clain (spécialisé dans les médicaments antipaludiques : modes d'action, et mécanismes et évolution des résistances) et co-encadrée par le docteur Audrey Sabbagh (spécialisée dans la génétique des populations et la phylogénie). Maria Miteva (INSERM U1268 MCTR, CiTCom UMR 8038 CNRS) a représenté une aide précieuse dans la bonne interprétation des simulations de dynamique moléculaire réalisées sur le domaine KREP de PfK13 ; et Isabelle Callebaut (UMR 7590, CNRS) a apporté de nombreux conseils dans la prédiction et l'interprétation de la structure tertiaire de PfCRT.

1.4.1. Problématique générale

Les pressions de sélection exercées sur les parasites *P. falciparum* par l'utilisation massive d'antipaludiques ont été à l'origine de phénomènes d'adaptation du parasite afin de résister aux médicaments. Diverses mutations parasitaires de résistance aux antipaludiques ont ainsi augmenté en fréquence dans les populations naturelles de *P. falciparum* en commençant le plus souvent par celles d'Asie du sud-est. Dans le cadre de ce travail de thèse, nous nous sommes spécifiquement intéressés à deux protéines impliquées dans la résistance : la protéine PfK13, dont les mutations localisées sur le domaine KREP confèrent une résistance aux ARTDs¹⁰⁴ ; et la protéine PfCRT, associée à la résistance à plusieurs antipaludiques à base de quinoléine incluant la CQ et la PPQ^{51,89}, cette dernière étant actuellement utilisée en combinaison avec la DHA³⁹.

Malgré les nombreux travaux conduits sur PfK13 et PfCRT, leurs fonctions physiologiques respectives chez le parasite demeurent globalement peu caractérisées. Par exemple, leur interactome – c'est-à-dire l'ensemble des molécules pouvant interagir avec ces protéines – n'a été que partiellement identifié. En effet, PfPI3K, avec laquelle PfK13 a été immunoprécipitée, représente aujourd'hui le seul substrat proposé de PfK13¹²⁶, et l'on ignore même si l'interaction entre ces deux protéines se fait de façon directe ou non. Concernant PfCRT, une multitude d'investigations basées sur la propriété de transport de cette protéine ont révélé plusieurs substrats potentiels^{164–166}, mais aucun – mis à part le glutathion – n'a été analysé directement chez le parasite¹⁶⁵. Pour ce dernier, son transport ne serait assuré que par les parasites mutants, il ne constitue donc probablement pas un substrat physiologique majeur. En outre, hormis les positions associées aux mutations de résistance, les régions fonctionnelles de ces protéines n'ont

pas été clairement mises en évidence : nous ignorons quels sont les sites impliqués dans des interactions avec les substrats, protéiques et non protéiques. Les identifier apporterait une contribution majeure à la caractérisation de la fonction de ces protéines. Enfin, dans un contexte de résistance, peu de données existent sur les altérations structurales et/ou physico-chimiques de PfK13 et PfCRT induites par les mutations de résistance. Mieux comprendre l'impact de ces mutations au niveau protéique pourrait permettre de formuler de nouvelles hypothèses sur leur fonctionnalité et/ou de conforter des propositions existantes sur les mécanismes de résistance aux antipaludiques.

Ce travail de thèse a ainsi pour objectif d'approfondir nos connaissances sur PfK13 et PfCRT, deux protéines d'intérêt clinique majeur. Deux questions principales ont été investiguées :

- Quelles sont les régions et sites potentiellement impliqués dans la fonction physiologique des protéines PfK13 et PfCRT ?
- Quel est l'impact des mutations de résistance sur la structure tertiaire et/ou les propriétés physico-chimiques de ces protéines ?

Ces questions ont été explorées par diverses approches exclusivement bioinformatiques basées sur l'évolution des séquences et la structure des protéines. Des expériences biochimiques de validation ont été initiées pour PfK13 et sont toujours en cours.

1.4.2. Quelles sont les régions et sites potentiellement impliqués dans la fonction physiologique de PfK13 et PfCRT ?

Au sein d'une même espèce, la vitesse d'évolution des gènes codant pour des protéines varie considérablement selon le type de protéine : celles fortement exprimées et/ou remplissant une ou plusieurs fonctions critiques pour le bon fonctionnement de l'organisme ont tendance à évoluer significativement plus lentement que les autres protéines¹⁸⁸. En plus de cette variabilité inter-protéines, les taux d'évolution varient naturellement au sein même des protéines entre les sites d'une séquence donnée (pour des raisons de recombinaison, de différents taux de mutation, d'importance fonctionnelle, *etc.*)¹⁸⁹. Cette variation inter-sites peut être imputable à une sélection positive, généralement liée à des changements des conditions environnementales induisant un fort stress biotique ou abiotique, et à une sélection purificatrice en éliminant les mutations délétères en raison de leur possible impact sur le fonctionnement de l'organisme qui les porte (**Encadré 5**). Ainsi, une grande partie de ce travail de thèse reposera sur la détection des signatures de sélection positive ou purificatrice afin de mettre en évidence des sites fonctionnels candidats. Plus spécifiquement, nous nous sommes concentrés sur les sites évoluant sous sélection purificatrice, en suivant l'hypothèse définie par le théoricien de l'évolution Motoo Kimura et la biologiste et généticienne Tomoko Ohta : « Les molécules ou parties de molécules moins importantes évoluent (en termes de changement d'acides aminés)

plus rapidement que les plus importantesⁱ »¹⁹⁰ au cours de l'évolution. Ainsi, certaines mutations sur des sites d'une protéine remplissant une ou plusieurs fonctions critiques pour la survie et/ou la reproduction de l'organisme seront éliminées par la sélection purificatrice. Ces sites sont donc plus conservés que le reste de la protéine (**Encadré 5**)¹⁸⁹, tels que des sites catalytiques ou ceux impliqués dans la fixation d'un substrat (**Figure 25**). De fait, des sites potentiellement fonctionnels peuvent être identifiés par des approches basées sur l'évolution de séquences codantes. Nous appellerons de tels sites, dans la suite de ce manuscrit, des sites fonctionnels *candidats*.

Encadré 5 : Définition des termes relatifs aux principes de l'évolution.

Sélection positive : Se produit lorsqu'une mutation bénéfique survient dans une population et augmente en fréquence, généralement en raison d'un avantage sélectif qu'elle confère à l'individu qui la porte dans un environnement donné. On parle également de sélection adaptative car elle permet l'adaptation d'un organisme à de nouvelles conditions environnementales.

Sélection purificatrice (ou négative) : Elle élimine les mutations délétères en raison de leur impact sur le fonctionnement de l'organisme qui les porte. La grande majorité des variations de séquence apparues au cours de l'évolution des espèces a été éliminée de manière à conserver la séquence protéique (et par là même, la fonction de la protéine).

Évolution neutre : Des mutations sont neutres du point de vue sélectif quand elles ne confèrent ni avantage ni désavantage à leur porteur. Conceptualisée par le généticien japonais Motoo Kimura, la « Théorie neutraliste de l'évolution » suggère que la plupart des changements affectant l'évolution de notre génome sont neutres et donc non adaptatifs.

Mutation non-synonyme : Mutation nucléotidique qui entraîne un changement de l'acide aminé sur la séquence protéique.

Mutation synonyme : Mutation nucléotidique qui n'altère pas la séquence protéique (mutation silencieuse).

Contrainte structurale : Évolution sous sélection purificatrice de sites importants pour le bon repliement structural et/ou le maintien de la structure de la protéine.

Contrainte fonctionnelle : Évolution sous sélection purificatrice de sites importants pour la fonction de la protéine (par exemple, des sites impliqués dans une interaction protéine-protéine).

ⁱ Citation originale : « Functionally less important molecules or parts of a molecule evolve (in terms of mutant substitutions) faster than more important ones ».

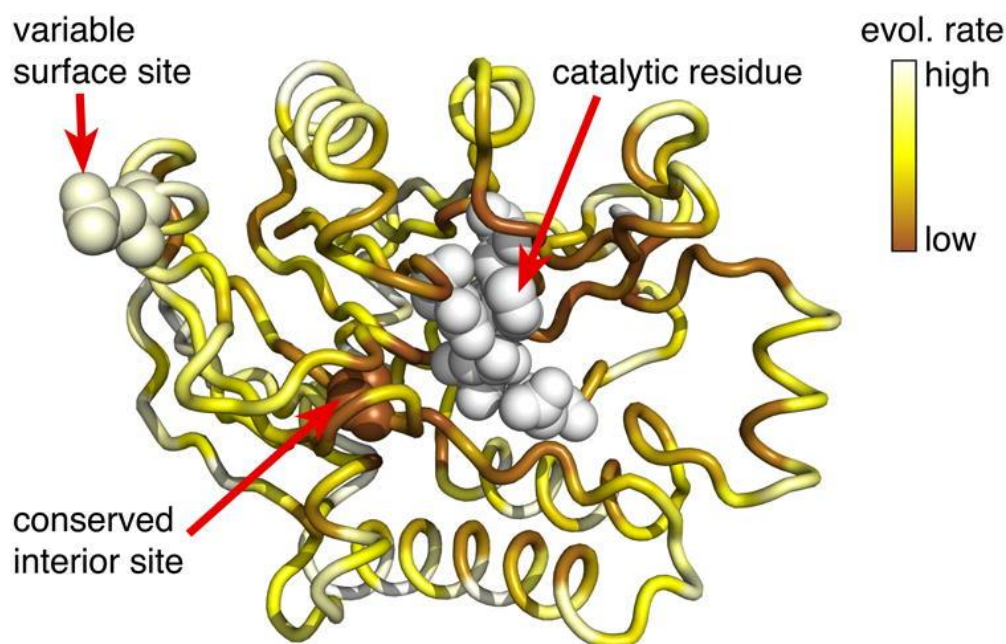


Figure 25 : Influences des contraintes structurales et/ou fonctionnelles sur les taux d'évolution spécifiques aux sites. À partir de la structure de l'exonuclease III de *E. coli* (PDB : 1AKO), différents taux d'évolution sont observés, allant du blanc pour les sites très conservés au marron pour les sites évoluant rapidement. Le site actif évolue lentement, tout comme les sites liés aux contraintes structurales, tandis que ceux à la surface de la protéine ont tendance à évoluer plus rapidement. La figure est tirée du travail d'Echave *et al.* (2016)¹⁸⁹.

Les méthodes d'inférence des taux de substitution spécifiques aux sites, généralement dédiées à la recherche de sites évoluant sous sélection positive, requièrent deux types de données : un alignement multiple de séquences orthologues – protéiques ou nucléotidiques – couvrant un nombre suffisant d'espèces pour être informatif (un lien phylogénétique entre les espèces plus ou moins fort sera privilégié en fonction du taux d'évolution de la protéine étudiée), et un arbre phylogénétique non enraciné construit à partir de cet alignement multiple¹⁸⁹. De récentes méthodes incluent également les informations relatives à la structure tertiaire afin de prédire des *patches* fonctionnels, se traduisant par un groupe de sites très conservés (c'est-à-dire soumis à une forte sélection purificatrice) et spatialement proches dans la structure tertiaire^{191,192}. Ces méthodes nécessitent des jeux de données volumineux en termes de nombre de séquences, ou au moins diversifiés (séquences présentant un certain nombre de différences entre elles), pour être informatifs et obtenir une inférence fiable. Préférentiellement, la plupart des sites pour un alignement donné doivent avoir connu plusieurs événements de substitution au cours de l'évolution des espèces pour qu'une estimation du taux d'évolution raisonnablement précise soit effectuée. Cependant, le niveau de divergence des séquences ne doit pas être trop élevé pour permettre de repérer facilement les régions d'homologie et de réaliser un alignement non ambigu des séquences (positionnement des événements d'insertion / délétion).

Outre une séquence (sauvage si l'on s'intéresse à la fonction native) de la protéine d'intérêt et une phylogénie construite à partir d'un alignement multiple de séquences orthologues de cette protéine dans plusieurs espèces, la méthodologie peut nécessiter une structure tertiaire

permettant d'étudier la corrélation spatiale des taux de conservation des sites. Si pour PfK13, la structure cristallographique des domaines BTB et KREP a été résolue aux rayons X en 2015 par le *Structural Genomics Consortium*, bien que non publiée dans un article scientifique à comité de lecture¹⁰⁹, celle de PfCRT demeure inconnue. Ici, la difficulté fut de prédire par homologie une structure tertiaire de PfCRT, sachant que peu de structures relatives aux transporteurs de la famille DMT ont été déterminées, avec des taux d'identité de séquence protéique par rapport à PfCRT extrêmement faibles (< 15%). Un long travail de modélisation couplant prédiction structurale, affinement et minimisation énergétique (afin d'obtenir une conformation dite stable) a été réalisé pour produire un modèle PfCRT selon des états conformationnels différents (*inward-facing* ou ouvert, et *occluded* ou fermé, comme la grande majorité des transporteurs). Une seconde difficulté résidait en la détection de *patches* fonctionnels candidats dans des protéines extrêmement conservées, comme c'est le cas pour la protéine PfK13 et dans une moindre mesure pour PfCRT. Dès lors, un objectif fut de concevoir des jeux de données suffisamment riches en information et d'intégrer des données structurales pour pouvoir détecter de façon précise et robuste ces *patches* fonctionnels.

Les différents sites fonctionnels candidats seront par la suite étudiés via des approches biochimiques et génétiques pour une validation expérimentale de leur rôle fonctionnel. Dans cet objectif, en fin de thèse, nous avons commencé à produire des domaines recombinants de PfK13 sauvages et mutés dans la bactérie *Escherichia coli* en vue de réaliser des expériences de *pull-down* différentiels et ainsi étudier ses interactions physiques protéine-protéine. Cette technique, en complément avec d'autres outils comme la transfection et l'édition du génome chez le parasite, permettrait de confirmer le rôle des sites candidats dans la fonction de PfK13, mais également d'identifier des substrats protéiques. Les premiers résultats de ces différentes expériences biochimiques que nous avons initiées seront présentés ici, bien que ces investigations soient maintenant menées à temps plein par la doctorante Laurine Rondepierre (UMR 261). En parallèle, des *docking* ont été réalisés sur le *patch* fonctionnel prédit du domaine KREP de PfK13 en utilisant les informations disponibles dans la littérature concernant des molécules ciblant des sites fonctionnels d'autres protéines KREP et similaires à ceux que nous avons identifiés.

1.4.3. Quel est l'impact des mutations de résistance sur la structure tertiaire des protéines PfK13 et PfCRT ?

Si les mutations localisées sur PfK13 et PfCRT ont été montrées depuis plusieurs années comme conférant respectivement la résistance aux ARTDs et à diverses molécules à base de quinoléine^{51,89,104}, leur impact sur certaines propriétés structurales et/ou physico-chimiques n'a été que peu étudié. Par conséquent, un second objectif de ce travail de thèse fut de mieux comprendre l'effet de quelques mutations de résistance sur la structure tertiaire de ces protéines.

La dynamique moléculaire constitue l'une des approches bioinformatiques les plus fiables pour évaluer l'impact d'une mutation sur une structure protéique. Elle consiste à étudier via des simulations la trajectoireⁱ et l'évolution du comportement de la protéine au cours du temps dans un solvant artificiel – généralement de l'eau. Cette approche a été conduite dans le cas de la protéine PfK13 pour quelques mutations impliquées dans la résistance aux ARTDs (R539T et C580Y). Nous n'avons toutefois pas réalisé ce travail pour PfCRT pour plusieurs raisons : outre la difficulté à générer la bicouche lipidique et formuler les contraintes structurales, les modèles de structure tertiaire que nous avons générés demeurent prédictifs et incomplets. En revanche, nous avons conduit des mutagénèses *in silico* et étudié certains paramètres structuraux tels que le potentiel électrostatique, qui pourrait jouer un rôle majeur étant donné que la CQ est évacuée de la vacuole digestive du parasite dans un état di-protoné^{171,172}.

1.4.4. Congrès et publications

Les travaux de recherche relatifs à la protéine PfK13 ont été présentés sous forme d'un poster accompagné d'un *flash-talk* « *Delineating functional regions of the K13 artemisinin resistance protein through evolutionary and structural analyses* » au congrès BioMalPar XIV en mai 2018 (Heidelberg, Allemagne) et d'une présentation orale « *Evolutionary insights into Kelch13 and its relationships with artemisinin resistance in Plasmodium falciparum malaria parasites* » au congrès Young Researchers in Life Science en mai 2017 (Paris, France). Enfin, les résultats bioinformatiques les plus importants ont été publiés dans le journal *Scientific Reports* en juillet 2019, sous le titre « *Comparative structural and evolutionary analyses predict functional sites in the artemisinin resistance malaria protein K13* »¹⁹³.

Un manuscrit sur les résultats de nos recherches sur la protéine PfCRT, intitulé « *Structural and evolutionary analyses of the Plasmodium falciparum chloroquine resistance transporter* » a été soumis mi-août 2019 pour publication dans une revue scientifique internationale à comité de lecture (*Scientific Reports*), et est actuellement en révision.

ⁱ Succession des conformations de la structure tertiaire.

- La thèse s'est déroulée au sein de l'UMR 261 MERIT sous la direction de Jérôme Clain et a été co-encadrée par Audrey Sabbagh.
- Deux questions majeures ont été investiguées par des approches bioinformatiques :
 - Quelles sont les sous-régions les plus conservées, et probablement fonctionnelles, des protéines PfK13 et PfCRT ?
 - Quel est l'impact des mutations de résistance sur la structure tertiaire de ces protéines ou des domaines protéiques concernés ?
- Un volet d'exploration biochimique a été initié afin de valider une zone présumée d'interaction protéine-protéine dans le domaine KREP de PfK13, actuellement poursuivi par Laurine Rondepierre (doctorante, UMR 261).
- Le travail de thèse a débouché sur :
 - Une publication en premier auteur dans le journal *Scientific Reports* ;
 - Un second article en premier auteur soumis et actuellement en révision ;
 - Deux congrès internationaux (un poster et un *flash-talk*, et une présentation orale).

Partie 2

Méthodologie



*On résiste à l'invasion des armées, on
ne résiste pas à l'invasion des idées.*

— **Victor Hugo** (1802 – 1885)

Ce travail de thèse s'appuie principalement sur des méthodes bioinformatiques. Dans cette section, nous développerons les concepts théoriques des différentes méthodes mises en œuvre pour répondre aux problématiques biologiques exposées en première partie (cf. section 1.4.1.). Nous détaillerons les protocoles expérimentaux permettant de détecter les sites sous sélection naturelle à partir de séquences d'ADN codant ou de protéines, et en particulier celui couplant évolution et structure des protéines. Puis, nous présenterons les différents paramètres structuraux qui ont été analysés. Enfin, les notions de prédiction structurale, de dynamique et de *docking* moléculaires seront également explicitées.

2.1.1. Mesure des pressions de sélection agissant sur les gènes *pfk13* et *pfcr1*

Dans un contexte de séquences nucléotidiques codant pour des protéines, les taux d'évolution sont généralement estimés par le rapport $\omega = d_N/d_S$, où d_N correspond au taux de substitution non-synonyme, et d_S au taux de substitution synonyme¹⁹⁴. Le ratio ω est principalement utilisé pour détecter des sites évoluant sous sélection positive, c'est-à-dire lorsqu'une mutation bénéfique survient dans une population et augmente en fréquence, souvent en raison d'un avantage sélectif. Typiquement, un rapport $\omega > 1$ suggère une sélection positive ; un rapport $\omega = 1$ reflète une évolution neutre ; et un rapport $\omega < 1$ reflète l'action d'une sélection purificatrice (dite aussi sélection négative ; **Encadré 5**)^{194,195}. Si plusieurs approches ont été développées pour estimer les taux d'évolution spécifiques aux sites, la plus robuste est l'approche *Random-Effects Likelihood* (connue sous l'abréviation REL) qui considère le taux comme une variable aléatoire le long de la séquence protéique^{196,197}. Dans ce travail de thèse, cette approche a été suivie pour *pfk13* et *pfcr1* en utilisant l'outil codeml du package PAML v4.8¹⁹⁵. L'outil inclut un certain nombre de modèles d'évolution ayant chacun ses spécificités (**Tableau 5**). L'hétérogénéité du rapport d_N/d_S entre les lignées ou branches de la phylogénie peut être testée par comparaison du modèle « *free-ratio* », qui suppose autant de valeurs du paramètre ω que de nombre de branches dans la phylogénie, avec le modèle nul M0, lequel suppose une unique valeur de ω pour toutes les branches de l'arbre. La variation du rapport d_N/d_S entre les différents codonsⁱ le long d'une séquence codante est quant à elle évaluée par différents modèles dits « modèles de sites » : M1a, M2a, M3, M7 et M8^{198,199}. Le modèle M1a considère uniquement deux classes de sites possibles : l'une avec un $\omega = 1$ et l'autre avec un $\omega < 1$, en excluant donc la présence possible de sites sous sélection positive. Le modèle M2a reprend les bases du modèle M1a, en ajoutant une troisième classe ω pouvant inclure des sites évoluant sous sélection positive ($\omega > 1$). Le modèle M3 inclut une distribution discrète de rapports ω indépendants, dont le nombre et les valeurs sont estimés à partir du jeu de données. Le modèle M7 assume une loi bêta de dix ratios ω compris dans un intervalle [0 ; 1], tandis que le modèle M8 reprend les bases du modèle M7, en ajoutant la possibilité d'avoir un rapport $\omega > 1$ (de façon similaire au modèle M2a). L'hétérogénéité du rapport ω entre les codons – suggérant que les sites (ou codons) de la protéine (ou séquence codante) évoluent à des taux différents – est testée en comparant le modèle nul M0 avec le modèle M3. La comparaison des modèles M1a et M7 respectivement avec les modèles M2a et M8 constituent deux tests statistiques permettant de détecter la présence de sélection positive (avec éventuellement la détection des codons spécifiquement ciblés par cette sélection)¹⁹⁸. Les comparaisons de modèles sont réalisées à l'aide d'un test statistique basé sur un rapport de vraisemblance, le test LRT (pour *Likelihood Ratio Test*)²⁰⁰.

ⁱ Triplet de nucléotides (désigné par les initiales des noms des trois bases respectives) qui détermine la synthèse cellulaire des acides aminés.

Tableau 5 : Les différents modèles de substitution des sites proposés par l'outil codeml de PAML.

Modèle	Particularités	Nombre de classes de ω	Nb Par. ^a	Paramètres
M0	ω unique le long de la séquence	1	1	ω
M1	Modèle neutre	2	2	$p_0, p_1 (= 1 - p_0)$ $\omega_0 < 1$ et $\omega_1 = 1$
M2	Modèle de sélection positive	3	4	$p_0, p_1, p_2 (= 1 - p_0 - p_1)$ $\omega_0 < 1, \omega_1 = 1$ et $\omega_2 > 1$
M3	Modèle discret, ω variable	≥ 2	5	$p_0, p_1, (p_2 (= 1 - p_0 - p_1))$ ω_0, ω_1 (et ω_2)
M7	Modèle neutre	10	2	p et q (paramètres de la loi)
M8	Modèle de sélection positive	10	4	$p_0, p_1 (= 1 - p_0)$ p, q et $\omega_{10} > 1$

Note – ^a Nb Par. : nombre de paramètres p impliqués dans les différents modèles ; celui-ci varie selon le nombre de classes ω . Les paramètres p_0, p_1 et p_2 représentent les proportions de sites associées à ω_0, ω_1 et ω_2 respectivement.

L'une des particularités de l'outil codeml implémenté dans PAML est de pouvoir réaliser des partitionnements au sein de la séquence de référence dans l'alignement multiple des séquences nucléotidiques. Autrement dit, il est possible de tester si des groupes de codons (appelées partitions ci-après) correspondants à différentes régions d'une protéine évoluent de façon variable²⁰¹. L'analyse de partitions, que nous avons menés pour le gène *pfprt*, est basée sur une suite de modèles, baptisés A à E (**Tableau 6**). Le modèle A suppose une homogénéité de tous les paramètres évolutifs entre les partitions. Les modèles B à E se complexifient par ajout progressif de paramètres évolutifs différents entre les partitions : un taux différent de substitution nucléotidique (rs , modèle B), une variation dans la fréquence des codons (πs , modèle C), un rapport d_N/d_S différent (ω , modèle D), et une hétérogénéité du rapport transitionⁱ/transversionⁱⁱ (κ , modèle E). Tout comme les modèles M0 à M8, ces modèles à partitionnement sont comparés par paire à l'aide de la statistique LRT. La confrontation des modèles A et B teste si les taux de substitution nucléotidique (rs) diffèrent significativement entre les partitions ; tandis que les comparaisons des modèles B:D et C:E évaluent la variation des rapports $\pi s, \kappa, \omega$ entre les partitions²⁰¹. Une plus haute vraisemblance des modèles D et E par rapport à B et C suggère que les codons relatifs aux différentes partitions n'évoluent pas à la même vitesse.

Tableau 6 : Modèles de substitution à partitionnement proposés par l'outil codeml de PAML.

Modèle	Paramètres de partition	Paramètres
A	Mêmes taux de κ, ω et πs	$b + 2 + 9$
B	Différents rs , mêmes κ, ω et πs	$b + (g - 1) + 2 + 9$
C	Différents rs et πs , mêmes κ et ω	$b + (g - 1) + 2 + g * 9$
D	Différents rs, κ et ω , même πs	$b + (g - 1) + g * 2 + 9$
E	Différents rs, κ, ω et πs	$b + (g - 1) + g * 2 + g * 9$

Note – Les modèles sont spécifiés par l'option G dans le fichier de l'alignement et par la variable Mgene du fichier de contrôle PAML. κ , rapport transition/transversion ; $\omega, d_N/d_S$; πs , fréquence des codons ; rs , taux de substitution nucléotidique ; b, nombre de branches de l'arbre phylogénétique ; g, nombre de partitions.

ⁱ Remplacement d'une purine (A ou G) par une purine, ou d'une pyrimidine (T ou C) par une pyrimidine.

ⁱⁱ Remplacement d'une purine par une pyrimidine ou d'une pyrimidine par une purine.

2.1.2. Mesure de la vitesse d'évolution dans un contexte structural de Pfk13 et PfcRT

La plupart des méthodes mesurant les taux d'évolution spécifiques aux sites (à l'échelle des codons ou des acides aminés) – incluant PAML – ont été conçues pour rechercher des traces de sélection positive (**Tableau 7**). Même si PAML, par exemple, peut être utilisé pour rechercher les sites sous sélection purificatrice, en extrayant les valeurs ω individuelles pour chaque codon, aucune statistique n'est prévue pour cette application. En outre, ces méthodes supposent une indépendance des taux d'évolution entre les sites²⁰², suggérant que les sites fonctionnels sont localisés aléatoirement dans la structure tertiaire. Or, les régions fonctionnelles – comme un site catalytique d'une enzyme – sont souvent constituées de plusieurs sites. Pour prendre en compte cette dimension spatiale, des méthodes avec fenêtre glissante (*sliding-window*) appliquée à la structure tertiaire ont été développées, mais la taille de la fenêtre doit être fixée *a priori*. Pour éviter cet écueil, l'outil FuncPatch a été conçu pour rechercher spécifiquement des *patches*ⁱ de sites très conservés en estimant les taux de substitution spécifiques aux sites dans la structure tertiaire ainsi que leur corrélation spatiale^{191,192}. Si une corrélation spatiale des taux de substitution est détectée pour la protéine d'intérêt, alors cette méthode est plus informative que les autres méthodes, comme PAML¹⁹¹. Outre l'alignement des séquences protéiques orthologues et la phylogénie correspondante, cette méthode requiert donc une structure tertiaire de la protéine – qu'elle ait été déterminée expérimentalement ou inférée par des méthodes *in silico* (**Tableau 7**). La finalité de cette méthode est de rechercher des *patches* de sites extrêmement conservés durant l'évolution des espèces et qui pourraient donc avoir un rôle fonctionnel et/ou structural^{191,192}. Ici, les taux de substitution (à l'échelle des acides aminés d'une protéine) estimés sont utilisés comme indicateurs de la fonctionnalité : de faibles taux de substitution suggèrent l'action d'une sélection purificatrice visant à conserver des sites critiques pour la fonction et/ou la structure de la protéine. À titre d'exemple, les auteurs de FuncPatch ont montré que les taux de substitution spécifiques aux sites étaient spatialement corrélés au sein de la protéine humaine B7-1 (impliquée dans un signal permettant l'activation des lymphocytes T, essentiels dans la réponse immunitaire secondaireⁱⁱ chez les mammifères) et ont détecté un *patch* de sites très conservés au niveau du domaine IgC, qualifié de fonctionnel car correspondant au domaine de liaison avec CTLA-4 et CD28 (**Figure 26, panel gauche**)²⁰³. À l'inverse, une estimation des taux de substitution sans l'apport de la structure tertiaire (avec PAML par exemple) révèle une distribution hétérogène des sites les plus conservés dans la structure, correspondant aux acides aminés signatures de la famille structurale considérée (**Figure 26, panel droite**)¹⁹².

ⁱ Groupe de sites spatialement proches dans la structure tertiaire de la protéine.

ⁱⁱ Lors d'un premier contact avec un antigène, le système immunitaire déclenche des mécanismes de défense : la réponse adaptative primaire, se traduisant par la production d'anticorps et par la multiplication des lymphocytes spécifiques de l'antigène. Lors d'un second contact avec le même antigène, le système immunitaire déclenche une réponse immunitaire plus rapide et plus efficace : la réponse secondaire.

Tableau 7 : Comparaison des méthodes PAML et FuncPatch.

Méthode	Objectif premier de la méthode	Type de séquences	Nécessite une structure tertiaire
PAML	Recherche de sites (codons) évoluant sous sélection positive Recherche de groupes de sites (acides aminés) très conservés et spatialement proches dans la structure tertiaire d'une protéine	Nucléotidiques	Non
FuncPatch		Protéiques	Oui

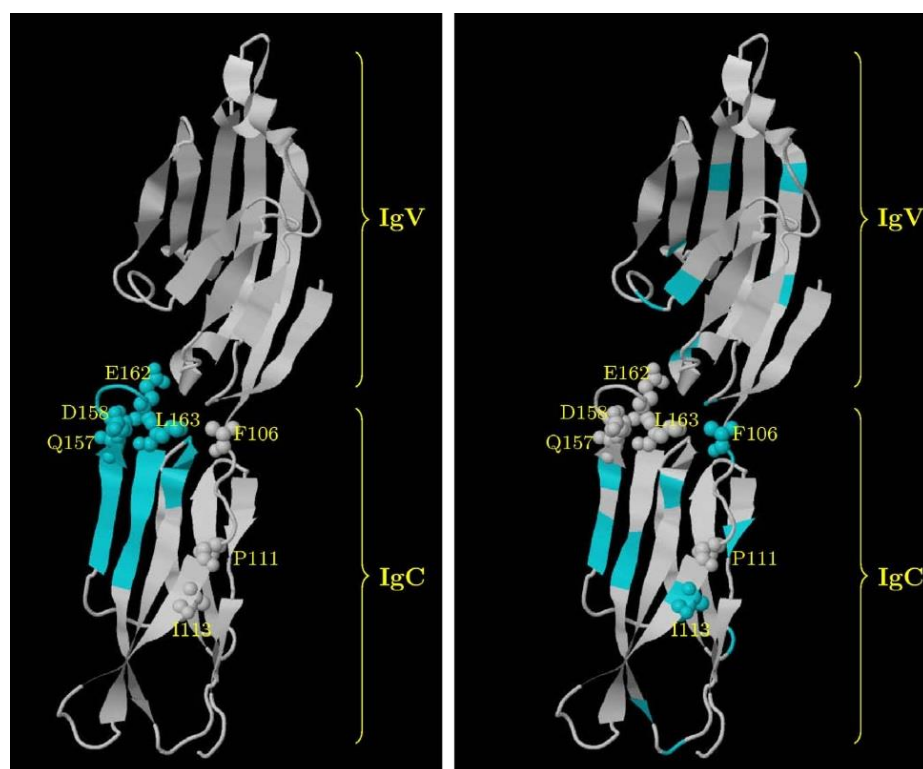


Figure 26 : Localisation dans la structure tertiaire des sites les plus conservés (au seuil de 10%) dans la protéine humaine B7-1 en considérant (panel gauche) ou non (panel droite) l'environnement spatial des sites pour l'estimation de leur taux de substitution. GP4Rate a été utilisé pour estimer les taux de substitution en prenant en compte la structure tertiaire¹⁹² ; Rate4Site a été utilisé pour estimer les taux de substitution sans considérer les informations de la structure tertiaire. Les sites les plus conservés sont représentés en bleu cyan ; les sites clés dans la fonction de B7-1 sont indiqués en lettres jaunes. La figure est tirée du travail de Huang et Golding (2014)¹⁹².

Si les sites fonctionnels sont très conservés et que plusieurs forment des clusters dans la structure tertiaire, leurs taux de substitution doivent être positivement corrélés. Brièvement, le principe de l'estimation des taux de substitution spécifiques aux sites réalisés par FuncPatch est le suivant : les taux sont calculés à l'échelle protéique (alignement de séquences d'acides aminés) par une approche similaire à la méthode REL (précédemment citée)¹⁹². La proximité physique des acides aminés dans la structure tertiaire utilisée (correspondant à une des séquences de l'alignement multiple) est calculée à partir des distances euclidiennes par paire d'acides aminés, basées sur les atomes de carbone α (Ca). La détection d'une corrélation spatiale positive des taux de substitution se fait par la comparaison de modèles bayésiens. Un modèle nul – sans

corrélation spatiale des taux de substitution – est construit, puis comparé au modèle alternatif qui considère une telle corrélation. À partir de multiples jeux de données (alignements multiples de séquences protéiques orthologues, phylogénies et structures tertiaires), il a été proposé empiriquement qu'un facteur bayésien différentiel entre les modèles supérieur à une valeur conservative égale à 8 se traduit par la positivité d'une corrélation spatiale des taux de substitution¹⁹¹. Les sites les plus conservés et spatialement proches (*patch*) dans la structure tertiaire est révélé par la mise en évidence des sites les plus conservés à un seuil de 10%, fixé par les auteurs de FuncPatch¹⁹¹. Cette approche a notamment servi dans ce travail de thèse à délimiter les régions possiblement impliquées dans la fonction physiologique de PfK13 et PfCRT.

2.1.3. Approche comparative évolutive et structurale pour la recherche de sites fonctionnels candidats

Une fois un *patch* de sites conservés identifié, l'étape suivante consistait à cibler plus précisément au sein de celui-ci les sites pouvant jouer un rôle clé dans la fonction de la protéine. Cela peut être réalisé en s'appuyant sur les sites fonctionnels, expérimentalement validés, de protéines ou domaines protéiques appartenant à la même famille structurale que la protéine d'intérêt (nommées ensuite protéines ou structures de *référence*). Par superposition des structures tertiaires de *référence* avec la structure d'intérêt, nous pouvons identifier des sites conservés du *patch* de la protéine d'intérêt qui s'alignent structuralement avec des sites fonctionnellement validés des protéines de *référence*. Par transfert, nous pouvons supposer leur possible rôle fonctionnel dans la protéine d'intérêt. Cette stratégie est présentée sous forme de graphe en **figure 27**.

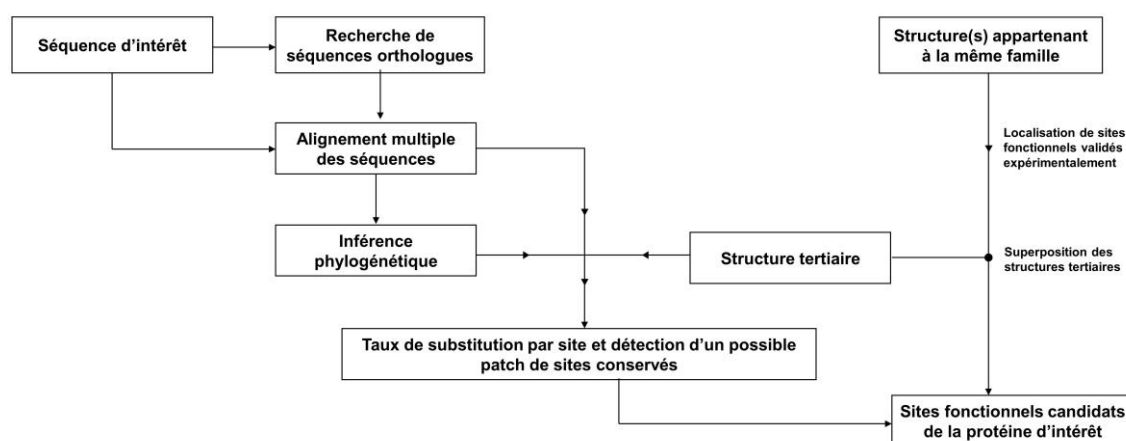


Figure 27 : Recherche de sites fonctionnels candidats par une approche comparative évolutive et structurale. À partir de la séquence d'une protéine d'intérêt dont on ignore les sites impliqués dans sa fonction, une analyse évolutive et structurale peut permettre de délimiter les zones soumises à une forte contrainte structurale et/ou fonctionnelle. En réalisant la même approche sur des domaines structuraux fonctionnellement annotés et appartenant à la même famille structurale, il peut être possible d'extraire au sein des *patches* de sites conservés des candidats ayant un possible rôle fonctionnel.

2.1.4. Prédiction de la structure tertiaire des protéines

La structure tertiaire d'une protéine représente son repliement dans l'espace. Bien que ces dernières années, on assiste à une augmentation exponentielle du nombre de structures déterminées par des techniques expérimentales (rayons X, RMN, *etc.* ; <https://www.rcsb.org/stats/growth/overall>), pour la très grande majorité des protéines, une structure tertiaire déterminée expérimentalement n'est pas disponible. Ceci est d'autant plus vrai pour les protéines transmembranaires (comme PfCRT) dont l'expression hétérologue et l'extraction des membranes sont plus délicates. À titre d'exemple, la *Protein Data Bank* (PDB) inclut à peine 0,2% des séquences répertoriées dans la base de données de séquences UniProt²⁰⁴. Cependant, diverses méthodologies ont été développées pour prédire la structure tertiaire d'une protéine en se basant sur des structures tertiaires connues.

Pour une protéine homologue à une protéine de structure connue (cette dernière sera nommée ensuite structure *parente*), une approche comparative, appelée modélisation par homologie, peut être réalisée²⁰⁵. Inversement, pour une protéine dont la structure semble totalement inconnue, on optera pour une méthode de reconnaissance structurale en utilisant une bibliothèque de motifs structuraux²⁰⁶. Dans ce travail de thèse, la structure tertiaire du domaine coiled-coil de PfK13 étant non déterminée, celle-ci a été prédite par une méthode de reconnaissance structurale puisqu'aucune structure parente proche n'a été détectée dans les banques de données. Concernant PfCRT, dont la structure est non déterminée à ce jour, nous avons opté pour la modélisation par homologie pour prédire différents états conformationnels du transporteur, étant donné la disponibilité de structures tertiaires de protéines *parentes* appartenant à la même superfamille (DMT). Les différentes étapes de la stratégie suivie pour PfCRT sont présentées en **figure 28**.

La modélisation par homologie comprend un certain nombre d'étapes incontournables, commençant par la recherche d'un repliement 3D connu jugé similaire à la protéine (ou au domaine) cible (ici, PfCRT), et se terminant par l'optimisation et l'affinement de la prédiction structurale²⁰⁵. La première étape consiste en la recherche d'une ou plusieurs structures existantes dans la PDB à l'aide de l'outil BLAST. Un certain nombre de résultats sera retourné, détaillant les alignements de séquences protéiques par paire (utile pour cibler les régions conservées) et des scores sur la significativité statistique de la similarité entre les deux séquences d'acides aminés (à noter qu'un faible pourcentage d'identité de séquence n'empêche pas la similarité structurale). Afin d'établir une relation entre la séquence d'intérêt et la structure tertiaire pour des molécules divergentes, des techniques de « *threading* » ont été conçues, consistant à mesurer l'adéquation entre un repliement 3D et une séquence peptidique. Le *threading* permet d'identifier des homologies structurales, et ce même lorsque le pourcentage d'identité entre deux séquences est inférieur à 20%^{207,208}. L'outil Phyre2 se base notamment sur cette approche, permettant d'identifier des structures connues pour lesquelles un très faible pourcentage d'identité de séquence avec la protéine cible est observé, mais présentant une haute

probabilité dans la similarité des repliements structuraux²⁰⁹. Cet outil nous a notamment servi à identifier différents transporteurs appartenant à la même superfamille que la protéine PfCRT, malgré un pourcentage d'identité de séquence d'acides aminés compris entre 12 et 15%.

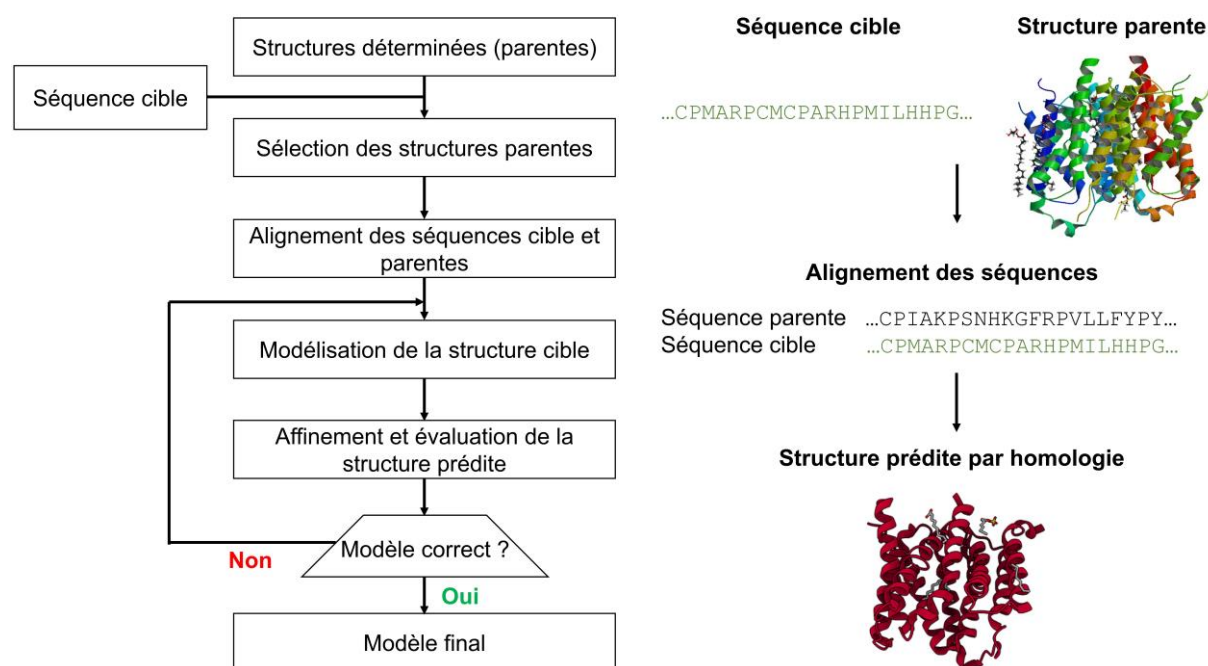


Figure 28 : Différentes étapes de la prédiction de la structure tertiaire d'une protéine par modélisation par homologie. La figure a été adaptée du travail de Sliwoski et ses collaborateurs (2014)²¹⁰.

Une fois la (ou les) structure(s) parente(s) identifiée(s), un alignement de la séquence d'acides aminés de la protéine cible avec celle de chaque structure parente est réalisé²⁰⁵ (plusieurs alignements si plusieurs structures parentes ont été sélectionnées). Il s'agit de l'opération la plus délicate et nécessitant le plus d'attention dans la modélisation par homologie. Une erreur dans l'alignement peut provoquer des erreurs notamment dans les structures secondaires, pour des raisons d'incompatibilité de chaînes latérales. Pour éviter ce type d'erreurs, une première étape consiste à identifier les invariants structuraux, c'est-à-dire les acides aminés communs parmi les membres de la famille structurale, souvent importants pour le bon repliement de la protéine²¹¹. Ces invariants servent de points d'ancrage pour l'alignement, que nous avons réalisé en utilisant une méthode progressive qui considère la structure secondaire et le profil hydrophatiqueⁱ de PfCRT²¹².

L'alignement des paires de séquences d'acides aminés (parente et cible) étant réalisé, la prédiction de la structure tertiaire de la protéine cible (PfCRT) peut débuter. Dans ce travail de thèse, nous avons utilisé l'outil MODELLER²¹³, lequel se base sur une suite d'algorithmes pour

ⁱ Mesure permettant de connaître le caractère hydrophile ou hydrophobe de la chaîne latérale d'un acide aminé (à partir de la séquence d'acides aminés).

calculer automatiquement des modèles structuraux en satisfaisant les contraintes spatiales incluant :

- des contraintes relatives aux distances inter- et intra-atomiques et aux angles dièdresⁱ à appliquer sur les acides aminés de la séquence cible, extraites à partir de son alignement et avec les informations structurales des séquences parentes ;
- des contraintes stéréochimiques incluant les longueurs des liaisons inter-atomiques ;
- des préférences statistiques pour les angles dièdres et les interactions inter-atomiques non liées (électrostatiques et van der Waals) en se basant sur un ensemble représentatif de structures protéiques connues.

Ces différentes contraintes spatiales, individuellement exprimées en tant que fonction de densité de probabilité, sont combinées dans une fonction objective, optimisée par une combinaison de gradients conjuguésⁱⁱ et de courtes dynamiques moléculaires associées à un recuit simulé^{iii,213}. Un nombre k de modèles prédits peut être généré (ici, nous avons fixé la valeur $k = 1\,000$). Ces modèles sont ensuite classés sur la base d'un score énergétique tel que le DOPE (pour *Discrete Optimized Protein Energy*), lequel calcule le potentiel énergétique basé sur l'ensemble des interactions entre les paires d'atomes dans un rayon donné. Plus l'énergie est négative, plus stable est la conformation prédite²¹³.

Le meilleur modèle de structure tertiaire obtenu a ensuite été optimisé à l'aide de différents outils avant d'être utilisé pour des analyses structurales approfondies. À titre d'exemple, nous avons appliqué des méthodes d'affinement des boucles – dont la structure tertiaire est souvent plus difficile à modéliser – et des minimisations d'énergie (avec l'outil YASARA²¹⁴) des modèles PfCRT pour générer les structures les plus stables possibles²¹⁵. Une méthode de vérification systématique consiste à analyser la conformation du squelette polypeptidique des protéines²¹⁶. Chaque acide aminé est associé à des angles dièdres ϕ et ψ ^{iv} compris entre -180 et $+180$, sachant que certaines valeurs ne sont pas possibles car certaines distances inter-atomiques sont énergétiquement défavorables pour la stabilité de la structure (diagramme de Ramachandran ; **Figure 29**). Un modèle de qualité présente généralement plus de 98% d'acides aminés dont les angles ϕ et ψ sont situés dans les régions favorables^{v,216}. Enfin, d'autres méthodes – certaines dédiées aux protéines transmembranaires telles que QMEANBrane²¹⁷ utilisée ici – permettent de mesurer l'énergie dégagée par la conformation prédite, puis de la comparer à celles de toutes les structures de la PDB. Si le modèle de structure tertiaire est associé à des profils énergétiques

ⁱ Angle entre deux plans.

ⁱⁱ Permet de résoudre des systèmes d'équations linéaires dont la matrice est symétrique et définie positive (c'est-à-dire si toutes ses valeurs propres sont strictement positives).

ⁱⁱⁱ Méthode empirique dans laquelle on alterne des cycles de refroidissement lent et de réchauffage afin de minimiser l'énergie.

^{iv} Les angles dièdres ϕ et ψ entre deux résidus n et $n+1$ dans un peptide sont respectivement définis par les quatre atomes $C_n-(N-C\alpha-C)_{n+1}$ et $(N-C\alpha-C)_n-N_{n+1}$.

^v Combinaisons d'angles ϕ et ψ tolérées pour la structure secondaire de la liaison peptidique.

éloignés de ceux présents dans la PDB, de nouvelles optimisations sont nécessaires pour produire un modèle utilisable et interprétable.

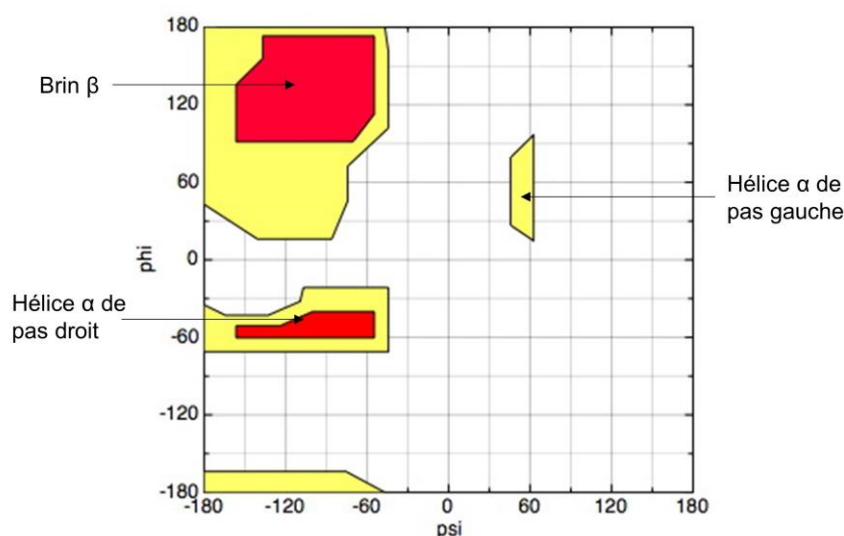


Figure 29 : Diagramme de Ramachandran. Chaque acide aminé dans la structure tertiaire est associé à des angles ϕ (ϕ) et ψ (ψ) formés par les liaisons peptidiques. Toutes les paires d'angles ϕ et ψ ne sont pas possibles car certaines conduisent à des distances inter-atomiques qui sont énergétiquement très défavorables. Trois régions sont qualifiées comme favorables. Les deux principales (en rouge) correspondent aux structures secondaires régulières majoritairement observées dans les protéines : les hélices α de pas droit et les brins β . La troisième région, plus petite, correspond à une conformation en hélice α de pas gauche ($\phi > 0$).

2.1.5. Mesures biophysiques à partir des structures tertiaires de PfK13 et PfCRT

Sous sa forme tertiaire, une protéine arbore un certain nombre de propriétés dites structurales, telles que son degré d'exposition au solvant, le nombre de contacts inter-atomiques par résidu d'acide aminéⁱ, ou encore son niveau de flexibilité. Analyser ces informations permet de mieux comprendre la conformation prise par la protéine, mais également, une fois mise en relation avec les taux de substitution préalablement estimés, de déterminer des zones à la fois conservées et présentant des caractéristiques structurales particulières.

Surface relative d'accessibilité au solvant (RSA) associée à un résidu d'acide aminé : Dans les premiers travaux s'intéressant au niveau de conservation dans un contexte structural, il a été montré que les résidus d'acides aminés participant au cœur des protéines – généralement non accessible au solvant – sont nettement plus conservés que les autres^{190,218}. Pour un résidu d'acide aminé donné d'une protéine, sa surface d'accessibilité au solvant (en Å³) est généralement normalisée par la valeur maximale d'accessibilité au solvant pour ce même acide aminé obtenu dans des échelles publiées : le paramètre obtenu est la surface relative d'accessibilité au

ⁱ Un acide aminé engagé dans une ou deux liaisons peptidiques au sein d'un polypeptide est un résidu d'acide aminé.

solvant (RSA) du résidu considéré²¹⁹. Le RSA varie de 0% pour les résidus d'acides aminés totalement enfouis dans la protéine à 100% pour ceux intégralement exposés. Il a initialement été supposé que l'accessibilité au solvant était la contrainte structurale dominante sur le taux d'évolution des sites. Par exemple, un travail basé sur un jeu de 11 protéines distinctes a divisé les résidus en fonction de leur RSA et de la structure secondaire dans laquelle ils sont localisés (hélices, brins, *etc.*), puis démontré que les sites exposés évoluaient plus rapidement que les sites enfouis quelle que soit la structure secondaire²²⁰. Le RSA d'un site peut être calculé à l'aide d'outils dédiés tels que DSSP²²¹.

Taux de contacts inter-atomiques lié à l'entourage d'un résidu d'acide aminé (appelé WCN pour *Weighted Contact Number* en anglais) : Les résidus d'acides aminés ayant une forte densité de contacts inter-atomiques avec les atomes des résidus d'acides aminés voisins ont tendance à évoluer plus lentement²²². Cette corrélation est encore plus forte si la mesure est effectuée sur les chaînes latérales des résidus, plutôt que sur les Ca uniquement comme ce fut le cas dans les premières études²²³. Le paramètre WCN basé sur les chaînes latérales de chaque résidu d'acide aminé peut être calculé à l'aide d'algorithmes dédiés tels que celui proposé par Sydykova et ses collègues (2017)²²⁴.

Fluctuations quadratiques moyennes associées à un résidu d'acide aminé (ou RMSF pour *Root-Mean-Square Fluctuations* en anglais) : Les protéines ne sont pas structuralement figées. Chaque résidu d'acide aminé fluctue de façon plus ou moins forte selon sa localisation et sa fonction. Par exemple, certaines protéines de transport modifient leur état conformationnel pour exposer le site de liaison aux molécules substrats. De surcroît, les changements conformationnels pourraient contrôler la tolérance mutationnelle d'un site : un site situé sur une région fortement flexible serait plus tolérant à la mutation qu'un site situé dans une région rigide²²⁵⁻²²⁷. La quantification de la dynamique des sites au cours du temps peut être faite par des simulations de dynamique moléculaire.

Potentiel électrostatique à la surface d'une protéine ou région protéique : Ce paramètre ne présente aucune relation avérée avec les taux de substitution spécifiques aux sites, mais révèle des informations essentielles sur les surfaces d'interaction protéine-protéine²²⁸. Ce potentiel, dont l'unité est généralement exprimée en kT/e (où k représente la constante de Boltzmann, T la température désirée, et e la charge d'un électron), est influencé par la charge des chaînes latérales et des liaisons ioniques. Il peut être calculé à l'aide du logiciel APBS (pour *Adaptive Poisson-Boltzmann Solver*)²²⁹, dans lequel l'équation de Poisson-Boltzmann est linéarisée à une température et à des constantes diélectriques de la protéine et du solvant fixées. La surface de la protéine est ensuite cartographiée de façon cubique selon un rayon donné, suivant une échelle souvent comprise entre $-10 kT/e$ à $+10 kT/e$.

2.1.6. Simulations de dynamique moléculaire du domaine KREP de Pfk13

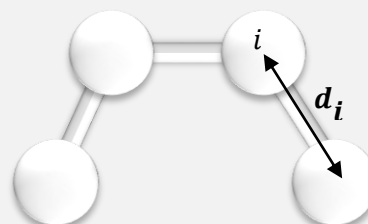
Les simulations de dynamique moléculaire consistent à étudier la trajectoire d'une molécule (par exemple une protéine ou un domaine protéique ; ici, des versions sauvage et mutantes du domaine KREP de Pfk13) en lui appliquant les lois du mouvement de Newton. Dans de telles simulations, une molécule est représentée par un assemblage d'atomes dont les positions spatiales fluctuent au cours du temps. Chacun des atomes constitutifs de la molécule peut être considéré comme une masse dont le mouvement est déterminé en résolvant la seconde loi de Newton²³⁰ : « Tout corps persévère dans l'état de repos ou de mouvement uniforme en ligne droite dans lequel il se trouve, à moins que quelque force n'agisse sur lui, et ne le contraigne à changer d'état ». Les paramètres relatifs aux simulations de dynamique moléculaire étant extrêmement nombreux – et souvent complexes –, nous détaillerons ici uniquement les principales notions à connaître.

Pour réaliser une simulation de dynamique moléculaire, il est nécessaire de disposer d'une structure tertiaire – déterminée expérimentalement ou par des méthodes de *modeling* – dont on veut étudier la trajectoire. Les interactions relatives à chaque atome sont décrites par un champ de force, que l'on pourrait assimiler à une équation modélisant l'énergie potentielle d'un ensemble d'atomes selon divers paramètres. Il existe plusieurs champs de force (GROMOS, AMBER, CHARMM, OPLS) incluant plus ou moins de paramètres, chacun étant amélioré régulièrement par les chercheurs afin de produire les simulations les plus réalistes possibles²³¹. Dans cette étude, nous avons considéré le champ de force AMBER99SB-ILDN en raison de son utilisation courante dans l'étude des trajectoires de domaines KREP^{232–234}. Tout champ de force considère deux types d'interactions atomiques : les interactions entre atomes liés de façon covalente, auxquelles sont associées les énergies de liaison, d'angle de valence et de torsion des angles dièdres ; et les interactions entre atomes non liés de façon covalente, correspondant aux interactions de van der Waals, aux interactions électrostatiques et aux liaisons hydrogènes (**Encadré 6**)^{235,236}.

Encadré 6 : Types d'énergies et d'interactions mises en jeu dans un champ de force.

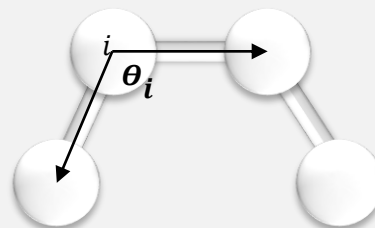
Énergie d'élongation des liaisons (ou vibrations d'allongement dites « *stretching* ») : correspond à l'élongation d'une liaison covalente entre deux atomes d'une molécule. Le calcul de cette énergie se traduit par l'équation suivante :

$$E_{\text{Elongation}} = \frac{1}{2} \sum_n K_{d,i} (d_i - d_{0,i})^2$$
, où d_i est la distance entre les deux atomes liés, $d_{0,i}$ la distance de la liaison i à l'équilibre, et $K_{d,i}$ une constante de force d'élongation de la liaison i (exprimée en $\text{kJ} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$). Les paramètres $d_{0,i}$ et $K_{d,i}$ sont dépendants de la nature de la liaison entre les deux atomes liés. Pour une molécule (par exemple une protéine), la sommation porte sur toutes les liaisons covalentes n au sein de la molécule.



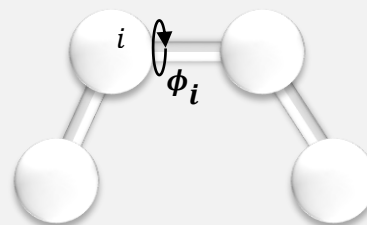
Énergie de déformation des angles (ou vibrations de déformation dites « *bending* ») : représente la variation d'un angle relatif à trois atomes liés de façon covalente d'une molécule. Le calcul de cette énergie se traduit par :

$E_{Déformation} = \frac{1}{2} \sum_n K_{\theta,i} (\theta_i - \theta_{0,i})^2$, où θ_i est l'angle de liaison relatif impliquant trois atomes, $\theta_{0,i}$ l'angle de liaison à l'état d'équilibre (pour le type d'atome considéré), et $K_{\theta,i}$ une constante de force de flexion (exprimée en $\text{kJ.mol}^{-1}.\text{deg}^{-2}$). Les paramètres $\theta_{0,i}$ et $K_{\theta,i}$ sont dépendants de la nature de l'atome central (impliqué dans les deux liaisons). Pour une molécule, la sommation porte sur tous les angles n au sein de la molécule.



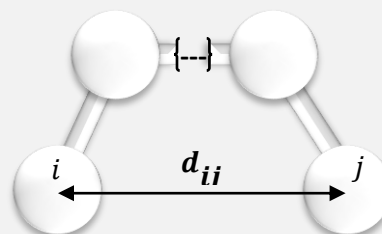
Énergie de torsion des angles dièdres : correspond à la capacité de rotation (ou de torsion) autour d'une liaison covalente, nécessitant la définition de l'angle dièdre entre deux plans, eux-mêmes définis par trois atomes d'une molécule. L'énergie se traduit par la formule :

$E_{Torsion} = \frac{1}{2} \sum_n V_{p,i} [1 + \cos(p_i \phi_i - \phi_{0,i})]$ où ϕ_i représente l'angle dièdre considéré, $\phi_{0,i}$ l'angle de phase, p_i la périodicité de la rotation, et $V_{p,i}$ une constante de force de courbure (exprimée en kJ.mol^{-1}). Les paramètres $\phi_{0,i}$ et $V_{p,i}$ sont dépendants de la nature de la liaison et des atomes liés. Pour une molécule, la sommation porte sur tous les angles n au sein de la molécule.



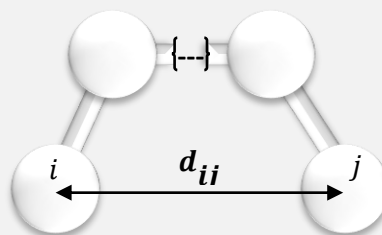
Interaction de van der Waals : représente une liaison non covalente, électrique et de faible intensité, qui s'exerce entre plusieurs atomes. Ces interactions sont souvent modélisées par un potentiel de Lennard-Jones, lequel décrit l'interaction entre deux atomes par mesure de l'énergie potentielle, et se traduit par :

$E_{vdW} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{d_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{d_{ij}} \right)^6 \right]$, où ϵ_{ij} représente le paramètre de Lennard-Jones (constant pour l'étude d'un seul type d'atomes), σ_{ij} la distance constante entre deux atomes où forces répulsives et attractives sont supposées identiques, d_{ij} la distance entre les deux atomes considérés. La sommation porte sur toutes les paires n d'atomes non liés au sein de la molécule.



Interaction électrostatique : correspond à l'interaction électrique entre deux atomes chargés électriquement. Plus la distance entre deux atomes est élevée, plus l'énergie électrique entre eux est faible. Les interactions électrostatiques se traduisent par la formule :

$E_{Electrostatique} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{q_i q_j}{4\pi\epsilon_0 d_{ij}}$, où q représente la charge partielle de l'atome, d_{ij} la distance entre les atomes considérés, $\frac{1}{4\pi\epsilon_0}$ la constante diélectrique du vide. La sommation porte sur toutes les paires n d'atomes non liés au sein de la molécule.



Les coordonnées initiales relatives à chaque atome de la structure tertiaire sont déterminées par une minimisation d'énergie du système, consistant à optimiser la géométrie de la structure jusqu'à l'obtention d'un minimum local²³⁵. Cette étape est nécessaire et permet d'éviter qu'une énergie mécanique trop importante n'influe sur l'assemblage moléculaire dès le début de la simulation. Ainsi, à partir d'un jeu de coordonnées initiales (typiquement, une structure tertiaire au format .pdb), on cherche à déterminer les coordonnées des atomes telles que la sommation des différents types d'interaction – précédemment décrites – soit la plus faible possible. Parmi la multitude de méthodes disponibles, nous avons utilisé la méthode « *Steepest descent* », communément utilisé pour minimiser l'énergie du système²³⁵. À partir d'une énergie initiale préalablement calculée, chaque atome est déplacé individuellement, à tour de rôle, dans une direction afin de calculer la nouvelle énergie. Dès lors, la direction prise par l'atome n'est conservée que si l'énergie diminue. On cherche donc, au sein de la direction donnée, une coordonnée spatiale qui minimise cette énergie. Un pas peut être inclus dans l'algorithme, correspondant à la distance à parcourir pour arriver au minimum. À noter que si la méthode est rapide lors des premiers cycles – et donc efficace pour supprimer tout mauvais contact –, la convergence finale est très lente (car les directions de recherche ne sont pas optimisées).

Pour connaître le mouvement d'un atome, il suffit de dériver l'énergie du système par rapport à la position de l'atome donné, puis de l'associer à la seconde loi de Newton. Les positions à un temps t donné sont utilisées pour estimer les positions au temps $t + \delta t$ (δt étant le pas d'intégration), elles-mêmes utilisées pour obtenir les positions au temps $t + 2\delta t$, et ainsi de suite (série de Taylor). Le pas d'intégration δt n'est pas choisi de façon aléatoire et doit être préférentiellement le plus petit possible (généralement, inférieur à 10 femtosecondesⁱ puisque cette durée correspond au mouvement atomique le plus rapide : la liaison carbone-hydrogène). Plusieurs algorithmes ont été développés pour simuler ces mouvements atomiques au cours du temps (Verlet, Leap-frog, Velocity-Verlet...) ²³⁷.

L'évolution du système dans le temps est corrélée à la température et à la pression. En effet, la vitesse relative aux atomes est directement influencée par la température et la pression dans un système. En conséquence, la réalisation d'une simulation de dynamique moléculaire doit se dérouler dans un environnement à température et pression constantes. Le choix des constantes de température et de pression dépend de la question scientifique posée : dans ce travail, nous avons fixé la température à 300 K et une pression de 1 bar, des valeurs communément utilisées en dynamique moléculaire. Le contrôle de la pression et de la température se fait dans des ensembles utilisant des thermostats et barostats tels que Berendsen et Parrinello-Rahman respectivement^{238,239}.

Une grande majorité des systèmes moléculaires évoluent dans un solvant, aussi faut-il « solvater » le système lors d'une simulation de dynamique moléculaire²³⁵. Pour ce faire, une

ⁱ 1 femtoseconde = 1 fs = 10^{-15} seconde.

« boîte » est créée – dont la taille est définie par le chercheur – et contiendra l'ensemble du système (la structure d'intérêt ainsi que les molécules d'eau). La forme d'une structure tertiaire de protéine pouvant par ailleurs être très variable, les boîtes existent sous différentes formes : parallélépipédiques, octaédriques, dodécaédriques, *etc.* Afin de diminuer les temps de calcul, qui augmentent proportionnellement au nombre d'atomes inclus dans le système, il faudra donc choisir la forme de la boîte la plus en adéquation avec la structure moléculaire d'intérêt²³⁵. Pour éviter que la structure d'intérêt ne touche les parois de la boîte, une distance minimale est définie (généralement, de 10 Å si on ne s'attend pas à un « débobinage » de la structure) entre les atomes de la structure les plus proches des parois de la boîte et les extrémités de celle-ci. Diverses stratégies ont été développées pour optimiser les temps de calcul. Entre autres, lorsqu'une molécule d'eau sort de la boîte, celle-ci est repositionnée du côté de la boîte opposé à sa sortie, en conservant la même vitesse²⁴⁰. On appelle cette représentation la condition limite périodique (ou PBC pour *Periodic Boundary Condition*), utilisée dans cette étude. Sachant l'utilisation de conditions périodiques aux bornes, des algorithmes supplémentaires (que nous ne détaillerons pas ici) ont été introduits tels que celui basé sur la sommation d'Ewald (PME pour *Particle Mesh Ewald*) afin de calculer plus efficacement les interactions – notamment électrostatiques – entre atomes. À noter qu'un seuil est défini afin de ne les calculer que pour les proches voisins d'un atome donné.

La structure d'intérêt doit évoluer dans un système électriquement neutre. Le niveau électrique est dépendant de la structure étudiée et doit être ajusté par l'ajout de contre-ions (généralement, le cation sodium Na⁺ ou l'anion chlorure Cl⁻).

Dans le cadre de cette thèse, nous avons effectué un certain nombre de simulations de dynamique moléculaire sur des structures tertiaires sauvage et mutantes du domaine KREP de PfK13 sur une durée de 100 nanosecondesⁱ afin d'analyser les altérations structurales engendrées par les mutations de résistance (R539T et C580Y). Les simulations ont été réalisées avec le logiciel GROMACS²⁴¹, relativement simple d'utilisation et offrant un certain nombre d'outils d'analyse une fois la simulation terminée. Les différentes étapes de production et d'analyse – dans le cadre de l'outil GROMACS – sont présentées dans la **figure 30**.

ⁱ 1 nanoseconde = 1 ns = 10⁻⁹ seconde.

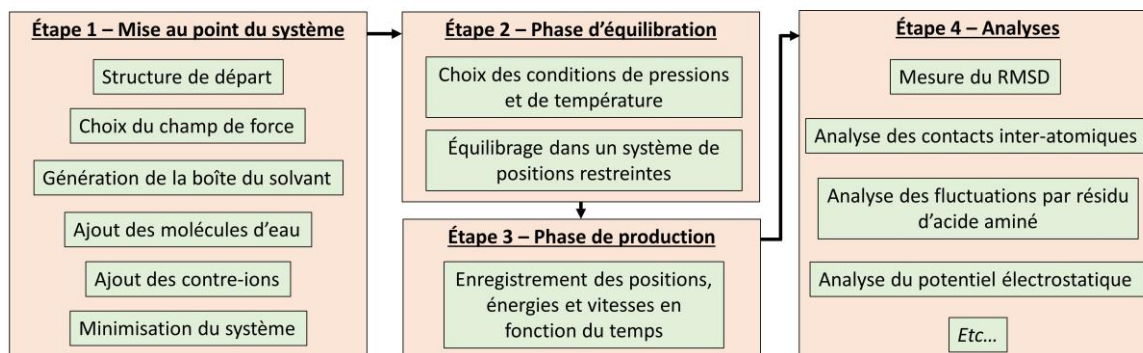


Figure 30 : Étapes de préparation, de production et d'analyses relatives aux simulations de dynamique moléculaire via le logiciel GROMACS. Dans un premier temps, le système est mis au point : la structure d'intérêt – au format .pdb – est préparée puis placée dans un milieu aqueux. Une fois le système neutralisé, une minimisation d'énergie est réalisée pour obtenir un système stable en supprimant la majorité des contacts non favorables. Ensuite, le système est soumis à un équilibration suivant des conditions de pression et de température. Puis, la phase de production débute durant une durée plus ou moins longue fixée par le chercheur. Enfin, un certain nombre d'analyses sont réalisées afin d'étudier la trajectoire de la protéine, ainsi que certaines de ses caractéristiques globales ou locales.

2.1.7. Docking moléculaire sur le domaine KREP de Pfk13

Nos recherches évolutives et structurales nous ont amené à identifier une possible surface d'interaction dans le domaine KREP de Pfk13. Il peut donc être intéressant d'étudier sa pharmaco-modulation par des petites molécules ou ligands, en analysant le processus de liaison moléculaire (qu'on appelle communément *docking*, se traduisant par « amarrage moléculaire ») avec ces ligands. Ainsi, le *docking* vise à déterminer les orientations dans l'espace et des conformations favorables pour la fixation d'un ligand à une surface protéique²⁴².

Rechercher un ligand pouvant se lier à une protéine de façon non covalenteⁱ (puisque l'on ignore les particularités biochimiques de Pfk13) est une tâche complexe. Les ligands disponibles à ce jour sont organisés en chimiothèques de tailles variables et pouvant contenir plusieurs millions de molécules²⁴³. Des criblages virtuels peuvent être conduits afin de tester l'affinité de liaison entre une protéine ou une région protéique et des milliers de molécules, quoique consommateurs en termes de temps et d'infrastructure informatique. Une alternative consiste à sélectionner un ensemble de molécules déjà testées expérimentalement et qui interagissent avec un domaine de la même famille que le domaine d'intérêt.

Une simulation de *docking* peut se résumer en deux étapes (**Figure 31**) :

- le *docking* à proprement parlé, au cours duquel un ligand est placé dans le site de liaison supposé du domaine d'intérêt. Un échantillonnage de ligands, fixés au domaine d'intérêt selon plusieurs conformations (qu'on appelle poses), est ensuite produit ;

ⁱ Qui n'altère pas les propriétés biochimiques de la protéine.

- le *scoring*, qui évalue l'affinité de liaison entre le ligand et la protéine d'un point de vue énergétique. Ce score permet *in fine* de retenir la meilleure pose parmi celles proposées. Ce score est calculé en considérant entre autres les interactions hydrophobes et électrostatiques et les liaisons hydrogènes. Plus l'énergie de liaison est négative, et plus l'affinité entre le ligand et le domaine d'intérêt est importante.

Le *docking* peut être réalisé manuellement en plaçant le ligand au niveau du site d'interaction supposé de la protéine et préalablement défini par le chercheur. Une boîte artificielle, rectangulaire, est générée pour cibler la région où l'interaction est supposée avoir lieu. Dans ce travail de thèse, la définition de la boîte a été réalisée à l'aide de l'outil Autodock²⁴⁴, tandis que les *docking* et les calculs des scores ont été générés à l'aide du logiciel Vina²⁴⁵.

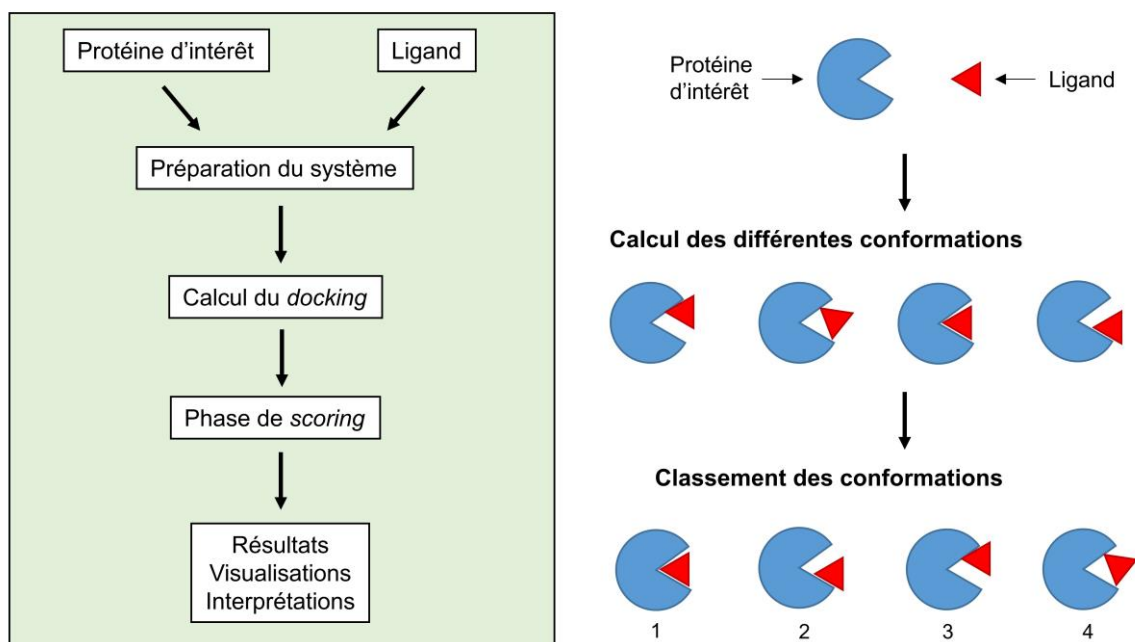


Figure 31 : Présentation des différentes étapes du *docking* moléculaire. À partir de la protéine d'intérêt et d'un ou plusieurs ligands préalablement sélectionnés, le chercheur cible la région où l'interaction protéine-ligand est supposée avoir lieu. Le *docking* est ensuite réalisé selon différentes poses, puis un classement (*scoring*) est généré sur la base de l'énergie associée à l'affinité de liaison.

- Les sites au sein d'une protéine ayant un rôle moins important sur le plan fonctionnel évoluent plus rapidement (en termes de changement d'acides aminés) que ceux ayant un rôle plus important.
- Les sites supposés fonctionnels de PfK13 et PfCRT ont été recherchés en combinant les données évolutives et les informations de structure tertiaire, puis comparés à des protéines ou domaines protéiques de la même famille.
- En cas d'absence de structure tertiaire, un modèle peut être prédit par homologie structurale (PfCRT) ou par reconnaissance structurale (domaine coiled-coil de PfK13).
- Diverses propriétés structurales telles que le nombre de contacts inter-atomiques, le niveau de flexibilité ou le degré d'exposition au solvant peuvent être étudiées pour chaque résidu d'acide aminé afin de mieux caractériser la protéine.
- Des simulations de dynamique moléculaire du domaine KREP de PfK13 sauvage et mutants (conférant la résistance aux ARTDs) ont été réalisées afin d'étudier les variations de conformation au cours du temps.
- Des *docking* protéine-ligand ont été réalisés afin d'identifier des molécules capables d'interagir de façon non covalente avec le domaine KREP de PfK13, et ainsi d'étudier sa pharmaco-modulation.

”
*Celui qui a besoin d'un protocole n'ira
jamais loin ; les génies lisent peu,
pratiquent beaucoup et se font d'eux-
mêmes.*

— **Denis Diderot** (1713 – 1784)

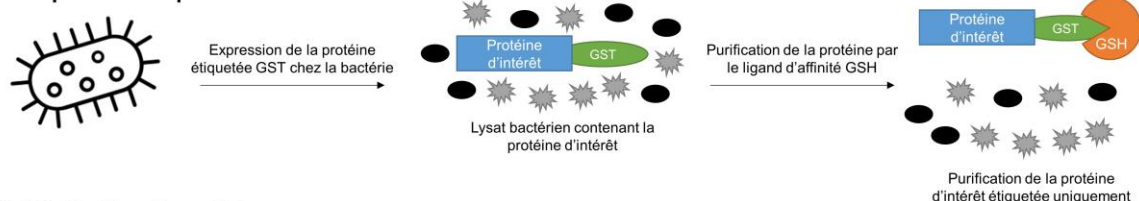
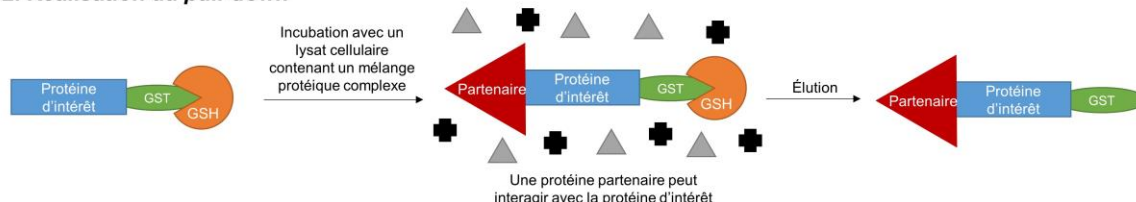
Les analyses bioinformatiques peuvent mettre en évidence de possibles zones fonctionnelles au sein des protéines. Celles-ci doivent toutefois être validées par des approches expérimentales. La technique du *pull-down* différentiel peut représenter une première validation *in vitro* : des protéines (ou domaines) recombinantes sauvage(s) et mutante(s) invalidant la zone d'interaction supposée sont produites afin de mesurer une quelconque différence dans l'affinité de liaison avec des protéines partenaires. Dans cette section, nous présenterons le concept et l'objectif d'un *pull-down* à partir de domaines recombinants PfK13 exprimés chez la bactérie *E. coli*. Nous détaillerons l'ensemble des conditions réalisées et les différents tampons préparés. La finalité première de ce travail fut de proposer un protocole d'expression et de purification de différents domaines de PfK13, et réaliser un premier essai de *pull-down*.

2.2.1. Introduction au *pull-down*

Dans la cellule, les protéines participent à de vastes réseaux d'interactions entre protéines. Ces protéines s'assemblent et se désassemblent de concert selon des conditions intra, inter et extracellulaires. Une étape préliminaire dans la compréhension de la structure et de la fonction des protéines consiste à déterminer quelles protéines interagissent les unes avec les autres, identifiant ainsi les voies biologiques pertinentes. Parmi les différents protocoles disponibles, le *pull-down* représente un outil *in vitro* précieux pour étudier les voies cellulaires via l'étude des interactions physiques protéine-protéine²⁴⁶. Les analyses de *pull-down* sont utiles à la fois pour confirmer l'existence d'une interaction protéine-protéine prédite et comme analyse de dépistage initiale pour l'identification d'interactions protéine-protéine jusque-là inconnues²⁴⁶. Plusieurs autres techniques permettent de réaliser ce type d'expériences comme la co-immunoprécipitation²⁴⁷ ; le choix du *pull-down* sera justifié ci-après. La technique du *pull-down* correspond à une forme de purification par affinité dans laquelle la protéine dont on cherche les partenaires est utilisée comme appât (ici, les différents domaines de Pfk13)²⁴⁶. La protéine d'appât est préalablement marquée avec une étiquette (aussi appelée tag) en N- ou C-terminal, exprimée chez la bactérie *E. coli* puis purifiée à l'aide d'un ligand d'affinité dépendant de l'étiquette²⁴⁶. La protéine purifiée peut ensuite être incubée avec une source de protéines contenant des protéines « proies » présumées telle qu'un lysat cellulaire²⁴⁶. La composition des protéines « proies » est déterminée en fonction de la question biologique, qui peut être : i) soit de confirmer une interaction suspectée, qui aurait été mise en évidence par une autre approche ; ii) soit d'identifier une interaction totalement inconnue. Si des protéines interagissent avec la protéine appât, celles-ci peuvent être observées après électrophorèse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) et coloration, mais aussi et surtout identifiées par spectrométrie de masse²⁴⁶.

Parmi la multitude d'étiquettes existantes, deux sont principalement utilisées à des fins de *pull-down* : l'étiquette Glutathion S-transférase (GST) et l'étiquette poly-histidine (généralement constituée d'au moins six résidus histidine)²⁴⁸. GST est souvent privilégiée car elle améliore la solubilisation de la protéine recombinante lors de l'expression et la lyse bactérienne, et est fortement exprimée²⁴⁸. Toutefois, les expérimentations avec différentes étiquettes sont souvent menées en parallèle car la GST, comparée à d'autres tags, peut influencer certaines propriétés de la protéine d'intérêt en raison de sa capacité à dimériser et de sa taille (240 acides aminés, 26 kDa). De plus, la GST étant très exprimée, elle peut entraîner une accumulation de protéines agrégées dans les corps d'inclusion, rendant la solubilisation des protéines recombinantes difficile²⁴⁸. Dans ce travail de thèse, seule l'étiquette GST a été utilisée, bien que les constructions avec étiquette poly-histidine seront également testées ultérieurement. Un schéma récapitulant l'approche de *pull-down* (avec étiquette GST) est présenté en **figure 32**.

1. Expression et purification

2. Réalisation du *pull-down*

3. Révélation des protéines partenaires

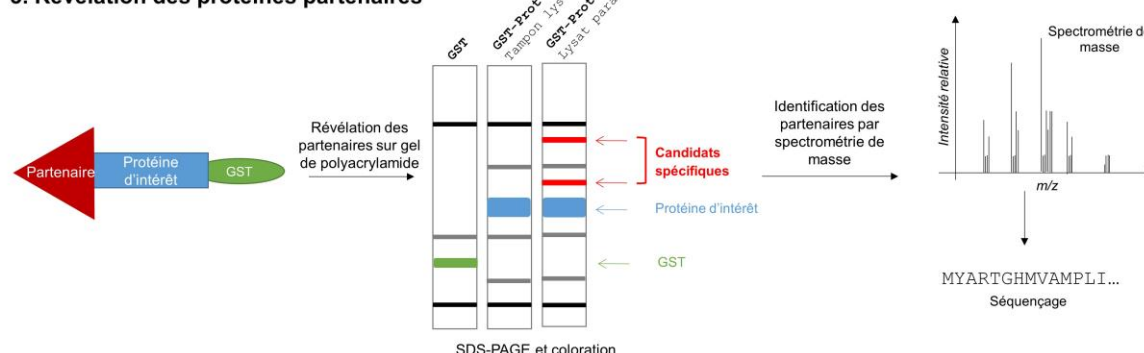


Figure 32 : Approche de *pull-down* dans la recherche de protéines partenaires à partir d'une protéine d'intérêt (ou appât) étiquetée GST. La bande bleue sur le gel correspondant à la protéine d'intérêt contient en fait la protéine d'intérêt fusionnée à la GST.

Comme explicité dans la partie introductive de cette thèse, les sites fonctionnels de PfK13 ne sont pas caractérisés, et ses substrats demeurent mal connus. L'approche de *pull-down* aurait ainsi deux avantages : *i*) identifier de possibles substrats protéiques, mis à part le candidat principal qui est la PfPI3K¹²⁶ ; et *ii*) valider expérimentalement les sites fonctionnels de PfK13 identifiés par nos analyses bioinformatiques (cf. sections 2.1.2. et 2.1.3.) par des approches différentielles sauvage(s) *versus* mutante(s)ⁱ. Un prérequis au *pull-down* différentiel est donc de détecter par notre méthode un substrat protéique interagissant avec une construction sauvage. Différentes constructions correspondant en une concaténation des domaines PfK13 ont été conçues : GST-KREP (GST-K), GST-BTB-KREP (GST-BK) et GST-coiled-coil-BTB-KREP (GST-CBK ; **Figure 33**). À noter que le tag GST a été positionné en N-terminal des domaines puisque la majorité des études ayant réalisées des protéines recombinantes impliquant un domaine BTB ou KREP ont placé leur tag en N-terminal^{129,249,250}. Dans cette thèse, nous nous sommes pour l'instant uniquement intéressés à des domaines sauvages de PfK13, bien qu'*in fine*, des comparaisons avec des domaines mutants, soit avec des mutations de résistance aux ARTDs (par exemple, la mutation C580Y), soit avec des mutations artificielles invalidant des sites

ⁱ La mutation non-synonyme, pas nécessairement retrouvée dans les populations naturelles, aurait pour rôle d'invalider la zone d'intérêt.

d'interaction candidats, seront réalisées (*pull-down* différentiels). Ces éléments seront davantage expliqués dans la partie *Discussion*.

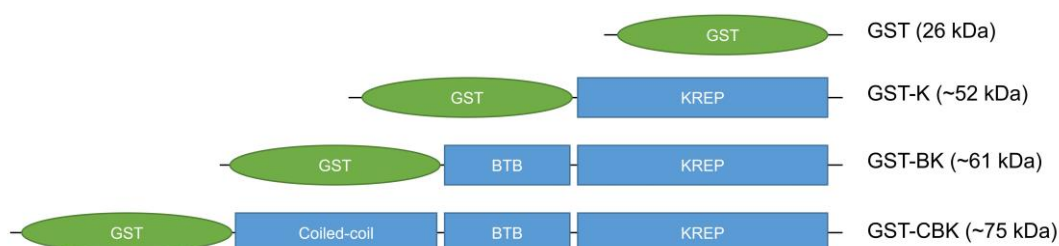


Figure 33 : Schématisation des différentes constructions Pfk13 étiquetées GST. La première construction correspond au seul domaine GST qui servira de contrôle lors des *pull-down*. Les autres constructions représentent une suite concaténée des différents domaines annotés de Pfk13. L'ajout des deux autres domaines (coiled-coil et BTB), étant *a priori* eux aussi impliqués dans des interactions protéine-protéine, pourrait favoriser une interaction avec certains ligands.

2.2.2. Construction des plasmides

La séquence nucléotidique codant Pfk13 a été synthétisée par la compagnie privée GenScript. Préablement, la séquence nucléotidique a été optimisée en modifiant certaines bases sans modifier la séquence protéine (substitutions synonymes) afin d'améliorer le rendement d'expression chez la bactérie *E. coli* BL21-CodonPlus. En effet, les séquences codantes de *P. falciparum* sont exceptionnellement riches en nucléotides A et T (**Tableau 1**), ce qui peut entraîner des différences importantes dans l'usage des codons entre les deux espèces. La séquence codant les domaines d'intérêt de Pfk13 a ensuite été clonée dans le plasmide pGEX-6P-1 (sélection ampicilline (gène *ampR*), promoteur TAC inducible par l'isopropyl β -D-1-thiogalactopyranoside ou IPTG ; **Figure 34**) en C-terminal du tag GST, de sorte que le domaine KREP de Pfk13 soit complètement libre à son extrémité C-terminale. La préparation des plasmides a été faite par Jérôme Clain (directeur de thèse, UMR 261) et Laurine Rondepierre (doctorante, UMR 261).

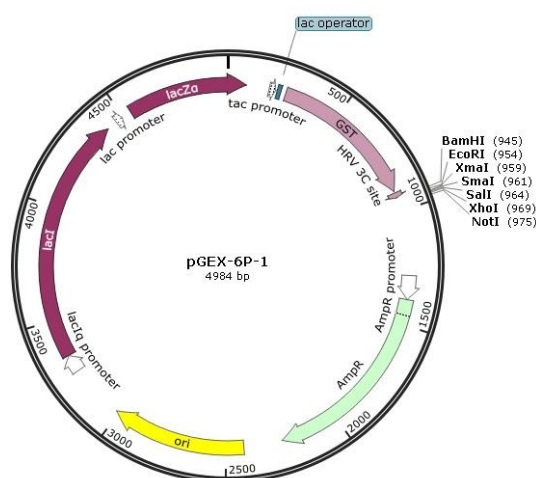


Figure 34 : Carte du plasmide pGEX-6P-1. Les ADN codant pour les protéines / domaines d'intérêt sont insérés dans ce plasmide en 3' et en fusion de la séquence codant pour la GST. La culture des bactéries se fera sous sélection ampicilline. Un promoteur TAC est inclus permettant l'induction par l'IPTG. GST,

Glutathion S-transférase ; AmpR, gène de résistance à l'ampicilline ; ori, origine de réplication ; lac operator, opéron lactose ; HRV 3C site, site de reconnaissance et de clivage pour la protéase rhinovirus 3C ; bp, paire de bases.

2.2.3. Transformation par choc thermique dans les souches d'expression

Les différents plasmides réalisés ont été transformés dans des bactéries *E. coli* BL21-CodonPlus. Cette souche est communément utilisée pour l'expression de routine de protéines recombinantes (**Figure 34**). L'étape de transformation, consistant à faire pénétrer un ADN étranger (ici les plasmides) dans les bactéries, a été effectuée par choc thermique en suivant un protocole standard.

2.2.4. Optimisation des conditions d'expression des domaines recombinants

Les conditions optimales d'expression des domaines recombinants ont été identifiées par variation de la concentration de l'inducteur (IPTG), de la température et du temps d'induction. Les bactéries transformées ont été inoculées dans 5 mL de milieu LB (Luria-Bertani) + ampicilline. Les précultures ont été incubées sous agitation à 160 rotations par minute (RPM) sur la nuit à 37°C. Celles-ci ont ensuite été utilisées pour préparer des cultures de 15 mL supplémentées d'ampicilline, puis incubées à 37°C sous agitation jusqu'à atteindre une densité optique comprise entre 0,6 et 0,8 mesurée à 600 nm, correspondant à la phase exponentielle de croissance bactérienne. L'expression de la protéine a été induite par l'ajout d'IPTG à une concentration de 0,3 et 1 mM (suivant les essais), à des températures de 18°C, 30°C et 37°C pendant 4 heures (30°C et 37°C) ou 14 à 16 heures (18°C ; **Tableau 8**). Au terme de l'expression, les cultures ont été centrifugées à 5 000 *g* pendant 10 minutes à 4°C afin de collecter les bactéries, puis conservées à -80°C jusqu'à leur utilisation ultérieure. Pour chaque condition, 1 mL de culture fut aliquoté avant et après induction afin de mesurer la densité optique et confirmer rapidement l'expression des domaines recombinants par SDS-PAGE.

Tableau 8 : Conditions testées pour l'expression de la protéine de fusion GST-K.

Souche	Construction	Température	Concentration en IPTG	Temps d'induction
<i>E. coli</i> BL21- CodonPlus	GST-K	18°C	0,3 mM	14-16 heures
			1 mM	
		30°C	0,3 mM	4 heures
			1 mM	
		37°C	0,3 mM	4 heures
			1 mM	

Note – L'ensemble de ces conditions a uniquement été testée pour la construction GST-K. Nous avons fait l'hypothèse que les résultats seraient sensiblement similaires pour les constructions GST-BK/CBK.

Pour éviter la multiplication des protocoles, la meilleure condition pour la construction GST-K a été choisie pour les autres constructions GST-BK/CBK. Une vérification sur ces dernières a toutefois été réalisée afin de confirmer la bonne expression pour la condition choisie.

Une fois l'ensemble des tests relatifs à l'expression des domaines recombinants effectués – ainsi que ceux relatifs à leur purification détaillés ci-après – des productions en grand volume ont été cultivées en adaptant les différents protocoles (1 litre de culture par construction).

2.2.5. Solubilisation des domaines recombinants

Les domaines recombinants ont ensuite été purifiés à partir du lysat bactérien issu de ~15 mL de culture. Les culots bactériens précédemment obtenus ont été repris dans un tampon de lyse dont le volume dépend de la densité optique de la culture préalablement mesurée. Divers tampons furent testés pour optimiser la solubilisation des domaines recombinants, nommés TLb1 à TLb9, dont les compositions sont données dans l'**encadré 7**.

Encadré 7 : Composition des tampons de lyse (TLb) testés pour lyser les bactéries et solubiliser les domaines recombinants.

TLb1 : PBS 1X, pH 7,4 (NaCl 0,138 M ; KCl⁻ 0,0027 M ; Na₂HPO₄ 0,010 M ; KH₂PO₄ 0,0018 M) ; EDTA 1 mM, pH 8 ; Triton X-100 0,1% ; Complete™ Protease Inhibitor Cocktail (Roche) 1X ; lysozyme 100 µg/mL.

TLb2 : Tris-HCl 10 mM, pH 8 ; NaCl 150 mM ; EDTA 1 mM ; lysozyme 100 µg/mL ; DTT 5 mM ; Complete™ Protease Inhibitor Cocktail (Roche) 1X ; Triton X-100 0,1%.

TLb3 : TLb2 en augmentant la concentration de Triton X-100 à 1% final.

TLb4 : TLb2 en remplaçant le Triton X-100 par du NP-40 1% final.

TLb5 : TLb2 en remplaçant le Triton X-100 par du CHAPS 1% final.

TLb6 : TLb2 en remplaçant le Triton X-100 par du Tween-20 1% final.

TLb7 : TLb2 en remplaçant le Triton X-100 par du DDM 1% final.

TLb8 : TLb2 en remplaçant le Triton X-100 par du sarkosyl 0,5% final. Après sonication, rajout de Triton X-100 2% final.

TLb9 : TL8 en augmentant la concentration de sarkosyl à 1,5% et de triton X-100 à 4% final.

Les temps d'incubation sur glace furent variables en fonction du tampon utilisé (allant de 15 minutes pour TLb1 à TLb7 jusqu'à 1 heure pour TLb8 et TLb9). Les culots resuspendus ont été lysés par sonication sur glace pendant une minute (six cycles de 10 secondes de pulse – 10 secondes de repos ; 12 cycles furent réalisés dans le cas des gros volumes de lysat). En raison des difficultés de solubilisation de nos domaines recombinants, une autre durée de sonication a été investiguée (trois cycles de 5 secondes de pulse – 15 secondes de repos). La fraction soluble du lysat a été isolée par centrifugation à 15 000 *g* pendant 15 minutes à 4°C. Le surnageant

contenant la fraction soluble des protéines des bactéries a été prélevé, dont un aliquot a été testé par SDS-PAGE puis coloration au bleu de Coomassie pour vérifier la solubilisation des domaines recombinants. Les choix définitifs du tampon et des détails protocolaires seront présentés dans la partie *Résultats* (cf. section 3.2.2.).

2.2.6. Purification des domaines recombinants

Afin de purifier les domaines recombinants, une résine glutathion(GSH)-agarose (aussi appelée « billes ») a été utilisée pour sa capacité de liaison à la GST (GSH est un substrat de la GST). La résine, préalablement lavée trois fois avec un tampon STE (Tris-HCl 10 mM, pH 8 ; NaCl 150 mM ; EDTA 1 mM), a été incubée avec la fraction soluble du lysat bactérien pendant une heure sur roue (15 tours par minute) à 4°C. Étant donné que la capacité de liaison est comprise entre 5 et 10 mg de GST par mL de résine, la quantité de résine dépend de la quantité de protéines totales solubilisées (laquelle est estimée à l'aide d'une gamme BSA). La résine a ensuite été récupérée par centrifugation à 500 g pendant une minute à 4°C. Une portion de la fraction soluble (non fixée à la résine) fut aliquotée pour analyse ultérieure. La résine a ensuite été lavée trois fois, pour éliminer les protéines bactériennes contaminantes à l'aide du tampon STE à froid (10 volumes de STE pour un volume de résine), par incubation sur roue (15 tours par minute) pendant 10 minutes à 4°C puis centrifugation à 500 g pendant deux minutes à 4°C. À noter que divers composants ont été rajoutés individuellement au tampon STE pour optimiser les lavages et tenter de supprimer la présence de contaminantsⁱ, tels que l'ATP, une plus forte concentration en NaCl, ou encore la température lors des lavages. Ces différents éléments seront présentés dans la partie *Résultats* (cf. section 3.2.5.).

Afin de tester le repliement correct de la partie GST de nos protéines de fusion ainsi que leur solubilité, une élution par compétition avec du glutathion a été réalisée. La résine a été incubée avec un tampon d'élution (Tris-HCl 50 mM, pH8 ; NaCl 150 mM ; DTT 5 mM ; Triton X-100 0,1% ; GSH réduit 10 mM) sur roue (15 tours par minute) pendant une heure à 4°C. L'éluat (fraction soluble) a été isolé de la résine par centrifugation à 500 g pendant deux minutes à 4°C et testé par SDS-PAGE.

2.2.7. Préparation du lysat parasite

Le lysat parasite a été préparé à partir de culots de parasites *P. falciparum*, produits *in vitro* par Jérôme Clain (UMR 261) puis conservés à une température de -80°C. La culture des parasites était asynchrone mais contenait une majorité de stade trophozoïte avec vacuole digestive et hémozoïne bien formés.

ⁱ Protéines présentes mais non-souhaitées après l'étape de purification.

Pour éviter les processus de congélation et de décongélation qui pourraient dégrader les protéines, le culot parasite a été brièvement sorti du congélateur, placé sur glace, puis une petite quantité du culot fut prélevé par grattage à l'aide d'un scalpel – préalablement nettoyé à l'éthanol. Le reste du culot fut rapidement remis dans le congélateur. La portion de culot prélevée a été décongelée progressivement sur glace, puis resuspendue avec 9 volumes de tampon de lyse. Tout comme pour les bactéries, différents tampons de lyse des parasites (TLp) ont été testés (**Encadré 8**) pour déterminer la condition optimale de solubilisation des protéines parasites après incubation sur roue pendant 1 heure à 4°C. À noter que pour TLp1 et TLp2, une étape de sonication (TLp1 : six cycles de 10 secondes de pulse – 50 secondes de repos ; TLp2 : deux cycles de 2 secondes de pulse – 2 secondes de repos) a été réalisée en raison de l'absence de DNase (qui permet de rendre le lysat moins visqueux). Les protéines solubles du parasite ont été obtenues après centrifugation à 20 000 g pendant 15 minutes à 4°C, puis testées par SDS-PAGE.

Encadré 8 : Composition des tampons de lyse (TLp) testés pour lyser et solubiliser les protéines parasites.

TLp1 (RIPA) : Tris-HCl 20 mM, pH 8 ; NaCl 150 mM ; EDTA 1 mM ; Triton X-100 0,1% ; NP-40 1% ; Complete™ Protease Inhibitor Cocktail (Roche) 1X.

TLp2 : HEPES 50 mM, pH 7,5 ; NaCl 150 mM, CHAPS 0,5% ; fluoride de sodium (NaF) 1 mM ; orthovanadate de sodium 100 µM.

TLp3 : Tris-HCl 10 mM pH 7,2 ; NaCl 150 mM ; MgCl₂ 1 mM ; DNase 0,6 U/µL ; dodecylmaltoside (DDM) 1% ; Complete™ Protease Inhibitor Cocktail (Roche) 1X.

2.2.8. Réalisation du *pull-down*

Pour le *pull-down*, nous avons utilisé les domaines purifiés fixés à la résine agarose-GSH (donc non élués). Après estimation des concentrations de protéines fixées sur les billes (GST, qui sert de contrôle, et GST-K/BK/CBK), une homogénéisation des concentrations (en molarité) a été réalisée en ajoutant de la résine agarose-GSH afin de manipuler des volumes de résine équivalents lors de l'expérience. En parallèle, la résine fixée aux domaines recombinants a subi une étape optionnelle de blocage à la BSA, permettant en théorie de limiter des interactions protéine-protéine aspécifiques. Chaque construction (résine agarose-GSH-GST et les trois résines agarose-GSH-GST-K/BK/CBK) a été incubée sur la nuit à 4°C avec un volume de lysat parasite dépendant de la concentration en protéines parasites, estimée par la méthode Bradford²⁵¹. Des contrôles supplémentaires ont été ajoutés où le lysat parasite a été remplacé par le tampon de lyse parasite afin de repérer facilement les protéines contaminantes bactériennes. Le lendemain, la suspension a été centrifugée à 500 g pendant deux minutes à 4°C, et le surnageant contenant la fraction du lysat parasite non fixée à la résine mise de côté. La résine fut lavée trois fois avec 10 volumes de tampon STE pour éliminer les contaminants. Les protéines

associées à la résine en fin d'expérience ont été analysées par SDS-PAGE. Dans la mesure où les quantités peuvent être relativement faibles, deux types de coloration ont été investiguées : une coloration rapide au Coomassie standard (0,25% bleu de Coomassie, complété par de l'éthanol à 96%) et une coloration plus sensible et compatible avec la spectrométrie de masse (0,08% Coomassie Brilliant Blue G250, 10% d'acide citrique, 8% de sulfate d'ammonium et 20% de méthanol). Le choix de la coloration a eu un impact important sur les analyses, détaillées dans la partie *Résultats* (cf. section 3.2.5.).

- Le *pull-down* est une technique expérimentale permettant d'étudier l'interactome d'une protéine (ou d'un fragment protéique), préalablement exprimée chez la bactérie *E. coli* puis purifiée.
- Un *pull-down* a été appliqué à plusieurs domaines de la protéine Pfk13, dont l'interactome demeure aujourd'hui mal connu, par incubation avec un lysat parasitaire.
- Différentes constructions ont été préparées, basées sur une concaténation des domaines annotés de Pfk13, en utilisant l'étiquette GST en N-terminal : GST, GST-KREP, GST-BTB-KREP, GST-coiled-coil-BTB-KREP.
- Diverses conditions d'expression, de solubilisation et de purification des domaines recombinants ont été testées, puis examinées par SDS-PAGE.
- Le choix des protocoles définitifs est présenté dans la partie *Résultats*.

Partie 3

Résultats

Études bioinformatiques et expérimentales sur la protéine PfK13

Section 3.1



*Ce n'est pas la force, mais la persévérance,
qui fait les grandes œuvres.*

— **Samuel Johnson** (1709 – 1784)

PfK13 est supposée participer à un (ou plusieurs) complexe(s) protéique(s). Cependant, ses zones d'interaction n'ont pas encore été caractérisées. Dès lors, notre objectif fut de détecter des régions soumises à l'action de la sélection naturelle (positive ou purificatrice), cette action reflétant une possible importance fonctionnelle dans la structure et/ou la fonction de la protéine. Ces différents résultats bioinformatiques, basés sur l'évolution et la structure, seront suivis par quelques travaux expérimentaux (*pull-down*, *docking*), initiés en fin de thèse.

3.1.1. Recherche des zones conservées de PfK13

3.1.1.1. La haute conservation de *k13* au cours de l'évolution des *Apicomplexa*

Dans un premier temps, nous avons souhaité étudier les pressions de sélection agissant sur le gène *k13* et quantifier son niveau de conservation à l'échelle du protéome. Quarante-trois séquences *k13* – supposées « sauvages », ont été obtenues par des stratégies de type BLAST sur des banques de données telles que la NCBI ou la plateforme EuPathDB²⁵², couvrant 21 espèces de *Plasmodium*, 7 espèces de *Cryptosporidium*, 7 espèces de *Piroplasmida* et 8 espèces de *Coccidia*. Un alignement multiple de ces séquences a été produit avec l'outil MAFFT²⁵³, puis nettoyé en supprimant les positions peu informatives (par exemple, la région N-terminale, faiblement conservée). L'alignement nettoyé a ensuite été considéré pour inférer la phylogénie du gène *k13* avec le programme PhyML²⁵⁴, en ayant préalablement estimé le meilleur modèle d'évolution moléculaire avec le package SMS²⁵⁵. Les pressions de sélection agissant sur le gène *k13* ont été quantifiées à l'aide du rapport ω , estimé via l'outil codeml du package PAML¹⁹⁵ qui utilise l'alignement *k13* et la phylogénie correspondante, établis préalablement. Ces analyses ont montré que le gène *k13* est soumis à une extrême sélection purificatrice, suggérant que la grande majorité des variations de séquence apparues au cours de l'évolution des espèces a été éliminée de manière à conserver la séquence protéique (et par là même, la fonction de la protéine). Aucun site n'a été détecté comme évoluant sous sélection positive. À l'échelle du protéome, grâce à des données de Jefferas *et al.* (2015) que nous avons réanalysées, nous avons montré que K13 fait partie des 5% des protéines les plus conservées parmi les gènes codant des protéines pour lesquelles un orthologue a été identifié parmi cinq espèces de *Plasmodium*. Cette analyse était basée sur le rapport ω estimé pour 3 256 protéines, sous le modèle M0²⁵⁶. De par l'absence de sélection positive, nous avons par la suite analysé plus finement les niveaux de sélection purificatrice au sein de la protéine K13.

3.1.1.2. Un domaine coiled-coil fortement conservé

À l'échelle des *Apicomplexa*, le domaine coiled-coil de K13 s'avère extrêmement conservé¹⁹³. Structuralement, un coiled-coil est généralement composé de deux à sept hélices enroulées les unes autour des autres (les dimères et trimères étant les types les plus courants). Puisque la structure tertiaire de ce domaine chez PfK13 n'a pas encore été déterminée expérimentalement et en l'absence de structures homologues dans les banques de données, celle-ci a été prédite par une méthode *ab initio* dans laquelle divers motifs structuraux sont recherchés, comparés à la séquence cible puis assemblés. Par cette approche, il apparaît que le coiled-coil de PfK13 serait composé de deux longues hélices qui s'enroulent ensemble (**Figure 35**). Les protéines incluant un coiled-coil sont souvent impliquées dans des fonctions biologiques importantes, participant à l'oligomérisation et pouvant interagir avec l'ADN^{257,258}. Puisque la structure résolue des domaines BTB et KREP de PfK13 forme un homodimère, nous ne pouvons donc écarter la

possibilité d'un état homo-oligomérique de Pfk13 plus complexe en présence du coiled-coil. Ignorant le rôle de ce domaine dans Pfk13, nous avons vérifié s'il présente un motif capable de se lier avec l'ADN via l'algorithme de prédiction DNABind²⁵⁹. D'après cet outil, la probabilité de liaison entre le coiled-coil isolé de Pfk13 et l'ADN est égale à 99.86%. Cependant, cette probabilité diminue à 53.8% lorsque la séquence Pfk13 est soumise intégralement. Tester expérimentalement une interaction du coiled-coil de Pfk13 ou de Pfk13 entière avec l'ADN nous semble donc pertinent pour mieux comprendre le rôle de ce domaine.

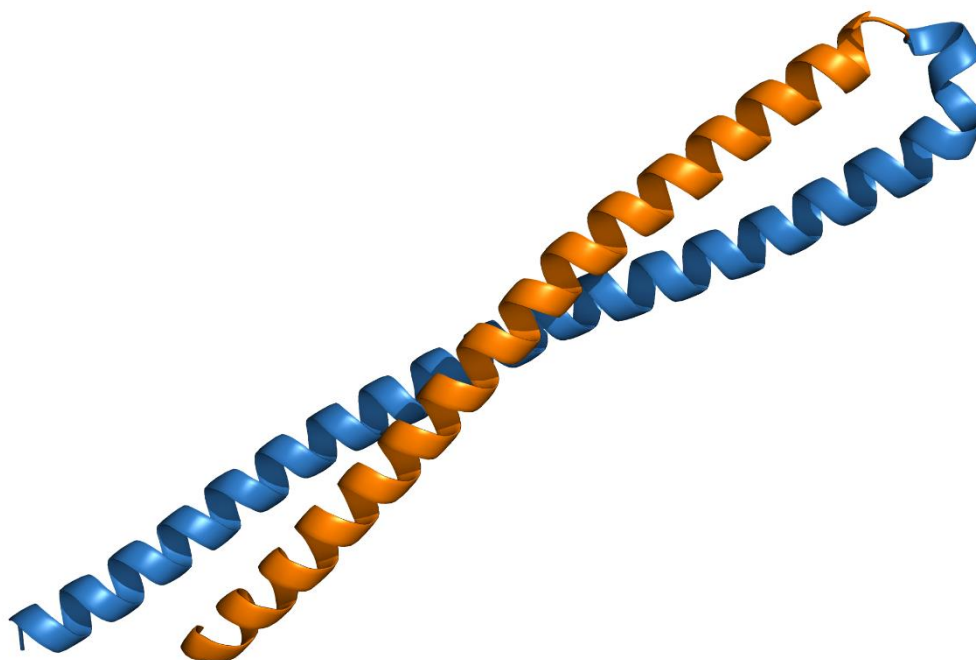


Figure 35 : Structure tertiaire du domaine coiled-coil de Pfk13 prédite par une approche *ab initio*. En l'absence de structures homologues empêchant une modélisation par approche comparative, la structure tertiaire du coiled-coil de Pfk13 a été prédite par une méthode *ab initio* dans laquelle des motifs structuraux sont comparés à la séquence cible puis assemblés. Chaque couleur correspond à une hélice α (bleu : positions 214 – 284 ; orange : positions 285 – 342).

3.1.1.3. Les similitudes du domaine BTB de Pfk13 avec celui des protéines KCTD

L'étude du domaine BTB de Pfk13 a révélé plusieurs caractéristiques inattendues :

- celui-ci est plus court en termes d'éléments structuraux que ceux issus de différentes familles protéiques contenant un BTB (BTB-ZF, BTB-Kelch, RhoBTB, BTB-NPH3, MATH-BTB, Elongin C, KCTD, KCNA), similairement à la protéine Elongin C, elle-même impliquée dans un complexe d'ubiquitination²⁶⁰ ;
- il partage un pourcentage d'identité de séquence plus élevé avec le BTB de la famille protéique *Potassium (K+) Channel Tetramerization Domain* (KCTD). Les protéines KCTD demeurent mal caractérisées à ce jour mais sont généralement impliquées dans des complexes d'ubiquitination lors d'états tétramérique ou pentamérique du domaine BTB²⁶¹.

L'outil FuncPatch a ensuite été utilisé pour rechercher l'existence d'une corrélation spatiale des taux de substitution spécifiques aux sites et pour détecter d'éventuels *patches* de sites conservés (soumis à une forte sélection purificatrice) au sein de la structure tertiaire. Un tel *patch* a été identifié dans le domaine BTB, correspondant à une possible zone d'interaction protéine-protéine par comparaison avec des domaines BTB issus d'autres familles protéiques. De façon notable, le *patch* de sites conservés couvre partiellement la surface d'interaction : i) avec une protéine Cullin (comme observé avec le complexe Elongin C-Cullin 2²⁶⁰), nécessaire pour la construction d'un complexe d'ubiquitination ; et ii) avec un autre domaine BTB dans le cas d'assemblage tétramérique ou pentamérique des KCTDs²⁶¹.

3.1.1.4. La poche du domaine KREP de Pfk13, une possible surface d'interaction protéine-protéine

Des analyses similaires réalisées sur le domaine KREP de Pfk13 ont révélé la présence d'un *patch* de sites très conservés sur la face inférieure du KREP, correspondant à une poche peu profonde exposée à la surface (Figure 5a de l'article 1). Par comparaison structurale, l'équivalent de cette poche chez les protéines BTB-Kelch correspond souvent à une zone d'interaction avec leur(s) substrat(s) protéique(s)^{129,131,262}. Pour confirmer la robustesse de notre approche dans le contexte de domaines KREP, le même procédé évolutif / structural a été conduit pour des domaines KREP dont la poche a été expérimentalement validée comme surface d'interaction avec des substrats dans des complexes d'ubiquitination / dégradation. L'analyse a inclus la protéine KEAP1 qui régule le facteur de transcription Nrf2¹³¹ ; les protéines KLHL2 et KLHL3, lesquelles ciblent des kinases²⁶² ; et enfin la protéine KLHL12, impliquée dans la régulation de protéines *Dishevelled*²⁶³. Les analyses évolutives et structurales menées sur de larges échantillons (incluant les données de séquence pour environ 150 espèces distinctes par protéine) ont systématiquement révélé la poche du domaine KREP comme enrichie en sites très conservés (Figure 5d de l'Article 1). Étant donné la répétabilité de cette observation au sein de structures KREP fonctionnellement annotées, nous avons conclu que la poche située sur le domaine KREP de Pfk13 représente une possible zone d'interaction protéine-substrat. La comparaison des différentes poches a permis d'identifier des arginines et sérines comme candidates pour l'interaction entre Pfk13 et son substrat : R529, S576, S577, R597 et S623. En effet, ces acides aminés ont été démontrés comme essentiels dans l'interaction de KEAP1 avec Nrf2¹³¹, et de KLHL2/KLHL3 avec leurs kinases respectives²⁶².

3.1.2. Les mutations de résistance aux ARTDs n'altèrent pas les propriétés structurales de la poche du domaine KREP de Pfk13

Nous avons émis l'hypothèse que les mutations de résistance aux ARTDs, presque toutes localisées sur le domaine KREP de Pfk13, modifieraient certaines propriétés structurales ou

physico-chimiques de la poche, limitant ainsi son interaction avec un ou plusieurs substrats. Cependant, les mutations de résistance sont localisées sur une large partie du KREP, et aucune n'est directement située à la surface de la poche. Dès lors, nous avons supposé que ces mutations pourraient induire des modifications à longue distance pour altérer la poche. Des simulations de dynamique moléculaire de domaines KREP mutants C580Y (mutation la plus répandue en Asie du sud-est) et R539T (associée à un très haut degré de résistance) ont révélé une conservation du repliement structural du domaine, des structures secondaires et du nombre total moyen de ponts hydrogènes, par rapport aux résultats obtenus avec le domaine KREP sauvage. En revanche, des hausses de fluctuations locales des acides aminés au sein des systèmes mutants conférant la résistance aux ARTDs, relativement au système sauvage, ont été observées pour quelques sites distants de la poche. Ces fluctuations coïncident avec la production de cartes de corrélation croisées (appelée DCCM pour *Dynamical cross correlation map*), lesquelles montrent une hausse significative de mouvements anticorrélésⁱ dans les systèmes mutants R539T et C580Y par rapport au système sauvage. Ces analyses ont par la suite été réalisées sur la mutation A578S de KREP, démontrée expérimentalement comme non associée à la résistance aux ARTDs¹¹⁶. Les fluctuations étaient globalement similaires à celles observées dans le système sauvage, avec une absence de mouvements anticorrélés. Enfin, l'étude des contacts relatifs aux positions 539 et 580 a révélé plusieurs modifications notables dans les systèmes mutants : R539T perd entre deux et trois ponts hydrogène qui sont eux systématiquement retrouvés chez le sauvage, tandis que C580Y est supposée perdre un pont disulfure avec C532. De façon remarquable, aucune altération structurale ou physico-chimique relative à la poche n'a été observée. L'ensemble de ces observations est cohérent avec l'hypothèse selon laquelle les mutations R539T et C580Y, plutôt que d'altérer la structure de la poche où l'interaction avec un substrat est supposée avoir lieu, occasionneraient une instabilité structurale locale autour de la position mutée. Le mécanisme conduisant *in fine* à la résistance aux ARTDs reste toutefois inconnu ; nous discuterons plusieurs hypothèses dans l'article et dans la partie *Discussion* (cf. section 4.2.1.3.).

3.1.3. Article 1

L'article ci-après, paru dans le journal *Scientific Reports*, présente de façon détaillée nos résultats sur l'évolution et la structure de Pfk13. Les résultats et matériels supplémentaires publiés avec cet article sont fournis dans la partie *Annexe* (section Matériel supplémentaire – Article 1).

ⁱ Paires d'acides aminés dont les atomes de Ca se déplacent dans des directions opposées.

SCIENTIFIC REPORTS

OPEN

Comparative structural and evolutionary analyses predict functional sites in the artemisinin resistance malaria protein K13

Romain Coppée¹, Daniel C. Jeffares², Maria A. Miteva³, Audrey Sabbagh¹ & Jérôme Clain^{1,4}

Numerous mutations in the *Plasmodium falciparum* Kelch13 (K13) protein confer resistance to artemisinin derivatives, the current front-line antimalarial drugs. K13 is an essential protein that contains BTB and Kelch-repeat propeller (KREP) domains usually found in E3 ubiquitin ligase complexes that target substrate protein(s) for ubiquitin-dependent degradation. K13 is thought to bind substrate proteins, but its functional/interaction sites and the structural alterations associated with artemisinin resistance mutations remain unknown. Here, we screened for the most evolutionarily conserved sites in the protein structure of K13 as indicators of structural and/or functional constraints. We inferred structure-dependent substitution rates at each amino acid site of the highly conserved K13 protein during the evolution of *Apicomplexa* parasites. We found two solvent-exposed patches of extraordinarily conserved sites likely involved in protein-protein interactions, one in BTB and the other one in KREP. The conserved patch in K13 KREP overlaps with a shallow pocket that displays a differential electrostatic surface potential, relative to neighboring sites, and that is rich in serine and arginine residues. Comparative structural and evolutionary analyses revealed that these properties were also found in the functionally-validated shallow pocket of other KREPs including that of the cancer-related KEAP1 protein. Finally, molecular dynamics simulations carried out on Pfk13 R539T and C580Y artemisinin resistance mutant structures revealed some local structural destabilization of KREP but not in its shallow pocket. These findings open new avenues of research on one of the most enigmatic malaria proteins with the utmost clinical importance.

Current efforts to control malaria are threatened by the spread in Southeast Asia (SEA) of *Plasmodium falciparum* parasites that are resistant to artemisinin derivatives (ARTs)¹. Treatment failures are now reported in some geographic areas of SEA for the current front-line ART-based combination therapies^{1–3}. Artemisinin resistance (ART-R) is defined as parasites exhibiting *in vivo* a delayed clearance time following an ART-based treatment¹ and *in vitro* an increased survival rate following a brief exposure to a high dose of ART⁴. This phenotype is primarily conferred by single non-synonymous mutations in the *P. falciparum* *k13* (*pfk13*, also named *pfkelch13*) gene^{5,6}. Multiple *pfk13* ART-R alleles have emerged concomitantly in the early 2000s⁷, until a specific, multidrug-resistant lineage carrying the *pfk13* C580Y allele spread and nearly reached fixation, especially in the East Thailand-Cambodia-Lao PDR-Vietnam region^{5,7–10}.

pfk13 encodes a 726-amino acid protein (Pfk13) consisting of a poorly conserved *Apicomplexa*-specific N-terminal region and three annotated, highly conserved domains⁹ (Fig. 1a): a coiled-coil-containing (CCC; amino acids 212–341), a BTB (Broad-complex, tramtrack and bric-à-brac; also known as BTB/POZ; amino acids 350–437) and a C-terminal Kelch-repeat propeller (KREP; amino acids 443–726) which harbors nearly all *pfk13* alleles associated with ART-R^{5,7}. Using large scale mutagenesis screens, the *k13* gene was reported as essential for the intraerythrocytic growth of both *P. falciparum* and *P. berghei* asexual parasites^{11,12}. Removal of Pfk13,

¹Université de Paris, UMR 261 MERIT, IRD, F-75006 Paris, France. ²Department of Biology and York Biomedical Research Institute, University of York, Wentworth Way, York, YO10 5DD, UK. ³Université de Paris, Inserm U1268 MCTR, CITCom UMR 8038 CNRS, Paris, France. ⁴Centre National de Référence du Paludisme, Hôpital Bichat-Claude Bernard, Assistance Publique des Hôpitaux de Paris, F-75018 Paris, France. Correspondence and requests for materials should be addressed to A.S. (email: audrey.sabbagh@parisdescartes.fr) or J.C. (email: jerome.clain@parisdescartes.fr)

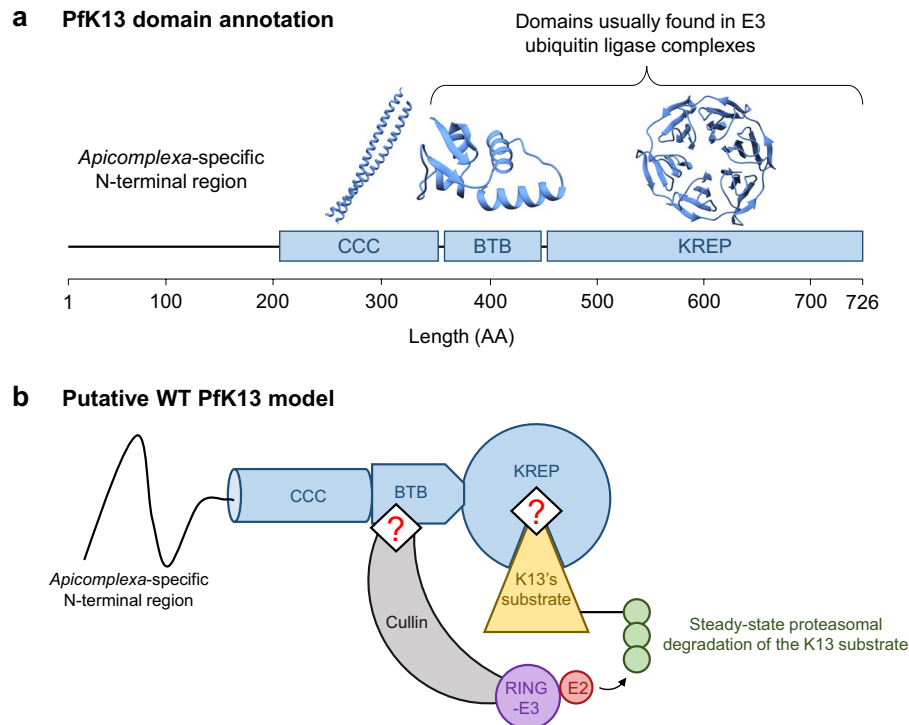


Figure 1. Schematic and structural representation of PfK13 and its putative function as substrate adaptor. **(a)** PfK13 domain annotation. Three domains of PfK13 are annotated in databases: coiled-coil-containing (CCC), BTB, and Kelch-repeat propeller (KREP). The *Apicomplexa*-specific N-terminal domain is predicted to exhibit a random-coil conformation. The crystal structure of BTB and KREP domains was solved; the CCC domain is expected to form two helices coiling together. **(b)** Proposed, simplified model of the protein complex containing PfK13, based on PfK13 domain annotation and co-immunoprecipitation experiments^{17,21,22,89,90}. The BTB domain of PfK13 is expected to bind a scaffold Cullin protein, while the KREP domain likely binds to the substrate molecule(s) further ubiquitinated and possibly degraded by the proteasome. Importantly, the regions and sites of PfK13 involved in binding are unknown (represented as red ‘?’ symbol in the figure). For ease of representation, PfK13 was shown as a monomer although the crystal structure of PfK13 BTB-KREP was solved as a dimer.

through conditional mislocalization of the protein or diCre-based gene deletion, leads to an arrest at the early intraerythrocytic ring stage followed by a slow transition to condensed parasite forms¹³. The timing of the growth arrest in PfK13-inactivated parasites is consistent with the ART-R phenotype being expressed at the early ring stage¹⁴. In addition, parasites carrying *pfk13* ART-R mutations exhibit an altered intraerythrocytic cell cycle, with a lengthened ring stage and a shortened trophozoite stage, relative to wild type (WT)¹⁵. At the molecular level, naturally-occurring *pfk13* mutants associate with several key features: an increased expression of unfolded protein response pathways, detected by the *in vivo* transcriptomics analysis of 1,043 *P. falciparum* isolates from patients with acute malaria response¹⁶; lower levels of ubiquitinated proteins^{17,18}; 2-fold lower abundance of PfK13 protein, as measured by quantitative dimethyl-based proteomics analysis of Cambodian isogenic strains¹⁹ (note that a different result was reported for two African isogenic strains²⁰); and a basal, constitutive phosphorylation of the parasite eukaryotic initiation factor-2 α (eIF2 α) at the early intraerythrocytic stage, which controls the repression of general translation and ART-induced latency^{15,21}. Altogether, it seems that the parasite acquires ART-R by altering its cell cycle such as it spends more time at a developmental stage that is less sensitive to ART drugs (early ring), while at the same time mounting a response mitigating ART-induced damages. Some indications of the interactors with PfK13 partially clarify its function. For example, PfK13 was immunoprecipitated with the phosphatidylinositol 3-kinase (PfPI3K), an enzyme producing the lipid phosphatidylinositol 3-phosphate (PI3P) involved in protein export from the *P. falciparum* endoplasmic reticulum and which cellular levels associate with ART-R¹⁷. Ubiquitination of PfPI3K and its subsequent proteasomal degradation are decreased in *pfk13* mutant parasites. The increased abundance of PfPI3K, and its lipid product PfPI3P, in mutant parasites may influence host cell remodeling and neutralize ART toxic proteopathy^{17,20}.

These experimental data and the Gene Ontology annotation of PfK13, which contains BTB and KREP domains, are consistent with the hypothesis that PfK13 may function as an E3 ubiquitin ligase substrate adaptor (Fig. 1b)^{17,21–23}. Many BTB- and/or KREP-containing proteins are found in multi-subunit RING-type E3 ligase complexes in which substrate protein(s) will be ubiquitinated^{24–27}. In those complexes, the BTB domain mediates both varying oligomerization architectures and the recruitment of a scaffold Cullin protein which interacts with RING E3 protein, whereas the KREP domain serves as a receptor for substrate protein(s) to be ubiquitinated. The E3 ligase activity of RING then catalyzes the transfer of ubiquitin from the ubiquitin-conjugating enzyme

(E2) to the protein substrate(s)^{24,28}. Some KREP domains contain at their bottom side a shallow pocket involved in the binding of substrate protein(s)²⁴. For example, the shallow pocket of KEAP1 and KLHL3, from the large BTB-Kelch family of proteins, directly binds to the transcription factor Nrf2 and some WNK kinases, respectively, and controls their ubiquitination^{29,30}. Importantly, PfK13 is reported to be the sole BTB-KREP-containing protein in *Plasmodium* while mammals include ~60 members, each targeting specific substrate(s)³¹. Therefore, PfK13 may bind several substrates, in addition to PfPI3K.

Although it is expected that PfK13 may act similarly to other BTB-Kelch proteins involved in ubiquitination activities¹⁷, the putative binding regions and functionally important sites of PfK13 and the structural alteration(s) associated with PfK13 ART-R mutations remain uncharacterized (Fig. 1b). Because of its high conservation⁹ and essentiality across *Apicomplexa* species^{11–13,32}, the function of K13 likely remained unchanged during *Apicomplexa* evolution. As a consequence, we would expect that amino acid replacements at the functional sites of K13 would be strongly detrimental to protein function, and would be purged by purifying selection. To find the most conserved amino acid sites of K13, we inferred substitution rates for each amino acid site of the WT K13 sequence from 43 *Apicomplexa* species, taking into account both the spatial correlation of site-specific substitution rates in the protein tertiary structure and the species phylogeny³³. We identified in BTB and KREP two solvent-exposed patches of extraordinarily conserved sites, likely involved in protein-protein interactions. As controls, we found similarly evolutionarily conserved and located patches in several BTB- and KREP-containing proteins found in mammals, including KEAP1. Finally, we tested through molecular dynamics simulations whether the R539T and C580Y ART-R mutations in PfK13 altered some structural properties of the KREP domain and its putative binding sites.

Results

K13 sequence sampling, multiple alignment and phylogenetic reconstruction. Forty-three amino acid sequences from distinct *Apicomplexa* species were unambiguously identified as orthologous to PfK13 in sequence databases, encompassing 21 *Plasmodium* and 22 other *Apicomplexa* K13 sequences (*Cryptosporidia*, *n* = 7; *Piroplasmida*, *n* = 7; and *Coccidia*, *n* = 8). The length of K13 protein sequences ranged from 506 (*Eimeria brunetti*) to 820 (*Hammondia hammondi*) amino acids (Supplementary Table S1). By visual inspection, the three annotated domains of K13 (CCC, BTB and KREP) were very conserved (see the K13 sequence alignment in Supplementary Fig. S1), whereas the N-terminal region preceding the CCC domain appeared much more variable in length and amino acid composition, to a point of not being even detected in databases for many non-*Plasmodium* *Apicomplexa* species. Since K13 sequences aligned poorly in that region, the first 234 amino acid positions of the K13 multiple alignment were removed, along with other positions showing a high divergence level and/or gap enrichment among sequences (32 amino acid positions; Supplementary Fig. S1). The final K13 multiple sequence alignment contained 514 amino acid positions which covered the whole CCC, BTB and KREP domains. The average pairwise sequence identity in that cleaned K13 sequence alignment was 64.6%, ranging from 48.6% for the *Babesia bigemina*-*Cryptosporidium ubiquitum* comparison to 99.2% for the *P. chabaudi* spp. pair. This is well below the average pairwise sequence conservation found among *Plasmodium* species (84.9%, with a minimum value of 78.3% for the *P. vivax*-*P. chabaudi adami* pair).

The maximum-likelihood phylogenetic tree built from the curated alignment of the corresponding *Apicomplexa* *k13* cDNA sequences revealed four monophyletic groups: *Cryptosporidia*, *Plasmodium*, *Piroplasmida* and *Coccidia*, all supported by high bootstrap values ($\geq 98\%$; Supplementary Fig. S2). The group of *Hematozoa k13* sequences appeared as paraphyletic (bootstrap value = 100%), with *Piroplasmida* unexpectedly clustering with *Coccidia* (Supplementary Fig. S2). The phylogenetic relationships of *Plasmodium k13* sequences were largely consistent with the acknowledged phylogeny of *Plasmodium* species, except for bird-infecting *Plasmodium k13* sequences (*P. gallinaceum* and *P. relictum*), which appeared related to human-infecting *P. ovale* spp. sequences, although this grouping was poorly supported (bootstrap value = 47%; Supplementary Fig. S2).

The *k13* sequence has evolved under strong purifying selection. To evaluate the selective pressure acting on *k13* which reflects its level of conservation during evolution, we used codon substitution models to estimate the rate of non-synonymous to synonymous substitutions, $\omega = d_N/d_S$, across codon sites of the *k13* sequence (site models) and branches of the *k13* phylogeny (branch models). A series of nested likelihood ratio tests (LRTs) using different sets of site and branch models were carried out using the codeml tool from the PAML package³⁴. When applied to the *k13* codon alignment, LRTs of codon and branch models indicated varying substitution rates ω both across codon sites of the *k13* sequence (M0-M3 comparison, $p = 3.3 \times 10^{-225}$) and among branches of the *k13* phylogeny (M0-FR, $p = 1.9 \times 10^{-53}$; Tables 1 and S2). This suggests that *k13* has evolved under a variable selective regime both across codon sites and among lineages. No evidence of positive selection was found in any of the tree branches (Supplementary Fig. S3). Similarly, site models incorporating positive selection (M2a and M8) provided no better fit to the data than those including only purifying selection and neutral evolution (M1a and M7: $2\Delta\ell = 0$ in both cases; Table 1), thus supporting an absence of detectable adaptive selection events at any *k13* codon site over the long time scale of *Apicomplexa* evolution. Altogether, the data indicate that much of the K13 protein, except the N-terminal region which was not studied here, has been strongly conserved over evolutionary time.

When considering the one-ratio PAML value for 3,256 protein-coding genes previously estimated by Jeffares *et al.*³⁵ using six *Plasmodium* species, *k13* ranked among the 5% most conserved protein-coding genes of the *Plasmodium* proteome (rank: 140/3,256; Fig. 2a). Since a significant correlation between protein length and one-ratio PAML value was evidenced in the whole dataset (Spearman's rank correlation: $p = 9.2 \times 10^{-82}$, $r = 0.33$; Supplementary Fig. S4), we repeated the analysis by considering only those protein sequences whose length was included in an interval of ± 100 amino acids centered on the PfK13 protein length (Spearman's rank correlation: $p = 0.83$, $r = 0.01$). Again, *k13* ranked among the most conserved protein-coding genes of the *Plasmodium*

Models	Purpose of the test	$2\Delta\ell$	df	<i>p</i> value
M0-FR	Heterogeneity of ω among lineages of the phylogenetic tree	457.0	82	1.9×10^{-53}
M0-M3	Heterogeneity of ω across codon sites	1,046.3	4	3.3×10^{-225}
M1a-M2a	Detection of sites evolving under positive selection	0	2	1
M7-M8	Detection of sites evolving under positive selection	0	2	1

Table 1. Lack of detectable positive selection in the *k13* gene using codon substitution models. ω , non-synonymous (d_N) to synonymous (d_S) substitution rate ratio (d_N/d_S); $2\Delta\ell$, log-likelihood ratio of the two tested models; df, degrees of freedom. $2\Delta\ell$ was compared to a chi-squared table to determine the significance of the likelihood ratio tests. The M1a-M2a comparison is reported to be more stringent than the M7-M8 comparison³⁴.

proteome (sized-rank: 6/393), whereas four other five- or six-bladed KREP protein-coding sequences showed much less intense levels of conservation than *k13* (Figs. 2a and S5).

Variable levels of amino acid conservation between the annotated domains of K13. We then compared the conservation level between the annotated domains of K13 (CCC, BTB and KREP) using ω estimates obtained under the best fitted PAML model M3 that indicates a variable selective regime among codon sites (Supplementary Table S2). First, we noted that the three domains have evolved under strong purifying selection with most codon sites being highly conserved during evolution (Fig. 2b). BTB was however found to evolve under more intense purifying selection than either CCC ($p = 1.6 \times 10^{-4}$, Mann-Whitney *U* test) or KREP ($p = 1.0 \times 10^{-3}$), and no difference in ω estimates was detected between CCC and KREP ($p = 0.75$; Fig. 2c). To confirm these results, we inferred site-specific substitution rates at the protein level using the FuncPatch server which considers their spatial correlation in protein tertiary structure (hereafter called λ substitution rates)³³. λ could not be inferred for the CCC domain because of the lack of a resolved 3D structure. The analysis confirmed that BTB was more conserved than KREP during *Apicomplexa* evolution ($p = 5.4 \times 10^{-5}$, Mann-Whitney *U* test).

The BTB domain of K13 resembles that of KCTD proteins and exhibits a predicted functional patch.

We then performed a more extensive study of the BTB and KREP domains of K13 because of their likely role in mediating K13 functions and the availability of their tertiary structures. To detect patches of slowly evolving amino acid sites in the BTB-KREP structure, we focused on the site-specific substitution rates λ at the amino acid level mentioned above. λ , estimated using FuncPatch, has been shown to provide a more reliable estimation of the conservation level at amino acid sites compared to standard substitution estimates, especially in the case of highly conserved proteins^{33,36}.

Although K13 shares similarities with the BTB-Kelch family of proteins that are commonly found in E3 ubiquitin ligase complexes⁵, the BTB domain of K13 exhibited atypical features compared to this protein family. First, the K13 BTB fold appeared shortened, lacking the A6 helix and the N- and C-terminal extensions²⁶ (Fig. 3a), similarly to the Elongin C protein which acts as an E3 ubiquitin ligase adaptor mediating the ubiquitination of substrates³⁷. Second, the primary BTB sequence of K13 grouped with those of the potassium (K⁺) channel tetramerization domain (KCTD) protein family (Fig. 3b). Finally, the BTB of K13 exhibited a higher similarity in tertiary structure with that of KCTD17 (used as reference for KCTD proteins) compared to those of Elongin C and KEAP1 (the later used as reference for BTB-Kelch proteins): the root-mean-square deviations (RMSDs) of atomic positions for pairs of superimposed BTB domains were 1.13 Ångström (Å) for K13-KCTD17, 2.33 Å for K13-Elongin C and 2.17 Å for K13-KEAP1. In all, these results suggest that K13 BTB is related in sequence and structure to that of KCTD proteins. KCTDs are soluble non-channel proteins containing a BTB domain (but no KREP) involved in variable oligomerization architectures and often acting as versatile scaffold in E3 ubiquitin ligase complexes²⁷.

To identify putative functional sites in K13 BTB, we examined whether a spatial correlation of the site-specific substitution rates λ was present in the K13 BTB-KREP tertiary structure. Despite a low standard deviation of substitution rates across amino acid sites, a significant spatial correlation was found (Table 2). The 10% most conserved sites predicted by FuncPatch formed one clearly bounded patch on the BTB surface and consisting of sites located at the B2-B3 and B4-A4 loops and at the A4 helix (positions/residues 355–358/NVGG, 397–399/DRD and 402–403/LF using the Pfk13 sequence numbering; Fig. 3c). To test whether a similarly located, conserved patch was also found in the BTB domain of KCTD proteins – to which K13 BTB is the most similar – we inferred site-specific substitution rates λ from 124 and 139 orthologous sequences of SHKBP1 and KCTD17, acting as E3 ubiquitin ligase adaptors mediating the ubiquitination of substrates²⁷. For both proteins, the 10% most conserved BTB positions formed a patch that partially overlapped with the one of K13, with positions 357 (B2-B3 loop), 397 and 398 (B4-A4 loop) being shared between the three patches (Pfk13 sequence numbering; Fig. 3d and Table 2). These positions are usually involved in BTB-BTB interactions in tetrameric or pentameric states of KCTDs²⁷ or in BTB-Cullin interactions in some other BTB-containing protein families in monomeric or dimeric states²⁶, as in the X-ray structure of the Elongin C-Cullin2 complex (Supplementary Fig. S6)³⁷. This suggests that these K13 BTB sites may be involved in some protein-protein interactions.

The KREP domain of K13 exhibits a conserved, rigid, solvent-exposed shallow pocket. In KREP-containing proteins, the KREP domain usually serves as the receptor for substrate(s) further ubiquitinated^{24,28}. Before examining the conservation level of the K13 KREP domain, we first re-evaluated its architecture using the resolved 3D structure of Pfk13 KREP (PDB code 4YY8, chain A; unpublished results)³⁸. The Pfk13

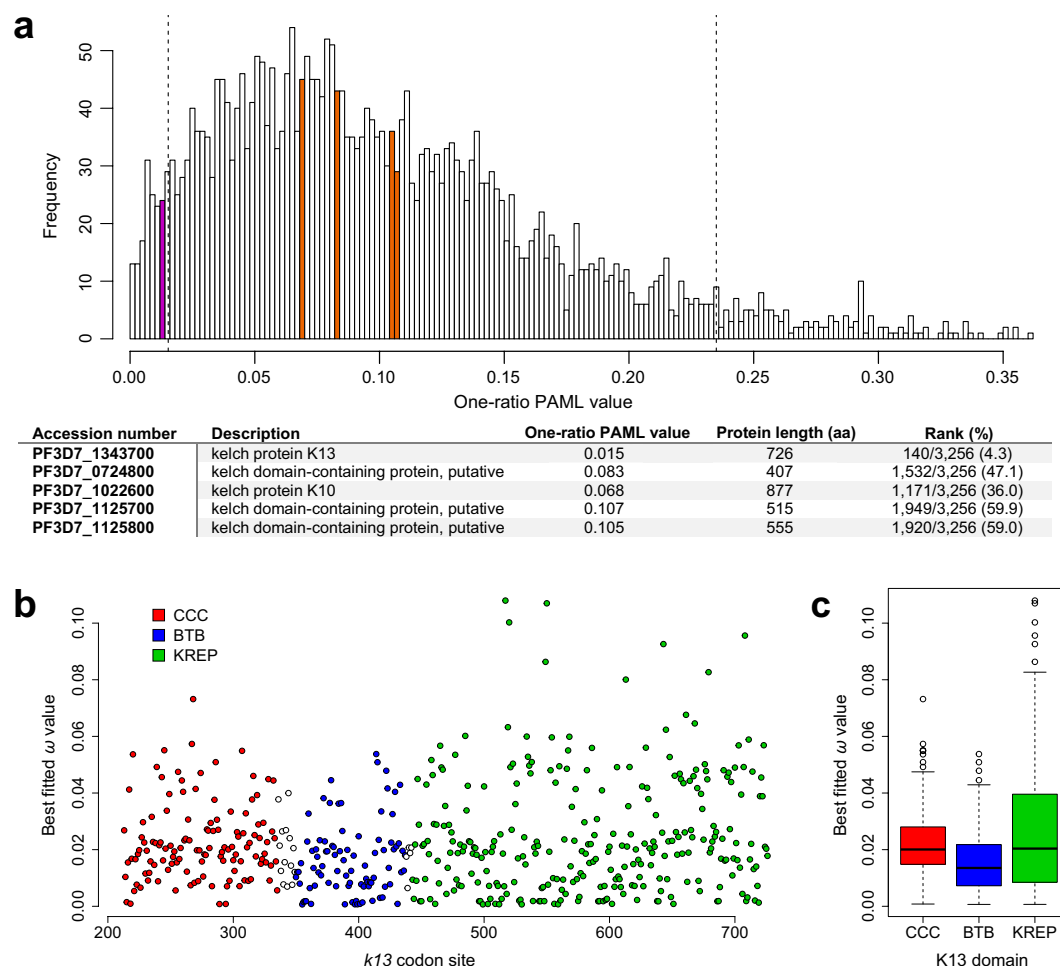


Figure 2. Conservation level of the K13 protein, domains and codon sites. **(a)** Conservation level of *k13* compared to those of all protein-coding genes of *Plasmodium*. The one-ratio PAML values, calculated for 3,256 orthologous genes among six *Plasmodium* species (*P. falciparum*, *P. berghei*, *P. chabaudi*, *P. vivax*, *P. yoelii* and *P. knowlesi*), were retrieved from the study of Jeffares and colleagues³⁵. Magenta and orange bars indicate the location of *k13* and four other KREP protein-coding genes (PlasmoDB accession numbers: PF3D7_1022600, PF3D7_1125800, PF3D7_0724800, PF3D7_1125700) respectively. Vertical dashed lines show the 5% cutoff of the most and less conserved protein-coding genes. Among KREP protein-coding genes, only *k13* belonged to the top 5% of the most conserved protein-coding genes of the *Plasmodium* proteome. The table shown below the histogram provides the rank of each KREP-containing protein-coding sequence in the whole dataset. **(b)** Conservation level of the *k13* codon sites. The scatter plot shows the conservation level of *k13* codon sites using the *pfk13* sequence as reference for the codon numbering (starting at codon 213). White circles correspond to inter-domain positions. All codon sites were reported to evolve under strong purifying selection, with ω drastically < 1 . We used the ω estimates obtained under the best fitted PAML model M3 that indicates a variable selective regime among codon sites. **(c)** Conservation level of the annotated K13 domains. The box-whisker plot shows that BTB evolves under more intense purifying selection compared to either CCC or KREP, using non-parametric Mann-Whitney *U* test ($p < 0.05$). Box boundaries represent the first and third quartiles and the length of whiskers correspond to 1.5 times the interquartile range.

KREP structure is composed of six repeats of the kelch motif/blade^{5,39}. As expected, each blade is a β -sheet secondary structure involving four twisted antiparallel β -strands (numbered A to D). The innermost A strands line the central channel of the KREP fold whereas the outermost D strands are part of the KREP outside surface. The top face of the domain, containing the central channel opening, is formed by residues from several strands and from AB and CD loops. Although we cannot exclude that the central channel can carry some functions, it was reported to be mainly a structural consequence of kelch motif assembly⁴⁰. The bottom face of PfK13 KREP is composed of residues from the DA and BC loops and contains a shallow pocket, similarly to other KREP structures²⁴. Since there is no conventional definition for the shallow pocket delineation in KREP fold, we defined it as the amino acids forming the surface plan of the pocket and protruding out of the plan ($n = 19$ positions; Figs. 4a and S7).

To characterize the pattern of conservation within the domain, we superimposed the site-specific substitution rates λ inferred by FuncPatch onto an amino acid sequence alignment of the six blades (Fig. 4b), custom-produced

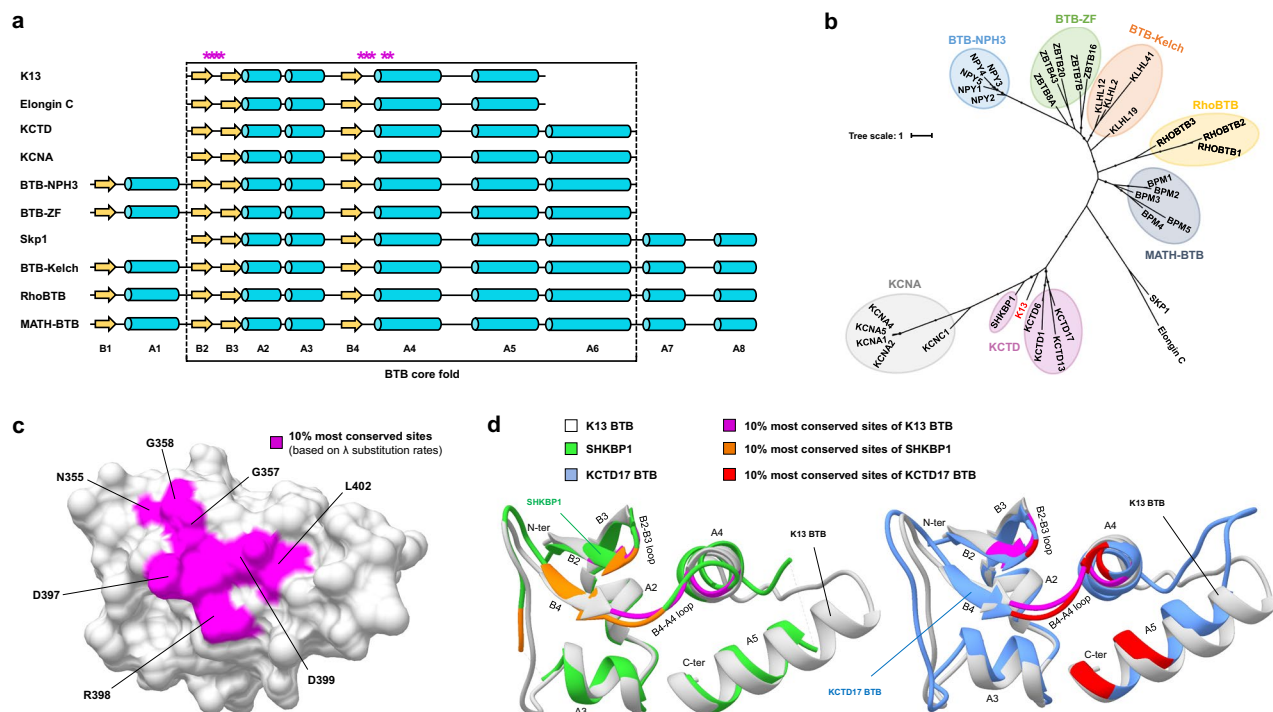


Figure 3. Conservation and structure homology of the K13 BTB fold. **(a)** Linear schematic representation of the BTB fold of some BTB-containing protein families. Yellow arrows and cyan cylinders represent strands and helices, respectively. Stars in magenta correspond to the 10% most conserved K13 BTB amino acid sites (based on the ranking of the λ substitution rates, FuncPatch analysis; Supplementary Dataset S1). Structural elements are labelled. **(b)** Maximum-likelihood unrooted phylogenetic tree of the BTB core fold using a few reference sequences per BTB-containing protein family. Each BTB-containing protein family forms a monophyletic group, identified with a colored background. K13 BTB is written in red color and clusters with the KCTD protein family of BTB-containing proteins. **(c)** Patch of slowly evolving amino acid sites in a three-dimensional view of PfK13 BTB. The amino acid sites are labelled using the PfK13 sequence as reference. **(d)** Superposition of the BTB fold of K13 with that of two members of the KCTD protein family. The most conserved amino acid sites for each protein was based on FuncPatch analysis and are shown in magenta, orange and red for K13, SHKBP1 and KCTD17, respectively. Other amino acid sites are shown in white, green, and cornflower blue for K13, SHKBP1 and KCTD17, respectively. Three common positions were identified in the BTB conserved patches of K13, SHKBP1 and KCTD17: 357 on the B2-B3 loop and 397–398 on the B4-A4 loop.

Protein domain	l	σ	log Bayes factor
K13 _{BTB-KREP}	5	0.9	30.84
K13 _{KREP}	3	0.9	20.67
SHKBP1 _{BTB}	11	1.5	13.27
KCTD17 _{BTB}	7	1.3	9.69
KEAP1 _{KREP}	5	1.7	44.06
KLHL2 _{KREP}	7	1.5	93.54
KLHL3 _{KREP}	7	1.3	79.47
KLHL12 _{KREP}	7	1.5	25.18

Table 2. Strength of the spatial correlation of λ substitution rates in the BTB and KREP domains of K13 and other proteins. l , strength of the spatial correlation of site-specific substitution rates λ over protein tertiary structure, in Å; σ , signal standard deviation measuring the variability of λ substitution rates across amino acid sites; log Bayes factor, statistical significance of the spatial correlation of site-specific substitution rates λ . A log Bayes factor ≥ 8 identifies a significant spatial correlation of site-specific substitutions rates λ in protein tertiary structure (conservative cutoff as suggested by FuncPatch' authors)³³.

from a structural alignment of the six blades using PyMOL (Figs. 4a and S8). We found that: (i) the conservation level significantly differed between the six blades ($p = 3.0 \times 10^{-3}$, Kruskal-Wallis H test), with blade VI exhibiting the lowest level of conservation (Figs. 4b and S9); (ii) loops were surprisingly more conserved than strands ($p = 5.0 \times 10^{-3}$, Mann-Whitney U test; Fig. 4b); (iii) the solvent-exposed A and D strands were less conserved



Using the PFK13 KREP structure as reference, we also identified three remarkable features of the conserved patch predicted by FuncPatch. First, it overlapped with a region of the shallow pocket harboring a slightly

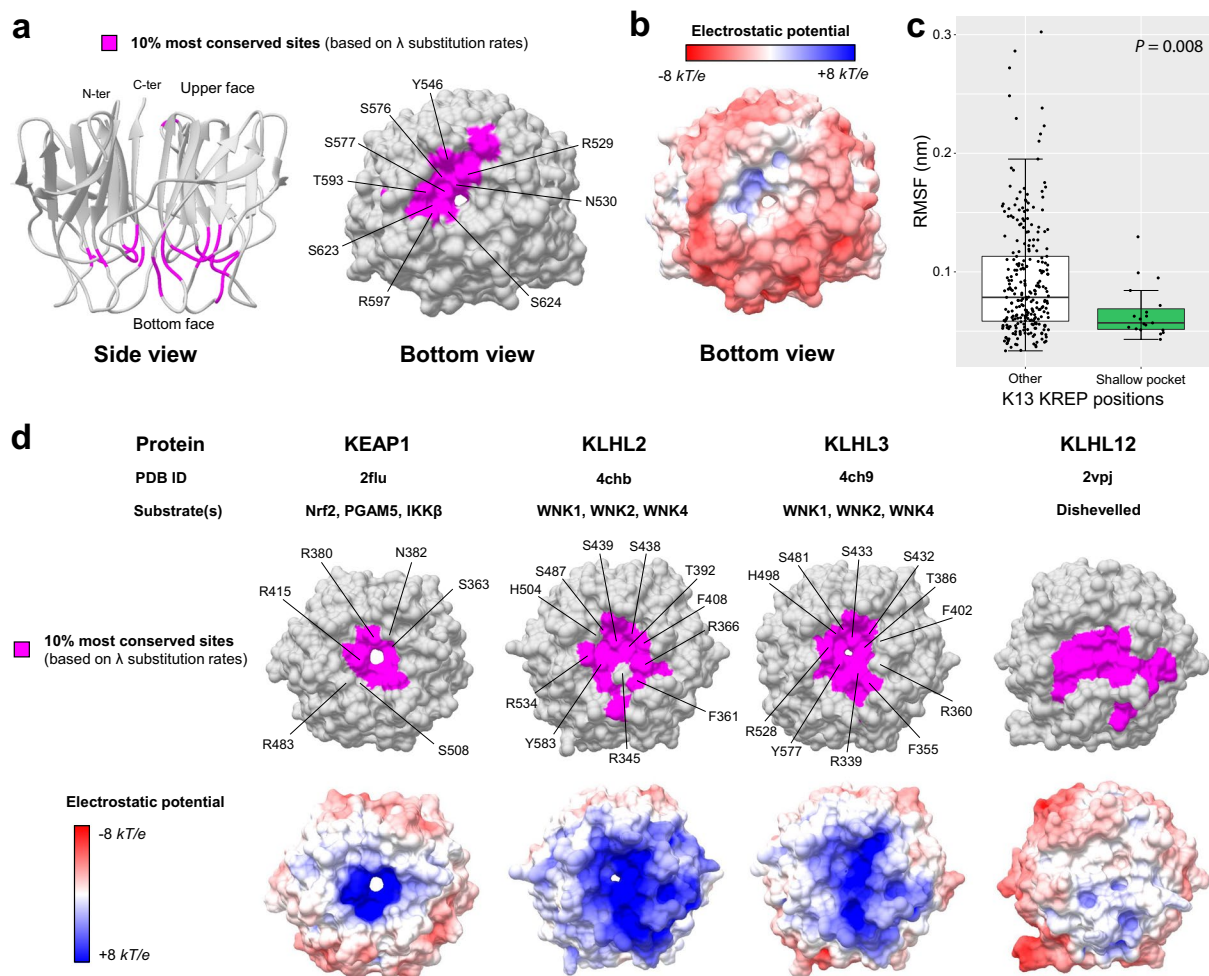


Figure 5. Conservation level and electrostatic potential across the KREP structures of K13 and other BTB-Kelch proteins. **(a)** Location of the 10% most conserved amino acid sites (magenta) on the three-dimensional structure of PfK13 KREP. The conservation level of positions was defined using the site-specific substitution rates λ estimated with FuncPatch (Supplementary Dataset S1). The KREP structure is shown from the side view as cartoon (left structure) and from the bottom view as surface (right structure). The amino acid sites forming the surface of the shallow pocket and belonging to the 10% most conserved sites are labelled. **(b)** Electrostatic surface potential of the PfK13 KREP structure, estimated with the APBS method. Electrostatic potential values are in units of kT/e at 298 K, on a scale of $-8 kT/e$ (red) to $+8 kT/e$ (blue). White color indicates a neutral potential. The missing charges were added using the *Add Charge* function implemented in UCSF Chimera. **(c)** Box plots showing the distribution of root-mean-square fluctuations (RMSFs; Supplementary Dataset S2) for the PfK13 KREP shallow pocket positions (*shallow pocket* group, green) and the remaining PfK13 KREP positions (*other* group, white). RMSFs were calculated through a molecular dynamics simulation for a duration of 100 ns. Box boundaries represent the first and third quartiles and the length of whiskers correspond to 1.5 times the interquartile range. The difference between groups was evaluated by non-parametric Mann-Whitney U test. **(d)** Location of the 10% most conserved amino acid sites and electrostatic potential for KEAP1, KLHL2, KLHL3 and KLHL12 KREP structures. The color code and structure orientation are the same as for PfK13 in panels a and b. For KEAP1, KLHL2 and KLHL3, the key amino acids interacting with their respective protein substrates are labelled. The PDB codes and protein substrates are provided above each KREP structure.

electropositive surface potential, in contrast to the overall electronegative one of the KREP bottom surface (Fig. 5b). Second, it contained several evolutionarily conserved arginine and serine residues (PfK13 R529, S576, S577, R597 and S623; Fig. 5a and Table 3), which are known to mediate protein-protein interactions in the pocket of other KREP domains^{41,42}. And third, the shallow pocket, although being at the domain surface, was more rigid compared to the rest of the KREP structure ($p = 8.0 \times 10^{-3}$, Mann-Whitney U test), as measured by the root-mean-square fluctuation values (RMSF, which measures to what extent a given residue changes its position over time) during a molecular dynamics simulation of 100 nanosecond (ns) at a temperature of 300 K (Fig. 5c; two other replicates produced similar results, data not shown). Altogether, our analyses of K13 KREP revealed that the shallow pocket is extremely conserved, poorly flexible, and differentially charged compared to the remaining KREP bottom face.

AA ^a position	Ref. AA ^a (<i>P. fal</i>)	PfK13 mutant ^b	Sequence variation across <i>Apicomplexa</i> species		λ -based KREP rank n, (%) ^c
			<i>Plasmodium</i> (n = 21)	Non- <i>Plasmodium</i> (n = 22)	
451	F	—	F (21)	H (16); C (4); T (2); S (1); N (1)	132 (46.48)
456	Y	—	Y (21)	H (12); I (7); P (2); Y (1)	125 (44.01)
482	Y	S ^{AFR}	Y (21)	Y (17); H (5)	37 (13.03)
498	N	I ^{AFR}	N (21)	Q (13); Y (7); H (2)	44 (15.49)
529	R	K ^{AFR}	R (21)	R (22)	1 (0.35)
530	N	K ^{AFR} /Y ^{AFR}	N (21)	N (22)	9 (3.17)
546	Y	—	Y (21)	F (22)	13 (4.58)
551	I	—	I (21)	I (19); V (3)	86 (30.28)
576	S	I ^{AFR}	S (21)	S (22)	3 (1.06)
577	S	p ^{AFR}	S (21)	S (22)	10 (3.52)
593	T	S ^{AFR}	T (15); A (6)	T (22)	19 (6.69)
597	R	—	R (21)	R (22)	12 (4.23)
623	S	C ^{SEA}	S (21)	S (22)	2 (0.70)
624	S	—	S (21)	A (15); G (7)	23 (8.10)
640	I	V ^{AFR}	I (21)	I (19); M (2); V (1)	42 (14.79)
671	M	—	M (21)	M (17); L (3); I (1); F (1)	73 (25.70)
672	N	—	N (21)	D (22)	47 (16.55)
688	E	—	E (21)	Q (22)	45 (15.85)
717	F	—	F (21)	Y (19); S (3)	33 (11.62)

Table 3. Conservation level of the protein sites constituting the K13 KREP shallow pocket. ^aAA: amino acid position. Ref. AA indicates the wild type amino acid in PfK13. ^bNone of the PfK13 mutant positions located at the K13 KREP shallow pocket was reported to confer the ART-R phenotype. All PfK13 mutants presented here have been described in one or two parasite samples from large population surveys^{5,8,9,91,92}. AFR (Africa) and SEA (Southeast Asia) refer to the geographic origin of the PfK13 mutant parasite isolate. ^cThe rank attributed to the shallow pocket positions was based on the site-specific substitution rates λ (Supplementary Dataset S1) estimated for the 284 studied KREP positions using FuncPatch³³. The lower the rank, the more conserved was the position. Positions in bold are those belonging to the 10% most conserved positions of KREP.

The conserved K13 KREP patch is related to KREP binding activities in well-characterized BTB-Kelch proteins. To evaluate the reliability of FuncPatch to infer conserved functional sites in the context of KREP folds, we studied four other BTB-Kelch proteins found in mammals and which are functionally and structurally well-characterized and involved in E3 ubiquitin ligase complexes (KEAP1, KLHL2, KLHL3 and KLHL12). All these proteins are known to bind substrate proteins through validated binding residues located in their KREP shallow pocket: Nrf2 for KEAP1⁴², WNK for KLHL2 and KLHL3⁴¹, and Dishevelled for KLHL12⁴³. Using large sets of orthologous amino acid sequences (ranging from 129 sequences for KLHL12 to 162 sequences for KLHL2), a statistically significant spatial correlation of site-specific substitution rates λ was detected for each KREP fold (Table 2). In each case, the 10% most conserved KREP positions clustered in their respective shallow pocket (highest $p = 1.4 \times 10^{-7}$ for KEAP1, chi-squared test; Fig. 5d and Supplementary Table S3) and included some critical amino acid residues for substrate binding. The shallow pocket of KEAP1, KLHL2 and KLHL3 KREP structures also showed a markedly electropositive surface potential while that of KLHL12 was more neutral and similar to that of PfK13 (Fig. 5d). Therefore, these results demonstrate the reliability of our structural-evolutionary approach to detect functional sites in a context of KREP folds.

K13 KREP positions involved in ART-R are not associated with basic structural and long-term evolutionary parameters. Several non-synonymous mutations in PfK13 KREP have been reported to confer ART-R to *P. falciparum* parasites from SEA^{5–9}. Here, we tested whether these ART-R positions have specific evolutionary or structural properties. Therefore, K13 KREP positions were classified as associated (n = 28) or not (n = 256) with an ART-R mutation, on the basis of the last World Health Organization (WHO) status report on ART-R (listed in Supplementary Table S4)⁴⁴. First, we observed a wide distribution of ART-R positions across the KREP fold (Fig. 6a), but none located at the shallow pocket, although this trend was not statistically confirmed ($p = 0.23$, Fisher's exact test; 0/19 ART-R mutations for the shallow pocket positions, and 28/265 ART-R mutations for the remaining PfK13 KREP positions). Second, no difference in the inter-species spatially correlated site-specific substitution rates λ was observed between the two groups of KREP positions ($p = 0.96$, Mann-Whitney *U* test; Supplementary Fig. S11). Of note, ten positions associated with an ART-R mutation exhibited a substantially higher substitution rate than position 580 – associated with the *pfk13* C580Y allele which has outcompeted other alleles in the East Thailand-Cambodia-Lao PDR-Vietnam region⁷ (Supplementary Table S4). Finally, three structural parameters were also estimated for each position using the WT PfK13 KREP structure⁴⁵: the relative solvent accessibility (RSA) which measures to what extent an amino acid is accessible to

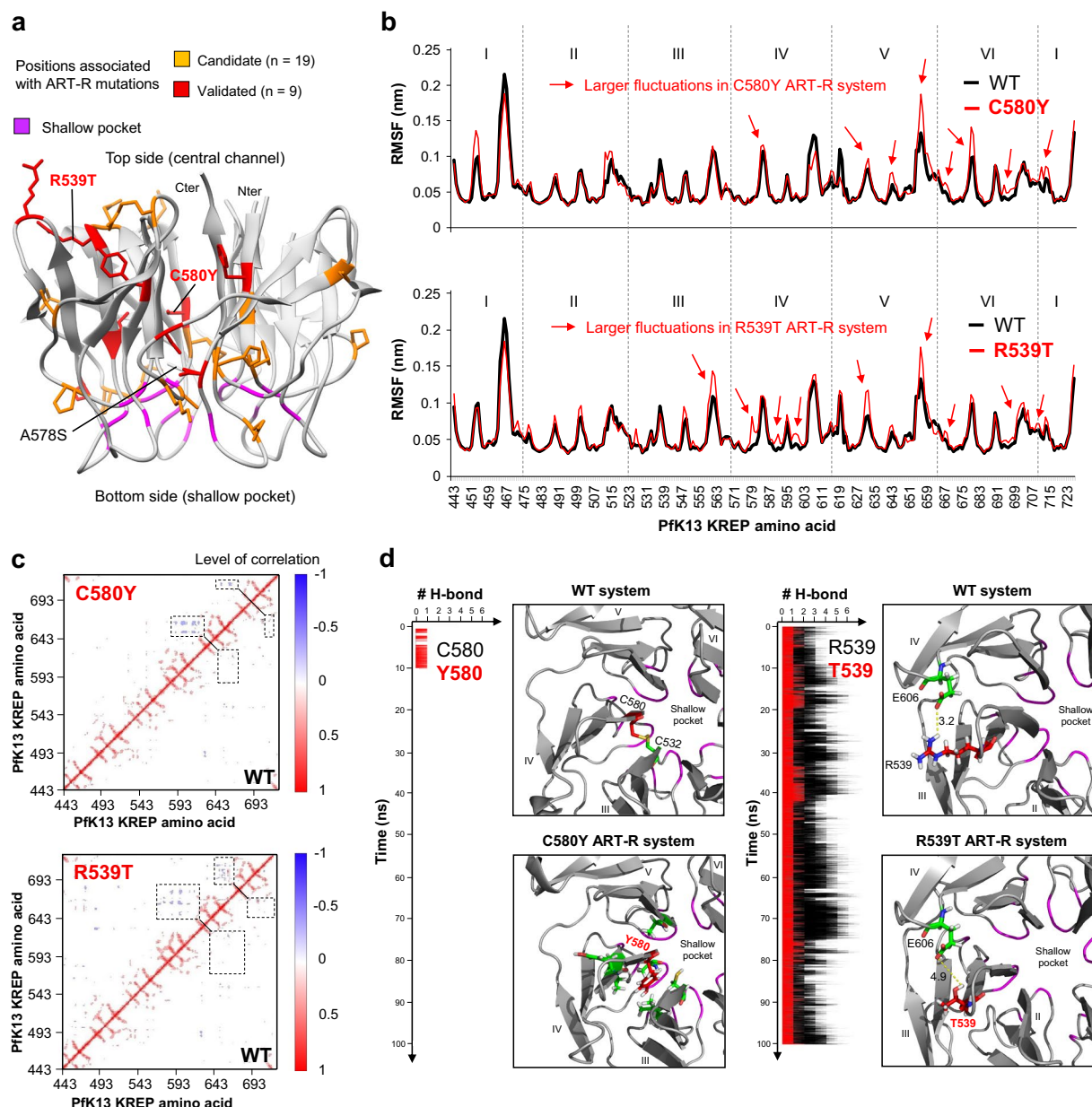


Figure 6. Results of molecular dynamics simulations for WT and ART-R mutant PfK13 KREP structures. Molecular dynamics simulations were carried out on the KREP structure using GROMACS during 100 ns at a temperature of 300 K in an all-atom system. The first ns (0 to 5 ns) correspond to an equilibration phase. (a) Mapping of PfK13 ART-R mutations onto the KREP structure. The positions associated with a validated or a candidate ART-R mutation are colored in red and yellow respectively. Validated and candidate ART-R mutations are defined on the basis of the last WHO status report on ART-R⁴⁴. Amino acids forming the shallow pocket are colored in purple. The mutations studied by molecular dynamics simulations are labelled, including the A578S, which is not associated with ART-R. (b) Root-mean-square fluctuation (RMSF) values of WT (black) and mutant ART-R (red) PfK13 KREP structures. RMSF per position was calculated on the backbone C α atoms (excluding the first five ns, corresponding to the equilibration phase). (c) Dynamical cross-correlation maps (DCCMs) of WT (*bottom right*) and mutant ART-R (*top left*) systems. In DCCMs, positive correlation for a pair of residues (red) implies that the two residues move in the same direction, while negative correlation (blue) indicates that the two residues move in opposite directions. Dashed boxes correspond to the differential movements between the WT and mutant systems. Maps were generated using Bio3D in R⁸⁷. (d) Local impacts of C580Y and R539T ART-R mutations on PfK13 KREP structure. Blade number and location of the shallow pocket are indicated. Inter-atomic distances are expressed in Å.

the solvent, the side-chain weighted contact number (WCN_{sc}) which estimates how densely packed a residue is within the protein tertiary structure, and the RMSF mentioned above (which measures to what extent a given residue changes its position over time). No difference was detected for these structural parameters between the

KREP positions associated or not to ART-R (RSA: $p = 0.44$; WCN_{sc}: $p = 0.46$; RMSF: $p = 0.20$, Mann-Whitney U test; similar results were obtained with other replicates for RMSF; Supplementary Fig. S11). Altogether, these results suggest that K13 KREP positions involved in ART-R are not associated with basic structural and long-term evolutionary parameters.

PfK13 ART-R mutations cause local structural destabilization of KREP. Next, we tested whether the introduction of amino acid changes that confer ART-R would alter the PfK13 KREP structure, and more specifically the putatively functional shallow pocket. Here, we focused on the ART-R mutations C580Y and R539T because they are the most prevalent and associated with very high level of ART-R respectively. The mutations were introduced individually *in silico*, then the mutant structures were subjected to molecular dynamics simulations during 100 ns (namely C580Y ART-R and R539T ART-R systems). As controls, we studied the WT structure and an ART-sensitive mutant structure (A578S, commonly found in Africa; A578S ART-S system). In all the mutant systems, we did not detect any striking global defect of the KREP structure, as assessed by the total number of hydrogen bonds (H-bonds) in KREP (Supplementary Fig. S12a) and by both the RMSD (Supplementary Fig. S12b) and solvent accessible surface area (Supplementary Fig. S12c) that were averaged across all KREP positions. Also, the electrostatic surface potential of the WT and mutant structures were similar (Supplementary Fig. S12d).

We then looked at individual positions across the molecular dynamics simulations of the WT and mutant systems by examining the spatial fluctuations of KREP backbone C α atoms (RMSF data). Larger fluctuations of amino acid residues from blades IV, V and VI were observed in the mutant ART-R systems compared to the WT system (Fig. 6b). This contrasts with the A578S ART-S system in which fewer fluctuations were found compared to the WT system (Supplementary Fig. S13a). The fluctuations were similar for the amino acid residues forming the shallow pocket in the mutant systems compared to the WT system (Supplementary Fig. S14a). We further calculated the dynamical cross-correlation maps (DCCMs) which measure for each pair of residues whether they move in the same direction (correlated motions) or in opposite directions (anti-correlated motions), using the backbone C α atoms along the molecular dynamics simulations. Increased anti-correlated motions were observed in the mutant ART-R systems – especially for R539T – compared to the WT system (Fig. 6c), involving amino acid residues located on blades IV, V and VI, which is consistent with RMSF profiles. No difference in correlated and anti-correlated motions of KREP residues was found between the A578S ART-S and WT systems (Supplementary Fig. S13b).

Finally, we looked at the local environment of the mutant residues. In the C580Y ART-R system, a putative disulfide bridge C532-C580 (blade III-blade IV) was lost (Fig. 6d). In addition, H-bonds involving C580Y (blade IV) with amino acid residues from blades III, IV and V were found only during the first ten ns of the simulation whereas they were not found at all in the WT system (Fig. 6d). The gain and then early, definitive loss of these H-bonds suggested a local rearrangement of the amino acids close to C580Y and that the loss of the disulfide bridge destabilizes the C580Y structure. In the R539T ART-R system, the R539T residue (blade III) was involved in a drastically lower number of H-bonds, including the complete loss of the salt bridge with E606 (blade IV), which was systematically found during the 100 ns simulation in the WT system (Fig. 6d). The loss of an inter-blade H-bond may affect structure stability. In the A578S ART-S system, one new intra-blade H-bond with G591 (blade IV) was recurrently gained and lost during simulations compared to the WT system (Supplementary Fig. S13c), suggesting that it is not critical for the structure stability. Interestingly, none of the tested mutations was found to alter the H-bond network at the shallow pocket (Supplementary Fig. S14b).

Altogether, these results suggest that *pfk13* R539T and C580Y ART-R mutations cause local structural destabilization of KREP rather than directly altering the putative binding region (shallow pocket).

Discussion

Combining evolutionary and tertiary structure information provides a powerful and efficient way to gain insight into the functionality of protein sites⁴⁵. People usually search for amino acid sites that have evolved more rapidly than expected under a neutral model and interpret them as a signature of adaptive evolution corresponding to a gain of new function(s)^{46,47}. Here, we focused on the most slowly evolving, patches of amino acid sites in 3D structure to identify sub-regions of K13 that are likely to play a function conserved during the long time-period of *Apicomplexa* evolution. Detecting highly conserved sites in extremely conserved genes, such as *pfk13*, through population genetics would require a very large sampling of gene sequences⁴⁸, which has not yet been reached for *pfk13*⁹.

In the context of a sustained and intense purifying selection operating in all annotated domains of K13, our inter-species analysis of K13 sequence evolution coupled to the BTB-KREP tertiary structure of PfK13 identified two patches of particularly slowly evolving sites. The most striking conserved patch is found in the K13 KREP domain and is formed by specific positions from inter- and intra-blade loops. This conserved patch largely overlaps with a solvent-exposed, shallow pocket located at the bottom face of the KREP structure (opposite to the face with the central channel entry). Extreme conservation of amino acid sites could be due to constraints linked to the protein structure – for proper folding and stability – or function⁴⁵. Several lines of evidence suggest that the extremely conserved shallow pocket of K13 contains functional sites. First, the pocket is highly enriched in conserved positions, whereas solvent-exposed sites in proteins are usually evolutionarily less conserved than buried ones⁴⁵. Second, a similar shallow pocket is found in the KREP domain of well-characterized BTB-Kelch proteins (KEAP1, KLHL2, KLHL3 and KLHL12) which act as E3 ubiquitin ligase substrate adaptors. In those proteins structurally related to K13, the pocket directly mediates KREP's binding activities^{24,41,42} and was also predicted as a conserved patch in our analyses. The KREP shallow pockets of K13 and other BTB-Kelch were found to share interesting properties: they display a differential electrostatic surface potential, as compared to the rest of the KREP bottom face, and they are rich in highly conserved arginine and serine residues (PfK13 R529, R597, S576,

S577 and S623; Table 3). In KEAP1, KLHL2 and KLHL3, the corresponding conserved arginine and serine residues directly bind acidic peptides derived from their substrates, the transcription factor Nrf2 and kinases WNK respectively^{41,42}. The electrostatic surface potential of the shallow pocket of K13 is slightly positive, therefore the K13 substrate molecule(s) may also harbor an electronegative binding motif. Altogether, these results indicate that the shallow pocket of K13 KREP exhibits several properties of a conserved, binding surface and we speculate that it may be critical for the recognition of K13 substrate molecule(s). We propose that conserved arginine and serine residues are putative candidate binding sites in the shallow pocket of Pfk13, a hypothesis that could be tested by biochemical experiments using recombinant K13 KREP domains. In *P. falciparum*, PfPI3K is a likely partner candidate as it is immunoprecipitated with full-length Pfk13, and its ubiquitination and proteasomal degradation are altered by the Pfk13 C580Y mutation on the KREP domain¹⁷. Another candidate may be the PK4 kinase which phosphorylates eIF2 α , a key mediator of translation-mediated latency involved in ART-R¹⁵.

The *pfk13*-mediated ART-R mechanism in *P. falciparum* is related to the binding activity of Pfk13¹⁷, presumably through the KREP domain since ART-R mutations cluster in this Pfk13 domain^{5,7,8}. In KEAP1, cancer-related missense mutations in the KREP domain were reported to be pathogenic through diverse molecular effects such as reduced binding of Nrf2 (mainly mutations at Nrf2 binding sites in the shallow pocket) and decreased protein stability or abundance (mutations outside the shallow pocket)^{49–51}. In Pfk13, it is interesting to note that the 28 validated or candidate *pfk13* ART-R alleles are largely distributed across the Pfk13 KREP structure and are all located outside the shallow pocket (Fig. 6a): some validated ART-R alleles have a preferential localization at positions proximal to the A and B strands (F446I, Y493H, R539T, I543T, C580Y)⁹ and some other candidates are found in the underlayer of the pocket. In addition, polymorphisms which have an uncharacterized phenotype were reported at 9 out of the 19 positions forming the shallow pocket, but found in only one or two parasite samples from very large population surveys (Table 3). Therefore, one may speculate that amino acid changes at the shallow pocket positions, that we predict to be directly involved in substrate binding, are too functionally damaging to provide a long-term competitive advantage. Rather, many *pfk13* ART-R alleles may alter other properties of Pfk13 such as its abundance, for example through altered protein synthesis or folding or stability. In this scenario, a lesser abundance of mutant – but still functional – Pfk13 protein would result in decreased Pfk13 cellular activity. In addition to the location of ART-R mutations mentioned above, several data support this *stability* scenario. First, the C580Y ART-R mutation leads to an approximately 2-fold lower abundance of Pfk13 protein, as measured by quantitative dimethyl-based proteomics analysis of several strains derived from clinical ART-resistant Cambodian parasites¹⁹ (note that a different result was reported for two African isogenic strains using western blot analysis²⁰). Second, several mutations in the KREP domain of the BTB-Kelch protein Gigaxonin, which cause giant axonal neuropathy in humans^{52,53}, were shown to decrease Gigaxonin protein abundance through altered stability, as evidenced by pulse chase experiments⁵³. Interestingly, the Gigaxonin mutation C464Y, which is located in the KREP central channel such as C580Y in Pfk13, was identified in a compound heterozygous patient and was also associated with a decreased Gigaxonin protein abundance. Third, molecular dynamics simulations on the BTB-Kelch KLHL3 protein revealed that some pathogenic mutations localized in the central channel of KREP (such as C580Y in Pfk13), have no effect on KREP-ligand interaction⁵⁴. And finally, the molecular dynamics simulations we carried out on Pfk13 R539T and C580Y ART-R KREP structures (at a temperature of 300 K during 100 ns) showed that R539T and C580Y mutations cause structural destabilization on blades IV, V and VI, at amino acid positions that are not located at the shallow pocket.

BTB appeared here as the most conserved domain of K13 over *Apicomplexa* evolution, and therefore likely carries critical activities. One surprising result of our study is that it most resembles the BTB domain of the KCTD protein family in primary sequence, tertiary structure and short domain size. KCTD proteins are often found in E3 ubiquitin ligase complexes, such as BTB-Kelch proteins, but lack a KREP domain²⁷. The shortened BTB domain of KCTDs can mediate protein oligomerization^{26,27,55}, consistent with the dimer observed in the solved BTB-KREP crystal structures of Pfk13 (PDB codes 4YY8 and 4ZGC)^{38,56}. The BTB domain of K13 harbors a predicted, functional patch – located at the B2-B3 and B4-A4 loops and at the A4 helix – that overlaps with the patch we found on the BTB domain of KCTDs. In KCTDs, these sites make BTB-BTB contacts in tetrameric or pentameric assemblies when BTB is solved as an isolated domain²⁷. However, the Pfk13 BTB-KREP structure forms a dimer and none of the highly conserved K13 BTB positions makes BTB-BTB contacts (PDB codes 4YY8 and 4ZGC)^{38,56}. In several solved BTB complexes, amino acids of the B4-A4 loop (corresponding to positions 397–399 in Pfk13) are exposed at the BTB-Cullin binding interface²⁶. These discrepancies in the role of the predicted BTB conserved patch could be due to the fact that additional domains (such as KREP) or partner proteins might constraint the folding of the BTB domain into oligomers or complexes. Altogether, data from the literature however support that the BTB predicted patch of K13 may mediate protein-protein interactions, possibly with a Cullin protein.

Interestingly, K13 also contains a highly conserved CCC domain which therefore likely carries critical activities. Consistent with this hypothesis, *pfk13* R239Q, E252Q and D281V alleles located in CCC confer ART-R, although at a moderate level compared to KREP mutations^{3,7,57}. Coiled-coils are ubiquitous protein-protein interaction domains composed of two or more α -helices coiling together⁵⁸. A CCC domain was reported in a few KREP-containing proteins (including some KCTDs) involved in cell morphogenesis but these CCC have a different domain organization than the one of K13³⁹. The CCC of K13 may participate in K13 oligomerization and/or serve as a binding interface with other molecules⁵⁹.

In conclusion, by comparative structural and evolutionary analyses, we identified the shallow pocket of the K13 KREP domain as a likely candidate surface for binding substrate molecule(s). Importantly, we observed that C580Y and R539T ART-R mutations cause local structural destabilization of the KREP structure rather than directly altering the shallow pocket. We also detected in the K13 BTB domain a conserved patch of sites that are involved in protein-protein interactions in known BTB-Cullin and BTB-BTB complexes. Efforts should now

focus on the validation of the binding properties of the K13 KREP shallow pocket and identify its binding partner(s). This may help to clarify the structure-function relationship in K13.

Materials and Methods

Collection of *k13* orthologous sequences from genomic databases. The amino acid sequence of PfK13 (PlasmoDB code PF3D7_1343700) was queried against the specialized eukaryotic pathogen database EuPathDB (release 33)⁶⁰ and the NCBI non-redundant protein database using blastp and tblastn searches (BLOSUM62 scoring matrix)⁶¹. A protein was considered as a likely orthologous sequence if the sequence identity was $\geq 30\%$ and the e-value below the 10^{-3} cutoff. Forty-three K13 sequences – and corresponding *k13* cDNA sequences – were retrieved from distinct *Apicomplexa* species including 21 *Plasmodium* species. A detailed bioinformatics analysis was performed on each protein sequence to confirm the presence of the three annotated domains of K13 (CCC, BTB and KREP) using InterPro⁶².

***k13* codon sequence alignment.** Considering the greater divergence of coding nucleotide sequences as compared to protein sequences due to the genetic code redundancy, a K13 protein sequence alignment was first generated using MAFFT version 7 (E-INS-I strategy with BLOSUM62 scoring matrix, gap opening penalty 2.0 and offset 0.1)⁶³. The output alignment was visually inspected and manually edited with BioEdit v7.2.5⁶⁴. The positions containing gaps in at least 30% of all sequences were removed for further evolutionary analyses, as suggested by PAML' author³⁴. Then, the *k13* nucleotide sequence alignment was generated with PAL2NAL⁶⁵ using the cleaned K13 amino acid alignment as template.

Phylogenetic analysis of *k13*. The phylogenetic relationships of *k13* nucleotide sequences were inferred using the maximum-likelihood method implemented in PhyML v3.0⁶⁶, after determining the best-fitting nucleotide substitution model using the Smart Model Selection (SMS) package⁶⁷. A general time-reversible model with optimized equilibrium frequencies, gamma distributed among-site rate variation and estimated proportion of invariable sites (GTR + G + I) was used, as selected by the Akaike Information Criterion. The nearest neighbor interchange approach was chosen for tree improving, and branch supports were estimated using the approximate likelihood ratio aLRT SH-like method⁶⁸.

Molecular evolutionary analysis of *k13*. To investigate the evolutionary regime that has shaped the *k13* protein-coding DNA sequence during species evolution, we analyzed the non-synonymous (d_N) to synonymous (d_S) substitution rate ratio ω ($=d_N/d_S$), estimated by maximum-likelihood using the codeml tool from PAML v.4.8^{34,69}. ω provides a sensitive measure of selective pressure at the amino acid level by comparing substitution rates with statistical distribution and considering the phylogenetic tree topology. Typically, $\omega < 1$ indicates purifying selection, while $\omega = 1$ and $\omega > 1$ indicate neutral evolution and positive selection respectively.

The variation of ω among lineages and codon sites was evaluated using codon models free-ratio (FR), M0, M1a, M2a, M3, M7 and M8^{70,71} and statistically compared using likelihood ratio tests (LRTs)⁷². The details of each model and PAML interpretations can be found in Supplementary Methods S1.

In addition to *k13*, four other five- or six-bladed KREP protein-coding sequences were considered to compare their ω with those estimated for the whole *Plasmodium* proteome. We used the ω values previously estimated by Jeffares *et al.*³⁵ with PAML under the one-ratio model for each of the 3,256 orthologous protein-coding genes among six *Plasmodium* species: *P. falciparum*, *P. berghei*, *P. chabaudi*, *P. vivax*, *P. yoelii* and *P. knowlesi*. A full description of the procedure is presented in the original paper³⁵.

Inferring site-specific substitution rates considering their spatial correlation in the K13 BTB-KREP tertiary structure.

Most methods – including PAML – assume that site-specific substitution rates are independently distributed across sites⁷³. However, it is widely acknowledged that amino acids located close to each other in protein tertiary structures are more likely to carry out similar functions, suggesting a site interdependence in amino acid sequence evolution attributed to tertiary structure^{73,74}. Consequently, the substitution rate at the protein level (named λ in this study) was inferred using the FuncPatch server³³. FuncPatch requires an amino acid sequence alignment, a phylogenetic tree and a protein tertiary structure to estimate the conservation level during species evolution and the characteristic length scale (in Å) of spatially correlated site-specific substitution rates λ . We used the X-ray structure at 1.81 Å resolution of WT PfK13 BTB-KREP as the reference structure which does not contain the conserved CCC domain (PDB code 4YY8, chain A)³⁸. Beforehand, a Ramachandran analysis was performed to validate the quality of the structure using MolProbity⁷⁵: 96.9% and 3.1% of the amino acids were in favored and allowed regions, respectively, and there were no outliers. FuncPatch only accepts monomeric proteins as input whereas BTB-KREP of PfK13 dimerizes in crystal structure. To take into account the dimeric organization of PfK13, its tertiary structure was edited using customized python scripts (Python v2.7.13) in order to merge the two monomers (chains A and B) and the K13 sequence was duplicated in the K13 protein sequence alignment. The analysis was also done using either one of the other monomeric BTB-KREP tertiary structure and also using a disulfide-bonded version of PfK13 BTB-KREP (PDB code 4ZGC, chain A)⁵⁶. All these control analyses yielded similar results (data not shown) as the one presented here. The spatial correlation of the site-specific substitution rates λ in the K13 tertiary structure was tested using a Bayesian model comparison, where a null model (model 0), in which no spatial correlation of site-specific substitution rates λ is present, is compared to the alternative model (model 1). As suggested by FuncPatch' authors, the spatial correlation was considered as significant if the estimated log Bayes factor (model 1 versus model 0) was larger than 8 in the dataset (conservative cutoff)³³.

Delineation of K13 KREP blades and secondary structures. The KREP domain of PfK13 is composed of six blades (or kelch motifs) having slightly different amino acid lengths. To get an accurate blade alignment at

the primary amino acid sequence level, we first sought to align the six blade structures. The PDB KREP structure was obtained from the Pfk13 BTB-KREP structure (PDB code 4YY8, chain A)³⁸ and was then divided into six parts, each one containing the atomic coordinates of one blade. The six blade structures were then aligned by minimizing the RMSD of atomic positions using the *align* function in PyMOL Molecular Graphics System (Schrödinger, LLC) so as to identify the amino acids from the six blades that are located at exactly the same blade position. This structure alignment was then used to align the six blades at the primary amino acid sequence level. The delineation of the strands and loops was obtained directly from the PDB file (PDB code 4YY8, chain A)³⁸.

Structural evolutionary analysis of the BTB and KREP domains in other BTB- and KREP-containing proteins.

To characterize the BTB domain of K13, we arbitrarily retrieved some members belonging to the main BTB-containing protein families (BTB-ZF, BTB-Kelch, RhoBTB, BTB-NPH3, MATH-BTB, KCTD, KCNA and SKP1 and Elongin C proteins; full list provided in Supplementary Table S5). A multiple protein alignment was generated using MAFFT version 7 (default parameters)⁶³ and was then manually edited with BioEdit v7.2.5⁶⁴ to retain only the region referring to the BTB core fold. The phylogenetic relationships were inferred with the PhyML program⁶⁶ using the best-fitting protein substitution model as determined by the SMS package⁶⁷. The model of K13-Cullin2 complex was obtained by a structural alignment using the *align* function in PyMOL (Schrödinger, LLC) of the K13 BTB domain (PDB code 4YY8)³⁸ with Elongin C using the X-ray structure at 3.2 Å resolution of Elongin C-Cullin2 complex (PDB code 4WQO)³⁷.

For further comparisons with the K13 BTB and KREP domains, site-specific substitution rates λ were inferred with FuncPatch for the BTB and KREP domains of several mammalian KCTD and BTB-Kelch proteins, respectively. In the present study, the proteins were selected on the basis of their sequence homology with K13, the availability of a solved 3D structure, and their known implication in a Cullin-RING E3 ligase complex as suspected for K13. In addition, only well-characterized ligand-binding function and the presence of a six-bladed KREP structure similar to the one of K13 were considered for BTB-Kelch proteins. After a careful review of the literature, we selected two KCTD proteins: SHKBP1 (UniProt code Q8TBC3) which regulates the epidermal growth factor receptor (EGFR) signaling pathway⁷⁶; and KCTD17 (Q8N5Z5) which mediates the ubiquitination and proteasomal degradation of the ciliogenesis down-regulation TCHP protein⁷⁷. Considering BTB-Kelch proteins, we focused on: KEAP1 (Q14145) which interacts with its client protein Nrf2 for the induction of cytoprotective responses to oxidative stress⁴²; KLHL2 (O95198) and KLHL3 (Q9UH77) which both participate in the ubiquitination and degradation of WNK substrates regulating blood pressure⁴¹; and KLHL12 (Q53G59) which negatively regulates the WNT-beta-catenin pathway through the degradation of Dishevelled proteins⁴³. The building of datasets is described in Supplementary Method S2, and the full list of orthologous sequences used for each mammalian KCTD and BTB-Kelch protein is provided in Supplementary Table S6. The phylogenetic relationships were inferred using PhyML⁶⁶ after determining the best-fitting protein substitution model with the SMS package⁶⁷. The 3D structures of KCTD BTB and BTB-Kelch KREP domains were retrieved from the PDBsum database under the following accession numbers: 4CRH for SHKBP1 (resolution: 1.72 Å)²⁷, 5A6R for KCTD17 (resolution: 2.85 Å)²⁷, 2FLU for KEAP1 (resolution: 1.50 Å, in complex with a Nrf2 peptide)⁷⁸, 4CHB for KLHL2 (resolution: 1.56 Å, in complex with a WNK4 peptide)⁴¹, 4CH9 for KLHL3 (resolution: 1.84 Å, in complex with a WNK4 peptide)⁴¹ and 2VPJ for KLHL12 (resolution: 1.85 Å)²⁴. Beforehand, the quality of each structure was validated using MolProbity⁷⁵: none of the structures had amino acids identified as outliers, and approximately 98% of the amino acids of each structure were in favored regions.

Molecular dynamics simulations and evaluation of structural properties.

Molecular dynamics simulations were carried out on the KREP domain of Pfk13 containing the C532-C580 disulfide bond (PDB code 4ZGC, chain A)⁵⁶ using GROMACS v. 5.0.7⁷⁹ for a duration of 100 ns at a temperature of 300 K. Although the C532-C580 disulfide bond is not present in the second resolved structure (PDB code 4YY8)³⁸, the two thiol side-chains face each other in the 4YY8 structure and the C532-C580 disulfide bond is recommended as assessed by the PROPKA program⁸⁰. Because no experimentally mutant Pfk13 structures are yet determined, we generated mutant Pfk13 KREP structures by individually introducing *in silico* the mutations R539T, C580Y or A578S: we used the *swapaa* function of UCSF Chimera to substitute the WT residue with the most probable rotameric conformation of the mutant residue⁸¹. Full parameters and options of the molecular dynamics simulations can be found in Supplementary Method S3.

Amino acid flexibility was investigated by calculating root-mean-square fluctuations (RMSFs) on side-chain residues using the inbuilt *rmsf* function in GROMACS. The number of distinct H-bonds formed within the protein during simulations was counted when the donor-acceptor distance is ≤ 3.5 Å and donor-hydrogen-acceptor angle $\leq 30^\circ$, using the *hbond* function in GROMACS. The electrostatic surface potential of each KREP structure was calculated using the Adaptive Poisson-Boltzmann Solver (APBS) method⁸². Beforehand, the required pqr input files were prepared using PDB2PQR v2.1.1⁸³. The missing charges were added using the *Add Charge* function implemented in UCSF Chimera⁸¹. A grid-based method was used to solve the linearized Poisson-Boltzmann equation at 298 K, with solute (protein) and solvent dielectric constant values fixed at 2 and 78.5, respectively. The contact surface selection was mapped using a radius of 1.4 Å in a scale of $-8 kT/e$ to $+8 kT/e$. The relative solvent accessibility (RSA), which quantifies the extent to which a given residue comes into contact with solvent, was estimated as its surface area accessible to water using the DSSP program⁸⁴, then normalized with the maximal accessible surface area of each amino acid⁸⁵. The side-chain weighted contact number (WCN_{sc}) of each amino acid, which quantifies the extent to which a given residue comes into contact with other residues in the protein (or packing density), was calculated using a customized python script provided by Sydykova and colleagues⁸⁶. Correlative motions of residues (based on C α atoms) and correlation degrees were calculated through dynamical cross-correlation maps (DCCMs). A positive correlation (>0) means the pair of residues move in the same direction; a negative correlation (<0) indicates that the pair of residues move in opposite directions. DCCMs were

produced using the Bio3D library implemented in R⁸⁷. Beforehand, GROMACS trajectories (in.xtc format) were converted into binary trajectories (.dcd) using the *catdcd* function from VMD⁸⁸, as required by Bio3D.

All structural properties were assessed using the aforementioned PDB files.

Structure visualization. All molecular drawings were generated using the UCSF Chimera software⁸¹ or PyMOL (Schrödinger, LLC).

References

- Dondorp, A. M. *et al.* Artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med.* **361**, 455–467 (2009).
- Duru, V. *et al.* *Plasmodium falciparum* dihydroartemisinin-piperaquine failures in Cambodia are associated with mutant K13 parasites presenting high survival rates in novel piperaquine *in vitro* assays: retrospective and prospective investigations. *BMC Med* **13**, 305 (2015).
- Phyo, A. P. *et al.* Declining Efficacy of Artemisinin Combination Therapy Against *P. falciparum* Malaria on the Thai-Myanmar Border (2003–2013): The Role of Parasite Genetic Factors. *Clin. Infect. Dis.* **63**, 784–791 (2016).
- Witkowski, B. *et al.* Novel phenotypic assays for the detection of artemisinin-resistant *Plasmodium falciparum* malaria in Cambodia: *in-vitro* and *ex-vivo* drug-response studies. *Lancet Infect Dis* **13**, 1043–1049 (2013).
- Ariey, F. *et al.* A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50–55 (2014).
- Straimer, J. *et al.* K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science* **347**, 428–431 (2015).
- Anderson, T. J. C. *et al.* Population Parameters Underlying an Ongoing Soft Sweep in Southeast Asian Malaria Parasites. *Mol Biol Evol* **34**, 131–144 (2017).
- Ménard, D. *et al.* A Worldwide Map of *Plasmodium falciparum* K13-Propeller Polymorphisms. *N Engl J Med* **374**, 2453–2464 (2016).
- MalariaGEN *Plasmodium falciparum* Community Project. Genomic epidemiology of artemisinin resistant malaria. *Elife* **5** (2016).
- Cerqueira, G. C. *et al.* Longitudinal genomic surveillance of *Plasmodium falciparum* malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. *Genome Biology* **18**, 78 (2017).
- Bushell, E. *et al.* Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260–272.e8 (2017).
- Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360** (2018).
- Birnbaum, J. *et al.* A genetic system to study *Plasmodium falciparum* protein function. *Nat. Methods* **14**, 450–456 (2017).
- Witkowski, B. *et al.* Reduced artemisinin susceptibility of *Plasmodium falciparum* ring stages in western Cambodia. *Antimicrob. Agents Chemother.* **57**, 914–923 (2013).
- Zhang, M. *et al.* Inhibiting the *Plasmodium* eIF2 α Kinase PK4 Prevents Artemisinin-Induced Latency. *Cell Host & Microbe* **22**, 766–776.e4 (2017).
- Mok, S. *et al.* Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* **347**, 431–435 (2015).
- Mbengue, A. *et al.* A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature* **520**, 683–687 (2015).
- Dogovski, C. *et al.* Targeting the Cell Stress Response of *Plasmodium falciparum* to Overcome Artemisinin Resistance. *PLOS Biology* **13**, e1002132 (2015).
- Siddiqui, G., Srivastava, A., Russell, A. S. & Creek, D. J. Multi-omics Based Identification of Specific Biochemical Changes Associated With PfKelch13-Mutant Artemisinin-Resistant *Plasmodium falciparum*. *J. Infect. Dis.* **215**, 1435–1444 (2017).
- Bhattacharjee, S. *et al.* Remodeling of the malaria parasite and host human red cell by vesicle amplification that induces artemisinin resistance. *Blood* **131**, 1234–1247 (2018).
- Bridgford, J. L. *et al.* Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome. *Nat Commun* **9**, 3801 (2018).
- Tilley, L., Straimer, J., Gnädig, N. F., Ralph, S. A. & Fidock, D. A. Artemisinin Action and Resistance in *Plasmodium falciparum*. *Trends in Parasitology* **32**, 682–696 (2016).
- The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
- Canning, P. *et al.* Structural Basis for Cul3 Protein Assembly with the BTB-Kelch Family of E3 Ubiquitin Ligases. *J. Biol. Chem.* **288**, 7803–7814 (2013).
- Geyer, R., Wee, S., Anderson, S., Yates, J. & Wolf, D. A. BTB/POZ Domain Proteins Are Putative Substrate Adaptors for Cullin 3 Ubiquitin Ligases. *Molecular Cell* **12**, 783–790 (2003).
- Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K. & Privé, G. G. Sequence and structural analysis of BTB domain proteins. *Genome Biology* **6**, R82 (2005).
- Pinkas, D. M. *et al.* Structural complexity in the KCTD family of Cullin3-dependent E3 ubiquitin ligases. *Biochem. J.* **474**, 3747–3761 (2017).
- Dhanoa, B. S., Cogliati, T., Satish, A. G., Bruford, E. A. & Friedman, J. S. Update on the Kelch-like (KLHL) gene family. *Hum Genomics* **7**, 13 (2013).
- Furukawa, M. & Xiong, Y. BTB Protein Keap1 Targets Antioxidant Transcription Factor Nrf2 for Ubiquitination by the Cullin 3-Roc1 Ligase. *Mol Cell Biol* **25**, 162–171 (2005).
- Shibata, S., Zhang, J., Puthumana, J., Stone, K. L. & Lifton, R. P. Kelch-like 3 and Cullin 3 regulate electrolyte homeostasis via ubiquitination and degradation of WNK4. *PNAS* **110**, 7838–7843 (2013).
- Gupta, V. A. & Beggs, A. H. Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skeletal Muscle* **4**, 11 (2014).
- Sidik, S. M. *et al.* A Genome-wide CRISPR Screen in *Toxoplasma* Identifies Essential Apicomplexan Genes. *Cell* **166**, 1423–1435.e12 (2016).
- Huang, Y.-F. & Golding, G. B. FuncPatch: a web server for the fast Bayesian inference of conserved functional patches in protein 3D structures. *Bioinformatics* **31**, 523–531 (2015).
- Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
- Jeffares, D. C., Tomiczek, B., Sojo, V. & dos Reis, M. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods Mol. Biol.* **1201**, 65–90 (2015).
- Huang, Y.-F. & Golding, G. B. Phylogenetic Gaussian Process Model for the Inference of Functionally Important Regions in Protein Tertiary Structures. *PLOS Computational Biology* **10**, e1003429 (2014).
- Nguyen, H. C., Yang, H., Fribourgh, J. L., Wolfe, L. S. & Xiong, Y. Insights into Cullin-RING E3 ubiquitin ligase recruitment: structure of the VHL-EloBC-Cul2 complex. *Structure* **23**, 441–449 (2015).
- Jiang, D. Q. *et al.* Crystal Structure Analysis of Kelch protein from *Plasmodium falciparum*. Available at, <https://www.rcsb.org/structure/4yy8>.

39. Adams, J., Kelso, R. & Cooley, L. The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol.* **10**, 17–24 (2000).
40. Prag, S. & Adams, J. C. Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. *BMC Bioinformatics* **4**, 42 (2003).
41. Schumacher, F.-R., Sorrell, F. J., Alessi, D. R., Bullock, A. N. & Kurz, T. Structural and biochemical characterization of the KLHL3–WNK kinase interaction important in blood pressure regulation. *Biochemical Journal* **460**, 237–246 (2014).
42. Canning, P., Sorrell, F. J. & Bullock, A. N. Structural basis of Keap1 interactions with Nrf2. *Free Radic Biol Med* **88**, 101–107 (2015).
43. Angers, S. *et al.* The KLHL12–Cullin-3 ubiquitin ligase negatively regulates the Wnt– β -catenin pathway by targeting Dishevelled for degradation. *Nature Cell Biology* **8**, 348–357 (2006).
44. WHO | Status report on artemisinin resistance and ACT efficacy. WHO Available at, <http://www.who.int/malaria/publications/atoz/artemisinin-resistance-august2018/en/> (August 2018).
45. Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109–121 (2016).
46. Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449 (2000).
47. Sawyer, S. L., Wu, L. I., Emerman, M. & Malik, H. S. Positive selection of primate TRIM5 α identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. USA* **102**, 2832–2837 (2005).
48. Ogden, K. K. *et al.* Molecular Mechanism of Disease-Associated Mutations in the Pre-M1 Helix of NMDA Receptors and Potential Rescue Pharmacology. *PLoS Genet* **13** (2017).
49. Khan, H. *et al.* Molecular effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1. *Biochem. J.* **467**, 141–151 (2015).
50. Hast, B. E. *et al.* Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination. *Cancer Res* **74**, 808–817 (2014).
51. Padmanabhan, B. *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol. Cell* **21**, 689–700 (2006).
52. Li, X., Zhang, D., Hannink, M. & Beamer, L. J. Crystal structure of the Kelch domain of human Keap1. *J. Biol. Chem.* **279**, 54750–54758 (2004).
53. Boizot, A. *et al.* The instability of the BTB-KELCH protein Gigaxonin causes Giant Axonal Neuropathy and constitutes a new penetrant and specific diagnostic test. *Acta Neuropathol Commun* **2**, 47 (2014).
54. Wang, L., Jiang, C., Cai, R., Chen, X.-Z. & Peng, J.-B. Unveiling the Distinct Mechanisms by which Disease-Causing Mutations in the Kelch Domain of KLHL3 Disrupt the Interaction with the Acidic Motif of WNK4 through Molecular Dynamics Simulation. *Biochemistry* **58**, 2105–2115 (2019).
55. Correale, S. *et al.* Molecular organization of the cullin E3 ligase adaptor KCTD11. *Biochimie* **93**, 715–724 (2011).
56. Jiang, D. Q. *et al.* Crystal structure of kelch protein with disulfide bond from *Plasmodium falciparum*. Available at, <https://www.rcsb.org/structure/4zgc>.
57. Boullé, M. *et al.* Artemisinin-Resistant *Plasmodium falciparum* K13 Mutant Alleles, Thailand–Myanmar Border. *Emerg Infect Dis* **22**, 1503–1505 (2016).
58. Hartmann, M. D. Functional and Structural Roles of Coiled Coils. *Subcell. Biochem.* **82**, 63–93 (2017).
59. de Moura, T. R. *et al.* Prp19/Pso4 Is an Autoinhibited Ubiquitin Ligase Activated by Stepwise Assembly of Three Splicing Factors. *Mol. Cell* **69**, 979–992.e6 (2018).
60. Aurecochea, C. *et al.* EuPathDB: the eukaryotic pathogen genomics database resource. *Nucleic Acids Res* **45**, D581–D591 (2017).
61. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
62. Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**, D211–D215 (2009).
63. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* **30**, 772–780 (2013).
64. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98 (1999).
65. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609–W612 (2006).
66. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
67. Lefort, V., Longueville, J.-E. & Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* **34**, 2422–2424 (2017).
68. Anisimova, M., Bielawski, J. P. & Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**, 1585–1592 (2001).
69. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
70. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998).
71. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573 (1998).
72. Vuong, Q. H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **57**, 307–333 (1989).
73. Rodrigue, N., Lartillot, N., Bryant, D. & Philippe, H. Site interdependence attributed to tertiary structure in amino acid sequence evolution. *Gene* **347**, 207–217 (2005).
74. Robinson, D. M., Jones, D. T., Kishino, H., Goldman, N. & Thorne, J. L. Protein Evolution with Dependence Among Codons Due to Tertiary Structure. *Mol Biol Evol* **20**, 1692–1704 (2003).
75. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).
76. Feng, L., Wang, J.-T., Jin, H., Qian, K. & Geng, J.-G. SH3KBP1-binding protein 1 prevents epidermal growth factor receptor degradation by the interruption of c-Cbl–CIN85 complex. *Cell Biochem. Funct.* **29**, 589–596 (2011).
77. Kasahara, K. *et al.* Ubiquitin-proteasome system controls ciliogenesis at the initial step of axoneme extension. *Nat Commun* **5**, 5081 (2014).
78. Lo, S.-C., Li, X., Henzl, M. T., Beamer, L. J. & Hannink, M. Structure of the Keap1:Nrf2 interface provides mechanistic insight into Nrf2 signaling. *EMBO J* **25**, 3605–3617 (2006).
79. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
80. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **7**, 525–537 (2011).
81. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
82. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).

83. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–667 (2004).
84. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
85. Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J. & Wilke, C. O. Maximum allowed solvent accessibilities of residues in proteins. *PLoS ONE* **8**, e80635 (2013).
86. Sydykova, D. K., Jack, B. R., Spielman, S. J. & Wilke, C. O. Measuring evolutionary rates of proteins in a structural context. *PLoS Res* **6**, 1845 (2017).
87. Skjærven, L., Yao, X.-Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399 (2014).
88. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**(33–38), 27–28 (1996).
89. Haldar, K., Bhattacharjee, S. & Safeukui, I. Drug resistance in Plasmodium. *Nat. Rev. Microbiol.* **16**, 156–170 (2018).
90. Paloque, L., Ramadani, A. P., Mercereau-Puijalon, O., Augereau, J.-M. & Benoit-Vical, F. Plasmodium falciparum: multifaceted resistance to artemisinins. *Malar. J.* **15**, 149 (2016).
91. de Laurent, Z. R. *et al.* Polymorphisms in the K13 Gene in Plasmodium falciparum from Different Malaria Transmission Areas of Kenya. *Am. J. Trop. Med. Hyg.* **98**, 1360–1366 (2018).
92. Kamau, E. *et al.* K13-Propeller Polymorphisms in Plasmodium falciparum Parasites From Sub-Saharan Africa. *J Infect Dis* **211**, 1352–1355 (2015).

Acknowledgements

We thank Y.-F. Huang and G.B. Golding for help with FuncPatch analysis. We thank O. Mercereau-Puijalon, I. Callebaut and R. Duval for helpful discussions. We thank A. Sissoko and F. Palstra for proofreading the manuscript. We thank F. Arieu for providing system resources for molecular dynamics simulations. R. Coppée is supported by PhD grant funding from the doctoral school MTCI n°563.

Author Contributions

R.C., A.S. and J.C. designed the research. A.S. and J.C. supervised the whole work, and M.A.M. supervised the molecular dynamics simulations. R.C. performed the building of datasets, statistical and evolutionary analyses, structural visualization and molecular dynamics simulations. D.C.J. provided additional data and input on the research design. All the authors participated in interpretations. R.C., A.S. and J.C. wrote the first draft of the manuscript. All the authors contributed for editing the manuscript and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-47034-6>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

3.1.4. Analyses expérimentales de la poche du domaine KREP de Pfk13

Des expériences *in vitro* sont nécessaires pour tester nos hypothèses sur le rôle fonctionnel de la poche du domaine KREP de Pfk13. Deux approches ont été initiées : le *pull-down* pour valider la capacité de liaison de la poche avec des protéines parasites ; et le *docking* moléculaire pour identifier de petites molécules capables de se lier *in silico* à la poche et tester leur effet sur la croissance parasitaire *in vitro*.

3.1.4.1. Analyses de *pull-down*

3.1.4.1.1. Expression des domaines recombinants Pfk13

Suite à la construction des plasmides et leur transformation dans la souche *E. coli* BL21-CodonPlus, différentes conditions ont été testées afin de maximiser l'expression des domaines recombinants (variation de la concentration de l'inducteur IPTG et de la température et du temps d'induction). L'analyse SDS-PAGE a montré que les domaines recombinants étaient systématiquement exprimés après induction dans les différentes conditions testées, et à des taux similaires : nous avons détecté une bande spécifiquement surexprimée dans chaque fraction induite (F_i) à un poids moléculaire correspondant à ceux attendus pour chacune des trois constructions (**Figure 36A à C**). Toutefois, une induction à plus faible température est reconnue pour favoriser le repliement des protéines et leur solubilisation lors de la lyse²⁶⁴. Par précaution, l'ensemble des analyses ultérieures est basé sur des domaines recombinants induits à une concentration en IPTG égale à 0,3 mM à une température de 18°C.

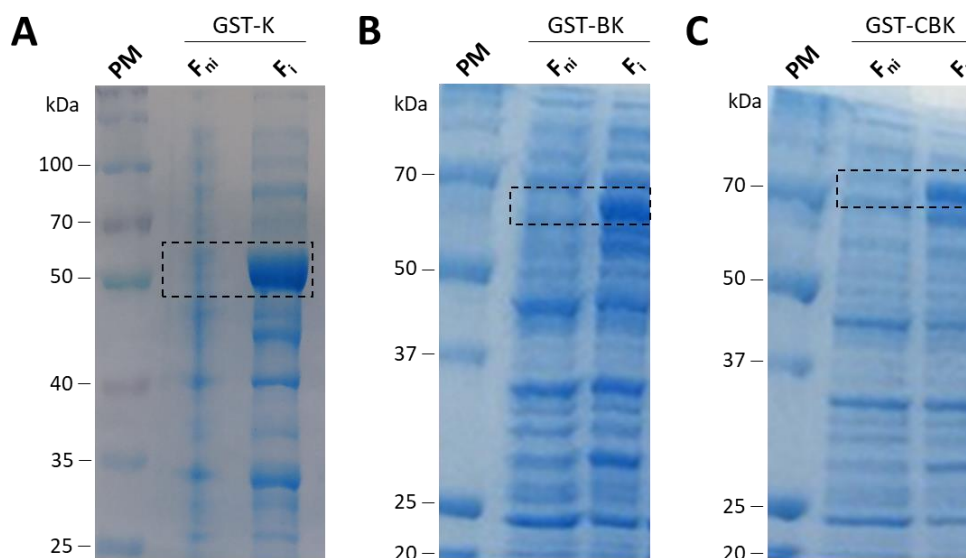


Figure 36 : Analyse sur gel SDS-PAGE 10% de l'expression des domaines recombinants GST-K (A), GST-BK (B) et GST-CBK (C). L'expression bactérienne présentée ici a été réalisée à 18°C durant 14 heures avec une concentration en IPTG égale à 0,3 mM. PM : marqueur de poids moléculaire ; F_{ni} : fraction non induite, F_i : fraction induite. Les zones encadrées correspondent aux poids moléculaires attendus des constructions GST-K (52 kDa), GST-BK (61 kDa) et GST-CBK (75 kDa). La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

3.1.4.1.2. Solubilisation des domaines recombinants PfK13

Des difficultés ont été rencontrées dans la solubilisation des protéines de fusion. Les différents tests présentés ci-après ont été uniquement conduits pour la construction GST-K. En utilisant un tampon à base de Triton X-100 (concentration finale : 0,1%) communément utilisé pour lyser les bactéries et solubiliser leurs protéines (TLb1 ; **Encadré 7**) couplé à une sonication durant une minute (six cycles, 10 secondes de pulse – 10 secondes de repos), les domaines recombinants se sont avérés totalement insolubles (**Figure 37**). Une autre durée de sonication a été testée (trois cycles, 5 secondes de pulse – 15 secondes de repos), mais elle n'a pas permis d'améliorer la solubilité du domaine recombinant GST-K.

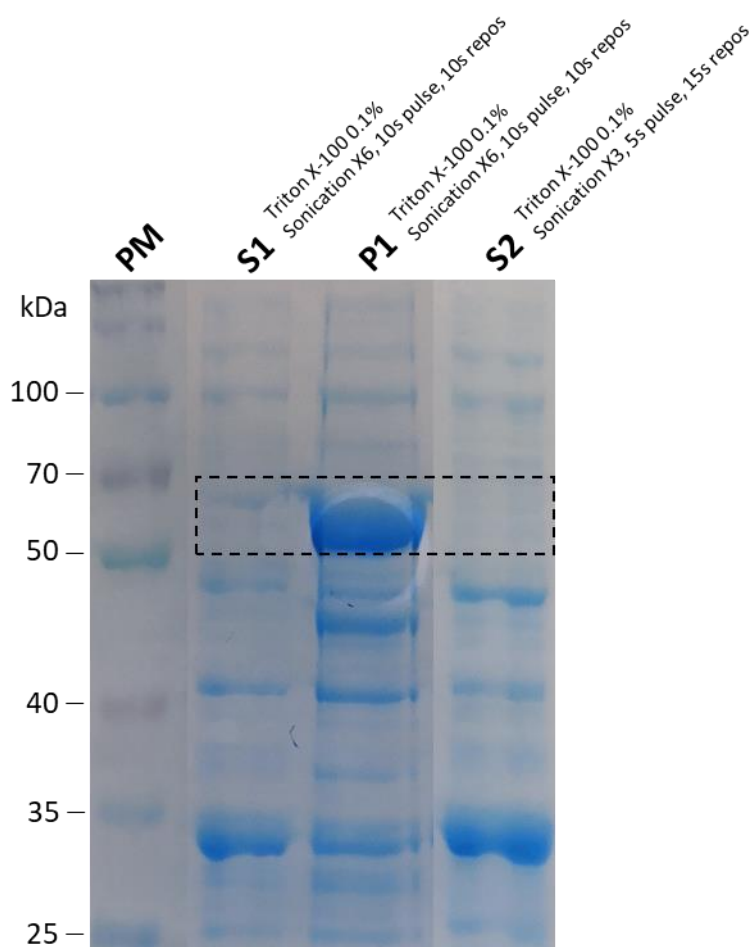


Figure 37 : Analyse sur gel SDS-PAGE 10% de la solubilité du domaine recombinant GST-K. Un aliquot de protéines solubles (S1 et S2) et du culot insoluble resuspendu (P1) ont été déposés sur gel après sonication et centrifugation du lysat bactérien. PM : marqueur de poids moléculaire. La zone encadrée correspond au poids moléculaire attendu de la construction GST-K (52 kDa). La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

Nous avons conclu que les difficultés de solubilisation rencontrées ne provenaient probablement pas de la sonication mais du tampon de lyse utilisé. Par conséquent, des tampons de lyse incluant différents détergents ont été préparés. La base du tampon est la suivante : Tris-HCl 10 mM, pH 8 ; NaCl 150 mM ; EDTA 1 mM ; lysosyme 100 µg/mL ; DTT 5 mM ; Complete™

Protease Inhibitor Cocktail (Roche) 1X. Chacun des détergents suivants a été ajouté séparément à une concentration finale de 1% : Triton X-100, NP-40, CHAPS, Tween-20 et DDM (le Triton X-100 a également été testé à une concentration de 0,1%). Malgré les différents types de détergents (non ioniques : Triton X-100, NP-40, Tween-20 et DDM ; zwitterionique : CHAPS), le domaine recombinant GST-K est demeuré totalement insoluble (**Figure 38**).

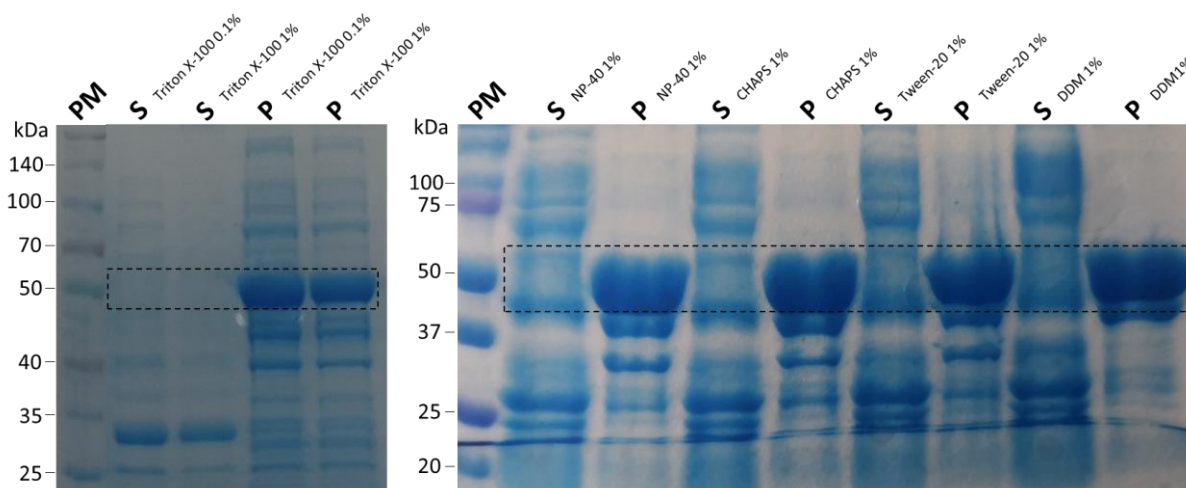


Figure 38 : Analyse sur gel SDS-PAGE 10% de la solubilité du domaine recombinant GST-K avec différents détergents. Après sonication et centrifugation du lysat bactérien, un aliquot de protéines solubles (S) et du culot insoluble resuspendu (P) ont été déposés sur gel. La solubilisation a été réalisée en utilisant différents types de détergent : Triton X-100 (concentration finale à 0,1 ou 1%), NP-40 (1%), CHAPS (1%), Tween-20 (1%) et DDM (1%). Un volume équivalent a été déposé dans les pistes solubles et non solubles. PM : marqueur de poids moléculaire. La zone encadrée correspond au poids moléculaire attendu de la construction GST-K (52 kDa). La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

Après une lecture approfondie de la littérature, il apparaît que certaines protéines recombinantes avec étiquette GST exprimées dans un plasmide pGEX peuvent être totalement insolubles²⁶⁵. Si l'urée permet de solubiliser des protéines jusqu'alors insolubles, elle présente un effet dénaturant important²⁶⁶. Une solution alternative consiste à utiliser le sarkosyl (ou sodium lauroyl sarcosinate), un détergent anionique très amphiphile en raison d'une longue chaîne hydrophobe composée de 12 carbones et d'une tête carboxylate hydrophile (**Figure 39**). Bien que son mécanisme d'action demeure mal compris, il est supposé que les molécules de sarkosyl encapsulent les protéines et empêchent la formation d'agrégation^{267,268}. Toutefois, le sarkosyl est rapporté comme perturbateur de l'interaction entre la GST et la résine agarose-GSH. Par conséquent, suite à la solubilisation des domaines recombinants et avant d'ajouter la résine, il est conseillé d'ajouter du Triton X-100. Ce détergent est supposé former des structures sous forme de micelles pour capturer les molécules de sarkosyl en plus de favoriser le bon repliement des protéines²⁶⁷. Il a été estimé que pour une concentration en sarkosyl de 0,5%, il faut rajouter 2% de Triton X-100 pour supprimer toutes les molécules de sarkosyl^{267,268}. Dans le cas de protéines très insolubles, les concentrations en sarkosyl et en Triton X-100 peuvent atteindre 1,5% et 4%, respectivement²⁶⁸. Ne sachant pas l'effet du sarkosyl sur le niveau de solubilisation du domaine recombinant GST-K, deux tampons de lyse ont été préparés en présence de sarkosyl

concentré à 0,5% et 1,5%. Après incubation et sonication (six cycles, 10 secondes de pulse – 10 secondes de repos), nous avons observé que la solubilisation du domaine recombinant GST-K augmentait avec la concentration en sarkosyl (**Figure 40**). À titre de contrôle et de comparaison, le même protocole en remplaçant le sarkosyl par du Triton X-100 à 1% final n'a pas permis de solubiliser la construction GST-K (**Figure 40**). Dans la mesure où il est conseillé de limiter la quantité de sarkosyl et que son effet est conservé lors de productions en grand volume²⁶⁷, nous avons décidé de conserver la condition contenant la plus faible concentration en sarkosyl (0,5%, soit le tampon de lyse TLb8 ; **Encadré 7**).

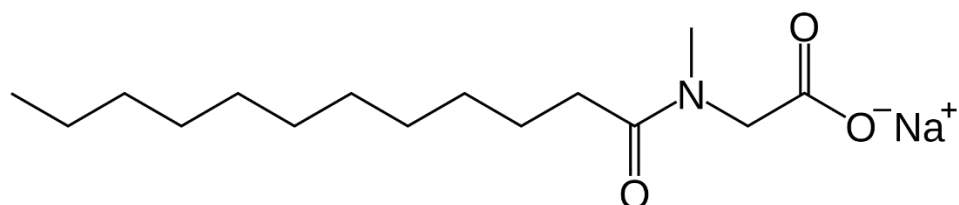


Figure 39 : Structure chimique du sodium lauroyl sarcosinate (sarkosyl). Le sarkosyl est très amphiphile en raison d'une longue chaîne hydrophobe composée de 12 carbones et d'une tête carboxylate hydrophile.

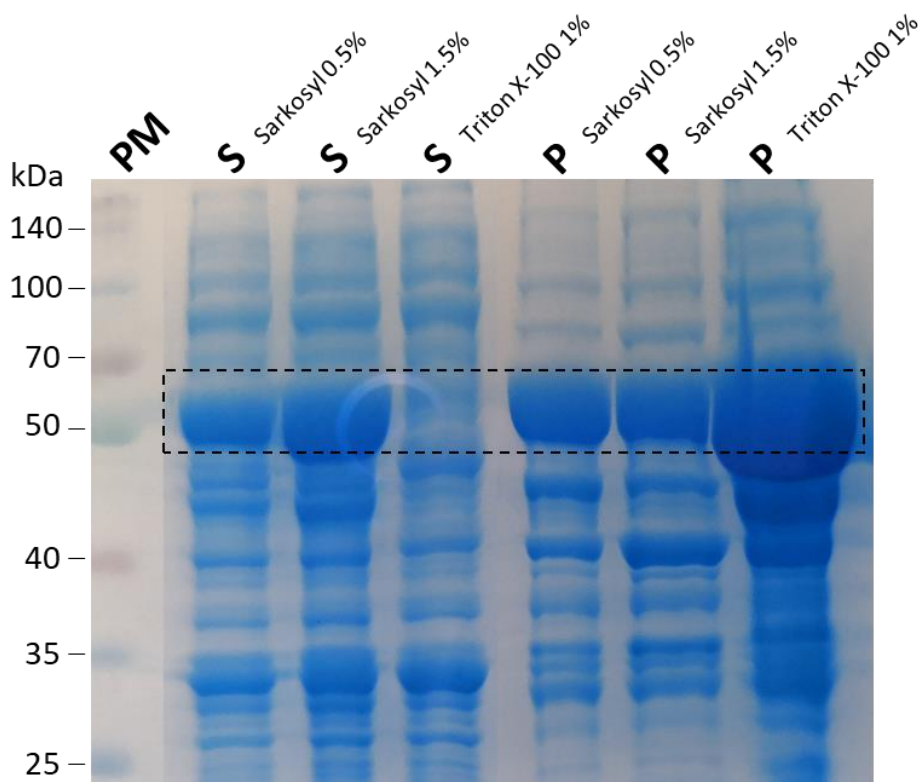


Figure 40 : Analyse sur gel SDS-PAGE 10% de la solubilité du domaine recombinant GST-K à l'aide du sarkosyl. Après sonication et centrifugation du lysat bactérien, un aliquot de protéines solubles (S) et du culot insoluble resuspendu (P) ont été déposés sur gel. La solubilisation des protéines bactériennes a été réalisée en utilisant différents types de détergent : Triton X-100 (concentration finale : 1%) et sarkosyl (0,5 ou 1,5%). Un volume équivalent a été déposé dans les pistes solubles et non solubles. PM : marqueur de poids moléculaire. La zone encadrée correspond au poids moléculaire attendu de la construction GST-K (52 kDa). La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

3.1.4.1.3. Purification des domaines recombinants PfK13

L'étape suivante a consisté à purifier les différents domaines recombinants PfK13 avec la résine agarose-GSH, puis de tester leur élution. Après incubation des domaines recombinants avec la résine, nous avons observé qu'une fraction ne s'était pas fixée à celle-ci (**Figure 41** ; exemple pour la construction GST-K). Cela suggère un possible repliement incorrect de la GST et/ou des domaines recombinants empêchant l'interaction GSH-GST, ou que le sarkosyl n'a pas été correctement éliminé, ou un excès de la quantité de domaines recombinants par rapport à la résine. À noter que dans les expérimentations qui suivront, nous n'avons pas réussi à augmenter la fraction fixée à la résine. Après trois étapes de lavage avec du tampon STE (sans détergent) pour supprimer les protéines contaminantes, une élution au glutathion a été effectuée (Tris-HCl 50 mM, pH8 ; NaCl 150 mM ; DTT 5 mM ; Triton X-100 0,1% ; GST réduit 10 mM). Nous avons observé une bande au sein des deux élutions au poids moléculaire attendu (**Figure 41** ; exemple pour la construction GST-K), suggérant qu'une fraction des domaines recombinants purifiés présente un domaine GST correctement replié (ou au moins son site actif). Cette analyse a aussi révélé qu'une fraction de protéines demeure accrochée à la résine après élution. Ce constat s'explique par la durée d'incubation de la résine avec le tampon d'élution : dans l'exemple présenté, une simple agitation à l'aide d'une pipette avait été réalisée. Cette méthode a par la suite été remplacée par une incubation sur roue à 4°C pendant 1 heure, permettant de décrocher la grande majorité des domaines recombinants liés à la résine.

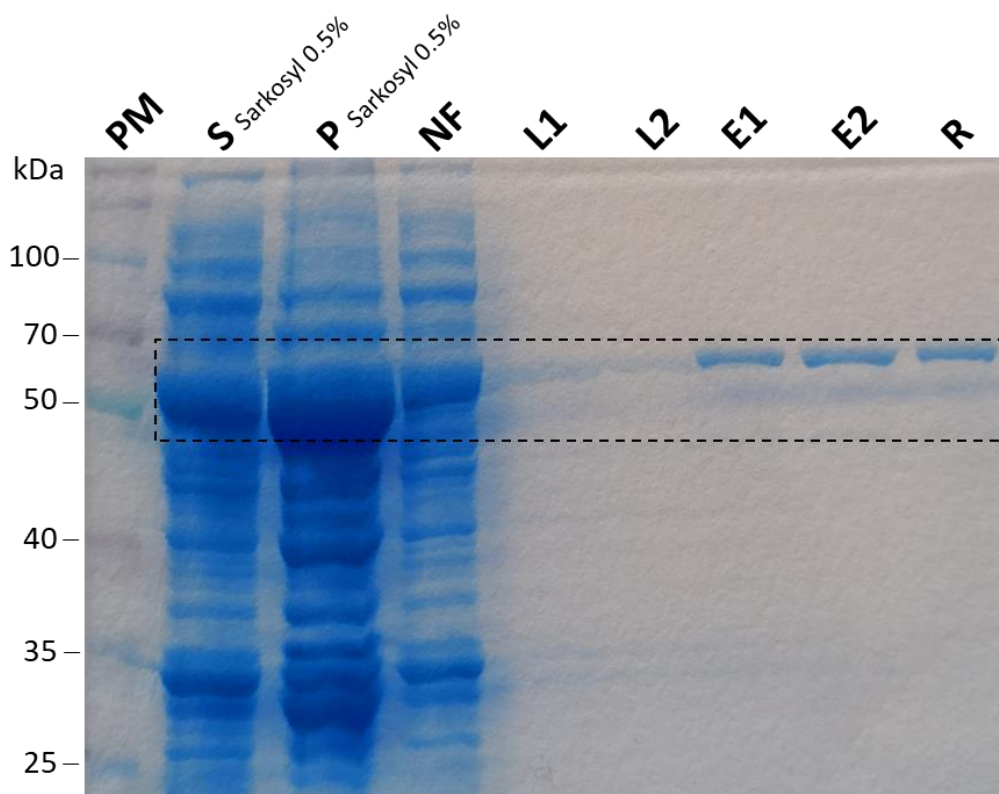


Figure 41 : Analyse sur gel SDS-PAGE 10% de la purification et de l'élution du domaine recombinant GST-K. La solubilisation du domaine recombinant a été réalisée avec le tampon de lyse à base de sarkosyl 0,5%. La résine a été incubée avec la fraction soluble pendant une heure sur roue (15 tours par minute) à

4°C. La résine a ensuite été lavée trois fois à l'aide du tampon STE (un aliquot des deux premiers lavages a été déposé). Une élution par compétition au glutathion a ensuite été réalisée. Un volume équivalent a été déposé dans les différentes pistes. PM : marqueur de poids moléculaire ; S : fraction soluble ; P : fraction insoluble ; NF : fraction protéique du lysat bactérien non fixée à la résine ; L1 et L2 : deux premiers lavages de la résine pour éliminer les protéines contaminantes ; E1 et E2 : fraction protéique obtenue après deux éluions successives avec du GSH. La zone encadrée correspond à la taille de la construction GST-K attendue (52 kDa). La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

3.1.4.1.4. La première analyse de *pull-down*

Suite à ces résultats, un lysat parasitaire (à partir d'une culture de parasites *P. falciparum*) a été préparé afin d'initier une première expérience de *pull-down*. Différents tampons TLp avec des détergents non dénaturants ont été utilisés pour solubiliser les protéines du parasite (**Encadré 8**). Nous n'avons pas observé de différences majeures dans les profils protéiques obtenus après solubilisation et révélés par SDS-PAGE (**Figure 42**), et avons décidé de conserver le tampon TLp1 (Tris-HCl 20 mM, pH 8 ; NaCl 150 mM ; EDTA 1 mM ; Triton X-100 0,1% ; NP-40 1% ; Complete™ Protease Inhibitor Cocktail (Roche) 1X) pour les analyses futures.

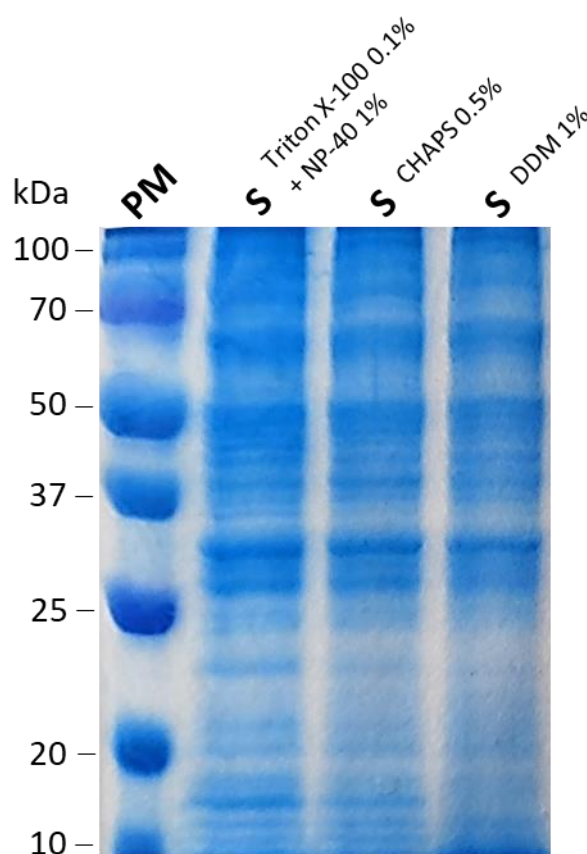


Figure 42 : Solubilisation des protéines de *P. falciparum*. Trois tampons de solubilisation à base de différents détergents ont été testés : Triton X-100 0,1% + NP-40 1% (TLp1) ; CHAPS 0,5% (TLp2) ; DDM 1% (TLp3). De multiples bandes couvrant l'échelle des poids moléculaires (PM) ont été observées, suggérant que de nombreuses protéines du parasite sont solubles. Une quantité similaire de protéines a été incubée avec les différents tampons : après centrifugation, un aliquot du surnageant a été analysé par SDS-PAGE. S : fraction soluble. La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

Avant la réalisation du *pull-down*, les protéines de fusion purifiées et fixées à la résine ont été divisées en deux volumes équivalents et ont subi ou non une étape de blocage à la BSA, permettant en théorie de limiter les interactions aspécifiques. Pour le premier test de *pull-down*, nous nous sommes d'abord concentrés sur les constructions GST (contrôle négatif) et GST-K pour ne pas multiplier les conditions. Une fraction de chaque construction purifiée et toujours fixée à la résine (concentrée en protéines à environ 4 µg/mL) a été incubée avec 4 volumes de lysat parasite (soit un ratio (volume : volume) billes pures : lysat parasite égal à 1 : 10) par rotation sur la nuit à 4°C. Des contrôles négatifs ont été préparés dans lesquelles les mêmes résines ont été incubées avec le tampon de lyse parasite à la place du lysat parasite. Suite aux différents lavages puis à la révélation par SDS-PAGE des protéines associées aux résines agarose-GSH-GST et agarose-GSH-GST-K, une bande majoritaire a été détectée à l'issue du *pull-down*, qui correspond comme attendu à la protéine GST ou GST-K utilisée comme appât (**Figure 43A et B**). Aucune protéine supplémentaire d'origine parasite n'a visuellement été détectée dans la piste GST-K avec lysat parasite par comparaison avec la condition sans lysat. À noter que l'étape de blocage à la BSA a nui à la bonne lisibilité des résultats sur gel, et avons préféré écarter son utilisation par la suite.

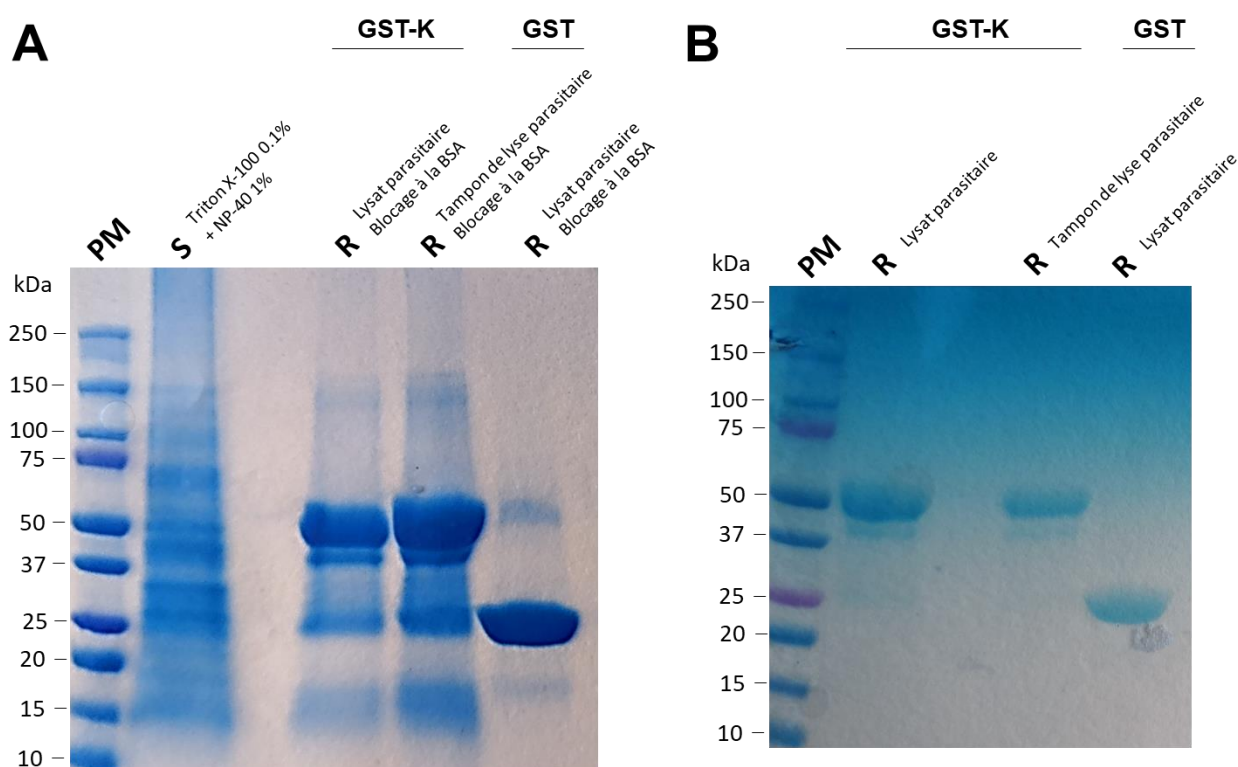


Figure 43 : Pull-down réalisé avec les constructions GST et GST-K. (A) Pull-down réalisé en ayant préalablement incubé la résine avec la BSA. (B) Pull-down réalisé sans phase de blocage à la BSA. PM : marqueur de poids moléculaire ; S : fraction soluble du lysat parasite ; R : résine. La résine (agarose-GSH-GST ou agarose-GSH-GST-K) a été incubée soit avec le lysat parasite, soit avec le tampon de lyse parasite (TLp, contrôle négatif). Un volume équivalent a été déposé dans les différentes pistes. La coloration des gels a été faite au Coomassie (0,25% bleu de Coomassie, complété par de l'éthanol à 96%).

3.1.4.1.5. Le problème des contaminants bactériens après purification

Nous avons souhaité tester une méthode de coloration plus sensible afin de révéler plus en détails les différentes protéines présentes dans les échantillons. Pour cela, la coloration rapide au Coomassie standard (0,25% bleu de Coomassie, complété par de l'éthanol à 96%) que nous utilisions a été remplacée par une coloration plus sensible et compatible avec la spectrométrie de masse (0,08% Coomassie Brilliant Blue G250, 10% d'acide citrique, 8% de sulfate d'ammonium et 20% de méthanol). Un nouveau *pull-down* a été réalisé suivant les mêmes conditions expérimentales, en incluant en outre les constructions GST-BK et GST-CBK. Pour chaque construction, exceptée pour agarose-GSH-GST seule, nous avons visuellement observé bien plus de protéines associées aux résines en fin de *pull-down*, mais aussi bien en présence qu'en absence de lysat parasitaire (**Figure 44**). Si cette nouvelle coloration au Coomassie permet effectivement de détecter un plus grand nombre de protéines, les résultats obtenus suggèrent l'existence d'une multitude de contaminants d'origine bactérienne pour l'ensemble des constructions exceptée GST. Ces contaminants protéiques pourraient potentiellement limiter les interactions entre les domaines PfK13 et leur(s) partenaire(s) en plus d'augmenter la difficulté à identifier des protéines parasitaires lors d'analyses par spectrométrie de masse. Cette nouvelle expérience n'a toutefois pas révélé de protéines supplémentaires d'origine parasitaire après incubation avec le lysat parasitaire par analyse SDS-PAGE. En revanche, une différence majeure fut observée avec la présence d'une bande intense à un poids moléculaire de 25 kDa uniquement dans les pistes avec lysat parasitaire (**Figure 44**). Il s'agit vraisemblablement de la GST, qui aurait été clivée des domaines PfK13, suggérant que les lysats parasitaires sont fonctionnels et que des protéases du parasite ont la capacité de cliver les constructions. À l'avenir, une plus grande concentration en inhibiteurs de protéases sera ajoutée afin de limiter cet effet (que nous n'avions toutefois pas constaté lors du premier *pull-down*), et nous diminuerons la quantité protéique des constructions qui était excessive dans cette expérience.

Dans la mesure où de nombreux contaminants étaient présents après purification des domaines recombinants (sur la base de la coloration sensible), nous avons tenté d'améliorer la qualité de la purification avec quelques tests simples. Une chromatographie d'exclusion a été réalisée par Laurine Rondepierre (doctorante, UMR 261) mais n'a pas permis d'éliminer les protéines bactériennes contaminantes. En parallèle, nous avons augmenté la stringence des lavages avec le tampon STE lors de la purification des domaines recombinants : *i*) en augmentant la concentration en NaCl (1 M) ; ou *ii*) en rajoutant de l'ATP (5 mM) et du MgCl₂ (15 mM), l'ATP servant de source d'énergie pour améliorer la renaturation des protéines tout en détachant les protéines contaminantes²⁶⁹. Cependant, la coloration sensible a de nouveau révélé une multitude de protéines contaminantes (**Figure 45** ; exemple avec la construction GST-CBK). Jusqu'à présent, aucune des conditions testées n'a permis d'éliminer les contaminants ; des tests supplémentaires sont donc nécessaires. Entre autres, il serait intéressant de vérifier si ces contaminants sont aussi présents dans des constructions basées sur un autre tag tel que le tag poly-histidine.

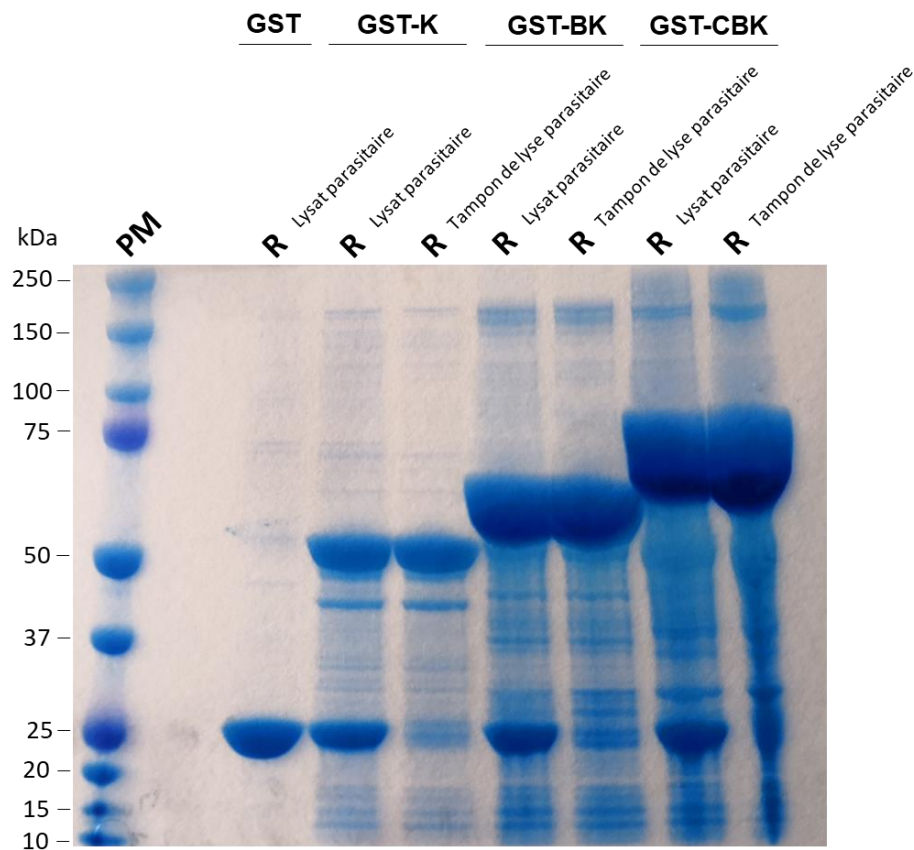


Figure 44 : Pull-down réalisé avec les constructions GST et GST-K/BK/CBK. PM : marqueur de poids moléculaire ; R : résine. La résine (agarose-GSH-GST-K/BK/CBK) a été incubée soit avec le lysat parasiteire, soit avec le tampon de lyse parasiteire (contrôle négatif). La coloration des gels a été faite au Coomassie (0,08% Coomassie Brilliant Blue G250, 10% d'acide citrique, 8% de sulfate d'ammonium et 20% de méthanol).

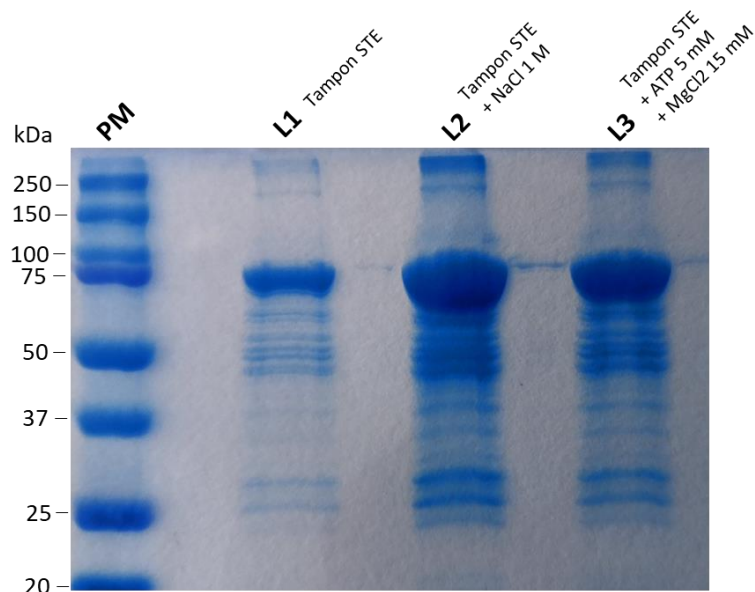


Figure 45 : Tests de suppression des contaminants pour la construction GST-CBK par différents tampons de lavage. L1, L2 et L3 : Fraction de la résine déposée après lavage selon différents tampons. Les différences de quantité dans le matériel protéique s'expliquent par une mauvaise manipulation de la résine. Cela n'a toutefois pas d'incidence sur nos conclusions. La coloration des gels a été faite au Coomassie (0,08% Coomassie Brilliant Blue G250, 10% d'acide citrique, 8% de sulfate d'ammonium et 20% de méthanol).

Puisque nous n'avons pas encore réussi à supprimer une partie des contaminants, aucun échantillon n'a pour l'heure été envoyé en spectrométrie de masse pour analyser le contenu en protéines. En revanche, nos travaux ont permis de mettre au point des conditions d'expression et de purification opérationnelles de différentes constructions Pfk13 chez *E. coli*. Deux axes de travail sont l'élimination des contaminants bactériens et une vérification plus poussée du bon repliement des domaines purifiés. Ces protocoles permettront *in fine* d'étudier l'interactome de Pfk13 par des approches de *pull-down* différentiels avec des domaines sauvage(s) et mutant(s) (mutation de résistance aux ARTDs et mutation non naturelle perte de fonction, c'est-à-dire invalidant des sites candidats d'interaction dans la poche du domaine KREP¹⁹³), mais aussi d'étudier des interactions avec des petites molécules.

3.1.4.2. Recherche d'inhibiteurs de la poche du domaine KREP de Pfk13 ?

Outre les *pull-down* différentiels, nous avons initié une approche pharmacologique pour essayer de trouver des preuves de la fonctionnalité de la poche du domaine KREP de Pfk13. Si le criblage virtuel à partir de milliers de molécules permettrait d'identifier de petites molécules qui auraient la meilleure affinité de liaison avec la poche, nous avons ici opté pour une première approche consistant à récupérer des inhibiteurs déjà connus pour leur liaison non covalente à d'autres domaines KREP que celui de Pfk13, et donc déjà testés expérimentalement. Une veille bibliographique (mots-clés : « inhibitor kelch », « propeller inhibitor », « keap1 inhibitor ») a permis de recenser 30 articles présentant des inhibiteurs de l'interaction de KEAP1 avec son substrat Nrf2¹³¹ au niveau de la poche du domaine KREP (KREP^{KEAP1}) et d'identifier 25 molécules inhibitrices. D'autres molécules ont été écartées car elles inhibent l'interaction entre KEAP1 et son substrat par d'autres mécanismes. Parmi les 25 inhibiteurs considérés, seuls 12 ont été retenus en raison de l'existence d'une structure tertiaire disponible. Les 13 autres inhibiteurs n'ont pas fait l'objet d'analyses supplémentaires, bien qu'il serait intéressant de les modéliser par la suite.

Afin de connaître l'affinité de liaison *in silico* entre la poche du domaine KREP de Pfk13 (KREP^{Pfk13}) et les différents inhibiteurs de l'interaction KREP^{KEAP1}-Nrf2, une série de *docking* a été conduite. À des fins de comparaison, les *docking* ont été réalisés en parallèle sur la poche de KREP^{KEAP1} (contrôle positif) et KREP^{Pfk13}. Trois molécules supplémentaires (des antipaludiques) ont été testées : l'atovaquone (ATV), la chloroquine (CQ) et la pyriméthamine (PYR). Nous avons supposé *a priori* qu'elles n'interagissent pas avec cette poche, servant ainsi de contrôles négatifs. Enfin, pour étudier la spécificité de ces interactions au niveau de la poche, les *docking* ont également été réalisés dans une région *a priori* non pertinente du domaine KREP, choisie aléatoirement en dehors de la poche. La préparation des différents fichiers a été faite avec l'outil Autodock²⁴⁴, tandis que les *docking* ont été produits avec Vina²⁴⁵. Pour chaque *docking*, neuf poses ont été générées, bien que l'on se soit uniquement focalisé ensuite sur la meilleure

(correspondant à la plus haute affinité de liaison protéine-ligand). Un récapitulatif du protocole suivi est présenté en **figure 46**.

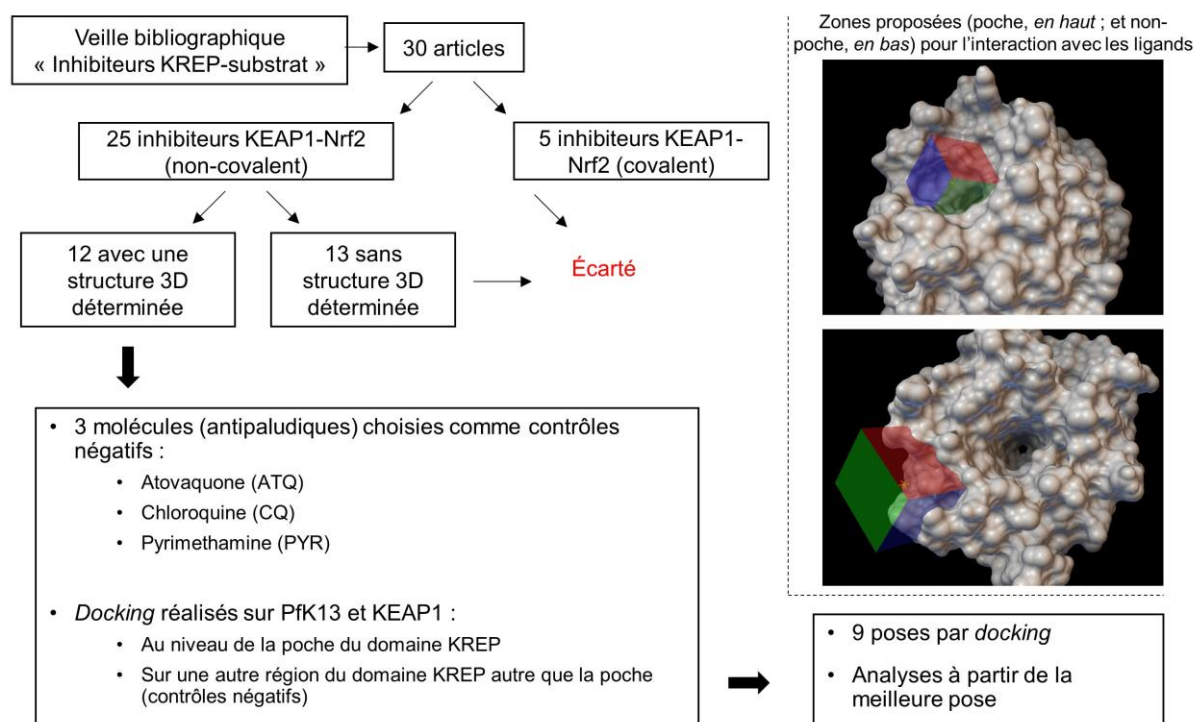


Figure 46 : Recherche d'inhibiteurs de la poche de KREP^{KEAP1} et docking moléculaires sur les poches de KREP^{Pfk13} et KREP^{KEAP1}.

Suite aux différents *docking*, les affinités de liaison des inhibiteurs testés étaient supérieures pour la poche de KREP^{KEAP1} que pour celle de KREP^{Pfk13}, sauf pour l'inhibiteur 4IN4 (**Figure 47**). Les affinités de liaison des inhibiteurs pour la poche de KREP^{Pfk13} étaient globalement fortes, avec trois inhibiteurs ayant une affinité en kcal/mol inférieure à -8 (4IN4, 5CGK et 5FNU ; **Figure 47**). Pour les trois molécules servant de contrôles négatifs, la CQ et la PYR étaient associées à une faible affinité de liaison (> -6 kcal/mol) tandis que l'ATV était associé à une affinité de liaison comprise entre -6 et -8 kcal/mol, quoique globalement inférieure à celles des différents inhibiteurs KREP^{KEAP1} testés (**Figure 47**). Les *docking* réalisés sur une région du domaine KREP autre que la poche pour KEAP1 et Pfk13 ont révélé des affinités de liaison nettement inférieures à celles observées pour les deux poches sauf pour les molécules contrôles CQ et PYR (**Figure 47**, *histogrammes jaunes et gris*).

Collectivement, ces résultats suggèrent que les inhibiteurs KEAP1 que nous avons étudiés sont des candidats intéressants à tester expérimentalement pour étudier la pharmaco-modulation de Pfk13.

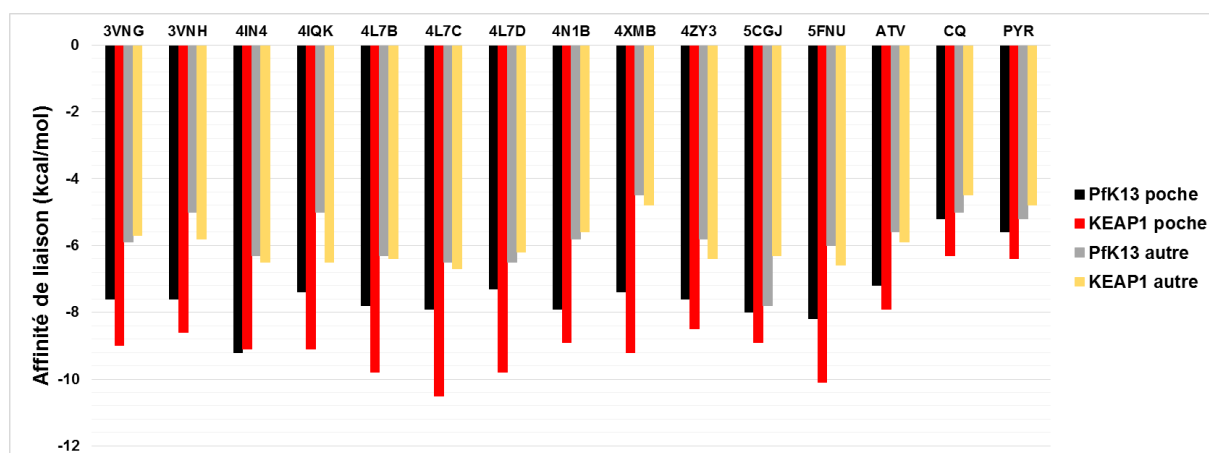


Figure 47 : Résultats des différents *docking* pour 15 molécules avec les poches KREP^{KEAP1} et KREP^{PfK13}. Seule la meilleure pose par molécule, correspondant à la plus haute affinité de liaison, est indiquée. L'affinité de liaison correspond à un échange énergétique, exprimée en kcal/mol. Généralement, on suppose qu'une énergie inférieure à -7 kcal/mol représente une forte affinité de liaison²⁷⁰. Les différentes molécules testées sont indiquées par leur identifiant PDB. Les trois derniers groupes d'histogrammes à droite exposent les résultats de *docking* pour trois antipaludiques servant de contrôles négatifs : ATV, atovaquone ; CQ, chloroquine ; PYR, pyriméthamine.

Parmi ces 12 inhibiteurs, quatre ont été achetés sur les plateformes commerciales Thermo Fisher ou Sigma-Aldrich (4IQ4, 4L7B, 4ZY3 et 5CGJ) puis testés pour mesurer leur effet sur la croissance de parasites *in vitro* par incubation avec l'inhibiteur durant 72 heures (à partir de cultures de parasites sauvages *P. falciparum*). La quantité d'ADN parasitaire en fin d'incubation a été mesurée par Laurine Rondepierre (doctorante, UMR 261) en utilisant l'agent intercalant SYBRGreen. Quatre concentrations de chaque inhibiteur ont été testées, diluées dans du diméthylsulfoxyde (DMSO)ⁱ : 0,1 μ M, 1 μ M, 10 μ M et 100 μ M. Des contrôles positif (500 nM de CQ) et négatifs (sans inhibiteur : globules rouges non infectés et globules rouges infectés) ont également été préparés. Après incubation sur trois jours puis ajout du tampon de lyse incluant le SYBRGreen, nous avons observé une faible inhibition de croissance, comprise entre 15 et 25% pour des concentrations finales en inhibiteurs égales à 0,1 μ M et 1 μ M (**Figure 48**). À une concentration de 10 μ M, l'inhibiteur 5CGJ a diminué la croissance des parasites d'environ 55%, les autres révélant un pourcentage de croissance diminué similaire aux concentrations 0,1 μ M et 1 μ M (**Figure 48**). Enfin, à une concentration très élevée de 100 μ M, les inhibiteurs ont diminué de plus de 50% de la croissance des parasites, 5CGJ étant là encore associée à la plus forte inhibition de croissance (90% ; **Figure 48**). Ces expériences nécessiteront d'être reproduites pour confirmation des effets observés, et d'autres inhibiteurs, incluant le meilleur candidat sur la base des *docking* moléculaires (4IN4), pourraient être testés (**Figure 47**).

ⁱ Utilisé pour sa capacité à solubiliser de nombreux composés organiques.

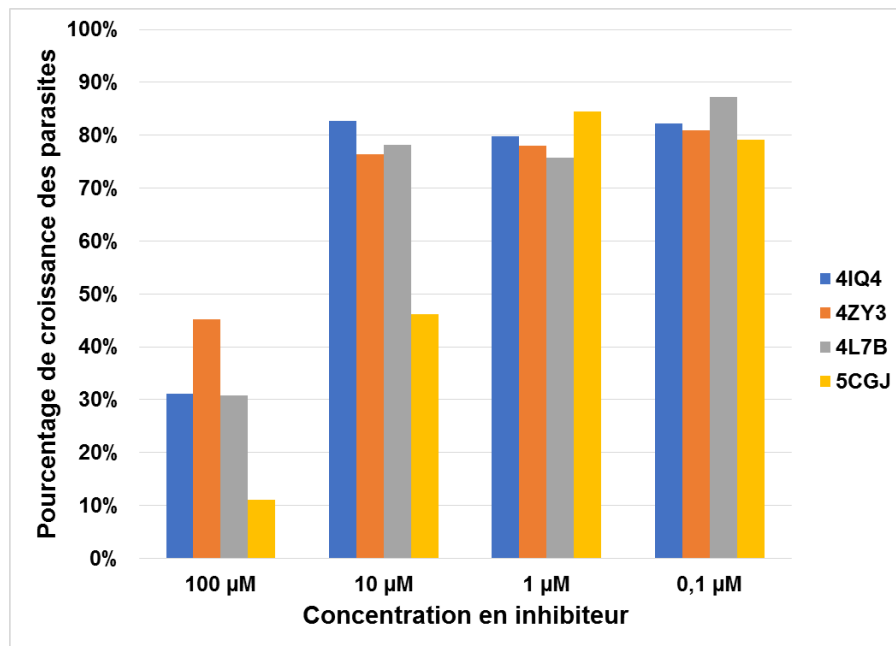


Figure 48 : Pourcentage de croissance des parasites *P. falciparum* après exposition à différentes concentrations en inhibiteurs KEAP1. Les noms des différentes molécules testées correspondent aux identifiants PDBs. La croissance des parasites est approximée par la mesure de la quantité d'ADN parasitaire détectée après 72 heures d'incubation en présence ou absence d'inhibiteur. La croissance des parasites en présence d'inhibiteur est exprimée en pourcentage de celle sans inhibiteur. Les expériences et analyses ont été réalisées par Laurine Rondepierre (doctorante, UMR 261).

- K13 fait partie des 5% des protéines les plus conservées du protéome de *Plasmodium*.
- K13 contient un domaine coiled-coil fortement conservé dont le rôle demeure obscur. Il pourrait participer à l'homo-oligomérisation de la protéine, servir de zone d'interaction avec d'autres protéines, ou encore se lier à l'ADN.
- Le domaine BTB est extrêmement conservé et partage une forte identité de séquence avec celui des protéines KCTD, souvent impliquées dans des complexes d'ubiquitination et associées à divers états homo-oligomériques. Nous avons prédit une zone d'interaction sur ce domaine, correspondant partiellement à la zone d'interaction avec une protéine Cullin.
- Le domaine KREP, impliqué dans la résistance aux ARTDs, est fortement conservé. Une poche peu profonde a été prédite comme possible surface d'interaction avec un substrat, avec un possible rôle majeur de résidus arginine et sérine, par comparaison avec des domaines KREP issus de protéines de la famille des BTB-Kelch impliqués dans des complexes d'ubiquitination.
- Les simulations de dynamique moléculaire révèlent que les mutations de résistance aux ARTDs déstabilisent localement la structure du KREP sans affecter la structure ni les propriétés physico-chimiques de la poche.
- Des analyses de *pull-down* ont été initiées pour valider le rôle fonctionnel de la poche du domaine KREP de PfK13. Trois constructions basées sur les trois domaines annotés de PfK13 (C : coiled-coil, B : BTB et K : KREP) ont été préparées avec un tag GST en N-terminal : GST-K, GST-BK et GST-CBK.
- Les différents domaines recombinants sont fortement exprimés chez la bactérie *E. coli*, et solubles avec utilisation d'un détergent particulier, le sarkosyl. Suite à leur purification, ils ont été convenablement élués par compétition avec le glutathion, suggérant que, au minimum, une fraction de la région GST est correctement repliée.
- Les premiers *pull-down* réalisés n'ont pas permis d'identifier des protéines partenaires parasitaires par SDS-PAGE, mais révèle la présence de protéines bactériennes contaminantes. Aucune méthode n'a pour l'heure été trouvée pour limiter leur présence. Des optimisations sont donc nécessaires avant de réaliser de nouveaux *pull-down*.
- De petites molécules se liant de façon non covalente à la poche du domaine KREP de KEAP1 sont associées à une forte affinité de liaison avec celle de PfK13. Ces molécules représentent de bons candidats pour étudier la pharmaco-modulation de PfK13.
- Quatre inhibiteurs ont été testés pour mesurer leur effet sur la croissance des parasites *P. falciparum*. À forte concentration (100 μ M), les inhibiteurs diminuent drastiquement la croissance des parasites.



Le temps de la réflexion est une économie de temps.

— **Publius Syrus** (vers -85 – après 43)

P fCRT est supposée transporter un ou plusieurs substrats physiologiques ainsi que des drogues. Cependant, les régions impliquées dans la reconnaissance et/ou interagissant avec ces molécules – en excluant les sites associés au transport de CQ – n’ont pas encore été clairement caractérisés. De façon similaire à PfK13, notre objectif fut de rechercher des régions soumises à une sélection naturelle particulière (positive ou purificatrice), qui peut refléter une possible importance fonctionnelle dans la structure et/ou la fonction de la protéine. L’ensemble des résultats bioinformatiques obtenus est présenté. À noter que pour l’heure, aucune expérimentation fonctionnelle n’a été entreprise pour valider nos observations.

3.2.1. À la recherche d'une région fonctionnelle de PfCRT

3.2.1.1. Forte conservation de la moitié vacuolaire de CRT

Tout d'abord, nous avons souhaité identifier et mesurer les pressions de sélection agissant sur le gène *crt*. Contrairement au gène *k13*, les séquences *crt* d'*Apicomplexa* n'appartenant pas au genre *Plasmodium* n'ont pas été incluses dans l'analyse car le pourcentage d'identité et de couverture de séquence au-delà du genre *Plasmodium* étaient relativement faibles (respectivement inférieurs à 30% et 80%, suggérant une possible fonction différente pour des homologues PfCRT non-*Plasmodium*). Au total, 24 séquences *crt* ont été obtenues par BLAST à partir des banques de données NCBI et EuPathDB²⁵². Suite à l'élaboration d'un alignement multiple des séquences *crt* et de l'arbre phylogénétique correspondant, l'analyse des pressions de sélection par le programme PAML a montré que le gène *crt* a été soumis à une forte pression de sélection purificatrice, suggérant que la grande majorité des mutations non-synonymes sont délétères et ont été éliminées durant l'évolution des *Plasmodium*. En utilisant l'outil dans PAML permettant de partitionner les codons selon des critères protéiques structuraux, nous avons constaté que la moitié vacuolaire de la protéine a évolué significativement plus lentement que la moitié cytoplasmique. De façon moins surprenante, les hélices transmembranaires se sont avérées être plus conservées que les boucles les reliant.

3.2.1.2. Prédiction de différents états conformationnels de la structure tertiaire de PfCRT

Pour prédire des sites fonctionnels, il est avantageux de disposer d'une structure tertiaire de la protéine d'intérêt, surtout dans le cas de séquences soumises à une forte sélection purificatrice comme nous l'avons montré pour K13. Celle de PfCRT n'a pas encore été déterminée. Depuis 2016, la structure tertiaire de quelques transporteurs composés de 10 hélices transmembranaires et appartenant à la même superfamille de protéines (DMT) que PfCRT a été déterminée expérimentalement^{271,272}. Ces structures tertiaires parentes permettent d'envisager la prédiction de la structure tertiaire de PfCRT par modélisation par homologie. De plus, la structure tertiaire de ces DMTs a été déterminée sous différents états conformationnels, offrant la possibilité de prédire la structure de PfCRT selon un état ouvert côté vacuolaire (ou *inward-facing*) et un état fermé (ou *occluded*).

Le modèle PfCRT à l'état ouvert a révélé la présence d'une cavité impliquant la grande majorité des dix hélices, tandis que la conformation fermée a mis en évidence une poche au cœur de la protéine, représentant vraisemblablement le site d'interaction avec le (ou les) substrat(s) physiologique(s) de PfCRT – par homologie avec les protéines DMTs. En outre, les modèles PfCRT arboraient plusieurs similitudes avec les autres DMTs, notamment un potentiel électrostatique à la surface très positif – incluant l'entrée de la cavité – tandis que le cœur de la protéine était globalement neutre.

3.2.1.3. Prédiction de sites fonctionnels de PfCRT

À l'aide de l'outil FuncPatch, nous avons mis en évidence qu'une partie de la cavité, et plus spécifiquement l'hélice 9 qui longe la cavité, est enrichie en sites particulièrement conservés et forment un possible *patch* fonctionnel. Cela suggère que la fonction de PfCRT serait bien liée à la cavité, incluant une poche de liaison comme c'est le cas pour la plupart des protéines DMT. Pour prédire plus spécifiquement certains sites qui auraient un rôle dans la fonction de transport, nous nous sommes concentrés sur ceux strictement conservés au cours de l'évolution des *Plasmodium* et impliqués dans l'architecture de la poche. En parallèle, les modèles PfCRT que nous avons prédits ont été alignés avec les structures DMT pour lesquelles les propriétés de liaison protéine-substrat étaient disponibles. Nous avons ainsi détecté chez PfCRT deux tyrosines (Y68, hélice 1 ; Y345, hélice 9), un aspartate (D329, hélice 8) et une serine (S349, hélice 9) comme candidats fonctionnels à tester en priorité. Trois de ces quatre sites étaient retrouvés dans le *patch* de sites conservés détecté par FuncPatch.

3.2.1.4. Un potentiel électrostatique altéré chez les structures PfCRT portant les mutations de résistance aux médicaments à base de quinoléine

D'après les modèles PfCRT générés, nous avons observé que la majorité des mutations de résistance à la CQ sont situées sur des positions impliquées dans l'architecture de la poche. À titre d'exemple, en nous basant sur l'haplotype Dd2 composé de huit mutations (M74I, N75E, K76T, A220S, Q271E, N326S, I356T et R371I), cinq des huit sites mutés sont localisés sur des positions formant la poche, incluant la position critique 76 retrouvée quel que soit l'haplotype de résistance. Dans la mesure où la CQ s'échappe de la vacuole digestive, son site d'action, sous un état di-protoné^{171,172}, nous avons émis l'hypothèse que les mutations de résistance altèrent le potentiel électrostatique à la surface de la protéine. Après avoir réalisé une série de mutagenèses *in silico* sur les modèles PfCRT pour produire des structures combinant les mutations de résistance à la CQ (haplotypes Dd2, 7G8 et GB4), nous avons remarqué que le potentiel électrostatique à la surface de la cavité et de la poche était devenu extrêmement négatif, alors qu'il était neutre voire positif à l'état sauvage.

Des mutagenèses additionnelles ont été faites sur les structures mutantes PfCRT Dd2 ou 7G8 pour produire des structures associées à la résistance à la PPQ. Chez les parasites, la résistance à la PPQ est associée à une atténuation importante mais non totale de la résistance à la CQ mesurée *in vitro* par des tests d'inhibition de croissance⁸⁹. Tout comme les mutations de résistance à la CQ, une majorité des mutations de résistance à la PPQ sont situées sur des positions impliquées dans la poche de liaison prédite (5 sur 8 : positions 97, 101, 145, 350 et 353). Le potentiel électrostatique est demeuré à un niveau similaire à celui des structures PfCRT mutantes conférant la résistance à la CQ, pouvant expliquer pourquoi la réversion de sensibilité à la CQ n'est pas totale.

3.2.2. Article 2

L'article ci-après, soumis et en cours de révision, présente l'ensemble de nos résultats de façon détaillé. Les résultats et matériels supplémentaires soumis avec cet article sont fournis dans la partie *Annexe* (section Matériel supplémentaire – Article 2).

Structural and evolutionary analyses of the *Plasmodium falciparum* chloroquine resistance transporter

Romain Coppée¹, Audrey Sabbagh^{1,*}, Jérôme Clain^{1,2,*}

¹ Université de Paris, UMR 261 MERIT, IRD, F-75006 Paris, France.

² Centre National de Référence du Paludisme, Hôpital Bichat-Claude Bernard, Assistance Publique des Hôpitaux de Paris, F-75018 Paris, France.

* Corresponding authors:

Jérôme Clain: jerome.clain@parisdescartes.fr

Audrey Sabbagh: audrey.sabbagh@parisdescartes.fr

Abstract

Mutations in the *Plasmodium falciparum* chloroquine resistance transporter (PfCRT) are related to resistances to several quinoline antimalarial drugs, chiefly to chloroquine (CQ) but also to partner molecules in current artemisinin-based combination therapies such as piperazine (PPQ). After two decades of research, the physiological function of the vacuolar PfCRT transporter still remains elusive: its tertiary structure is lacking, the amino acids involved in substrate binding are unknown, and the structural/functional alterations caused by the known drug resistance mutations are not characterized. Here, we produced models of the PfCRT wild-type (WT) tertiary structure in different conformational states and looked for the most conserved amino acid sites as putative indicators of structural and/or functional constraints. We found a hypothetical substrate-binding pocket in the core of the PfCRT membrane-spanning domain, partially structured by positions related to CQ and PPQ resistance mutations including the key K76T. Structural comparisons with drug/metabolite transporters phylogenetically close to PfCRT and previously characterized at the experimental level identified several amino acid sites possibly involved in PfCRT substrate transport. Finally, mutant-like CQ resistance structures of PfCRT revealed drastic changes in the electrostatic potential of the binding pocket. These changes might facilitate the escape of di-protonated CQ from the digestive vacuole into the cytoplasm. Altogether, our findings provide novel hypotheses that can be exploited for elucidating the role of PfCRT in the digestive vacuole physiology and drug resistance.

Introduction

Resistance to antimalarial drugs continues to threaten malaria control. Formerly used as first-line treatment for uncomplicated *Plasmodium falciparum* malaria in most endemic areas, chloroquine (CQ) has seen its efficacy reduced in the late 60s after a decade of use. CQ resistance (CQR) was initially reported in Southeast Asia and South America, and then in other endemic regions including Western Pacific and Africa¹. The widespread CQR has led to successive antimalarial drug replacements and since 2000 to the progressive adoption of

artemisinin-based combination therapies (ACTs). CQ remains highly effective to treat malaria caused by other human-infecting *Plasmodium* parasite species, especially the highly frequent *P. vivax*².

During the intraerythrocytic development of *Plasmodium* parasites, hemoglobin from the host erythrocyte is uptaken by the parasite and delivered into its digestive vacuole (DV), which is an acidic lysosome-like organelle. Vacuolar degradation of hemoglobin generates toxic Fe²⁺-heme molecules that spontaneously oxidize to form membrane-disrupting Fe³⁺-heme³. To avoid this, the Fe³⁺-heme molecules are biocrystallized in the DV into non-toxic, chemically inert crystals known as hemozoin.

Quinoline-containing antimalarial drugs such as CQ (including also piperaquine, amodiaquine and pyronaridine) are thought to interfere with the conversion of toxic heme to non-toxic hemozoin crystal. This inhibition leads to the accumulation of toxic Fe²⁺-heme and parasite death⁴. Although the precise mode of action of quinoline-containing drugs remains controversial, most studies have emphasized the essential role of increased concentrations of CQ within the DV⁵⁻⁷.

The main determinant of CQR is the *Plasmodium falciparum* chloroquine resistance transporter gene (*pfcr*t)^{8,9}. *Pfcr*t encodes a 424 amino acid protein (PfCRT), consisting of 10 putative transmembrane helices (TMs)⁹, which is essential to the intraerythrocytic development of the parasite¹⁰. Despite two decades of research, the physiological role of PfCRT in the parasite biology is still questioned. Its secondary structure, location in the DV membrane and phylogenetic relationships with the drug/metabolite transporter (DMT) superfamily – especially with nucleotide-sugar transporters and triose phosphate translocators – suggest its participation in the transport of various drugs and metabolites^{9,11}. Several transport activities have been reported for this protein, identified as a proton-coupled transporter recognizing some cationic substrates¹² and playing a role in iron and glutathione homeostasis¹³⁻¹⁵. Earlier PfCRT expression-based studies have also reported PfCRT as a Cl⁻ channel mediator, a proton pump regulatory and an activator of Na⁺/H⁺ exchangers¹⁶⁻¹⁸.

Molecular-epidemiological studies have identified at least six independent origins of CQR in different regions of the world¹⁹ with various mutant *pfcr*t haplotypes described, harboring 4 to 10 non-synonymous mutations such as the widely studied Dd2 haplotype which first occurred in Asia and then spread to Africa¹. The key K76T mutation is carried by all CQR haplotypes. It removes a positive charge in the TM 1, allowing the di-protonated form of CQ to be expelled outside the DV^{9,20}. However, the K76T mutation alone is not sufficient to confer CQR and additional mutations are needed to increase the CQR level and/or attenuate the fitness cost, highlighting the complex relationship between PfCRT mutant variants and the CQR phenotype^{21,22}. Although CQ is no longer used for the treatment of *P. falciparum* infections, PfCRT remains a protein of utmost clinical importance since both ancient and more recent PfCRT mutations are associated with altered *in vitro* antimalarial activity of different ACTs components, namely piperaquine (PPQ), the active amodiaquine metabolite, lumefantrine, and also artemisinin derivatives^{20,23-28}. However, for these drugs, the PfCRT-dependent pharmacomodulation might operate through different mechanisms than for CQ^{26,29,30}.

Because of the lack of an experimental tertiary structure of PfCRT, the location of the aforementioned combinational PfCRT mutations and their structural/physicochemical impact remain largely unknown. In addition, the functionally important amino acid sites of PfCRT in a wild-type (WT) context, that is those involved in the physiological transport activity of PfCRT, remain uncharacterized. Therefore, we report here models of different conformational states of the WT PfCRT structure obtained by homology modeling, based on the experimental tertiary structures of members of the DMT superfamily containing 10 TMs^{31,32} and close to PfCRT in the DMT phylogeny³³. The modeled PfCRT structures were then jointly studied with the CRT sequence conservation across *Plasmodium* species, examined at both the codon and amino

acid levels, leading to the identification of PfCRT candidate functional sites that may play a role in substrate trafficking. Finally, a series of *in silico* mutagenesis from the WT PfCRT models were performed to produce mutant-like PfCRT models associated with CQR or PPQR, revealing drastic alterations of PfCRT physicochemical properties compared to WT.

Results

CRT sequences and phylogenetic tree analysis.

Non-*Plasmodium* species have not been included in this study because we aimed to evaluate the CRT evolution at the *Plasmodium* level. Among the 24 *Plasmodium* CRT sequences retrieved from databases, only those of the bird-infecting *P. relictum* appeared incomplete with an absence of twenty amino acid positions in the C-terminal region of CRT. The other sequences showed high conservation in length, ranging from 423 to 425 amino acids (Supplementary Table S1), resulting in the presence of few indels (8 positions) in the CRT multiple sequence alignment. These indels were removed from subsequent analyses (Supplementary Fig. S1). The phylogenetic tree produced from the curated *crt* sequence alignment was highly congruent with the acknowledged phylogeny of *Plasmodium* species³⁴, with four main monophyletic groups (bird parasites, primate parasites '*Plasmodium*', rodent parasites and primate parasites '*Laverania*') supported by high bootstrap values ($\geq 70\%$; Fig. 1).

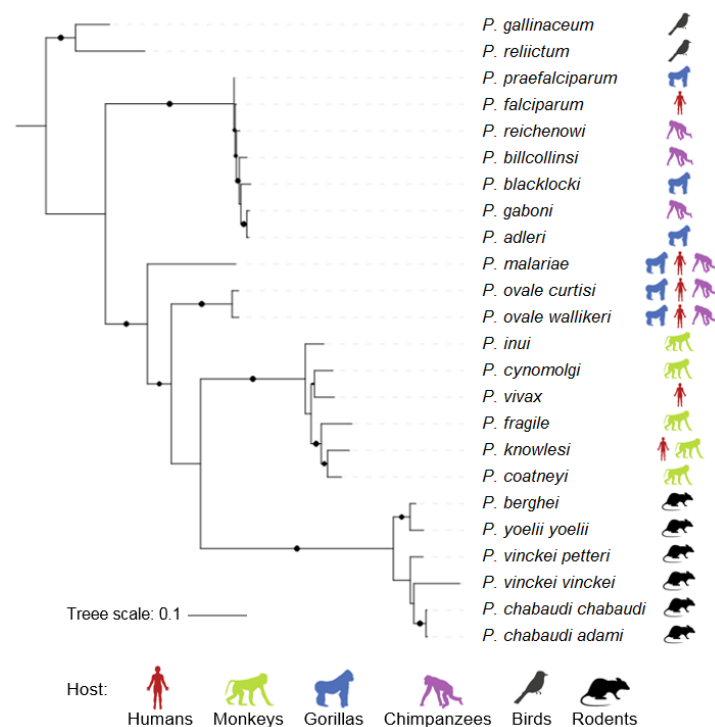


Figure 1. Maximum likelihood phylogenetic tree of orthologous *crt* protein-coding sequences from *Plasmodium*. Bootstrap values are shown as black circles ($>90\%$) using the approximate likelihood ratio aLRT SH-like method. Vertebrate host(s) during intraerythrocytic development is (are) provided for each *Plasmodium* species.

Variable levels of conservation across the *crt* gene sequence and phylogeny.

The selective pressures acting on the CRT-coding sequences were searched for by estimating the non-synonymous to synonymous substitution rate ratio ω ($= d_N/d_S$) across lineages (branch model) and codon sites (random-site models). We used branch and random-site models

implemented in the *codeml* program. The *crt* gene was found to evolve under a variable selective regime across *Plasmodium* phylogeny as the free-ratio model (that allows ω to vary along branches in the *crt* phylogenetic tree) had significant better fit to the data than the one-ratio null model M0 (which supposes only one ω value for all branches; $p = 1.85 \times 10^{-32}$; Table 1). Although purifying selection (ω or $d_N/d_S < 1$) was pervasive across the whole phylogeny, some specific branches showed evidence of positive selection ($d_N/d_S > 1$) or neutral evolution suggesting a few adaptive changes during *Plasmodium* evolution (see *P. vinckei vinckei* and also during the speciation event of *P. fragile* with *P. knowlesi* and *P. coatneyi*; Supplementary Fig. S2). The substitution rate ratio ω was also found to be heterogeneously distributed across *crt* codon sites as assessed by the M0:M3 comparison ($p = 1.45 \times 10^{-133}$; Table 1), but no specific codon site was detected to evolve under positive selection since the positive selection models M2a and M8 provided no better fits to the data than the neutral models M1a and M7, respectively ($p = 1.00$ in both cases; Table 1 and Supplementary Table S2). Using the best-fitted model M3 and by visual inspection of the multiple alignment, the sequence coding to the cytoplasmic N- and C-terminal extremities of CRT were found to be much less conserved than most of the remaining protein-coding sequence (Supplementary Data S1). Those extremities contained some phosphorylated sites in PfCRT³⁵, including T416 which however appeared highly conserved. Altogether, the data indicate that much of the CRT protein, except the N- and C-terminal regions, has been conserved over *Plasmodium* evolution.

Table 1. Results of PAML analyses

Models	2 $\Delta\ell$	df	<i>p</i> value	Conclusion
M0:free-ratio	260.20	44	1.85×10^{-32}	ω heterogeneous among lineages of the phylogenetic tree
M0:M3	623.24	4	1.45×10^{-133}	ω heterogeneous across codon sites
M1a:M2a	0.00	2	1.00	No codon sites detected to evolve under positive selection
M7:M8	0.00	2	1.00	No codon sites detected to evolve under positive selection
Partitions: cytoplasmic-half (n = 146 codon sites) versus vacuolar-half (n = 196 codon sites) sides				
MA:MB	14.59	1	1.33×10^{-4}	Different nucleotide substitution rates among partitions
MB:MD	19.71	2	5.24×10^{-5}	Heterogeneity of evolutionary parameters among partitions
MC:ME	18.87	2	7.97×10^{-5}	Heterogeneity of evolutionary parameters among partitions
Partitions: non-TM (n = 116 codon sites) versus TMs (n = 226 codon sites)				
MA:MB	18.57	1	2.22×10^{-14}	Different nucleotide substitution rates among partitions
MB:MD	2.57	2	1.87×10^{-3}	Heterogeneity of evolutionary parameters among partitions
MC:ME	0.92	2	4.58×10^{-3}	Heterogeneity of evolutionary parameters among partitions

Note – ω , non-synonymous (d_N) to synonymous (d_S) substitution rate ratio (d_N/d_S); 2 $\Delta\ell$, log-likelihood ratio of the two tested models; df, degrees of freedom; TM, transmembrane helix; non-TM, loop connecting two TMs. 2 $\Delta\ell$ was compared to a chi-squared table to determine the significance of the likelihood ratio tests. For the two series of partitioning tests shown here, codon sites were partitioned with the inward-facing state model of PfCRT and we discarded the poorly conserved N- and C-terminal regions, both located in the cytoplasm (this alteration was conservative). Similar results were obtained when codon sites were partitioned with the occluded state model than with the inward-facing state model (data not shown).

Predicting different conformational states of PfCRT tertiary structure.

In order to get a finer map of evolutionary constraints and functional sites in a structural context, modeling the PfCRT tertiary structure was required since no experimental structure is yet available for this specific DMT transporter. Since 2016, several DMT structures containing 10 TMs were experimentally resolved, allowing us to predict a PfCRT tertiary structure through homology modeling. We identified two DMT template structures based on the *confidence* criteria from Phyre2 which represents the probability that the match between the sequence and the template arises from a true relationship³⁶. The first DMT template, in an inward-facing state, is the nucleotide sugar GDP-mannose transporter Vrg4 (PDB ID: 5oge; 12.7% sequence

identity with PfCRT; [Supplementary Fig. S3](#) and [Table S7](#))³¹. The second DMT template, in an occluded state, is the triose-phosphate/phosphate translocator (TPT; PDB ID: 5y79; 14.5% sequence identity; [Supplementary Fig. S4](#) and [Table S7](#))³². PfCRT positions corresponding to the long vacuolar loop connecting TM 7 and TM 8 (PfCRT positions 269-313) and the cytoplasmic N- and C-terminal extremities (positions 1-55 and 399-424 respectively) were discarded because of insufficient amino acid coverage in DMT templates. In homology modeling, the paired target-template sequence alignment (here PfCRT-Vrg4 and PfCRT-TPT) is a critical step. Since hydrophobicity, which reflects the transmembrane topology, is typically conserved in TMs during evolution, the hydropathy profile can contain similar global features even in very distantly related proteins. Consequently, we first produced a superfamily-averaged hydropathy profile alignment with AlignMe³⁷ using the CRT multiple sequence alignment and an alignment of different members of the DMT superfamily. This profile-profile (CRTs-DMTs) alignment pinpointed some conserved amino acid positions that should be aligned in the final target-template sequence alignment. Finally, we produced a target-template alignment using the AlignMe PST algorithm which is designed for distantly related proteins (*i.e.* with a sequence identity < 15%)³⁷, then manually optimized it in order to align the putative important amino acid positions identified from the superfamily-averaged hydropathy profile alignment (see the PfCRT-TPT and PfCRT-Vrg4 final alignments in [Supplementary Figs. S3](#) and [S4](#) respectively). Two PfCRT conformational states were predicted using MODELLER: inward-facing state and occluded state using Vrg4 and TPT, respectively, as templates. After the refinement and minimization steps, local stereochemistry (MolProbity) validated the quality of the built PfCRT model, which exhibited similar quality indices than the templates as suggested by Ramachandran plots ([Table 2](#)). General (ERRAT, Prosa II) and transmembrane-specialized (QMEANBrane) 3D quality metrics also provided high confidence in structure conformation, energies and non-bonded interactions between atoms of the PfCRT models, when compared to the templates and other highly refined structures from the Protein Data Bank ([Table 2](#) and [Supplementary Fig. S6](#)). PfCRT aligned structurally with Vrg4 template with a Ca-based root-mean-square deviations (RMSD; which measures to what extent a PfCRT given residue changes its position compared to the aligned template residue) of 4.9 Å for the whole protein and 2.1 Å when TMs-connecting loops were discarded. In the occluded state model, PfCRT aligned structurally with TPT template with a Ca-based RMSD of 2.9 Å for the whole protein and 1.7 Å when TMs-connecting loops were discarded. In all, quality metrics indicated that the PfCRT models can be reliably used for further structural analyses.

Both inward-facing and occluded PfCRT models revealed symmetry-related structural repeats consisting of 10-TM bundle, approximately 30 Å in length. TM 1 to TM 5 and TM 6 to TM 10 surround a central cavity in anticlockwise and clockwise manners, respectively ([Fig. 2A](#)). The TMs boundaries were similar in the two states, except TM 1 and TM 3 that were shorter and interrupted by a short loop, respectively, in the inward-facing state ([Supplementary Fig. S6](#)). Of note, the key 76 position (K76T mutation) was embedded in the first vacuolar third of TM 1 ([Supplementary Fig. S6](#)). In the PfCRT inward-facing state, the structure is basket-shaped, similar to the Vrg4 template, with a deep cavity opened at the vacuolar side and composed of amino acids from most TM helices except TM 5 and TM 10 ([Fig. 2B](#)). In the PfCRT occluded state, a pocket can be observed in the core of the structure ([Fig. 2B](#) and full list of the 41 pocket positions in [Supplementary Table S3](#)), which corresponds to the substrate-binding pocket in the TPT template³². By estimating the electrostatic surface potential on the inward-facing state model, we observed that the cavity entry is slightly electropositive and the putative binding pocket is globally neutral (also observed in the occluded state model), a feature typically shared by some other DMTs including the YddG transporter ([Fig. 2C](#) and [Supplementary Fig. S7](#)).

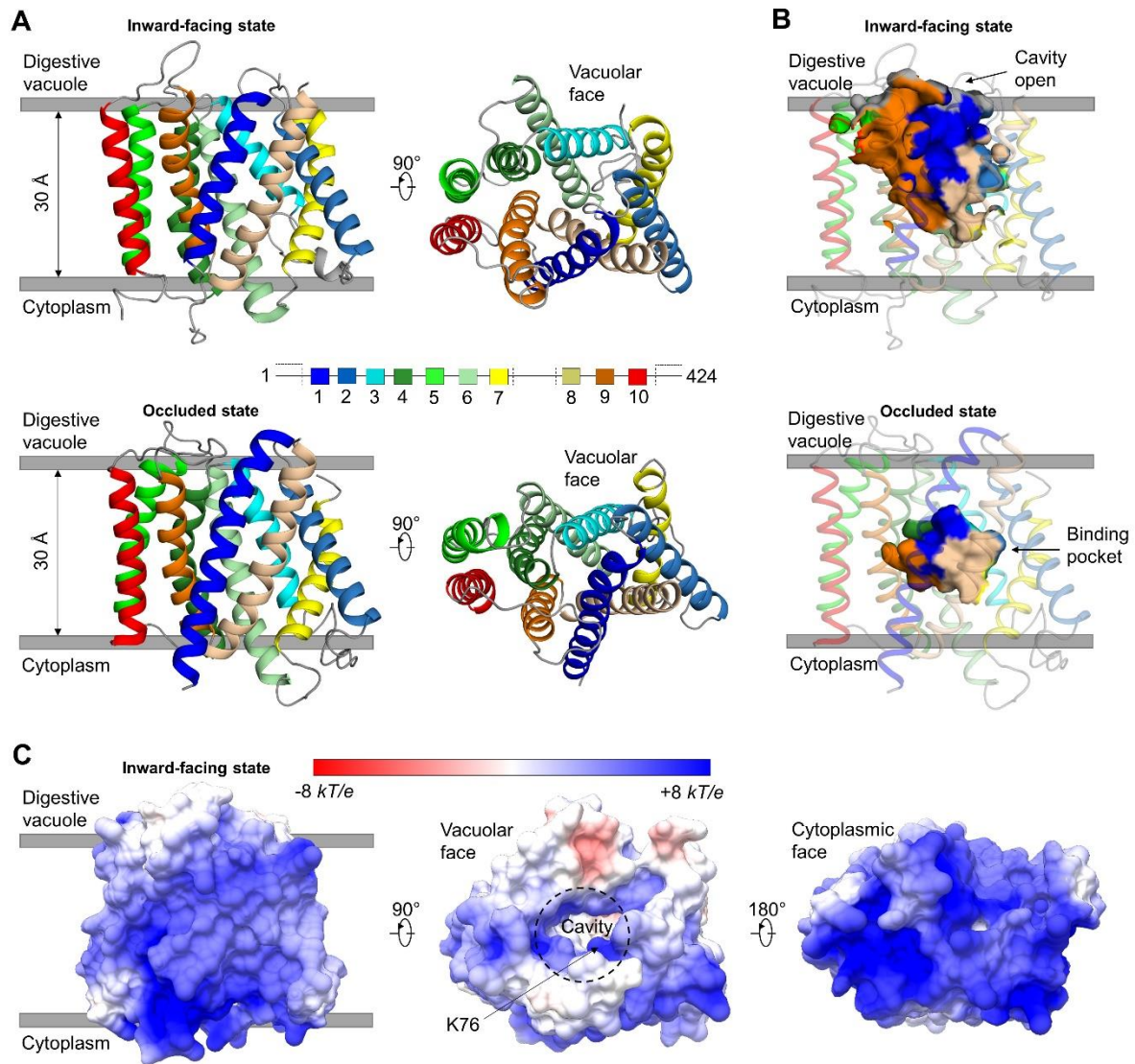


Figure 2. 3D-fold and structural features of the predicted inward-facing and occluded states of PfCRT. (A) Predicted tertiary structure of PfCRT in inward-facing (*top* panels) and occluded (*bottom* panels) states. Each TM is highlighted in a specific color. (B) Location of the cavity and the substrate-binding pocket (binding pocket) in the PfCRT models. In the inward-facing state, the cavity of PfCRT is open from the vacuolar face. In the occluded state, a substrate-binding pocket constitutes the core of the transporter. (C) Electrostatic surface potential of the inward-facing state of PfCRT. Values are in units of kT/e at 298 K, on a scale of $-8 kT/e$ (red) to $+8 kT/e$ (blue). White color indicates a neutral potential. The structure is shown from the side, vacuolar face and cytoplasmic face from the left to right respectively. The location of the key position K76 in CQR is indicated.

Vacuolar and TM sites are under strong purifying selection.

We used this new structural information to partition PfCRT sites and test whether some PfCRT regions, apart from the N- and C-ter, were on average differentially conserved. We used the fixed-site models and partitioning function implemented in the *codeml* program. We tested two different partitioning of *crt* codon sites: 1) cytoplasmic-half *versus* vacuolar-half sides of the transporter and 2) non-TMs (or loops) *versus* TMs. In the two different sets of partitioning, model B provided a better fit to the data than model A which assumes the same substitution pattern with identical parameters among partitions ($p = 10^{-4}$ and 2×10^{-14} for cytoplasm/vacuole and non-TM/TM respectively, [Table 1](#) and [Supplementary Table 4](#)). This indicates that the nucleotide substitution rates at vacuolar and TM sites were on average 0.8 and 0.62 times as low as at the cytoplasmic and non-TM sites respectively. More complex

models D and E also provided better fits to the data than comparison models B and C (Table 1), confirming a heterogeneity of evolutionary parameters in the two sets of partitioning (Supplementary Table 4). Vacuolar and TM sites were under stronger purifying selection than cytoplasmic and non-TM sites respectively (cytoplasm/vacuole: $\omega_1 = 0.121$ / $\omega_2 = 0.070$, best fitted D model; non-TM/TM: $\omega_1 = 0.111$ / $\omega_2 = 0.078$, best fitted E model; Supplementary Table 4). This is consistent with TMs being the hydrophobic core of the protein and the vacuolar side being the likely entry route for physiological substrates^{6,12,15,20,29,38}, respectively.

Table 2. Quality metrics for PfCRT models and templates

Protein ^a	ProS A II ^b	MolProbity ^c						ERRAT ^d [0-100]	QMEAN Brane ^e [0-1]
		Poor rotamers, n (%)	Favored rotamers, n (%)	Ramachandran outliers, n (%)	Ramachandran favored, n (%)	Bad bonds, n (%)	Bad angles, n (%)		
Vrg4 (5oge)	-1.76	21/268 (7.84)	218/268 (81.34)	2/298 (0.67)	278/298 (93.29)	0/2415 (0.00)	4/3278 (0.12)	92.32	0.80
TPT (5y79)	-3.17	5/251 (1.99)	239/251 (95.22)	0/303 (0.00)	300/303 (99.01)	0/2423 (0.00)	0/3313 (0.00)	96.61	0.84
PfCRT inward-facing	-2.72	4/279 (1.43)	269/279 (96.42)	0/301 (0.00)	287/301 (95.35)	0/2550 (0.00)	2/3446 (0.06)	98.7	0.79
PfCRT occluded	-2.11	0/272 (0.00)	270/272 (99.26)	0/294 (0.00)	287/294 (97.62)	4/2487 (0.16)	3/3366 (0.09)	99.0	0.81

Note – ^a The experimentally-determined tertiary structures of two DMTs (PDB IDs: 5oge and 5y79) were used as templates for homology modeling of PfCRT structure. ^b ProSA II generates an overall quality score compared with the ones from all experimentally determined protein chains available in the Protein Data Bank. ^c MolProbity indicates the accuracy of a macromolecular structure model and diagnoses local problems by all-atom contact analysis, complemented by updated versions of covalent-geometry and torsion-angle criteria. ^d ERRAT analyzes the statistics of non-bonded interactions between different atom types, then plots error values as a function of the position of a 9-residue sliding window, calculated by a comparison with statistics of non-bonded atom-atom interactions from highly refined structures. ^e QMEANBrane is a structure quality assessment tool dedicated to membrane proteins. It determines a local quality score per residue (the mean per-residue quality score is shown), by combining statistical potentials trained on membrane protein structures with a per-residue weighting scheme.

The cavity of PfCRT contains highly conserved amino acid sites.

We next searched for the most conserved amino acid positions in the PfCRT tertiary structures as putative indicators of structural and/or functional constraints. We used the FuncPatch server which jointly analyses phylogenetic, protein tertiary structure and conservation information. FuncPatch infers site-specific substitution rates at the protein level, hereafter called λ substitution rates, and is especially useful in the case of highly conserved proteins. The results obtained with the inward-facing and occluded states were very similar (Spearman's rank correlation: $p < 0.001$, $r = 0.99$; Fig. 3A and Supplementary Data S1), and only the data obtained with the PfCRT inward-facing state will be presented. We detected a significant spatial correlation of site-specific substitution rates in the PfCRT structure as evidenced by a log Bayes factor of 47.6. We then mapped on the tertiary structure the 10% ($n = 29$ positions) most conserved amino acid sites of the 297 PfCRT positions included in the FuncPatch analysis (the *Plasmodium*-specific TM 7-TM 8 vacuolar loop and the N- and C-terminal extremities of CRT were excluded from this analysis because of no tertiary structure information available for those positions). Remarkably, almost the whole TM 9 that lines the PfCRT cavity was highlighted (Fig. 3B and Supplementary Table S5). Other highly conserved positions were located on TM 4, TM 5, in the middle of TM 8 and on TM 10 (Fig. 3B). It is noteworthy that TM 9 has been reported to play a major role in substrate binding in other DMTs consisting of 10 TMs¹¹, and especially in those of the experimentally characterized Vrg4 and TPT proteins (PDB IDs: 5oge and 5y79; Supplementary Fig. S8)^{31,32}.

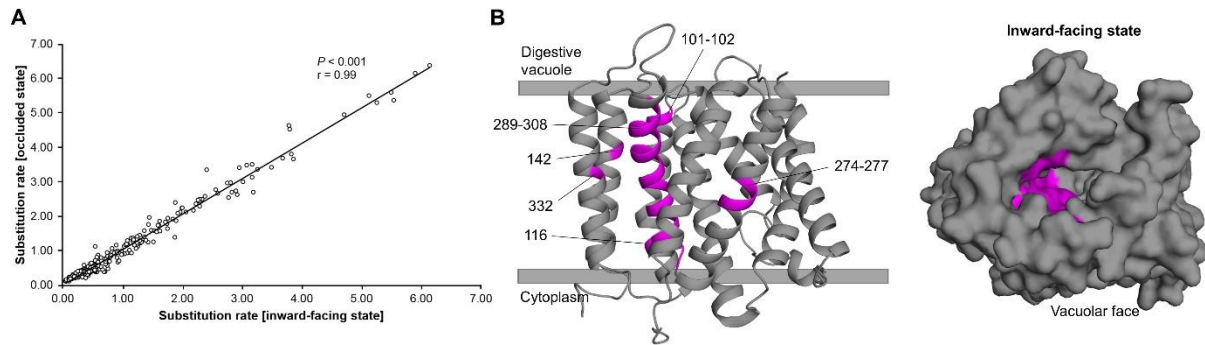


Figure 3. Location of the most conserved amino acid sites in the PfCRT tertiary structure. (A) Positive correlation in spatially correlated substitution rates between the occluded state and the inward-facing state of PfCRT. Each point corresponds to a substitution rate associated with a PfCRT position in the tertiary structure. (B) Location of the 10% most conserved amino acid sites (inferred with FuncPatch) in the tertiary structure of PfCRT inward-facing state. The most conserved sites are colored in purple. The PfCRT structure is shown as cartoon from the side (*left structure*) and as surface from the vacuolar face (*right structure*).

PfCRT amino acids Y68, D329, Y345 and S349 might play a role in physiological substrate trafficking.

Some of the most highly conserved amino acid sites of PfCRT might be under strong functional – rather than structural – constraints. Being a member of the DMT protein superfamily, PfCRT may bind a substrate through amino acids located in its central cavity, and more particularly in the binding pocket (occluded state). Consequently, we focused on the strictly conserved amino acid positions across *Plasmodium* species and whose side-chains in the PfCRT occluded model were involved in the architecture of the binding pocket. After these filtering steps, we listed 24 amino acid positions ([Supplementary Table S6](#)). We then performed paired structural alignments of the PfCRT models with three *reference* DMT structures: the previously described Vrg4 and TPT proteins and the YddG protein (another 10 TMs-containing DMT transporter, in an inward-facing state). We focused only on structurally aligned positions that have been shown experimentally to be important or essential for substrate trafficking in *reference* DMT proteins ([Supplementary Fig. S8](#))^{31,32,39}. Two tyrosines of PfCRT (Y68 in TM 1 and Y345 in TM 9) paired with two tyrosines of Vrg4 (Y28 in TM 1 and Y281 in TM 9; [Fig. 4A](#))³¹. One aspartic acid of PfCRT (D329 in TM 8) paired with a tyrosine of the TPT protein (Y339 in TM 8, [Fig. 4B](#))³². Finally, PfCRT S349 (TM 9) paired with both S244 from the bacterial YddG protein and G285 from Vrg4 ([Fig. 4C](#)), which play a role in the transport of threonine/methionine and guanosine monophosphate (GMP) respectively^{31,39}. Importantly, of those four candidate PfCRT positions, three belonged to the conserved patch evidenced by the FuncPatch server (D329, Y345 and S349; [Supplementary Table S5](#)). Overall, this comparative evolutionary and structural analysis identified several candidate amino acid sites likely involved in the physiological transport activity of PfCRT ([Fig. 4D](#)). Remarkably, none are reported to be associated with drug resistance mutations.

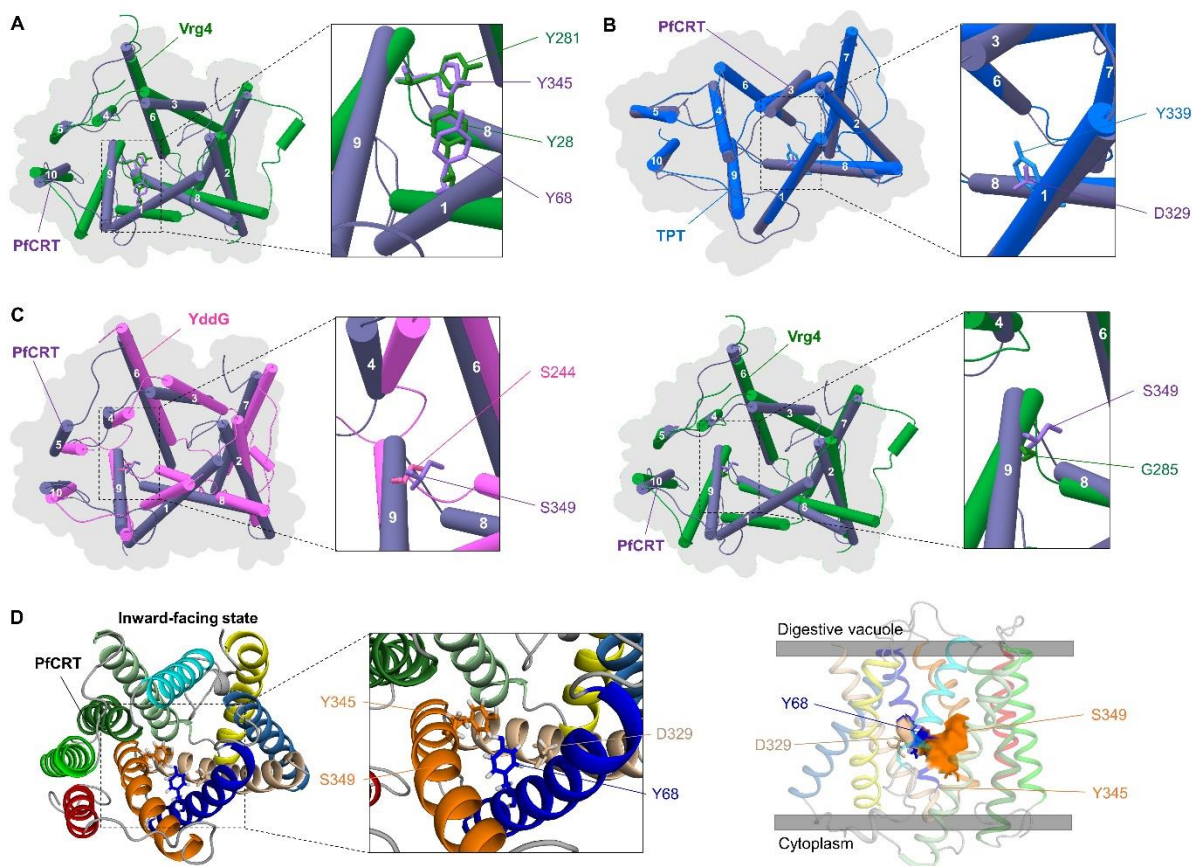


Figure 4. Comparative structure-based alignment of evolutionary conserved CRT positions located in the putative binding pocket of PfCRT. Structural alignment of PfCRT (grey purple) with (A and C) the Vrg4 (PDB ID: 5oge, green forest), (B) TPT (PDB ID: 5y79, blue) and (C) YddG (PDB ID: 5i20, pink) proteins using the *align* function in UCSF Chimera. The TMs helices are numbered. The PfCRT inward-facing state was aligned with Vrg4 and YddG, while the PfCRT occluded state was aligned with the TPT protein. (D) Location of the four putative PfCRT functional sites Y68, D329, Y345 and S349 in the PfCRT inward-facing state model. The structure is shown from the vacuolar face (left panel) with a zoom on the proposed functional positions, and from the side (right panel), as cartoon while candidate functional positions are shown in surface. These four amino acid positions cover a large part of the putative binding pocket.

Most PfCRT mutations conferring drug resistance are located in the cavity or binding pocket.

Mutations in PfCRT confer drug resistance, chiefly to CQ, but also alter sensitivities to other ACT drugs. To better understand the underlying resistance mechanisms, we first mutated PfCRT models *in silico* by jointly introducing several mutations to produce Dd2-, GB4- and 7G8-like mutant structures which confer CQR and correspond to Asian, African and South American CQR haplotypes respectively (Table 3). Note that the Q271E was not included in the mutant structures since it is located in the TM 7-TM 8 long vacuolar loop that was discarded from our PfCRT predicted models. From an evolutionary view, five of the nine CQR-related positions investigated here were strictly conserved during *Plasmodium* evolution including the position 76 related to the key CQR K76T mutation. We noted that the conservation level of positions related to CQR mutations M74I, N75E and K76T markedly diminished when site-specific substitution rates were computed by taking into account their spatial location in the structure (FuncPatch versus PAML data in Fig. 5A; Supplementary Data S1), indicating that these three amino acid sites were spatially close to less conserved ones. Remarkably, the CQR-related positions we analyzed were directly located in the cavity of PfCRT in the inward-facing state, except position 74 (mutation M74I) which lies in TM 1 but on the outer surface of the

transporter and position 371 (mutation R371I) in the TM 9-TM 10 vacuolar loop at the entry of the cavity (Fig. 5A). Based on the predicted PfCRT occluded state, six positions related to CQR mutations were predicted to be involved in the architecture of the putative binding pocket (C72S, N75E, K76T, A220S, N326S/D and I356T/L), including five of the eight CQR mutations carried by the canonical Dd2 haplotype (Table 3). As the key K76T mutation conferring CQR removes a positive charge in the cavity, we investigated *in silico* the changes in the electrostatic surface potential of the PfCRT^{Dd2}, PfCRT^{GB4} and PfCRT^{7G8} models. While the surface potential of the PfCRT^{WT} cavity and binding pocket was neutral (Fig. 5B), it turned highly electronegative in all CQR-like mutant structures (Fig. 5C and Supplementary Fig. S9A).

Table 3. PfCRT haplotypes investigated in this study

Origin region ^a	Parasite line ^b	Drug response ^c	TM 1	TM 1	TM 1	TM 1	TM 2	TM 2	TM 2	TM 3	TM 6	TM 6	17-8	TM 8	TM 9	TM 9	TM 9	TM 9	L9-10
			72	74	75	76	93	97	101	145	218	220	271	326	343	350	353	356	371
All regions	3D7	All drug sensitivity	C	M	N	K	T	H	C	F	I	A	Q	N	M	C	G	I	R
SE Asia	Dd2	CQR, PPQS	C	I	E	T	T	H	C	F	I	S	E	S	M	C	G	T	I
SE Asia	Dd2 ^{T93S}	CQS, PPQR	C	I	E	T	S	H	C	F	I	S	E	S	M	C	G	T	I
SE Asia	Dd2 ^{H97Y}	CQS, PPQR	C	I	E	T	T	Y	C	F	I	S	E	S	M	C	G	T	I
SE Asia	Dd2 ^{C101F}	CQS, PPQR	C	I	E	T	T	H	F	F	I	S	E	S	M	C	G	T	I
SE Asia	Dd2 ^{F145I}	CQS, PPQR	C	I	E	T	T	H	C	I	I	S	E	S	M	C	G	T	I
SE Asia	Dd2 ^{I218F}	CQS, PPQR	C	I	E	T	T	H	C	F	F	S	E	S	M	C	G	T	I
SE Asia	Dd2 ^{M343L}	CQS, PPQR	C	I	E	T	T	H	C	F	I	S	E	S	L	C	G	T	I
SE Asia	Dd2 ^{G353V}	CQS, PPQR	C	I	E	T	T	H	C	F	I	S	E	S	M	C	V	T	I
S America	7G8	CQR, PPQS	S	M	N	T	T	H	C	F	I	S	Q	D	M	C	G	L	R
S America	7G8 ^{C350R}	CQS, PPQR	S	M	N	T	T	H	C	F	I	S	Q	D	M	R	G	L	R
Africa	GB4	CQR, PPQS	C	I	E	T	T	H	C	F	I	S	E	N	M	C	G	I	I

Note – ^a Region of origin of the haplotype. SE Asia, Southeast Asia; S America, South America. ^b For Dd2 and 7G8 parasite lines, one additional mutation was found to remove CQR but confer PPQR. ^c CQS, chloroquine sensitivity; CQR, chloroquine resistance; PPQS, piperazine sensitivity; PPQR, piperazine resistance. Grey shading indicates residues that differ from the WT allele. In this study, the Q271E mutation was not introduced in the mutant structures and subsequent analyses since it is located in the long TM 7-TM 8 vacuolar loop that was discarded for tertiary structure prediction. The location in the structure (TM, transmembrane helix; L, loop) is indicated above each position. Positions and mutations written in red letters are part of the predicted substrate binding pocket (based on the predicted PfCRT occluded state model).

Next, the recently identified, additive PfCRT mutations conferring PPQR on a CQR genetic background (T93S, H97Y, C101F, F145I, I218F, M343L and G353V: all in isolation in PfCRT^{Dd2}; and C350R in PfCRT^{7G8})^{26,40,41} were individually introduced in the PfCRT CQR-like mutant models. As CQR-related positions, PPQR-related positions were almost all located in the cavity of PfCRT (inward-facing state), except I218F and M343L which were located at the outer surface and at the cytoplasmic face of the transporter respectively (Fig. 5A). In the occluded state model, positions related to PPQR mutations H97Y, C101R, F145I, C350R and G353V were involved in the architecture of the binding pocket. We observed a large electronegative surface potential in the cavity and binding pocket similarly to the mutant PfCRT structures conferring CQR (Fig. 5D and Supplementary Fig. S9B), except for C350R that rendered neutral a large part of the cavity except in the proximity of K76T (note that this was found only in the inward-facing state; Supplementary Fig. S9B). Six of eight PPQR-related positions were strictly conserved during *Plasmodium* evolution (T93, C101, I218, M343, C350, G353; Supplementary Fig. S1).

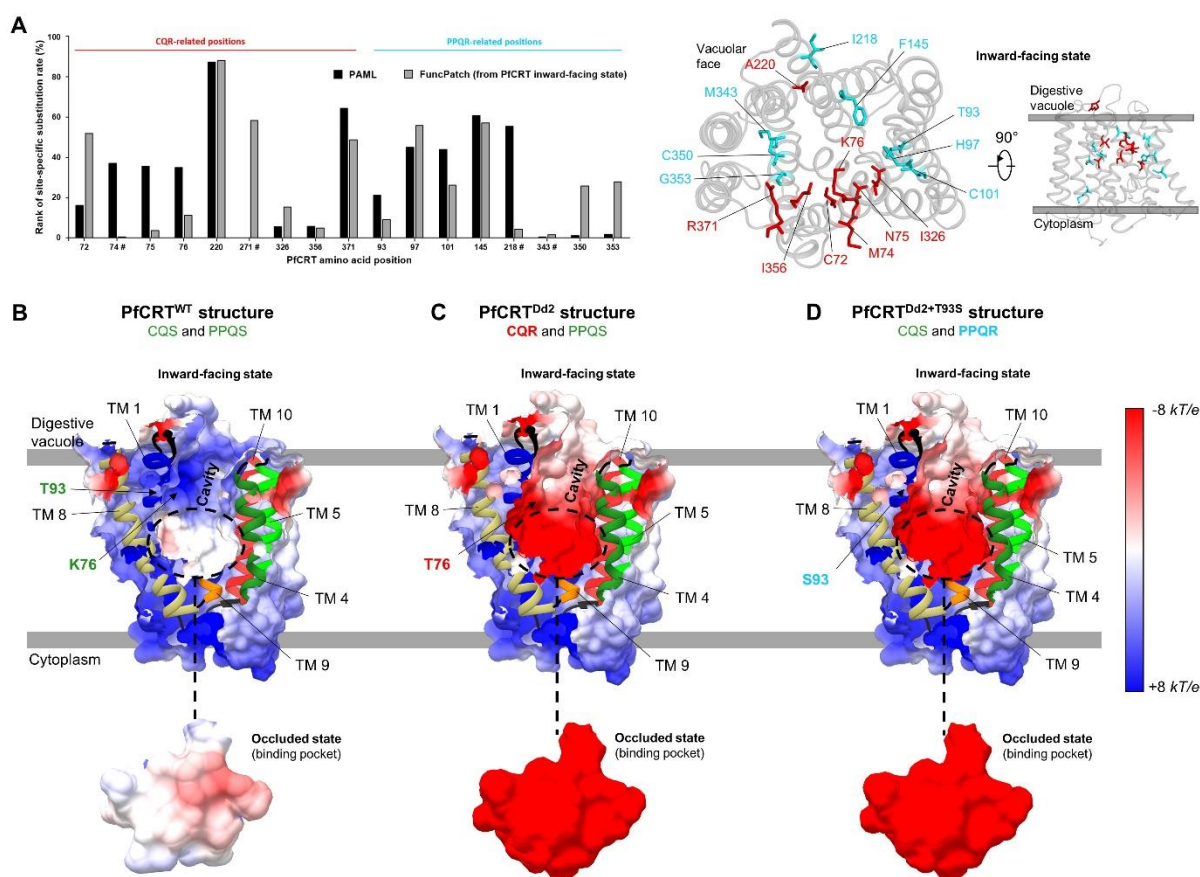


Figure 5. Location and physicochemical effects of drug resistance mutations on the PfCRT structure. (A) *Left*: Conservation level of PfCRT amino acid positions associated with drug resistance. The conservation level is expressed as the rank in percentage of site-specific substitution rates estimated using FuncPatch (inward-facing state model) and PAML. The lower the rank, the higher the conservation level. The # symbol after a PfCRT position indicates that the position is not oriented towards the cavity nor involved in the binding pocket, based on the two PfCRT models. *Right*: Location of PfCRT amino acid positions associated with drug resistance in the PfCRT inward-facing state model. CQR and PPQR mutations are shown in red and cyan, respectively. The electrostatic surface potential of the PfCRT cavity and binding pocket are shown for WT (B), CQR/PPQR (C) and CQS/PPQR (D) structures. Values are in units of kT/e at 298 K, on a scale of $-8 kT/e$ (red) to $+8 kT/e$ (blue). White color indicates a neutral potential. For each structure are indicated: the TMs (colored according to the Figure 2A), the phenotype (CQR, chloroquine resistance; CQS, chloroquine sensitivity; PPQR, piperaquine resistance; PPQS, piperaquine sensitivity) and the mutations introduced in the tertiary structure (only the K76T and T93S mutations are shown for the PfCRT^{Dd2} and PfCRT^{Dd2+T93S} structures for ease of representation).

Altogether, we conclude that PfCRT positions associated with mutations conferring resistance to quinoline-containing drugs are globally conserved, mostly located in the cavity/binding pocket where physiological substrate trafficking is supposed to take place, and make the surface potential of the cavity and binding pocket more electronegative.

Discussion

Experimental tertiary structures of PfCRT are currently lacking. Since 2016, a few DMTs containing 10 TMs have however been experimentally solved^{31,32,39}, which allowed us to predict a PfCRT tertiary structure through template-based homology. Importantly, the two DMTs we used as templates (TPT and Vrg4) are close to PfCRT in the DMT phylogeny³³, making comparative analyses suitable for this set of DMT proteins. Although sequence identity between PfCRT and each of the two DMT templates was $< 15\%$, we identified structural invariants shared by CRT and DMT superfamily profiles, rooting the PfCRT-TPT and PfCRT-

Vrg4 alignments. To our knowledge, few studies have focused on the modeling of PfCRT^{42,43}. Patel and colleagues have predicted a tertiary structure of PfCRT by homology modeling using MFS protein templates which belong to another superfamily of transporters not related to DMT, and report a cavity different in amino acid composition compared to our DMT-based predicted models⁴². Antony *et al.* used different computational methods from ours, but they do not present the architecture of the PfCRT cavity and binding pocket⁴³.

Based on evolutionary and homology-modeling results, PfCRT is on average more conserved on its vacuolar-half side and its membrane-spanning domain, and it contains a vacuolar cavity (in the inward-facing model) and a substrate binding pocket (in the occluded model) in which are located the most conserved positions of the transporter and most of the CQR mutations. The vacuolar side of PfCRT may therefore play an important role in substrate recognition and/or transport, which supports the hypothesis of a physiological substrate being expelled by PfCRT out of the DV. This is in line with the observation that recombinant PfCRT either reconstituted in proteoliposomes¹² or expressed at the surface of *Xenopus laevis* oocytes^{15,20} transports CQ, peptides, basic amino acids, and ferric/ferrous iron. In the parasite, these solutes are known to accumulate in the DV from which they could be transported out either solely by mutant PfCRT (CQ in its di-protonated form) or by WT and/or mutant PfCRT (peptides, basic amino acids, and ferric/ferrous iron).

As evidenced by our joint structural evolutionary analysis, the 10% most conserved amino acid sites were mainly located on TM 9 that runs alongside the open cavity of PfCRT (inward-facing state model) and were also involved in the architecture of the putative substrate-binding pocket (occluded state model). Remarkably, in other DMTs such as Vrg4, TPT and YddG, TM 9 positions contribute to substrate binding^{31,32,39}. By applying to PfCRT positions a very conservative set of criteria (see results), including experimental evidences of transport function in other DMTs, we identified four PfCRT candidate sites in the substrate-binding pocket that may be involved in the transport of physiological substrate(s): Y68, D329, Y345 and S349. None of these positions were found mutated in drug-resistant parasites. Mutations at those positions could be either too deleterious for the parasite physiology and survival or not involved in drug transport. Three other interesting sites in TM 8 (N326) and TM 9 (Q352 and I356) from the highly conserved patch are located in the PfCRT binding pocket. Experimental evidences support their role in PfCRT function. In a photolabeling study using partially purified recombinant PfCRT reconstituted in proteoliposomes, Q352 was shown to interact in concert with K76 and other positions to CQ⁴⁴. Regarding N326 and I356, they were initially mutated in the canonical Dd2 haplotype. Their removal does not significantly alter CQR but enhances parasite fitness in *pfcrt*-modified isogenic parasite lines²⁷. Interestingly, these N326/I356-revertant fitter haplotypes are now predominant in sub-Saharan Africa²⁷.

The modeled PfCRT structures predicted here provide some insights into quinoline-containing drug resistance. The most striking result is that most positions mutated in CQR, including the key K76T, are located in the cavity of PfCRT (in the inward-facing state model), and more particularly in its predicted substrate-binding pocket (in the occluded state model). Several of the mutated positions (C72S, M74I, N75E, N326S/D, I356T/L) are in close proximity with K76T, suggesting that all those mutations work together with K76T. We found that these amino acid changes jointly cause drastic alterations of the electrostatic surface potential in the vacuolar entry of the cavity and in the binding pocket that become highly electronegative in the CQR-like PfCRT mutant structures. This radical shift in electrostatic surface potential is very likely a major contributor to the dramatically increased uptake of di-protonated CQ in *X. laevis* oocytes expressing CQR versus WT PfCRT²⁰ and the decreased accumulation of CQ in the DV of CQR parasites^{6,45}.

In our PfCRT models, most of the novel PPQR mutations that occurred in Dd2 or 7G8 CQR genetic backgrounds^{26,28,40,41} are located in the cavity and binding pocket of the transporter but did not alter their electronegative surface potential compared to CQR mutant structures (Dd2, GB4 and 7G8 haplotypes). This was moderately surprising because some PPQR mutations (such as M343L or G353V) attenuate but do not fully reverse CQR²⁶. Remarkably, Dd2 parasites engineered to express PPQ-sensitive and PPQ-resistant *pfcr*t alleles accumulate similar levels of [³H]-PPQ after 1 h of incubation²⁶. These results contrast with similar earlier experiments for CQ^{9,20} and suggest a different PfCRT-mediated mechanism for PPQ resistance²⁶. For example, PPQR mutations may generate novel binding sites for PPQ, leading to a functional inhibition of certain PfCRT isoforms²⁶. Alternatively, in PPQ-resistant parasites, mutant PfCRT might facilitate or inhibit the transport of molecules that are deleterious or beneficial, respectively, for PPQ toxicity in the DV rather than transporting or binding PPQ itself. Such complementary mechanism has been proposed for CQR based on the selective transport of the protective tripeptide glutathione by mutant PfCRT¹⁴.

Our study has several limitations. First, we did not predict the tertiary structure of the TM 7-TM 8-connecting vacuolar loop (54 amino acid long, positions 264-317) neither the cytoplasmic N-terminal (55 amino acid long, positions 1-55) nor the C-terminal (26 amino acid long, positions 399-424) extremities, which all contain functional sites^{35,38,46-48}. Second, we cannot explain the dynamic evolution of PfCRT haplotypes. While our evolutionary data are based on long-term site-specific substitution rates, PfCRT haplotypes include simultaneously several mutations that act together to alter both drug resistance and parasite fitness. It is well known that the fitness cost and the resistance level may be different for a same amino acid position depending on other mutations in the same protein⁴⁹, including PfCRT²¹. And finally, we provide no novel functional tests to support the accuracy of our PfCRT models. However, our two independent PfCRT models provided convergent results with the identification of an open cavity and a substrate-binding pocket. Several functional studies, in genetically-modified parasites and in heterologous expression systems, already explored the impact of resistance mutations on PfCRT-mediated drug transport. The location of several drug resistance mutations in the open cavity/binding pocket of the PfCRT model provides further support to our results.

In conclusion, using recently solved tertiary structure of DMTs, we performed a comparative structural evolutionary analysis of PfCRT. We pinpointed several candidate amino acid sites in the putative cavity and binding pocket of PfCRT that are likely to play a critical role in substrate transport. The PfCRT models also revealed some features in quinoline-containing drug resistance, especially the likely importance of the electrostatic potential in the efflux of di-protonated CQ. Waiting for an experimental 3D structure of PfCRT, our study provides new testable hypotheses regarding the role of PfCRT in drug resistance and substrate recognition and transport.

Materials and methods

Collection of PfCRT orthologous sequences.

The PfCRT amino acid sequence (PF3D7_0709000) was queried against the specialized *Plasmodium* database (EuPathDB, release 38) and the non-redundant protein database of NCBI using blastp and tblastn searches (e-value cutoff = 10⁻⁶, BLOSUM62 scoring matrix)⁵⁰. Twenty four CRT protein and corresponding cDNA sequences from distinct *Plasmodium* species were retrieved ([Supplementary Table S1](#)). For *P. ovale* spp. and *P. inui* sequences, exon skipping was manually performed so as to reconstruct the full cDNA and amino acid sequences.

Multiple protein and codon sequence alignments.

The CRT protein sequence alignment was first generated with TM-aligner⁵¹ using the transmembrane-specific substitution matrix PHAT⁵² and with gap opening and extension penalties set to 10 and 1, respectively. The output alignment was visually inspected and manually edited using BioEdit v7.2.5⁵³. The positions of the CRT multiple alignment containing gaps in $\geq 30\%$ of all sequences were removed. Then, a *crt* nucleotide alignment was generated with PAL2NAL⁵⁴ using the cleaned CRT protein sequence alignment.

Phylogenetic analysis.

A phylogenetic tree was built from the multiple *crt* nucleotide alignment by using the maximum likelihood method implemented in PhyML v3.0⁵⁵, after determining the best-fitting nucleotide substitution model using the Smart Model Selection (SMS) package⁵⁶. A general time-reversible model with optimized equilibrium frequencies, gamma distributed among-site rate variation and estimated proportion of invariant sites (GTR + G + I) was used, as selected by the Akaike Information Criterion. The nearest neighbor interchange approach implemented in PhyML was chosen for tree improving, and branch supports were estimated using the approximate likelihood ratio aLRT SH-like method⁵⁷. The *crt* phylogenetic tree was rooted using the most distant bird-infecting *Plasmodium* species as outgroup and displayed using the iTOL server⁵⁸.

Analysis of selective pressures acting on *crt*.

To investigate the evolutionary regime that has shaped the CRT-coding DNA sequence during *Plasmodium* evolution, the nucleotide sequence alignment and the maximum likelihood phylogenetic tree were submitted to the *codeml* tool of PAML v.4.9h^{59,60}. The level of coding sequence conservation was measured by calculating the non-synonymous (d_N) to synonymous (d_S) substitution rate ratio ω ($= d_N/d_S$). Theoretically, positive selection at codon sites is indicated by $\omega > 1$, while codon sites associated with $\omega = 1$ and $\omega < 1$ suggest neutral evolution and purifying selection, respectively. We investigated the heterogeneity of ω among lineages and along the *crt* sequence; then we searched for codons evolving under positive selection using a set of random-site and branch models (free-ratio, M0, M1a, M2a, M3, M7 and M8)^{61,62}. Model M0 supposes the same d_N/d_S ratio for all branches of the phylogeny, while the “free-ratio” branch model assumes independent d_N/d_S ratio for every branch. Model M1a allows two classes of sites under neutral evolution ($\omega = 1$) and purifying selection ($0 \leq \omega < 1$), whereas model M2a adds a third class of sites to model M1a as positive selection ($\omega > 1$). In the discrete model M3, several classes ($k = [3, 4, 5]$ in this study) of independent ω were estimated, each of them being associated with a specific proportion. Model M7 assumes a β -distribution of ten ω ratios limited to the interval $[0, 1]$ with two shape parameters p and q . In model M8, an additional site class is estimated, with ω possibly > 1 as M2a does. The heterogeneity of ω among codon sites was tested by comparing model M0 to model M3, while paired models M1a:M2a and M7:M8 allow to detect positively selected sites⁶¹. Candidate sites for positive selection were identified in M2a and M8 models using the Bayes empirical Bayes inference (BEB)⁶³, which calculates the posterior probability that every codon site belongs to a site class affected by positive selection; and using the naïve empirical Bayes (NEB) inference in model M3 (in which no BEB approach is implemented yet), considered as less powered than the BEB inference. For models considering k classes of d_N/d_S ratio, we used the mean ω value at each codon site, calculated as the sum of the ω values of each k class weighted by their estimated probabilities (Supplementary Data S1)⁶⁰.

Statistical analyses based on structural information of the PfCRT models were performed using the partitioning function implemented in *codeml*⁶⁴ in order to test i) whether the

vacuolar-half side is more or less conserved than the cytoplasmic-half side; and *ii*) whether TMs are more or less conserved than non-TM (or loops). Partitionings are based on fixed-site models A to E. Model A – that corresponds to model M0 – assumes identical parameters for transition/transversion (κ) rate ratio, branch lengths and equilibrium nucleotide frequencies, and unique ω ratio for partitions. Model B allows homogeneity among partitions in codon frequencies, κ and ω ratios but different rates of evolution, whereas model C further accounts for potential variation in codon frequency. Model D uses different κ and ω ratios based on the same codon frequencies for partitions, and model E ultimately allows different κ , ω and codon frequencies between partitions. The A:B comparison tests for different nucleotide substitution rates among partitions, while both B:D and C:E comparisons evaluate the variation in κ , ω , and codon frequencies among partitions⁶⁴.

The comparisons of models were performed using Likelihood Ratio Tests (LRTs)⁶⁵. For every LRT, twice the log-likelihood difference between alternative and null model ($2\Delta\ell$) was compared to critical values from a chi-squared distribution with degrees of freedom equal to the difference in the number of estimated parameters between both models⁶⁶.

Modeling the PfCRT structure.

Since no experimental structure has yet been determined for PfCRT, we aimed at modeling it in different conformational states using a template-based protein structure modeling approach. The PfCRT sequence was searched against the Phyre2 server to find the appropriate templates. A nucleotide sugar GDP-mannose transporter Vrg4 (outward-facing state, PDB ID: 5oge)³¹ and a triose-phosphate/phosphate translocator (occluded state, PDB ID: 5y79)³² were identified as sharing putative similar structural folding with PfCRT with confidence > 98%. The very low sequence identity of these proteins with PfCRT, reaching only 15% with the triose-phosphate/phosphate translocator, makes homology modeling challenging but still feasible (Supplementary Table S7)⁶⁷. In homology modeling, the sequence alignment of the target (here PfCRT) and the template (here other DMT proteins) is a critical step. Because hydrophobicity reflects the transmembrane topology and is typically conserved in TMs during evolution, we first produced a superfamily-averaged hydropathy profile alignment with AlignMe³⁷ using the CRT multiple sequence alignment and an alignment of different members of the DMT superfamily. The profile (CRT)-profile (DMT) alignment pinpointed some amino acids which are supposed to be important to align in the final target-template sequence alignment (Supplementary Figs. S4 and S5). Finally, the target-template alignment was produced using the AlignMe PST algorithm, designed for distantly related proteins (*i.e.* with a sequence identity <15%)³⁷, then manually optimized it in order to align the putative important amino acid positions identified from the superfamily-averaged hydropathy profile alignment (Supplementary Figs. S4 and S5). The long TM 7-TM 8 vacuolar loop and the cytoplasmic N- and C-terminal extremities, were not considered for modeling because of the lack of sufficient amino acid positions from templates covering them. Initially, we built 1,000 3D models of each of PfCRT state satisfying the spatial restraints of the template structures using MODELLER 9.17⁶⁸. The best models among them were selected based on scores calculated from discrete optimized protein energy (DOPE) and GA341 functions. Then, the best PfCRT occluded and inward-facing state models were submitted to GalaxyRefine⁶⁹ for atomic-level, high-resolution refinement, which helped to achieve significant improvement in physical quality of the local structure. Finally, the refined models were subjected to an energy minimization to get the most stable conformations using the YASARA server⁷⁰.

The refined, energy-minimized PfCRT models were validated using MolProbity⁷¹ for local stereochemistry, and ERRAT⁷², ProSA II⁷³ and QMEANBrane⁷⁴ for tertiary folding and global 3D quality metrics. The PDB files of the refined, energy-minimized PfCRT

occluded and inward-facing state structures are provided in [Supplementary Files S1](#) and [S2](#), respectively.

Analyses of structural features.

Protein electrostatic surface potential was calculated using Adaptive Poisson-Boltzmann Solver (APBS)⁷⁵, after determining the per-atom charge and radius of the structure with PDB2PQR v.2.1.1⁷⁶. The Poisson-Boltzmann equation was solved at 298 K using a grid-based method, with solute and solvent dielectric constants fixed at 2 and 78.5, respectively. We used a scale of $-8\text{ kT}/e$ to $+8\text{ kT}/e$ to map the electrostatic surface potential in a radius of 1.4 Å.

Putative functional patches in the two PfCRT models, *i.e.* highly conserved amino acid positions that are in close physical proximity in the tertiary structure, were searched using the FuncPatch server by estimating site-specific substitution rates, possibly spatially correlated⁷⁷. We submitted to the server the PfCRT models along with the CRT protein sequence alignment and the *crt* phylogenetic tree. The spatial correlation of the site-specific amino acid substitution rates in the PfCRT predicted structures was tested using a Bayesian model comparison, where a null model (model 0), in which no spatial correlation of site-specific substitution rates is present, is compared to the alternative model (model 1). As suggested by FuncPatch' authors, a spatial correlation was considered as significant if the estimated log Bayes factor (model 1 versus model 0) was larger than the conservative cutoff value of 8⁷⁷.

Identifying candidate functional sites of PfCRT.

PfCRT amino acid positions involved in physiological substrate(s) trafficking remain unknown. In order to identify candidate functional amino acid sites, a comparative evolutionary and structural analysis was performed to filter step-by-step PfCRT amino acid sites. Because functional amino acids are often found very conserved over evolutionary time⁷⁸, only the strictly conserved amino acid positions of PfCRT during *Plasmodium* evolution were kept. From this subset of positions, we then focused on those involved in the putative substrate-binding pocket in the occluded state model of PfCRT since it is widely accepted that DMT proteins transport substrates through binding sites located in their pocket. Finally, the *reference* DMT template structures used to predict the conformational states of PfCRT tertiary structure (inward-facing state: Vrg4, PDB ID: 5oge; occluded state: TPT, PDB ID: 5y79)^{31,32}, in addition with another DMT protein (inward-facing state: YddG, PDB ID: 5i20)³⁹ were aligned with the PfCRT models. By focusing on the amino acids from these *reference* structures that have been shown to be important or essential for substrate trafficking, we listed PfCRT candidate positions that might contribute to PfCRT substrate trafficking ([Supplementary Fig. S10](#)).

In silico production of mutant PfCRT structures.

Some *in silico* mutagenesis from the refined, energy-minimized PfCRT models were performed. We generated three PfCRT mutant-like models per PfCRT conformational state based on the Dd2, GB4 and 7G8 haplotypes which are associated with CQR ([Table 3](#))^{9,79}. The CQ-resistant Dd2- and 7G8-like PfCRT structures were then subjected to individual additional mutations that attenuate CQR but confer PPQR ([Table 3](#))^{26,28,40,41}. Mutations were introduced using the *swapa* function of UCSF Chimera by substituting the residue with the most probable rotameric conformation⁸⁰.

Structure visualization

Molecular drawings were produced using UCSF Chimera⁸⁰.

References

1. Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**, 320–323 (2002).
2. WHO | World malaria report 2018. WHO Available at: <http://www.who.int/malaria/publications/world-malaria-report-2018/report/en/>. (Accessed: 5th August 2019)
3. Milani, K. J., Schneider, T. G. & Taraschi, T. F. Defining the morphology and mechanism of the hemoglobin transport pathway in *Plasmodium falciparum*-infected erythrocytes. *Eukaryotic Cell* **14**, 415–426 (2015).
4. Goldberg, D. E. Hemoglobin degradation in *Plasmodium*-infected red blood cells. *Semin. Cell Biol.* **4**, 355–361 (1993).
5. Yayon, A., Cabantchik, Z. I. & Ginsburg, H. Identification of the acidic compartment of *Plasmodium falciparum*-infected human erythrocytes as the target of the antimalarial drug chloroquine. *EMBO J.* **3**, 2695–2700 (1984).
6. Krogstad, D. J., Gluzman, I. Y., Herwaldt, B. L., Schlesinger, P. H. & Wellems, T. E. Energy dependence of chloroquine accumulation and chloroquine efflux in *Plasmodium falciparum*. *Biochem. Pharmacol.* **43**, 57–62 (1992).
7. Geng, Y., Kohli, L., Klocke, B. J. & Roth, K. A. Chloroquine-induced autophagic vacuole accumulation and cell death in glioma cells is p53 independent. *Neuro-oncology* **12**, 473–481 (2010).
8. Su, X., Kirkman, L. A., Fujioka, H. & Wellems, T. E. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant *P. falciparum* in Southeast Asia and Africa. *Cell* **91**, 593–603 (1997).
9. Fidock, D. A. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6**, 861–871 (2000).
10. Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360**, (2018).
11. Martin, R. E. & Kirk, K. The malaria parasite's chloroquine resistance transporter is a member of the drug/metabolite transporter superfamily. *Mol. Biol. Evol.* **21**, 1938–1949 (2004).
12. Juge, N. *et al.* *Plasmodium falciparum* chloroquine resistance transporter is a H⁺-coupled polyspecific nutrient and drug exporter. *Proc Natl Acad Sci U S A* **112**, 3356–3361 (2015).
13. Maughan, S. C. *et al.* Plant homologs of the *Plasmodium falciparum* chloroquine-resistance transporter, PfCRT, are required for glutathione homeostasis and stress responses. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2331–2336 (2010).
14. Patzewitz, E.-M. *et al.* Glutathione transport: a new role for PfCRT in chloroquine resistance. *Antioxid. Redox Signal.* **19**, 683–695 (2013).
15. Bakouh, N. *et al.* Iron is a substrate of the *Plasmodium falciparum* chloroquine resistance transporter PfCRT in *Xenopus* oocytes. *J. Biol. Chem.* **292**, 16109–16121 (2017).
16. Zhang, H., Howard, E. M. & Roepe, P. D. Analysis of the antimalarial drug resistance protein PfCRT expressed in yeast. *J. Biol. Chem.* **277**, 49767–49775 (2002).
17. Zhang, H., Paguio, M. & Roepe, P. D. The antimalarial drug resistance protein *Plasmodium falciparum* chloroquine resistance transporter binds chloroquine. *Biochemistry* **43**, 8290–8296 (2004).
18. Nessler, S. *et al.* Evidence for activation of endogenous transporters in *Xenopus laevis* oocytes expressing the *Plasmodium falciparum* chloroquine resistance transporter, PfCRT. *J. Biol. Chem.* **279**, 39438–39446 (2004).
19. Wellems, T. E., Hayton, K. & Fairhurst, R. M. The impact of malaria parasitism: from corpuscles to communities. *J. Clin. Invest.* **119**, 2496–2505 (2009).
20. Martin, R. E. *et al.* Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* **325**, 1680–1682 (2009).
21. Gabryszewski, S. J. *et al.* Evolution of Fitness Cost-Neutral Mutant PfCRT Conferring *P. falciparum* 4-Aminoquinoline Drug Resistance Is Accompanied by Altered Parasite Metabolism and Digestive Vacuole Physiology. *PLOS Pathogens* **12**, e1005976 (2016).
22. Lewis, I. A. *et al.* Metabolic QTL analysis links chloroquine resistance in *Plasmodium falciparum* to impaired hemoglobin catabolism. *PLoS Genet.* **10**, e1004085 (2014).

23. Sidhu, A. B. S., Verdier-Pinard, D. & Fidock, D. A. Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by *pfcr*t mutations. *Science* **298**, 210–213 (2002).
24. Cooper, R. A. *et al.* Alternative mutations at position 76 of the vacuolar transmembrane protein PfCRT are associated with chloroquine resistance and unique stereospecific quinine and quinidine responses in *Plasmodium falciparum*. *Mol. Pharmacol.* **61**, 35–42 (2002).
25. Valderramos, S. G. *et al.* Identification of a mutant PfCRT-mediated chloroquine tolerance phenotype in *Plasmodium falciparum*. *PLoS Pathog.* **6**, e1000887 (2010).
26. Ross, L. S. *et al.* Emerging Southeast Asian PfCRT mutations confer *Plasmodium falciparum* resistance to the first-line antimalarial piperazine. *Nature Communications* **9**, 3314 (2018).
27. Dhingra, S. K. *et al.* Global Spread of Mutant PfCRT and Its Pleiotropic Impact on *Plasmodium falciparum* Multidrug Resistance and Fitness. *mBio* **10**, e02731-18 (2019).
28. van der Pluijm, R. W. *et al.* Determinants of dihydroartemisinin-piperazine treatment failure in *Plasmodium falciparum* malaria in Cambodia, Thailand, and Vietnam: a prospective clinical, pharmacological, and genetic study. *Lancet Infect Dis* (2019). doi:10.1016/S1473-3099(19)30391-3
29. Bellanca, S. *et al.* Multiple drugs compete for transport via the *Plasmodium falciparum* chloroquine resistance transporter at distinct but interdependent sites. *J. Biol. Chem.* **289**, 36336–36351 (2014).
30. Richards, S. N. *et al.* Molecular Mechanisms for Drug Hypersensitivity Induced by the Malaria Parasite's Chloroquine Resistance Transporter. *PLoS Pathog.* **12**, e1005725 (2016).
31. Parker, J. L. & Newstead, S. Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature* **551**, 521–524 (2017).
32. Lee, Y. *et al.* Structure of the triose-phosphate/phosphate translocator reveals the basis of substrate specificity. *Nat Plants* **3**, 825–832 (2017).
33. Yen, M. R., Chen, J. S., Marquez, J. L., Sun, E. I. & Saier, M. H. Multidrug resistance: phylogenetic characterization of superfamilies of secondary carriers that include drug exporters. *Methods Mol. Biol.* **637**, 47–64 (2010).
34. Carlton, J. M. Evolution of human malaria. *Nat Microbiol* **3**, 642–643 (2018).
35. Kuhn, Y. *et al.* Trafficking of the phosphoprotein PfCRT to the digestive vacuolar membrane in *Plasmodium falciparum*. *Traffic* **11**, 236–249 (2010).
36. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858 (2015).
37. Stamm, M., Staritzbichler, R., Khafizov, K. & Forrest, L. R. AlignMe—a membrane protein sequence alignment web server. *Nucleic Acids Res.* **42**, W246-251 (2014).
38. Lee, A. H. *et al.* Evidence for Regulation of Hemoglobin Metabolism and Intracellular Ionic Flux by the *Plasmodium falciparum* Chloroquine Resistance Transporter. *Sci Rep* **8**, 13578 (2018).
39. Tsuchiya, H. *et al.* Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature* **534**, 417–420 (2016).
40. Dhingra, S. K. *et al.* A Variant PfCRT Isoform Can Contribute to *Plasmodium falciparum* Resistance to the First-Line Partner Drug Piperazine. *MBio* **8**, (2017).
41. Pelleau, S. *et al.* Adaptive evolution of malaria parasites in French Guiana: Reversal of chloroquine resistance by acquisition of a mutation in *pfcr*t. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11672–11677 (2015).
42. Patel, S. K. *et al.* Molecular interaction of selected phytochemicals under the charged environment of *Plasmodium falciparum* chloroquine resistance transporter (PfCRT) model. *J. Biomol. Struct. Dyn.* **34**, 290–303 (2016).
43. Antony, H. A. *et al.* In silico modeling of *Plasmodium falciparum* chloroquine resistance transporter protein and biochemical studies suggest its key contribution to chloroquine resistance. *Acta Trop.* **189**, 84–93 (2019).
44. Lekostaj, J. K., Natarajan, J. K., Paguio, M. F., Wolf, C. & Roepe, P. D. Photoaffinity labeling of the *Plasmodium falciparum* chloroquine resistance transporter with a novel perfluorophenylazido chloroquine. *Biochemistry* **47**, 10394–10406 (2008).
45. Sanchez, C. P., Stein, W. & Lanzer, M. Trans stimulation provides evidence for a drug efflux carrier as the mechanism of chloroquine resistance in *Plasmodium falciparum*. *Biochemistry* **42**, 9383–9394 (2003).
46. Sanchez, C. P. *et al.* Phosphomimetic substitution at Ser-33 of the chloroquine resistance transporter PfCRT reconstitutes drug responses in *Plasmodium falciparum*. *J. Biol. Chem.* (2019). doi:10.1074/jbc.RA119.009464

47. Pulcini, S. *et al.* Mutations in the Plasmodium falciparum chloroquine resistance transporter, PfCRT, enlarge the parasite's food vacuole and alter drug sensitivities. *Sci Rep* **5**, 14552 (2015).
48. Jones, M. L., Collins, M. O., Goulding, D., Choudhary, J. S. & Rayner, J. C. Analysis of protein palmitoylation reveals a pervasive role in Plasmodium development and pathogenesis. *Cell Host Microbe* **12**, 246–258 (2012).
49. Lozovsky, E. R. *et al.* Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12025–12030 (2009).
50. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
51. Bhat, B., Ganai, N. A., Andrabi, S. M., Shah, R. A. & Singh, A. TM-Aligner: Multiple sequence alignment tool for transmembrane proteins with reduced time and improved accuracy. *Sci Rep* **7**, 12543 (2017).
52. Ng, P. C., Henikoff, J. G. & Henikoff, S. PHAT: a transmembrane-specific substitution matrix. Predicted hydrophobic and transmembrane. *Bioinformatics* **16**, 760–766 (2000).
53. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98 (1999).
54. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* **34**, W609–W612 (2006).
55. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
56. Lefort, V., Longueville, J.-E. & Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* **34**, 2422–2424 (2017).
57. Anisimova, M., Bielawski, J. P. & Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**, 1585–1592 (2001).
58. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**, W242–W245 (2016).
59. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
60. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
61. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998).
62. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573 (1998).
63. Yang, Z., Wong, W. S. W. & Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118 (2005).
64. Yang, Z. & Swanson, W. J. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19**, 49–57 (2002).
65. Vuong, Q. H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **57**, 307–333 (1989).
66. Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449 (2000).
67. Nikolaev, D. M. *et al.* A Comparative Study of Modern Homology Modeling Algorithms for Rhodopsin Structure Prediction. *ACS Omega* **3**, 7555–7566 (2018).
68. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Bioinformatics* **54**, 5.6.1–5.6.37 (2016).
69. Heo, L., Park, H. & Seok, C. GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Res.* **41**, W384–388 (2013).
70. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114–122 (2009).
71. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).
72. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993).
73. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**, W407–410 (2007).

74. Studer, G., Biasini, M. & Schwede, T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* **30**, i505-511 (2014).
75. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
76. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665-667 (2004).
77. Huang, Y.-F. & Golding, G. B. FuncPatch: a web server for the fast Bayesian inference of conserved functional patches in protein 3D structures. *Bioinformatics* **31**, 523–531 (2015).
78. Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109–121 (2016).
79. Durrand, V. *et al.* Variations in the sequence and expression of the *Plasmodium falciparum* chloroquine resistance transporter (PfCRT) and their relationship to chloroquine resistance in vitro. *Mol. Biochem. Parasitol.* **136**, 273–285 (2004).
80. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

We are extremely grateful to I. Callebaut for her expert guidance and advices for the homology modeling analyses and for her critical review of the manuscript. We thank Y-F. Huang for help with FuncPatch analysis, R. Coucke for early work on the project, and R. Duval for discussions. R. Coppée is supported by PhD grant funding from the doctoral school MTCI n°563.

Author contributions

R.C., A.S. and J.C. designed the research. A.S. and J.C. supervised the whole work. R.C. performed the building of datasets, statistical and evolutionary analyses, homology modeling and structural visualization. All the authors participated in interpretations and wrote the manuscript.

Competing Interests

The authors declare no competing interests.

- CRT est très conservée à l'échelle des *Plasmodium*, et plus particulièrement la moitié vacuolaire de la protéine.
- Des structures expérimentales de protéines DMT ont été utilisées pour prédire la structure tertiaire de PfCRT selon deux états conformationnels différents : un état ouvert côté vacuolaire et un état fermé. Les modèles PfCRT arborent de fortes similitudes avec les structures DMT parentes, et sont composés de 10 hélices transmembranaires comme attendu.
- Les modèles PfCRT ont mis en évidence la présence d'une cavité, et plus spécifiquement d'une poche de liaison au cœur de la protéine où aurait lieu l'interaction avec un (ou plusieurs) substrat(s) physiologique(s), par homologie avec les DMTs.
- La cavité de PfCRT, et plus particulièrement l'hélice 9 (souvent impliquée dans la liaison au substrat dans d'autres DMTs) est enrichie en sites très conservés.
- Les sites Y68, D329, Y345 et S349 de PfCRT, strictement conservés à l'échelle des *Plasmodium* et impliqués dans l'architecture de la poche, sont prédits comme ayant un rôle clé dans la fonction de transport de substrats physiologiques de PfCRT.
- Les mutations de résistance à la CQ sont majoritairement localisées dans la cavité et impliquées dans l'architecture de la poche. Elles altèrent fortement le potentiel électrostatique à la surface de la cavité, passant d'un état neutre à un état très négatif.
- Les mutations de résistance à la PPQ sont principalement localisées dans la cavité et dans la poche. Elles conservent le potentiel électronégatif induit par les mutations de résistance à la CQ.

Partie 4

Discussion



La seule voie qui offre quelque espoir d'un avenir meilleur pour toute l'humanité est celle de la coopération et du partenariat.

— **Kofi Annan** (1938 – 2018)

Ce travail de thèse a tenté de répondre à des questions biologiques par des approches bioinformatiques. Au cours de nos analyses, nous nous sommes questionnés sur les méthodes employées et les algorithmes utilisés. Dans cette section, nous expliquerons d'abord comment nous avons interprété certains concepts ayant permis de détecter de possibles zones fonctionnelles au sein de PfK13 et PfCRT. Puis, nous proposerons quelques pistes méthodologiques bioinformatiques qui auraient pour objectifs d'optimiser et/ou compléter – à notre sens – les prédictions de sites fonctionnels. Les pistes expérimentales seront évoquées dans la seconde section.

4.1.1. Sélection purificatrice : contrainte structurale ou fonctionnelle ?

Un de nos objectifs majeurs fut de prédire les régions et sites fonctionnels de Pfk13 et PfcRT par détection des sites évoluant sous sélection purificatrice. Toutefois, la sélection purificatrice relative aux sites peut être liée dans notre contexte principalement à deux types de contrainte : i) des contraintes structurales si les sites ont un rôle essentiel dans le repliement structural et/ou le maintien de la structure ; et/ou ii) des contraintes fonctionnelles si les sites sont impliqués dans la fonction de la protéine (par exemple, ceux participant à l'interaction avec un substrat). De fait, un site évoluant sous une intense sélection purificatrice ne signifie pas nécessairement son implication dans la fonction de la protéine.

Initialement, nos travaux ont été uniquement menés avec l'outil codeml de PAML qui évalue les taux de substitution au niveau des codons. L'étude des sites les plus conservés (au seuil 10%) a révélé qu'il s'agissait principalement des acides aminés signatures de la famille structurale considéré. Par exemple, les sites les plus conservés au sein du domaine KREP de Pfk13 correspondaient principalement aux diglycines (GG) après le second brin (ou brin B), et les glutamates (E) sur le troisième brin (ou brin C), essentiels au bon repliement du KREP^{107,108} (**Figure 49**). Ce résultat révèle que l'outil PAML n'est pas nécessairement le plus adapté pour rechercher les sites les plus conservés, possiblement « fonctionnels », au sein de protéines fortement conservées comme peuvent l'être Pfk13 et PfcRT. Ce constat n'est toutefois pas surprenant puisque l'outil a été conçu pour détecter des sites évoluant sous sélection positive.

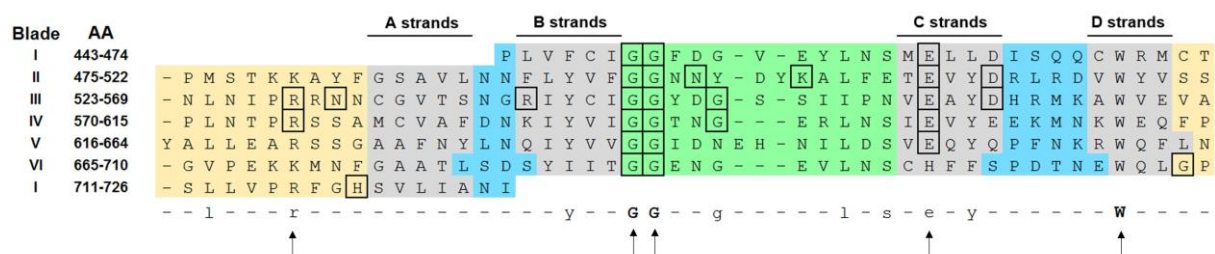


Figure 49 : Mise en évidence des sites les plus conservés du domaine KREP de Pfk13 avec l'outil codeml de PAML. Les motifs kelch (blades) ont été structuralement alignés pour produire l'alignement multiple des séquences. Une séquence consensus de l'alignement est indiquée en dessous de celui-ci. Les sites les plus conservés (au seuil 10%) d'après l'outil codeml sont entourés en noir. Les flèches indiquent les acides aminés importants pour le bon repliement d'un domaine KREP^{107,108}.

Par la suite, nous avons découvert l'outil FuncPatch, récemment développé¹⁹¹. À la différence de PAML, FuncPatch a été conçu pour rechercher les sites les plus conservés d'une protéine. Cet outil considère l'environnement spatial de chaque acide aminé dans la structure tertiaire pour estimer les taux de substitution spécifiques aux sites, là où d'autres programmes comme ConSurf ne font que calculer les taux de substitution spécifiques aux sites à partir de l'alignement multiple des séquences, puis les cartographier sur la structure tertiaire²⁷³. FuncPatch peut ainsi identifier des *patches* de sites conservés, c'est-à-dire un ensemble de sites spatialement proches dans la structure tertiaire et soumis à une forte contrainte structurale et/ou fonctionnelle.

FuncPatch est supérieur à des méthodes comme PAML uniquement s'il existe une corrélation spatiale des taux de substitution spécifiques aux sites au sein des protéines étudiées. Ce fut notamment le cas pour les protéines PfK13 et PfCRT, permettant une définition plus fine des sites sous sélection purificatrice. Pour renforcer l'hypothèse que le *patch* identifié est fonctionnel, nous avons réalisé des analyses évolutives et structurales au sein de domaines / protéines appartenant à la même famille structurale (protéines de *référence*) et pour lesquelles des données fonctionnelles étaient disponibles. Si les *patches* de sites conservés des protéines d'intérêt et de *référence* se superposent et que celui des protéines / domaines de *référence* contient des sites fonctionnels validés par des expériences, il est très probable que le *patch* de la protéine d'intérêt évoluant sous une forte sélection purificatrice ait un rôle dans la fonction de la protéine plutôt que dans son repliement.

4.1.2. Une nouvelle option pour la détection de *patches* de sites conservés

L'identification d'un *patch* de sites conservés au sein des protéines, telle que réalisée par l'outil FuncPatch, nous semble cependant perfectible. L'algorithme FuncPatch recherche l'existence d'une corrélation des taux de substitution spécifiques aux sites en utilisant les atomes de Ca de chaque acide aminé pour définir les distances physiques entre chaque paire d'acides aminés^{191,192}. De fait, la méthodologie ignore totalement l'orientation des chaînes latérales des acides aminés dans la structure tertiaire et leur possible proximité physique, alors que les chaînes latérales portent une partie de l'information physico-chimique des acides aminés. Des études ont montré que les chaînes latérales des acides aminés ont un lien avec les taux de substitution spécifiques aux sites. Entre autres, le taux de contacts inter-atomiques lié à l'entourage d'un acide aminé n'est corrélé à son taux de substitution qu'en considérant sa chaîne latérale²²³.

Si on considère par exemple les protéines transmembranaires, calculer les distances par paire d'acides aminés sur la base des atomes de Ca pourrait avoir un effet important sur le résultat final : au lieu de détecter une zone fortement conservée impliquant des chaînes latérales proches provenant d'acides aminés localisés sur différentes hélices plus ou moins distantes, l'algorithme actuel pourrait identifier préférentiellement des acides aminés issus d'une même hélice, car physiquement plus proches dans l'espace sur la base des atomes de Ca (**Figure 50A, panel gauche**). Comme analyse préliminaire, nous avons rédigé un algorithme qui calcule les distances physiques pour toutes les paires d'acides aminés (ici sur PfCRT) en se basant soit sur les atomes de Ca, soit sur les chaînes latérales. Pour simplifier les calculs, la chaîne latérale a été associée au centre de masse de l'acide aminé. En considérant uniquement les distances ≤ 15 Å, nous avons observé qu'une multitude de distances estimées sur les atomes de Ca étaient plus élevées que celles calculées à partir des chaînes latérales (**Figure 50B, encadré vert**), et inversement (**Figure 50B, encadré violet**). Une liste de ces paires d'acides aminés « discordantes » sera par la suite réalisée pour déterminer et analyser leur localisation dans la structure tertiaire.

Si par exemple plusieurs se trouvent dans la cavité ou la poche de PfCRT, cela pourrait peut-être avoir un effet sur la définition du *patch* identifié par FuncPatch. Pour confirmer *in fine* le possible gain de précision dans la détection et la délimitation de *patches* de sites conservés en considérant les chaînes latérales, plusieurs jeux de données pourraient être préparés couvrant divers types de protéines pour lesquelles des données fonctionnelles (sites critiques dans la fonction de la protéine) sont disponibles. Le contenu en sites fonctionnels pourrait être comparé entre les *patches* obtenus selon les deux approches *chaîne latérale* ou *Ca*, et ainsi vérifier quelle méthode est la plus précise (**Figure 50A**).

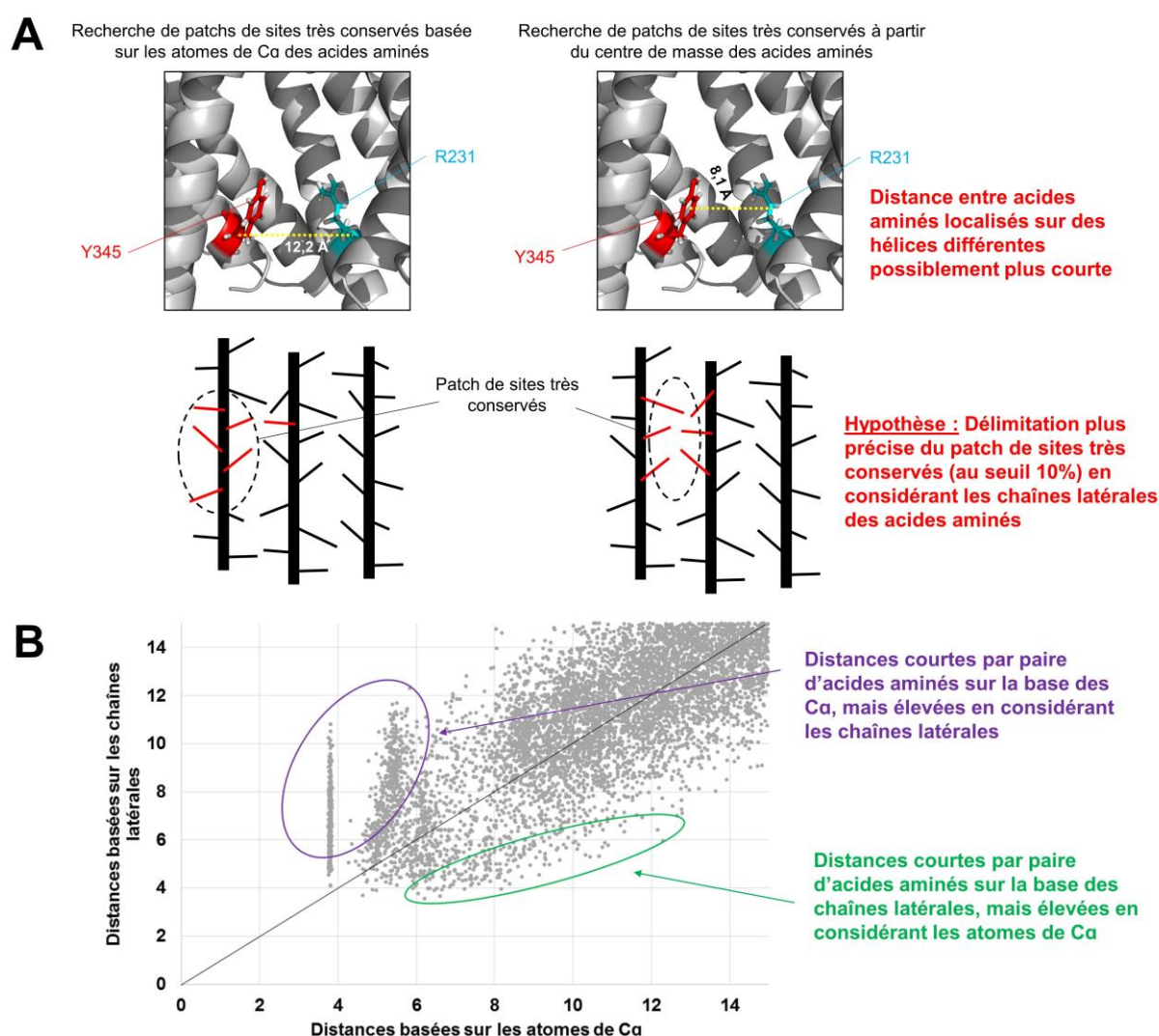


Figure 50 : Proposition d'une possible amélioration de l'outil FuncPatch dans la recherche de *patches* de sites conservés. Cet exemple est basé sur le modèle prédit de PfCRT (état *occluded*). (A) Sur le panel de gauche, le calcul de la distance basée sur les atomes de Ca entre les sites conservés Y345 et R230 (respectivement localisés sur les hélices 9 et 6) de PfCRT est égale à 12,2 Å. En revanche, si on considère l'orientation des chaînes latérales, la distance entre les deux sites est, dans cet exemple, estimée à 8,1 Å (panel de droite). (B) Variation des distances par paire d'acides aminés de PfCRT en considérant soit les chaînes latérales (ordonnées), soit les atomes de Ca (abscisses). Un point gris représente une paire d'acides aminés. Les distances ont été calculées avec un algorithme développé avec le langage python.

4.1.3. Étude comparative des niveaux de sélection purificatrice aux échelles inter- et intra-espèces

Dans ce travail de thèse, nous nous sommes focalisés sur des analyses évolutives inter-espèces (PfK13, *Apicomplexa* ; PfCRT, *Plasmodium*). Toutefois, nous avons également prévu d'investiguer les pressions de sélection agissant sur les gènes *pfk13* et *pfcr1* à l'échelle intra-espèce (*P. falciparum*). En effet, il aurait été intéressant de confirmer les résultats obtenus en inter-espèces avec des données intra-espèces – préférentiellement en absence de pression exercées par les médicaments – pour avoir un reflet de l'évolution naturelle de ces gènes directement au sein de l'espèce d'intérêt, ici *P. falciparum*. Une étude s'est notamment intéressée à analyser la variation des niveaux de sélection purificatrice au sein de récepteurs peu tolérants à la mutation et impliqués dans le potentiel post-synaptique excitateur, à partir de 141 352 séquences codantes du gène (*Homo sapiens*)²⁷⁴. En utilisant une fenêtre glissante de 31 codons le long de la séquence, il a été calculé par position le rapport entre le nombre observé de mutations non-synonymes et le nombre attendu de mutations non-synonymes. Par cette approche, les auteurs ont démontré quelles sont les régions des récepteurs qui ne tolèrent pas / peu la mutation, pouvant suggérer leur apport fonctionnel et/ou structural au sein de ces protéines.

Nous n'avons pas pu mettre en place ce système de façon robuste en raison du faible nombre de séquences *P. falciparum* *pfk13* / *pfcr1* disponibles publiquement. Nos premières investigations, sur *pfk13*, ont permis de recenser 3 247 séquences (obtenues à partir du projet communautaire *P. falciparum*¹¹⁵), parmi lesquelles 1 648 étaient d'origine africaine – où il est rapporté l'absence de sélection positive sur ce gène en Afrique¹⁴⁴ – couvrant seulement 22 évènements de substitution non-synonyme.

Alternativement, nous aurions pu directement analyser les taux de substitution intra-espèce spécifiques aux sites via PAML et FuncPatch. Cela n'a toutefois pas été investigué car ces méthodes requièrent des évènements de substitution aux sites suffisamment nombreux pour inférer des taux d'évolution robustes^{191,195}. Ces travaux sur la détection de sites très conservés au sein des populations de *P. falciparum* pourront être approfondis lorsque des échantillonnages beaucoup plus étoffés seront disponibles. Cela devrait être concevable étant donné la diminution des coûts relatifs à la production de données NGS.

- La sélection purificatrice relative aux sites peut s'expliquer par une contrainte fonctionnelle et/ou structurale. Nous pouvons privilégier l'une ou l'autre via l'apport de données fonctionnelles issues de protéines ou domaines appartenant à la même famille structurale.
- L'outil FuncPatch, utilisé pour délimiter des *patches* de sites conservés au sein d'une protéine, pourrait peut-être être enrichi en considérant l'orientation des chaînes latérales au lieu des atomes de Ca dans le calcul des distances par paire d'acides aminés.
- Une approche intra-espèce pour mesurer la variation du niveau de sélection purificatrice au sein d'une protéine est limitée chez *P. falciparum* par le nombre limité de séquences disponibles, comparativement à d'autres espèces.

Quel avenir pour les protéines PfK13 et PfCRT ?

Section 4.2

” *Il ne faut avoir aucun regret pour le
passé, aucun remords pour le présent, et
une confiance inébranlable pour l'avenir.*

— **Jean Jaurès** (1859 – 1914)

Nos analyses structurales et évolutives de PfK13 et PfCRT supportent, pour chacune de ces deux protéines, le rôle fonctionnel d'une poche très conservé dans la liaison avec un ou plusieurs substrats. Dans cette section, nous discuterons ces résultats à la lumière de la littérature. Nous proposerons également un certain nombre de pistes expérimentales ayant pour finalité de confirmer le rôle fonctionnel de ces poches, d'établir ou non leur relation avec la résistance aux antipaludiques, et d'identifier leurs possibles substrats.

4.2.1. Quelle(s) relation(s) entre la fonction de Pfk13, la poche du domaine KREP, et les mutations de résistance aux ARTDs ?

4.2.1.1. Pfk13 et sa possible participation dans un complexe d'ubiquitination

À notre connaissance, il n'existe pas de protéines arborant simultanément les trois domaines annotés de Pfk13 chez les mammifères ou les plantes (coiled-coil, BTB et KREP) : certaines présentent des domaines BTB et KREP (telles que les protéines BTB-Kelch comme KEAP1^{107,108,131}) ; d'autres incluent des domaines coiled-coil et KREP (comme la protéine Kel1p²⁷⁵) ; et quelques-unes arborent des domaines BTB et coiled-coil (par exemple, la protéine KCTD21²⁷⁶). Étant donné la présence des domaines BTB et KREP dans Pfk13, Arieu *et al.* (2014) ont classé cette protéine comme probable membre de la famille des BTB-Kelch¹⁰⁴. Nos travaux révèlent que cette classification est à considérer avec la plus grande prudence. En effet, les protéines BTB-Kelch présentent entre les domaines BTB et KREP une région supplémentaire d'environ 150 acides aminés nommée IVR (pour *Intervening Region*), absente dans Pfk13. Ce domaine IVR est essentiel pour interagir, de concert avec le BTB, avec une protéine Cullin. Par ailleurs, le domaine BTB de Pfk13 s'avère distant de celui des protéines BTB-Kelch : il présente un plus fort pourcentage d'identité de séquence et une meilleure superposition de structure avec celui des protéines KCTD. Bien que le rôle de ces dernières ne soit pas pleinement élucidé, la majorité d'entre elles forment divers complexes homo-oligomériques (majoritairement des tétramères et des pentamères), et participent à l'ubiquitination de molécules en recrutant à la fois une protéine Cullin et un substrat à ubiquitiner^{261,277}. Dans le cas de Pfk13, la résolution cristallographique de BTB-KREP a révélé un état homodimérique de cette région de la protéine. Toutefois, la présence d'un domaine coiled-coil, pouvant jouer un rôle dans l'homo-oligomérisation des protéines²⁵⁸, pourrait complexifier cet état.

D'après la littérature, un des rôles principaux des domaines KREP est de cibler une molécule qui sera ubiquitinée puis conduit en voie de dégradation par le protéasome¹⁰⁷. Nos travaux ont mis en évidence une poche peu profonde du domaine KREP de Pfk13 comme *patch* de sites très conservés. Cette poche, communément retrouvée dans les domaines KREP, représente souvent la zone d'interaction avec un substrat^{107,129,131,262,263}, bien que d'autres surfaces puissent servir de zone d'interaction protéine-substrat^{107,108}. Nous faisons l'hypothèse que cette poche a un rôle dans l'interaction de Pfk13 avec son substrat, et ce pour plusieurs raisons :

- Nos analyses évolutives et structurales sur plusieurs membres de la famille des BTB-Kelch, dont la poche a été validée au niveau expérimental, ont également révélé cette dernière comme enrichie en sites très conservés avec le même outil FuncPatch ;
- La poche contient plusieurs arginines et sérines strictement conservées à l'échelle des *Apicomplexa* (R529, S576, S577, R597 et S623), deux acides aminés essentiels dans l'interaction entre le domaine KREP et son substrat pour trois des quatre protéines BTB-Kelch étudiées : KEAP1, KLHL2 et KLHL3^{131,262} ;

- La poche est associée à un potentiel de surface neutre / électropositif, contrastant avec le reste du domaine, associé à un fort potentiel électronégatif. Ce différentiel électrostatique est également retrouvé chez les autres protéines BTB-Kelch investiguées dans cette étude (KEAP1, KLHL2, KLHL3 et KLHL12).

Toutefois, s'il est vraisemblable que Pfk13 soit impliquée dans un complexe d'ubiquitination contrôlant entre autres les niveaux cellulaires de PfPI3K¹²⁶, elle pourrait aussi avoir d'autres fonctions. Par exemple, chez les vertébrés, toutes les KCTDs ne participent pas à un complexe d'ubiquitination : certaines (comme KCTD1 et KCTD15) ont pour rôle d'inhiber l'activité de protéines de la famille AP-2 (pour *Activating Protein 2*)^{278,279}, lesquelles ont un rôle essentiel dans la régulation de l'expression des gènes²⁸⁰. Alternativement, Pfk13 pourrait avoir un rôle dans la régulation du cytosquelette d'actine (protéine importante dans l'architecture et les mouvements cellulaires) et dans le trafic vésiculaire, par homologie avec la protéine Kelch-like ECT2 interacting protein (KLEIP) qui contient un domaine BTB et un domaine KREP²⁸¹. Cette hypothèse est notamment confortée par le travail de Demas *et al.* (2018), qui a révélé que des mutations non-synonymes localisées sur un domaine similaire au KREP de Pfk13 de la protéine PfCoronin (supposée interagir avec l'actine^{108,137}), confèrent par leur seule présence la résistance aux ARTDs¹³⁷.

4.2.1.2. Vers une validation expérimentale du rôle de la poche du domaine KREP de Pfk13

4.2.1.2.1. Approches biochimiques : *pull-down* et *pull-down* différentiels

Des approches de *pull-down* ont été initiées afin de valider le rôle de la poche du domaine KREP de Pfk13 comme surface d'interaction, en plus de pouvoir étudier son interactome. Pour cela, différentes constructions consistant en une concaténation des domaines de Pfk13 (**Figure 33**) ont été préparées avec un tag GST localisé en N-terminal. Une première construction contenant uniquement le domaine KREP (noté K dans les constructions) avec l'étiquette GST a été définie afin de pouvoir détecter des partenaires qui interagiraient spécifiquement avec KREP. Les autres constructions GST-BK (BTB-KREP) et GST-CBK (coiled-coil-BTB-KREP) ont également été produites car il est possible que les domaines BTB et coiled-coil soient nécessaires pour permettre la bonne interaction entre les substrats et le domaine KREP. Par ailleurs, cela permettrait également de déterminer d'autres partenaires de Pfk13 spécifiques à ces domaines tels qu'une protéine Cullin.

Pour l'heure, aucun partenaire protéique n'a été détecté à partir des domaines recombinants sauvages avec un tag GST. De fait, aucun *pull-down* différentiel n'a encore été effectué, lequel aurait pour objectif de comparer des constructions sauvages et mutantes afin de mieux comprendre le lien entre la fonction de Pfk13, la poche du domaine KREP et les mutations de résistance aux ARTDs (**Figure 52**). Nous savons que cette approche *différentielle* est

opérationnelle, puisqu'elle a déjà été testée pour déterminer les protéines interagissant avec la protéase ST14 (pour *Suppression of Tumorigenicity 14*)²⁸² et la protéine SUMO (pour *small ubiquitin-like modifier*)²⁸³. Nos *pull-down* différentiels seront réalisés avec des domaines recombinants sauvages et mutants incorporant :

- soit la double mutation R529A/N530A, ayant pour objectif d'invalider la poche du domaine KREP (**Figure 51**). Ces deux positions ont été sélectionnées sur la base de nos travaux bioinformatiques : elles sont systématiquement conservées à l'échelle des *Apicomplexa*, sont situées à la surface de la poche, et s'alignent avec les sites critiques N414 et R415 du domaine KREP de KEAP1 dans l'interaction avec son substrat Nrf2 (**Figure 52**)¹³¹ ;
- soit la mutation C580Y, pour tester l'existence d'un lien entre les mutations de résistance et la possible fonctionnalité de la poche, par comparaison avec les constructions sauvage et double mutante R529A/N530A (**Figure 51**).

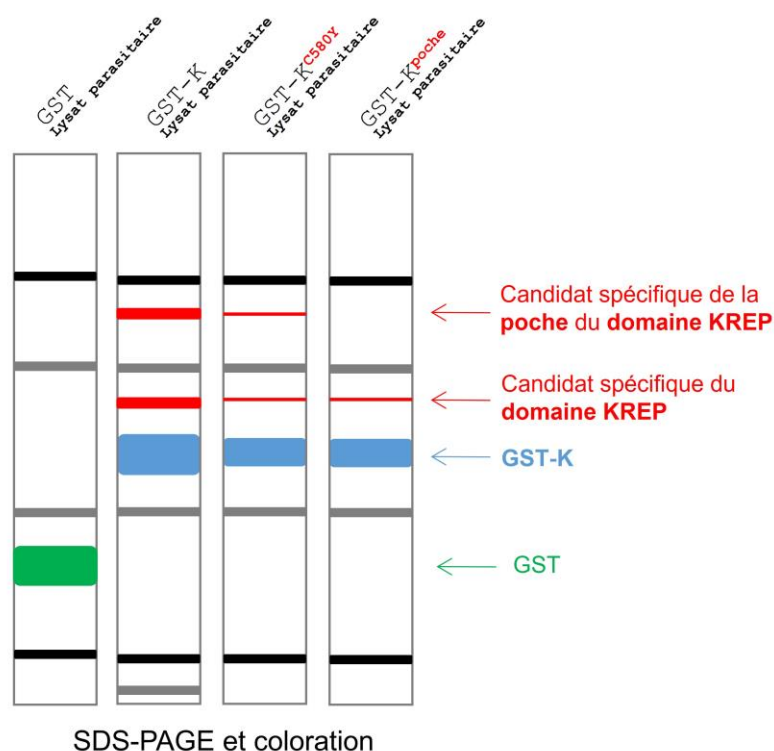


Figure 51 : Résultat hypothétique attendu à l'issu d'une expérience de *pull-down* différentiel révélé par SDS-PAGE. Dans cet exemple, on peut observer la présence d'une bande spécifique de la poche (*rouge, en haut*), car présente après incubation avec le lysat parasite pour les constructions sauvage GST-K et mutant conférant la résistance aux ARTDs GST-K^{C580Y}, mais absente pour GST-K^{Poche} (mutations R529A/N530A supposées invalider la fonction de la poche). Étant également absente avec la construction GST seule (contrôle), la protéine correspondante serait un candidat spécifique pour une interaction avec la poche. Dans cet exemple, la bande est bien moins intense sur la construction GST-K^{C580Y}, suggérant un possible lien entre le rôle de la poche et la mutation C580Y.

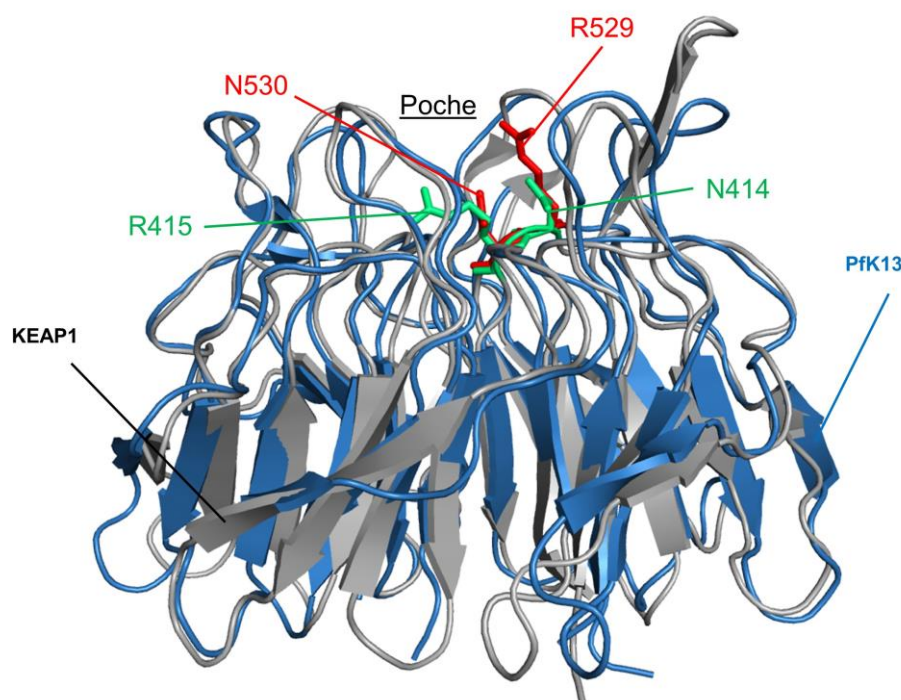


Figure 52 : Localisation des positions R529/N530 sur la structure tertiaire du domaine KREP de Pfk13. La structure tertiaire du KREP de Pfk13 (mode *cartoon*, bleu) a été superposée avec celle de KEAP1 (gris). Les positions R529/N530 de Pfk13 et N414/R415 de KEAP1 sont représentées en mode *stick*, respectivement en rouge et vert.

Avant d'envisager une expérience de *pull-down* différentiel, des optimisations sont à prévoir en raison de la présence de protéines contaminantes bactériennes sur nos différentes constructions, vraisemblablement liées aux domaines coiled-coil, BTB ou KREP, puisqu'elles ne sont pas retrouvées dans le contrôle GST. Les profils contaminants associés aux trois constructions GST-K/BK/CBK étaient similaires (**Figure 44**), suggérant que nombre des protéines contaminantes sont liées au domaine KREP. Il serait intéressant de comparer la construction GST-K sauvage et celle incluant la double mutation R529A/N530A afin de vérifier si certains de ces contaminants bactériens se lient spécifiquement à la poche. L'identification de tels contaminants, même bactériens, pourrait fournir des pistes pour le *Plasmodium*. Plusieurs protocoles ont été suivis pour tenter d'éliminer les contaminants, mais aucun n'a permis de purifier plus efficacement les domaines recombinants. De nouvelles perspectives sont toutefois envisagées telles que le rajout d'un lysat bactérien dénaturé (avec de l'urée) avant l'ajout de la résine²⁸⁴. Ces protéines dénaturées pourraient entrer en compétition dans la liaison avec d'autres protéines, incluant les possibles contaminants.

La présence des contaminants bactériens observés a considérablement nui à l'interprétation des résultats du *pull-down*, à l'issue duquel aucune protéine parasitaire n'a été détectée visuellement après SDS-PAGE. Plusieurs raisons peuvent expliquer ces résultats :

- Les partenaires parasites de Pfk13 n'auraient pas été convenablement solubilisés lors de la préparation du lysat parasitaire : d'autres détergents non dénaturants pourraient

être testés, comme le docéylmaltoside (DDM), mais les profils protéiques semblent similaires d'après nos tests avec trois détergents différents ; différents lysats parasitaires enrichis en différents stades érythrocytaires (*ring*, trophozoïtes et schizontes) pourraient être préparés ;

- Les domaines recombinants PfK13 seraient mal repliés, possiblement en raison de la propriété semi-dénaturante du sarkosyl : nous savons que l'élution par compétition avec le glutathion est efficace, confirmant que le site catalytique de la GST est intact ; pour évaluer spécifiquement le repliement des domaines de PfK13, des tests biophysiques (par exemple, le dichroïsme circulaire) pourraient être réalisés ;
- Les partenaires parasitaires de PfK13 seraient en trop faible quantité et donc non détectable sur gel : l'étape SDS-PAGE n'est pas indispensable, et les protéines associées à la résine agarose-GSH-GST-K/BK/CBK pourraient être analysées directement par spectrométrie de masse.

Bien que cela n'ait pas été fait, nous aurions pu envoyer quelques échantillons en spectrométrie de masse d'une part pour confirmer l'absence de protéines parasitaires tel qu'observé sur gel, mais également pour connaître les différents contaminants (type(s) de protéines) qui se seraient liées aux domaines de PfK13.

Outre les *pull-down*, d'autres expérimentations sont envisagées. Nous songeons notamment au dichroïsme circulaire, qui pourrait servir à étudier la déviation du repliement de PfK13 portant une mutation de résistance aux ARTDs (par exemple, la mutation C580Y) par rapport à la structure sauvage. Cette expérience servirait également à confirmer nos simulations de dynamique moléculaire, lesquelles ont suggéré un repliement structural très conservé du domaine KREP portant la mutation C580Y. En outre, nous pourrions étudier les stades homo-oligomériques de PfK13, soit en réalisant un gel de polyacrylamide en conditions non dénaturantes, soit en effectuant une filtration sur gel (ou chromatographie d'exclusion). Enfin, la structure tertiaire coiled-coil-BTB-KREP de PfK13 pourrait être déterminée afin de mieux comprendre le rôle du coiled-coil. Toutes ces expérimentations pourront être faites en coupant la GST des domaines PfK13 à l'aide de la protéase 2C, ou en utilisant un autre tag. D'ailleurs, des constructions sur la base d'un tag poly-histidine ont été préparées : en plus de fournir un élément de comparaison avec les constructions impliquant la GST, le tag poly-histidine ne présente pas de propriété de dimérisation contrairement à la GST, ce qui est probablement plus approprié pour les constructions incluant les domaines coiled-coil et BTB (ce dernier participant vraisemblablement à l'homodimérisation de PfK13). Les premiers résultats ont révélé que les domaines recombinants avec ce tag sont moins solubles que ceux avec le tag GST, et ce même en utilisant le sarkosyl (données non montrées).

4.2.1.2.2. Approche génétique : étude de parasites *P. falciparum* génétiquement modifiés

Une perspective envisagée – et complémentaire aux études avec des domaines recombinants purifiés – est d’analyser la protéine PfK13 directement chez *P. falciparum*. Ces analyses sont menées principalement par Laurine Rondepierre (doctorante, UMR 261). Pour cela, le gène *pfk13* sera modifié selon la méthode développée par l’équipe de Spielmann¹²¹, et mise au point au laboratoire ces derniers mois par Diana Roman (doctorante, UMR 261). L’objectif est de muter *pfk13* pour introduire dans la poche du domaine KREP la double mutation R529A/N530A que nous supposons être perte de fonction (**Figure 52**). Si les parasites mutants sont viables, nous pourrions étudier des phénotypes simples dans un premier temps, en nous basant sur ceux associés aux mutations de résistance : taux de croissance *in vitro*, distorsion du cycle intra-érythrocytaire (allongement du stade *ring*, raccourcissement du stade schizonte), sensibilité *in vitro* aux ARTDs (tests d’inhibition de croissance et RSA_{0-3h}) et taux de protéines ubiquitinées (étant une fonction attendue de PfK13).

4.2.1.2.3. Approche pharmacologique : recherche d’inhibiteurs de la poche

En parallèle des approches biochimique et génétique, nous avons commencé à rechercher des petites molécules qui pourraient interagir spécifiquement avec la poche du domaine KREP de PfK13. La logique première est d’apporter une preuve pharmacologique d’un rôle fonctionnel de cette poche. Dans ce travail de thèse, nous avons recensé 25 molécules se liant de façon non covalente à la poche du domaine KREP de KEAP1 (lesquels inhibent l’interaction avec son substrat cible Nrf2). Parmi elles, 12 ont fait l’objet de *docking* moléculaires : toutes ont révélé une forte affinité de liaison avec la poche du domaine KREP de PfK13, dont une avec une affinité supérieure à celle estimée avec la poche de KEAP1 (inhibiteur 4IN4). Nous avons commencé à tester quatre d’entre elles sur la souche sauvage 3D7 de *P. falciparum* en mesurant leur capacité à inhiber la croissance *in vitro* du parasite par mesure de la synthèse d’ADN après 72 heures d’incubation. À une concentration en inhibiteurs fixée à 0,1 μ M, toutes les molécules ont diminué d’environ 20% la croissance du parasite. Cependant, lorsque nous avons augmenté la concentration à 1 μ M et 10 μ M, la croissance du parasite est restée sensiblement similaire à 0,1 μ M, exceptée pour l’inhibiteur 5CGJ qui a diminué par deux cette croissance. Dès lors, nous pouvons nous demander si ces inhibiteurs ont eu un réel impact sur la croissance du parasite. Il est possible que ces derniers n’aient pas : *i*) été convenablement solubilisés ; ou *ii*) réussi à traverser les différentes membranes (globule rouge, parasite, etc.). Ces tests préliminaires seront répétés et complétés par des tests RSA_{0-3h} spécifiques de la résistance aux ARTDs. D’autres inhibiteurs seront inclus, potentiellement ceux pour lesquels nous n’avons pas encore étudié l’affinité de liaison par *docking* moléculaire.

Le phénotype attendu lié à l'inhibition de la poche du domaine KREP de Pfk13 avec de tels inhibiteurs demeure cependant incertain. Selon une première hypothèse, la présence de ces inhibiteurs pourrait conduire à la mort du parasite. En effet, puisque Pfk13 est une protéine essentielle¹¹¹, et en supposant que la poche du KREP joue un rôle clé dans cette fonction, inhiber la poche pourrait mener à des accumulations de protéines qui ne seraient pas dégradées. Alternativement, les inhibiteurs pourraient engendrer un phénotype de résistance aux ARTDs, dans le cas d'une relation entre la poche et les mutations de résistance. Pour tester ces phénotypes possibles, nous envisageons d'incuber les inhibiteurs avec des cultures de parasites sauvages et mutants résistants aux ARTDs (tels que les mutants C580Y). Des tests RSA_{0-3h} pourraient être menés afin d'évaluer le taux de parasites survivant à une forte concentration de DHA en présence ou en absence d'inhibiteurs. Enfin, nous pourrions également tester si ces différents inhibiteurs se lient aux domaines recombinants GST-K/BK/CBK, par exemple par microcalorimétrie ou dans nos expériences de *pull-down*.

4.2.1.3. Les mutations de résistance aux ARTDs : l'hypothèse d'une déstabilisation structurale du domaine KREP de Pfk13

Comme énoncé dans l'introduction (cf. section 1.2.6.), les mutations de résistance aux ARTDs localisées sur le domaine KREP de Pfk13 pourraient avoir divers effets, parmi lesquels : i) altérer la stabilité du domaine KREP et *in fine* de la protéine ; ou ii) modifier des zones d'interaction avec des protéines partenaires. Ces altérations occasionneraient une cascade de modifications biologiques chez le parasite, incluant une diminution de l'ubiquitination des protéines¹²⁵ dont PfPI3K¹²⁶, ou encore une phosphorylation du facteur d'initiation à la traduction PfcIF2α au stade *ring*¹²³. D'après la littérature, les changements d'acides aminés situés dans le domaine KREP de protéines BTB-Kelch peuvent avoir des conséquences diverses selon leur localisation sur la structure tertiaire (en plus du type des acides aminés mutants). Par exemple, les mutations localisées au niveau de la poche du domaine KREP de KEAP1 diminuent l'affinité de liaison avec le substrat cible (Nrf2), tandis que des mutations distantes / non situées à la surface de la poche peuvent diminuer l'abondance ou la stabilité de KEAP1²⁸⁵⁻²⁸⁷. En considérant les 31 mutations de résistance aux ARTDs répertoriées par l'OMS sur le domaine KREP de Pfk13⁵⁰, aucune n'est directement située à la surface de la poche : certaines sont inaccessibles au solvant (enfouies), d'autres sont localisées au niveau du tunnel central (comme la mutation C580Y), et d'autres encore se trouvent sur la face opposée à celle de la poche (comme la mutation R561H). Un recensement des mutations du gène *pfk13* identifiées lors de larges études d'épidémiologie moléculaire en Asie du sud-est a révélé que seuls quelques parasites (sur plusieurs milliers) portaient une mutation non-synonyme au niveau de la poche^{104,114-116,288}. Nous pouvons supposer que les sites impliqués dans la poche sont fonctionnellement trop importants pour fournir un avantage compétitif à long terme pour le parasite. Les mutations de résistance aux ARTDs pourraient modifier d'autres propriétés de Pfk13 telles que son abondance (synthèse diminuée, dégradation accélérée), mécanisme qui maintiendrait intacte la poche et sa

fonctionnalité. Dans un tel scénario, la protéine Pfk13 mutante serait toujours fonctionnelle mais, du fait d'une moindre abondance, aurait une activité cellulaire diminuée. Nous avons réanalysé les données disponibles dans la littérature concernant la quantité cellulaire de Pfk13 chez les parasites sauvages et les mutants. Notre analyse est résumée dans le **tableau 9** et décrite ci-dessous :

- Une étude de protéomique quantitative a révélé que les parasites résistants aux ARTDs porteurs de la mutation C580Y sur fond génétique Cambodgien, avaient approximativement deux fois moins de protéine Pfk13 que des souches sauvages de fond génétique comparable²⁸⁹ ;
- Dans deux autres études^{126,127}, l'équipe de Haldar a généré des parasites pseudo-diploïdes pour *pfk13* par intégration chromosomique d'une seconde copie du gène (taguée avec HA), sauvage ou mutante C580Y, dans deux souches différentes non asiatiques (3D7 : origine africaine ; et CS2 : origine brésilienne). La quantité de protéine Pfk13 mutante détectée par *Western Blot* (anticorps anti-tag HA) était franchement inférieure pour les deux souches^{126,127} par rapport à la quantité sauvage. Cette même équipe a également analysé la souche africaine NF54, dont 3D7 est une sous-population, et dont le gène *pfk13* endogène a été modifié par CRISPR/Cas9 pour y introduire la mutation C580Y¹⁰⁶. La quantité de protéine Pfk13 mutante était soit similaire à celle du sauvage¹²⁶, soit franchement inférieure (détection par *Western Blot* avec un anticorps anti-Pfk13)¹²⁷.

Tableau 9 : Différentes expériences mesurant la quantité de protéine Pfk13 mutante par rapport à leur souche de référence sauvage.

Origine	Lignées ^a	Modification génétique	Mutation testée	Valeur RSA _{0-3h} ^b	Protéine contrôle	Méthode de quantification Pfk13	Stade parasitaire	Ratio Pfk13 mutante / WT ^c	Ref ^d
Cambodge	PL2 et PL7	Aucune	C580Y	5% et 20 %	Protéome	Spectrométrie de masse	Trophozoïte	0,65	289
Cambodge	Cam3.II ^{rev}	Zing finger, remplacement	R539T	49%	Protéome	Spectrométrie de masse	Ring _{6-12h} Trophozoïte	0,25 0,6	289
Cambodge	Cam3.II ^{rev}	Zing finger, remplacement	C580Y	24%	Protéome	Spectrométrie de masse	Trophozoïte	0,7	289
Afrique	3D7	CRISPR/Cas9, pseudo-diploïde	C580Y	9%	BiP	<i>Western Blot</i> , anti-HA	ND	0,53	126
Afrique	NF54	CRISPR/Cas9	C580Y	13%	BiP	<i>Western Blot</i> , anti-Pfk13	ND	1,0	126
Afrique	NF54	CRISPR/Cas9	C580Y	13%	PfFKBP	<i>Western Blot</i> , anti-Pfk13	Ring _{0-3h}	0,66	127
Brésil	CS2	CRISPR/Cas9, pseudo-diploïde	C580Y	ND	BiP	<i>Western Blot</i> , anti-HA	ND	0,65	127
Brésil	CS2	CRISPR/Cas9, pseudo-diploïde	R539T	ND	BiP	<i>Western Blot</i> , anti-HA	ND	0,60	127

Note – a PL2, *pfk13* sauvage ; PL7 : *pfk13* C580Y. ^b ND, non décrit. ^c WT, Wild-type (sauvage). Pour les analyses de *Western Blot*, le ratio a été calculé en mesurant la différence d'intensité des bandes sauvage et mutante (préalablement normalisée grâce à la protéine contrôle). ^d Ref, référence.

Globalement, malgré les différences méthodologiques importantes, ces résultats n'invalident pas l'hypothèse d'une diminution, probablement faible, de la quantité de protéine Pfk13 pour

certaines mutations du domaine KREP. D'autres données de la littérature sur d'autres domaines KREP mutants et nos propres résultats sont également en cohérence avec cette hypothèse :

- Certaines mutations non-synonymes de la protéine Gigaxonine (membre de la famille des BTB-Kelch) diminuent son abondance en altérant sa stabilité structurale, générant des cas de neuropathie à axones géants chez l'Homme²⁹⁰. Parmi ces mutations se trouve C464Y (retrouvée chez un patient hétérozygote composite, donc portant aussi une autre mutation non-synonyme sur la seconde copie du même gène), qui est localisée au niveau du tunnel du domaine KREP de façon similaire à la principale mutation de résistance aux ARTDs C580Y sur Pfk13²⁹⁰ ;
- Des simulations de dynamique moléculaire réalisées sur la protéine KLHL3 (BTB-Kelch) ont révélé que les mutations non-synonymes localisées en dehors de la poche n'ont pas d'impact sur l'affinité de liaison avec son substrat (WNK), contrairement à celles situées à la surface de la poche²³⁴ ;
- Les simulations de dynamique moléculaire que nous avons réalisées sur les mutations C580Y et R539T n'ont révélé aucune modification dans l'architecture de la poche : l'ensemble des ponts hydrogène est préservé, tout comme le potentiel électrostatique à la surface. En revanche, des pertes notables de contacts inter-atomiques pouvant être importants pour la stabilité du domaine KREP ont été observées, par comparaison des structures mutantes C580Y et R539T avec les structures sauvage et mutante A578S (contrôle, sensible aux ARTDs^{50,116}) ;
- Une déstabilisation structurale a été prédite via des outils bioinformatiques pour la mutation F446I (retrouvée dans la partie ouest de l'Asie du sud-est) et la mutation V603E, encore non caractérisée cliniquement et récemment retrouvée en Asie du sud-est²⁹¹.

Pour tester cette hypothèse, outre le dichroïsme circulaire sur domaine recombinant purifié déjà énoncé précédemment (cf. section 4.2.1.2.1.), des expériences de *pulse-chase* pourraient être menées sur des cultures de parasites afin de mesurer la cinétique de synthèse et de dégradation de Pfk13.

Alternativement, les mutations de résistance aux ARTDs pourraient altérer la structure du KREP uniquement lors de conditions cellulaires particulières. Par exemple, Haldar a récemment proposé que les mutations pourraient agir de concert avec l'oxydation de cystéines voisines, douées de fonction *redox*, pour induire des changements conformationnels (comme c'est le cas pour la protéine KEAP1)²⁹². En effet, les principales positions associées aux mutations de résistance aux ARTDs sont souvent proches de cystéines. Cependant, alors que pour KEAP1, on compte 27 cystéines situées sur l'ensemble de la protéine (dont 9 sur le domaine KREP), il n'y en a que 7 sur Pfk13, toutefois toutes localisées sur le KREP. En nous basant sur la surface relative d'accessibilité au solvant des cystéines du domaine KREP de Pfk13, nous avons remarqué qu'une majorité d'entre elles étaient enfouies. Cependant, une analyse plus approfondie nous a permis de constater que le groupement thiol des cystéines, qui s'oxyde facilement, est

systématiquement exposé au solvant. L'hypothèse de Haldar est donc possible et pourrait être testée en adaptant des méthodes récemment développées couplant spectrométrie de masse et marquage des cystéines avec l'iodoacétamide (SICyLIA)²⁹³.

Au final, ce travail offre de nouvelles hypothèses et perspectives de travail concernant Pfk13. Les nouveaux objectifs fixés sont : i) de tester un rôle fonctionnel de la poche du domaine KREP ; ii) d'étudier l'existence d'un lien entre la poche et la résistance aux ARTDs ; et iii) de déterminer les molécules se liant à Pfk13 au niveau de la poche.

Bien que la résistance aux ARTDs n'ait pas encore émergée de façon notable en Afrique, des cas isolés ont été rapportés¹⁰¹ et *pfk13* continue d'être systématiquement génotypé afin de guetter toute émergence de mutations suspectes. Pour notre part, toute nouvelle mutation sur le domaine KREP de Pfk13 et ayant une fréquence croissante au cours du temps pourrait être évaluée par simulation de dynamique moléculaire afin de prédire son potentiel impact sur la structure tertiaire, en attendant les validations expérimentales qui prennent au minimum deux mois pour les approches génétiques type CRISPR/Cas9.

4.2.2. Quel lien entre la poche de PfCRT et sa fonction de transport ?

4.2.2.1. La poche de PfCRT, cœur de son activité ?

Dans la mesure où nous souhaitons prédire des sites fonctionnels du transporteur PfCRT, sa structure tertiaire, encore non déterminée expérimentalement, a été modélisée par une approche comparative en utilisant des structures résolues de protéines appartenant à la même superfamille de protéines (DMT). Cette modélisation a été réalisée pour deux états conformationnels : une structure ouverte côté vacuolaire (*inward-facing*) et une structure fermée (*occluded*). Les modèles PfCRT prédits révèlent la présence d'une large cavité dans la conformation ouverte côté vacuolaire et d'une poche située au centre de la protéine dans la conformation fermée. D'après ces différentes prédictions conformationnelles, nous pouvons supposer que les substrats physiologiques de PfCRT entrent dans le transporteur par cette cavité, puis seraient transitoirement localisés dans la poche. Un nouveau changement d'état conformationnel de PfCRT permettrait ensuite de les relâcher de l'autre côté de la membrane, comme cela a été proposé pour les protéines DMT Vrg4 et TPT^{271,272}. Le rôle fonctionnel de la poche de PfCRT est renforcée par :

- La détection d'un *patch* de sites très conservés au sein de PfCRT, principalement localisés au niveau de l'hélice 9, laquelle longe la cavité et est impliquée dans l'architecture de la poche ;
- La participation de sites localisés sur l'hélice 9 d'autres protéines DMT dans la liaison avec un ou plusieurs substrats^{158,271,272,294}.

Nos modèles de structure tertiaire de PfCRT présentent toutefois quelques limites. Tout d'abord, nous avons fait l'hypothèse que la structure de PfCRT est similaire aux structures parentes des protéines DMT choisies, alors que des différences sont évidentes telles que la présence sur PfCRT d'une longue boucle d'environ 50 acides aminés entre les hélices transmembranaires 7 et 8. Cette boucle ainsi que les extrémités N- et C-terminales n'ont pas pu être incluses dans nos prédictions en raison de leur spécificité à PfCRT, alors qu'elles pourraient avoir un rôle dans l'activité de PfCRT. De plus, les extrémités N- et C-terminales contiennent des sites de modification post-traductionnelle comme des phosphorylations sur les sites S33, S411, T416 et S420. Certaines de ces phosphorylations contrôlent l'adressage du transporteur à la membrane vacuolaire, et d'autres l'activité de transport de la CQ dans un contexte de résistance^{295,296}. Quant à la boucle connectant les hélices 7 et 8, elle est globalement conservée et localisée côté vacuolaire. En outre, puisqu'elle contient une mutation de résistance à la CQ (Q271E), elle pourrait participer au contrôle de l'accès à la cavité de PfCRT.

4.2.2.2. Validation des sites fonctionnels candidats

En comparant les sites impliqués dans la poche de PfCRT avec ceux similairement localisés pour diverses protéines DMT et caractérisés au niveau expérimental, nous avons sélectionné quatre sites potentiellement impliqués directement dans le transport d'un substrat physiologique de PfCRT : Y68, D329, Y345 et S349. Aucun de ces sites n'a été associé à une mutation conférant la résistance aux médicaments à base de quinoléine. Soit des mutations sur ces sites seraient trop délétères pour l'activité du transporteur, et peut-être la survie du parasite, soit elles n'auraient pas de lien avec le transport de médicaments. Trois autres sites intéressants sont Q352, N326, et I356. Le premier cité aurait, avec l'aide de l'hélice 1, un rôle dans la liaison avec des médicaments à base de quinoléine²⁹⁷. Concernant N326 et I356, ces sites furent initialement mutés pour conférer la résistance à la CQ ; leur suppression (haplotypes GB4 et 7G8, maintenant prédominants en Afrique¹⁷⁵) n'altère pas cette résistance, mais améliore significativement le taux de croissance érythrocytaire du parasite évalué *in vitro* avec des lignées parasitaires génétiquement modifiées¹⁷⁵. Leur implication dans une fonction physiologique de PfCRT est donc très probable.

Des expériences de transport de substrat(s) dans un système d'expression hétérologue de PfCRT et de modification du gène *pfCRT* chez le parasite permettraient de confirmer le rôle de Y68, D329, Y345 et S349 dans le transport des substrats de PfCRT. PfCRT étant une protéine

essentielle¹¹¹, un substrat physiologique de PfCRT devrait logiquement être transporté par la version sauvage du transporteur mais aussi par les formes mutantes conférant la résistance. Plusieurs substrats testés dans différents systèmes hétérologues satisfont ce critère : l'ion tétraéthylammonium (un substrat non naturel classiquement utilisé pour caractériser les transporteurs poly-spécifiques de cations organiques), des acides aminés basiques (arginine, lysine, histidine), des peptides courts, et les ions ferreux (Fe^{2+}) et ferrique (Fe^{3+})^{160,161,163,166}. Cependant, aucun de ces substrats n'a été validé par des expériences directement dans la cellule parasitaire. Parmi les quatre sites candidats à tester, trois sont des tyrosines (Y) ou aspartate (D), résidus souvent impliqués dans l'interaction avec des ions Fe^{2+} et/ou Fe^{3+} , à l'image des lactoferrines²⁹⁸. Un substrat prometteur à tester pourrait donc être le fer, qui a été très rigoureusement démontré comme transporté par les versions sauvage et mutante (Dd2) de PfCRT dans le système d'ovocytes de Xénope avec de nombreuses expériences et contrôles¹⁶⁶. Cette hypothèse pourrait être testée en collaboration avec l'équipe de Gabrielle Planelles (unité 1138, CNRS ERL8228)¹⁶⁶ avec laquelle nous sommes en contact.

En parallèle, le rôle de ces sites candidats pourrait être testé directement chez le parasite *P. falciparum* par modification du gène *pfCRT* afin de remplacer les tyrosines et aspartate décrites ci-dessus par une alanine (même stratégie que pour *pfk13*). Enfin, Jérôme Clain (directeur de thèse, UMR 261) a mis en place un système d'expression hétérologue de PfCRT à la surface de la levure *S. cerevisiae* (**Figure 53**), montrant une croissance diminuée des levures exprimant PfCRT mutante Dd2 par rapport à celles exprimant PfCRT sauvage, en présence d'une concentration élevée de CQ dans le milieu de culture (1 mM ou 4 mM ; **Figure 53**). Ces résultats suggèrent une accumulation augmentée, et toxique, de CQ chez les levures exprimant PfCRT mutante Dd2, résultat cohérent avec les expériences dans le système Xénope démontrant une accumulation augmentée de CQ radioactive dans l'ovocyte exprimant PfCRT mutante Dd2 à la surface. Nous pourrions reprendre cette méthodologie afin de tester si une modification de l'encombrement et/ou des propriétés physico-chimiques des chaînes latérales de ces sites candidats modifie le phénotype de croissance des levures en présence de CQ. Ces tests seraient réalisés sur la version Dd2 de PfCRT qui transporte la CQ.

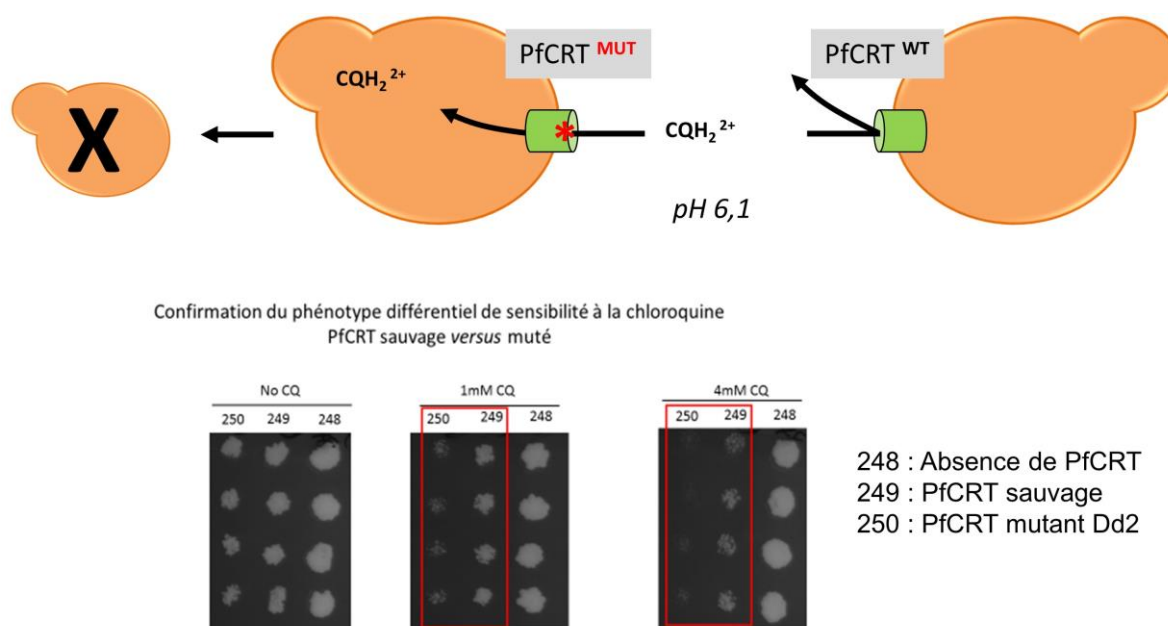


Figure 53 : Effet de PfCRT sur la croissance de levures en présence de CQ. En absence de la protéine PfCRT (souche 248), la présence de CQ perturbe peu la croissance de *S. cerevisiae* aux concentrations testées (pistes 248) sur un milieu standard (mais tamponné à pH 6) permettant la respiration et la fermentation. La croissance est moindre après introduction de CQ chez les levures exprimant une version sauvage de PfCRT (pistes 249), et encore plus diminuée chez celles exprimant une version mutante de résistance PfCRT^{Dd2} (pistes 250).

4.2.2.3. La relation entre la structure tertiaire de PfCRT et les mutations de résistance aux médicaments

Les modèles PfCRT prédits ont permis de mieux comprendre les altérations structurales engendrées par les mutations de résistance aux médicaments à base de quinoléine. Fait remarquable, une grande majorité des mutations de résistance à la CQ sont localisées dans la cavité de PfCRT, et plus particulièrement impliquées dans la poche de liaison au substrat supposée, incluant le site lié à la mutation clé K76T. De plus, plusieurs mutations de l'haplotype Dd2 (C72S, M74I, N75E, N326S/D et I356T/L) sont proches de K76T dans la structure tertiaire de PfCRT, suggérant que ces mutations modifient de concert certaines propriétés du transporteur pour conférer la résistance. Ces combinaisons de mutations provoquent des altérations drastiques du potentiel électrostatique à l'entrée de la cavité ainsi qu'au niveau de la poche de liaison, passant d'un état neutre chez le sauvage à un état électronégatif chez les structures mutantes conférant la résistance à la CQ. Dès lors, nous émettons l'hypothèse que le potentiel électrostatique de la poche représente un facteur clé permettant le transport de CQ à l'état di-protoné.

Des mutations individuelles apparues sur un fond génétique Dd2 ou 7G8 dans certaines populations naturelles de *P. falciparum* ont été récemment associées à une résistance à la PPQ, une des molécules partenaires utilisées dans les CTAs^{87,89,299}. Tout comme pour les mutations de résistance à la CQ, nous observons dans nos modèles PfCRT que celles conférant la résistance à

la PPQ sont majoritairement localisées dans la cavité et/ou dans la poche de liaison au(x) substrat(s) ; elles n'altèrent toutefois pas le potentiel électronégatif à la surface de la poche induit par les mutations de résistance à la CQ. Ce résultat fut plutôt surprenant car les mutations de résistance à la PPQ atténuent la résistance à la CQ. Cependant, il ne s'agit pas d'une réversion complète, comme le révèle le travail de Ross et ses collaborateurs (2018)⁸⁹. En outre, les parasites porteurs ou non de mutations conférant la résistance à la PPQ (sur fond génétique Dd2) accumulent des niveaux similaires de PPQ radioactives dans la cellule⁸⁹. Ce résultat contraste avec des expériences similaires sur la CQ^{51,170}, suggérant des mécanismes de résistance distincts pour ces deux molécules, néanmoins conférés par PfCRT. Une hypothèse plausible serait que ces nouvelles mutations pourraient favoriser le transport de molécules de détoxification dans la vacuole digestive, molécules qui neutraliseraient l'activité toxique de la PPQ dans la vacuole. Par exemple, la version mutante Dd2 de PfCRT au sein de lignées isogéniques *P. falciparum* améliore le transport de glutathion dans la vacuole digestive, réduisant la quantité d'hème libre disponible pour la CQ¹⁶⁵. Un mécanisme similaire pourrait donc être mis en place pour limiter l'activité de la PPQ. Alternativement, les mutations de résistance à la PPQ pourraient générer de nouveaux sites de liaison pour la PPQ, conduisant à une inhibition fonctionnelle de certains isoformes PfCRT comme c'est le cas pour la quinine³⁰⁰.

Deux perspectives *in silico* peuvent être envisagées pour compléter ces résultats. Tout d'abord, une série de *docking* moléculaires au niveau de la poche avec le VP (inhibiteur de canaux calciques et de PfCRT), la CQ et la PPQ donnerait des informations importantes concernant les propriétés de liaison de PfCRT. Nous avons récemment initié une première série de *docking* avec la CQ (aux états non- et di-protonés) sur PfCRT (sur la conformation *inward-facing*). Les premiers résultats révèlent que la CQ, quel que soit son état de protonation, se lie préférentiellement au niveau de la poche de liaison prédite. En outre, nous avons observé que la CQ di-protonée a une affinité de liaison meilleure avec les structures PfCRT sauvage et mutante Dd2 par rapport à la CQ non-protonée, selon les neuf meilleures poses prédites (**Figure 54**). Cette affinité serait sensiblement meilleure avec la structure mutante Dd2, relativement à la structure sauvage (**Figure 54**), suggérant que le transport de CQ di-protonée serait favorisé chez la structure mutante Dd2. Des tests supplémentaires sont toutefois nécessaires, par exemple en réalisant les *docking* avec la conformation fermée de PfCRT, mais aussi en étudiant plus en détails les modes de liaison de ces molécules sur les structures tertiaires.

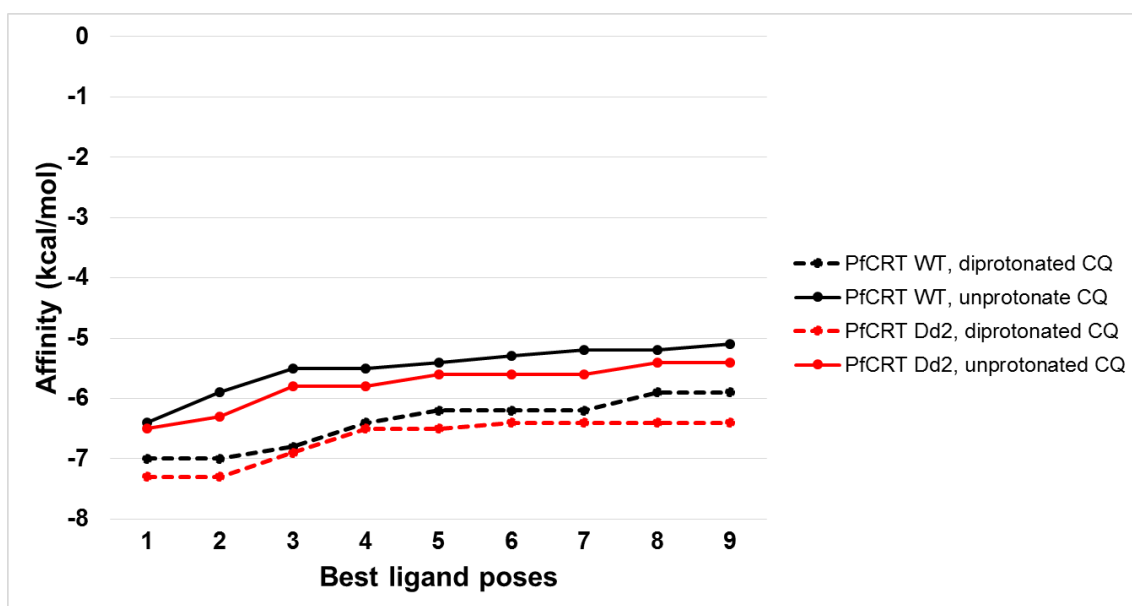


Figure 54 : Affinités de liaison de la CQ (non- et di-protonée) avec les structures tertiaires de PfCRT sauvage et mutante Dd2 (*inward-facing*). Les *docking* ont été réalisés avec l'outil Vina²⁴⁵. Seules les affinités des neuf meilleures poses sont représentées. Plus l'énergie en kcal/mol est faible, plus l'affinité de liaison est élevée.

Enfin, une suite logique serait d'étudier par des simulations de dynamique moléculaire l'évolution des interactions entre PfCRT et ces différentes molécules. En plus de détecter de possibles différences dans les affinités de liaison au cours du temps, l'approche fournirait des données supplémentaires concernant les altérations structurales engendrées par les mutations de résistance (tels qu'une modification de l'architecture de la poche, la perte de contacts importants, *etc.*). Cependant, nous émettons quelques réserves puisque les modèles de structure tertiaire prédits sont incomplets, et nous n'avons pas la certitude qu'ils reflètent la structure tertiaire telle qu'elle pourrait être déterminée par des expériences : les résultats devraient donc être interprétés avec la plus grande prudence.

En conclusion, ce second travail de thèse propose de nouvelles hypothèses et perspectives de recherche. Les nouveaux objectifs fixés sont de confirmer le rôle de la poche dans la fonction de transport de PfCRT et de tester son possible lien avec le transport de médicaments à base de quinoléine. Ces investigations pourraient être menées via des protéines PfCRT mutées au niveau des sites candidats à l'aide de systèmes d'expression hétérologues et des expériences de flux ou de croissance cellulaire.

4.2.3. PfK13 et PfCRT, de possibles cibles thérapeutiques

Les cancers peuvent être traités en recourant à des médicaments capables de cibler des molécules spécifiques comme des protéines, présentes à la surface ou à l'intérieur des cellules cancéreuses³⁰¹. Par exemple, la voie Nrf2, dont l'accumulation aberrante confère un avantage sélectif aux cellules cancéreuses et associée à leur transformation maligne, peut être modulée via l'utilisation d'inhibiteurs ciblant la poche du domaine KREP de KEAP1³⁰¹. Cette recherche de traitements par des petites molécules capables d'inhiber des interactions protéine-protéine est également investiguée du côté du paludisme.

Les poches de PfK13 et PfCRT, que nous prédisons comme fonctionnelles, pourraient représenter deux potentielles cibles thérapeutiques. En effet, ces deux protéines sont essentielles au moins pour le cycle érythrocytaire du parasite¹¹¹, qui est associé aux symptômes de la maladie. Altérer ou inhiber leur fonction pourrait empêcher / diminuer la croissance du parasite. Si on suppose que PfK13 contrôle les niveaux cellulaires d'un ou plusieurs substrats (lesquels pourraient se lier à la poche du domaine KREP de PfK13), déterminer des petites molécules qui cibleraient spécifiquement cette surface d'interaction permettrait de moduler directement la voie de dégradation des protéines. La voie métabolique du protéasome est extrêmement conservée chez *P. falciparum*¹¹¹ et représente une cible thérapeutique faisant l'objet de recherches récentes³⁰²⁻³⁰⁴, et est par ailleurs une des cibles des ARTDs⁶⁸. Concernant PfCRT, inhiber la poche aurait pour conséquence d'empêcher les substrats physiologiques d'accéder à leur destination finale, altérant ainsi la biologie du parasite. Une hypothèse circulant dans la littérature propose d'ailleurs que la CQ inhibe également PfCRT sauvage¹⁵⁵, un effet qui est plausible selon nos études préliminaires de *docking*. Une description fine de la poche du transporteur et de son éventuelle interaction avec la CQ permettrait d'explorer cette hypothèse.

- L'état homo-oligomérique de PfK13, qualifié jusque-là de dimère d'après la résolution cristallographique des domaines BTB et KREP, pourrait être plus complexe en raison du domaine coiled-coil.
- Diverses expérimentations sont proposées pour valider le rôle fonctionnel de la poche de PfK13 : *pull-down* différentiel, transfection chez le parasite *P. falciparum*, et étude de la pharmaco-modulation via les inhibiteurs KEAP1-Nrf2.
- Ces expérimentations auraient également pour objectif d'identifier des partenaires protéiques de PfK13, et d'évaluer l'existence d'un lien entre la résistance aux ARTDs et la poche du domaine KREP.
- Concernant le mécanisme de résistance aux ARTDs, l'hypothèse qui nous semble pour l'instant la plus cohérente avec les données disponibles serait que les changements d'acides aminés impliqués dans la résistance aux ARTDs diminueraient l'abondance de la protéine PfK13 plutôt que d'altérer sa zone d'interaction (supposément la poche).
- La poche située au centre du transporteur PfCRT servirait, par homologie avec les DMTs, à lier des substrats. Ces derniers seraient transportés de la vacuole digestive pour être libérés dans le cytoplasme du parasite.
- Quelques expérimentations sont proposées pour valider le rôle fonctionnel de la poche de PfCRT : transfection chez le parasite *P. falciparum* et systèmes d'expression hétérologues chez la levure et les ovocytes de Xénope.
- Le transport de CQ sous sa forme di-protonée serait lié à une plus grande électronégativité au niveau de la poche dans les formes mutantes de PfCRT.
- Le mécanisme de résistance à la PPQ, de par la conservation du potentiel électronégatif induit par les mutations de résistance à la CQ, devrait être différent de celui associé à la CQ.
- Des *docking* et simulations de dynamique moléculaire sont envisagés pour mieux comprendre les mécanismes et affinités de liaison au niveau de la poche avec le VP, la CQ et la PPQ.
- Les poches de PfK13 et PfCRT pourraient représenter deux potentielles cibles thérapeutiques.

Références bibliographiques

1. Cook, G. C. & Webb, A. J. Perceptions of malaria transmission before Ross' discovery in 1897. *Postgrad Med J* **76**, 738–740 (2000).
2. Hempelmann, E. & Krafts, K. Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malar. J.* **12**, 232 (2013).
3. Laval, G. *et al.* Recent Adaptive Acquisition by African Rainforest Hunter-Gatherers of the Late Pleistocene Sick-Cell Mutation Suggests Past Differences in Malaria Exposure. *Am. J. Hum. Genet.* **104**, 553–561 (2019).
4. Pappas, G., Kiriaze, I. J. & Falagas, M. E. Insights into infectious disease in the era of Hippocrates. *Int. J. Infect. Dis.* **12**, 347–350 (2008).
5. Joy, D. A. *et al.* Early origin and recent expansion of *Plasmodium falciparum*. *Science* **300**, 318–321 (2003).
6. Canali, S. Researches on thalassemia and malaria in Italy and the origins of the 'Haldane hypothesis'. *Med Secoli* **20**, 827–846 (2008).
7. Sallares, R., Bouwman, A. & Anderung, C. The spread of malaria to Southern Europe in antiquity: new approaches to old problems. *Med Hist* **48**, 311–328 (2004).
8. Yalcindag, E. *et al.* Multiple independent introductions of *Plasmodium falciparum* in South America. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 511–516 (2012).
9. Cox, F. E. History of the discovery of the malaria parasites and their vectors. *Parasit Vectors* **3**, 5 (2010).
10. Manson, P. Experimental proof of the mosquito-malaria theory. 1900. *Yale J Biol Med* **75**, 107–112 (2002).
11. The Nobel Prize in Physiology or Medicine 1902. <https://www.nobelprize.org/prizes/medicine/1902/ross/biographical/>.
12. Capanna, E. Grassi versus Ross: who solved the riddle of malaria? *Int. Microbiol.* **9**, 69–74 (2006).
13. Reed, W., Carroll, J., Agramonte, A. & Lazear, J. W. The Etiology of Yellow Fever—A Preliminary Note. *Public Health Pap Rep* **26**, 37–53 (1900).
14. Sutter, P. S. Nature's agents or agents of empire? Entomological workers and environmental change during the construction of the Panama Canal. *Isis* **98**, 724–754 (2007).
15. Votýpka, J., Modrý, D., Oborník, M., Šlapeta, J. & Lukeš, J. Apicomplexa. in *Handbook of the Protists* (eds. Archibald, J. M. *et al.*) 1–58 (Springer International Publishing, 2017).
16. Morrison, D. A. Evolution of the Apicomplexa: where are we now? *Trends Parasitol.* **25**, 375–382 (2009).
17. The Protozoan Phylum Apicomplexa: Volume 1. *CRC Press* Available at: <https://www.crcpress.com/The-Protozoan-Phylum-Apicomplexa-Volume-1/Levine/p/book/9781315897004>.
18. Mehlhorn, H., Peters, W. & Haberkorn, A. The formation of kinetes and oocyst in *Plasmodium gallinaceum* (Haemosporidia) and considerations on phylogenetic relationships between Haemosporidia, Piroplasmida and other Coccidia. *Protistologica* **16**, 135–154 (1980).
19. The Diversity of *Plasmodium* and Other Haemosporidians: The Intersection of Taxonomy, Phylogenetics and Genomics. Available at: <https://www.caister.com/hsp/abstracts/malaria/01.html>.
20. White, N. J. *et al.* Malaria. *Lancet* **383**, 723–735 (2014).
21. Gething, P. W. *et al.* A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar. J.* **10**, 378 (2011).
22. Gething, P. W. *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl Trop Dis* **6**, e1814 (2012).
23. Sabbatani, S., Fiorino, S. & Manfredi, R. The emerging of the fifth malaria parasite (*Plasmodium knowlesi*): a public health concern? *Braz J Infect Dis* **14**, 299–309 (2010).
24. WHO | World malaria report 2018. *WHO* Available at: <http://www.who.int/malaria/publications/world-malaria-report-2018/report/en/>.

25. Danis, K. *et al.* Autochthonous *Plasmodium vivax* malaria in Greece, 2011. *Eurosurveillance* **16**, 19993 (2011).
26. Prevention, C.-C. for D. C. and. CDC - Malaria - About Malaria - Biology. (2019). Available at: <https://www.cdc.gov/malaria/about/biology/index.html>.
27. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasit Vectors* **5**, 69 (2012).
28. Siciliano, G. & Alano, P. Enlightening the malaria parasite life cycle: bioluminescent *Plasmodium* in fundamental and applied research. *Front Microbiol* **6**, 391 (2015).
29. Lee, A. H., Symington, L. S. & Fidock, D. A. DNA repair mechanisms and their biological roles in the malaria parasite *Plasmodium falciparum*. *Microbiol. Mol. Biol. Rev.* **78**, 469–486 (2014).
30. Markus, M. B. Malaria: origin of the term 'hypnozoite'. *J Hist Biol* **44**, 781–786 (2011).
31. Wunderlich, J., Rohrbach, P. & Dalton, J. P. The malaria digestive vacuole. *Front Biosci (Schol Ed)* **4**, 1424–1448 (2012).
32. Krugliak, M., Zhang, J. & Ginsburg, H. 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. *Mol. Biochem. Parasitol.* **119**, 249–256 (2002).
33. Roepe, P. D. Molecular and physiologic basis of quinoline drug resistance in *Plasmodium falciparum* malaria. *Future Microbiol* **4**, 441–455 (2009).
34. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
35. Carlton, J. M. *et al.* Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* **455**, 757–763 (2008).
36. Pain, A. *et al.* The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799–803 (2008).
37. Rutledge, G. G. *et al.* *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. *Nature* **542**, 101–104 (2017).
38. Lindblade, K. A., Steinhardt, L., Samuels, A., Kachur, S. P. & Slutsker, L. The silent threat: asymptomatic parasitemia and malaria transmission. *Expert Rev Anti Infect Ther* **11**, 623–639 (2013).
39. WHO | Guidelines for the treatment of malaria. Third edition. WHO Available at: <http://www.who.int/malaria/publications/atoz/9789241549127/en/>.
40. TDR | Microscopy for the detection, identification and quantification of malaria parasites on stained thick and thin blood films in research settings. WHO Available at: https://www.who.int/tdr/publications/microscopy_detec_ident_quantif/en/.
41. WHO | Malaria rapid diagnostic test performance. Results of WHO product testing of malaria RDTs: round 7 (2015-2016). WHO Available at: <http://www.who.int/malaria/publications/atoz/978924151268/en/>.
42. Abba, K. *et al.* Rapid diagnostic tests for diagnosing uncomplicated non-falciparum or *Plasmodium vivax* malaria in endemic countries. *Cochrane Database Syst Rev* CD011431 (2014). doi:10.1002/14651858.CD011431
43. WHO | Guidelines for malaria vector control. WHO Available at: <http://www.who.int/malaria/publications/atoz/9789241550499/en/>.
44. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).
45. WHO | WHO Policy recommendation on intermittent preventive treatment during infancy with sulphadoxine-pyrimethamine (IPTi-SP) for *Plasmodium falciparum* malaria control in Africa.
46. WHO | Intermittent preventive treatment for infants using sulfadoxine-pyrimethamine (IPTi-SP) for malaria control in Africa: implementation field guide. WHO Available at: http://www.who.int/malaria/publications/atoz/whoivb11_07/en/.
47. Roser, M. & Ritchie, H. Malaria. *Our World in Data* (2013).

48. WHO | Management of severe malaria – A practical handbook. Third edition. *WHO* Available at: <http://www.who.int/malaria/publications/atoz/9789241548526/en/>.
49. WHO | Global report on antimalarial efficacy and drug resistance: 2000-2010. *WHO* Available at: <http://www.who.int/malaria/publications/atoz/9789241500470/en/>.
50. WHO | Status report on artemisinin resistance and ACT efficacy (August 2018). *WHO* Available at: <http://www.who.int/malaria/publications/atoz/artemisinin-resistance-august2018/en/>.
51. Fidock, D. A. *et al.* Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Mol. Cell* **6**, 861–871 (2000).
52. Sidhu, A. B. S., Verdier-Pinard, D. & Fidock, D. A. Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcr1 mutations. *Science* **298**, 210–213 (2002).
53. Wongsrichanalai, C., Pickard, A. L., Wernsdorfer, W. H. & Meshnick, S. R. Epidemiology of drug-resistant malaria. *The Lancet Infectious Diseases* **2**, 209–218 (2002).
54. Eklund, E. H. & Fidock, D. A. In vitro evaluations of antimalarial drugs and their relevance to clinical outcomes. *Int J Parasitol* **38**, 743–747 (2008).
55. Dondorp, A. M. *et al.* Artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med.* **361**, 455–467 (2009).
56. Miotto, O. *et al.* Emergence of artemisinin-resistant *Plasmodium falciparum* with kelch13 C580Y mutations on the island of New Guinea. *bioRxiv* 621813 (2019).
57. Jianfang, Z. *A Detailed Chronological Record of Project 523 and the Discovery and Development of Qinghaosu*. (Strategic Book Publishing & Rights Agency, LLC, 2013).
58. Miller, L. H. & Su, X. Artemisinin: Discovery from the Chinese Herbal Garden. *Cell* **146**, 855–858 (2011).
59. From *Artemisia annua* L. to Artemisinins - 1st Edition. Available at: <https://www.elsevier.com/books/from-artemisia-annua-l-to-artemisinins/tu/978-0-12-811655-5>.
60. Ansari, M. T. *et al.* Malaria and artemisinin derivatives: an updated review. *Mini Rev Med Chem* **13**, 1879–1902 (2013).
61. Meunier, B. & Robert, A. Heme as trigger and target for trioxane-containing antimalarial drugs. *Acc. Chem. Res.* **43**, 1444–1451 (2010).
62. Tilley, L., Straimer, J., Gnädig, N. F., Ralph, S. A. & Fidock, D. A. Artemisinin Action and Resistance in *Plasmodium falciparum*. *Trends in Parasitology* **32**, 682–696 (2016).
63. German, P. I. & Aweeka, F. T. Clinical pharmacology of artemisinin-based combination therapies. *Clin Pharmacokinet* **47**, 91–102 (2008).
64. Chotivanich, K. *et al.* Laboratory Detection of Artemisinin-Resistant *Plasmodium falciparum*. *Antimicrob Agents Chemother* **58**, 3157–3161 (2014).
65. Wang, J. & Lin, Q. Chemical proteomics approach reveals the direct targets and the heme-dependent activation mechanism of artemisinin in *Plasmodium falciparum* using an artemisinin-based activity probe. *Microb Cell* **3**, 230–231
66. Wang, J. *et al.* Haem-activated promiscuous targeting of artemisinin in *Plasmodium falciparum*. *Nat Commun* **6**, 10111 (2015).
67. Ismail, H. M. *et al.* Artemisinin activity-based probes identify multiple molecular targets within the asexual stage of the malaria parasites *Plasmodium falciparum* 3D7. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2080–2085 (2016).
68. Bridgford, J. L. *et al.* Artemisinin kills malaria parasites by damaging proteins and inhibiting the proteasome. *Nat Commun* **9**, 3801 (2018).
69. Fairhurst, R. M. & Dondorp, A. M. Artemisinin-Resistant *Plasmodium falciparum* Malaria. *Microbiol Spectr* **4**, (2016).

70. Buffet, P. A. *et al.* Ex vivo perfusion of human spleens maintains clearing and processing functions. *Blood* **107**, 3745–3752 (2006).
71. Pryce, J. & Hine, P. Pyronaridine-artesunate for treating uncomplicated *Plasmodium falciparum* malaria. *Cochrane Database Syst Rev* **1**, CD006404 (2019).
72. Nsanzabana, C. Resistance to Artemisinin Combination Therapies (ACTs): Do Not Forget the Partner Drug! *Tropical Medicine and Infectious Disease* **4**, 26 (2019).
73. Tse, E. G., Korsik, M. & Todd, M. H. The past, present and future of anti-malarial medicines. *Malaria Journal* **18**, 93 (2019).
74. Das, S., Saha, B., Hati, A. K. & Roy, S. Evidence of Artemisinin-Resistant *Plasmodium falciparum* Malaria in Eastern India. *New England Journal of Medicine* **379**, 1962–1964 (2018).
75. Das, S., Manna, S., Saha, B., Hati, A. K. & Roy, S. Novel p_{fk}elch13 gene polymorphism associates with artemisinin resistance in eastern India. *Clin. Infect. Dis.* (2018). doi:10.1093/cid/ciy1038
76. Rasmussen, C., Valecha, N. & Ringwald, P. Lack of convincing evidence of artemisinin resistance in India. *Clin. Infect. Dis.* (2019). doi:10.1093/cid/ciz166
77. Combrinck, J. M. *et al.* Insights into the role of heme in the mechanism of action of antimalarials. *ACS Chem. Biol.* **8**, 133–137 (2013).
78. Martin, R. E., Shafik, S. H. & Richards, S. N. Mechanisms of resistance to the partner drugs of artemisinin in the malaria parasite. *Curr Opin Pharmacol* **42**, 71–80 (2018).
79. Plucinski, M. M. *et al.* Efficacy of artemether-lumefantrine and dihydroartemisinin-piperaquine for treatment of uncomplicated malaria in children in Zaire and Uíge Provinces, Angola. *Antimicrob. Agents Chemother.* **59**, 437–443 (2015).
80. Okell, L. C. *et al.* Emerging implications of policies on malaria treatment: genetic changes in the P_fmdr-1 gene affecting susceptibility to artemether–lumefantrine and artesunate–amodiaquine in Africa. *BMJ Glob Health* **3**, (2018).
81. Sidhu, A. B. S. *et al.* Decreasing p_fmdr1 copy number in *Plasmodium falciparum* malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J. Infect. Dis.* **194**, 528–535 (2006).
82. Kaur, K., Jain, M., Reddy, R. P. & Jain, R. Quinolines and structurally related heterocycles as antimalarials. *Eur J Med Chem* **45**, 3245–3264 (2010).
83. Holmgren, G. *et al.* Amodiaquine resistant *Plasmodium falciparum* malaria in vivo is associated with selection of p_fcrt 76T and p_fmdr1 86Y. *Infect. Genet. Evol.* **6**, 309–314 (2006).
84. Desai, M. *et al.* Prevention of malaria in pregnancy. *The Lancet Infectious Diseases* **18**, e119–e132 (2018).
85. Amaratunga, C. *et al.* Dihydroartemisinin-piperaquine resistance in *Plasmodium falciparum* malaria in Cambodia: a multisite prospective cohort study. *Lancet Infect Dis* **16**, 357–365 (2016).
86. Leang, R. *et al.* Efficacy of dihydroartemisinin-piperaquine for treatment of uncomplicated *Plasmodium falciparum* and *Plasmodium vivax* in Cambodia, 2008 to 2010. *Antimicrob. Agents Chemother.* **57**, 818–826 (2013).
87. Dhingra, S. K. *et al.* A Variant P_fCRT Isoform Can Contribute to *Plasmodium falciparum* Resistance to the First-Line Partner Drug Piperaquine. *MBio* **8**, (2017).
88. Witkowski, B. *et al.* A surrogate marker of piperaquine-resistant *Plasmodium falciparum* malaria: a phenotype-genotype association study. *Lancet Infect Dis* **17**, 174–183 (2017).
89. Ross, L. S. *et al.* Emerging Southeast Asian P_fCRT mutations confer *Plasmodium falciparum* resistance to the first-line antimalarial piperaquine. *Nature Communications* **9**, 3314 (2018).

90. Looareesuwan, S. *et al.* Treatment of patients with recrudescence of falciparum malaria with a sequential combination of artesunate and mefloquine. *Am. J. Trop. Med. Hyg.* **47**, 794–799 (1992).
91. Cowman, A. F., Galatis, D. & Thompson, J. K. Selection for mefloquine resistance in *Plasmodium falciparum* is linked to amplification of the *pfmdr1* gene and cross-resistance to halofantrine and quinine. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1143–1147 (1994).
92. Wong, W. *et al.* Mefloquine targets the *Plasmodium falciparum* 80S ribosome to inhibit protein synthesis. *Nat Microbiol* **2**, 17031 (2017).
93. Price, R. N. *et al.* Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. *Lancet* **364**, 438–447 (2004).
94. Veiga, M. I. *et al.* Novel Polymorphisms in *Plasmodium falciparum* ABC Transporter Genes Are Associated with Major ACT Antimalarial Drug Resistance. *PLoS One* **6**, (2011).
95. Veiga, M. I. *et al.* Complex polymorphisms in the *Plasmodium falciparum* multidrug resistance protein 2 gene and its contribution to antimalarial response. *Antimicrob. Agents Chemother.* **58**, 7390–7397 (2014).
96. West African Network for Clinical Trials of Antimalarial Drugs (WANECAM). Pyronaridine-artesunate or dihydroartemisinin-piperaquine versus current first-line therapies for repeated treatment of uncomplicated malaria: a randomised, multicentre, open-label, longitudinal, controlled, phase 3b/4 trial. *Lancet* **391**, 1378–1390 (2018).
97. Leang, R. *et al.* Efficacy and Safety of Pyronaridine-Artesunate for Treatment of Uncomplicated *Plasmodium falciparum* Malaria in Western Cambodia. *Antimicrob. Agents Chemother.* **60**, 3884–3890 (2016).
98. Auparakkitanon, S., Chapoomram, S., Kuaha, K., Chirachariyavej, T. & Wilairat, P. Targeting of hematin by the antimalarial pyronaridine. *Antimicrob. Agents Chemother.* **50**, 2197–2200 (2006).
99. Madamet, M. *et al.* The *Plasmodium falciparum* chloroquine resistance transporter is associated with the ex vivo *P. falciparum* African parasite response to pyronaridine. *Parasit Vectors* **9**, 77 (2016).
100. Noedl, H. *et al.* Evidence of artemisinin-resistant malaria in western Cambodia. *N. Engl. J. Med.* **359**, 2619–2620 (2008).
101. Ikeda, M. *et al.* Artemisinin-Resistant *Plasmodium falciparum* with High Survival Rates, Uganda, 2014–2016. *Emerging Infect. Dis.* **24**, 718–726 (2018).
102. Doolan, D. L., Dobaño, C. & Baird, J. K. Acquired Immunity to Malaria. *Clin Microbiol Rev* **22**, 13–36 (2009).
103. Witkowski, B. *et al.* Novel phenotypic assays for the detection of artemisinin-resistant *Plasmodium falciparum* malaria in Cambodia: in-vitro and ex-vivo drug-response studies. *Lancet Infect Dis* **13**, 1043–1049 (2013).
104. Ariey, F. *et al.* A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* **505**, 50–55 (2014).
105. Straimer, J. *et al.* K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science* **347**, 428–431 (2015).
106. Ghorbal, M. *et al.* Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32**, 819–821 (2014).
107. Dhanoa, B. S., Cogliati, T., Satish, A. G., Bruford, E. A. & Friedman, J. S. Update on the Kelch-like (KLHL) gene family. *Hum Genomics* **7**, 13 (2013).
108. Adams, J., Kelso, R. & Cooley, L. The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol.* **10**, 17–24 (2000).
109. Jiang, D. Q. *et al.* Crystal Structure Analysis of Kelch protein from *Plasmodium falciparum*. *TO BE PUBLISHED* doi:10.2210/pdb4yy8/pdb

110. Jiang, D. Q. *et al.* Crystal structure of kelch protein with disulfide bond from *Plasmodium falciparum*. *TO BE PUBLISHED* doi:10.2210/pdb4zgc/pdb
111. Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360**, (2018).
112. Bushell, E. *et al.* Functional Profiling of a *Plasmodium* Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260-272.e8 (2017).
113. Anderson, T. J. C. *et al.* Population Parameters Underlying an Ongoing Soft Sweep in Southeast Asian Malaria Parasites. *Mol Biol Evol* **34**, 131-144 (2017).
114. Kamau, E. *et al.* K13-Propeller Polymorphisms in *Plasmodium falciparum* Parasites From Sub-Saharan Africa. *J Infect Dis* **211**, 1352-1355 (2015).
115. MalariaGEN *Plasmodium falciparum* Community Project. Genomic epidemiology of artemisinin resistant malaria. *Elife* **5**, (2016).
116. Ménard, D. *et al.* A Worldwide Map of *Plasmodium falciparum* K13-Propeller Polymorphisms. *N Engl J Med* **374**, 2453-2464 (2016).
117. Lu, F. *et al.* Emergence of Indigenous Artemisinin-Resistant *Plasmodium falciparum* in Africa. *N. Engl. J. Med.* **376**, 991-993 (2017).
118. Tun, K. M. *et al.* Spread of artemisinin-resistant *Plasmodium falciparum* in Myanmar: a cross-sectional survey of the K13 molecular marker. *The Lancet Infectious Diseases* **15**, 415-421 (2015).
119. Nair, S. *et al.* Fitness Costs and the Rapid Spread of kelch13-C580Y Substitutions Conferring Artemisinin Resistance. *Antimicrob. Agents Chemother.* **62**, (2018).
120. Straimer, J. *et al.* *Plasmodium falciparum* K13 Mutations Differentially Impact Ozonide Susceptibility and Parasite Fitness In Vitro. *MBio* **8**, (2017).
121. Birnbaum, J. *et al.* A genetic system to study *Plasmodium falciparum* protein function. *Nat. Methods* **14**, 450-456 (2017).
122. Witkowski, B. *et al.* Reduced artemisinin susceptibility of *Plasmodium falciparum* ring stages in western Cambodia. *Antimicrob. Agents Chemother.* **57**, 914-923 (2013).
123. Zhang, M. *et al.* Inhibiting the *Plasmodium* eIF2 α Kinase PK4 Prevents Artemisinin-Induced Latency. *Cell Host & Microbe* **22**, 766-776.e4 (2017).
124. Mok, S. *et al.* Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* **347**, 431-435 (2015).
125. Dogovski, C. *et al.* Targeting the Cell Stress Response of *Plasmodium falciparum* to Overcome Artemisinin Resistance. *PLOS Biology* **13**, e1002132 (2015).
126. Mbengue, A. *et al.* A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature* **520**, 683-687 (2015).
127. Bhattacharjee, S. *et al.* Remodeling of the malaria parasite and host human red cell by vesicle amplification that induces artemisinin resistance. *Blood* (2018). doi:10.1182/blood-2017-11-814665
128. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331-D338 (2017).
129. Canning, P. *et al.* Structural Basis for Cul3 Protein Assembly with the BTB-Kelch Family of E3 Ubiquitin Ligases. *J. Biol. Chem.* **288**, 7803-7814 (2013).
130. Geyer, R., Wee, S., Anderson, S., Yates, J. & Wolf, D. A. BTB/POZ Domain Proteins Are Putative Substrate Adaptors for Cullin 3 Ubiquitin Ligases. *Molecular Cell* **12**, 783-790 (2003).
131. Canning, P., Sorrell, F. J. & Bullock, A. N. Structural basis of Keap1 interactions with Nrf2. *Free Radic Biol Med* **88**, 101-107 (2015).

132. Jung, B.-J., Yoo, H.-S., Shin, S., Park, Y.-J. & Jeon, S.-M. Dysregulation of NRF2 in Cancer: from Molecular Mechanisms to Therapeutic Opportunities. *Biomol Ther (Seoul)* **26**, 57–68 (2018).
133. Sutherland, C. J. *et al.* pfk13-Independent Treatment Failure in Four Imported Cases of Plasmodium falciparum Malaria Treated with Artemether-Lumefantrine in the United Kingdom. *Antimicrob. Agents Chemother.* **61**, (2017).
134. Muwanguzi, J. *et al.* Lack of K13 mutations in Plasmodium falciparum persisting after artemisinin combination therapy treatment of Kenyan children. *Malar. J.* **15**, 36 (2016).
135. Bayih, A. G. *et al.* A Unique Plasmodium falciparum K13 Gene Mutation in Northwest Ethiopia. *Am. J. Trop. Med. Hyg.* **94**, 132–135 (2016).
136. Witkowski, B. *et al.* Increased tolerance to artemisinin in Plasmodium falciparum is mediated by a quiescence mechanism. *Antimicrob. Agents Chemother.* **54**, 1872–1877 (2010).
137. Demas, A. R. *et al.* Mutations in Plasmodium falciparum actin-binding protein coronin confer reduced artemisinin susceptibility. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12799–12804 (2018).
138. Hunt, P. *et al.* Experimental evolution, genetic analysis and genome re-sequencing reveal the mutation conferring artemisinin resistance in an isogenic lineage of malaria parasites. *BMC Genomics* **11**, 499 (2010).
139. Hunt, P. *et al.* Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. *Mol Microbiol* **65**, 27–40 (2007).
140. Borrmann, S. *et al.* Genome-wide screen identifies new candidate genes associated with artemisinin susceptibility in Plasmodium falciparum in Kenya. *Sci Rep* **3**, 3318 (2013).
141. Henriques, G. *et al.* Artemisinin resistance in rodent malaria--mutation in the AP2 adaptor μ -chain suggests involvement of endocytosis and membrane protein trafficking. *Malar. J.* **12**, 118 (2013).
142. Henriques, G. *et al.* The Mu subunit of Plasmodium falciparum clathrin-associated adaptor protein 2 modulates in vitro parasite response to artemisinin and quinine. *Antimicrob. Agents Chemother.* **59**, 2540–2547 (2015).
143. Adams, T. *et al.* Prevalence of Plasmodium falciparum delayed clearance associated polymorphisms in adaptor protein complex 2 mu subunit (pfap2mu) and ubiquitin specific protease 1 (pfubp1) genes in Ghanaian isolates. *Parasit Vectors* **11**, (2018).
144. Amambua-Ngwa, A. *et al.* Major subpopulations of Plasmodium falciparum in sub-Saharan Africa. *Science* **365**, 813–816 (2019).
145. Siddiqui, F. A. *et al.* Plasmodium falciparum Falcipain-2a Polymorphisms in Southeast Asia and Their Association With Artemisinin Resistance. *J. Infect. Dis.* **218**, 434–442 (2018).
146. Wang, Z. *et al.* Genome-wide association analysis identifies genetic loci associated with resistance to multiple antimalarials in Plasmodium falciparum from China-Myanmar border. *Scientific Reports* **6**, 33891 (2016).
147. Breglio, K. F. *et al.* A single nucleotide polymorphism in the Plasmodium falciparum atg18 gene associates with artemisinin resistance and confers enhanced parasite survival under nutrient deprivation. *Malar. J.* **17**, 391 (2018).
148. Nainwal, L. M. *et al.* Green recipes to quinoline: A review. *European Journal of Medicinal Chemistry* **164**, 121–170 (2019).
149. Kaufman, T. S. & Rúveda, E. A. The quest for quinine: those who won the battles and those who won the war. *Angew. Chem. Int. Ed. Engl.* **44**, 854–885 (2005).
150. Leslie Gunatilaka, A. A. Chapter 1 - Alkaloids from Sri Lankan Flora**This chapter is dedicated with great admiration to the memory of Professor Sir Derek H. R. Barton, a mentor who provided me with an unforgettable introduction to the world of natural

- products. in *The Alkaloids: Chemistry and Biology* (ed. Cordell, G. A.) **52**, 1–101 (Academic Press, 1999).
151. Singh, S. K. & Singh, S. A brief history of quinoline as antimalarial agents. *International Journal of Pharmaceutical Sciences Review and Research* **25**, 295–302 (2014).
152. Bhattacharjee, M. K. *Chemistry of Antibiotics and Related Drugs*. (Springer, 2016).
153. Bray, P. G. *et al.* PfCRT and the trans-vacuolar proton electrochemical gradient: regulating the access of chloroquine to ferriprotoporphyrin IX. *Mol Microbiol* **62**, 238–251 (2006).
154. Foley, M. & Tilley, L. Quinoline antimalarials: mechanisms of action and resistance and prospects for new agents. *Pharmacol. Ther.* **79**, 55–87 (1998).
155. Ecker, A., Lehane, A. M., Clain, J. & Fidock, D. A. PfCRT and its role in antimalarial drug resistance. *Trends Parasitol.* **28**, 504–514 (2012).
156. Tanner, M. *et al.* Malaria eradication and elimination: views on how to translate a vision into reality. *BMC Medicine* **13**, 167 (2015).
157. Wellems, T. E., Walker-Jonah, A. & Panton, L. J. Genetic mapping of the chloroquine-resistance locus on *Plasmodium falciparum* chromosome 7. *PNAS* **88**, 3382–3386 (1991).
158. Martin, R. E. & Kirk, K. The malaria parasite's chloroquine resistance transporter is a member of the drug/metabolite transporter superfamily. *Mol. Biol. Evol.* **21**, 1938–1949 (2004).
159. Yen, M. R., Chen, J. S., Marquez, J. L., Sun, E. I. & Saier, M. H. Multidrug resistance: phylogenetic characterization of superfamilies of secondary carriers that include drug exporters. *Methods Mol. Biol.* **637**, 47–64 (2010).
160. Juge, N. *et al.* *Plasmodium falciparum* chloroquine resistance transporter is a H⁺-coupled polyspecific nutrient and drug exporter. *Proc Natl Acad Sci U S A* **112**, 3356–3361 (2015).
161. Zhang, H., Howard, E. M. & Roepe, P. D. Analysis of the antimalarial drug resistance protein PfCRT expressed in yeast. *J. Biol. Chem.* **277**, 49767–49775 (2002).
162. Zhang, H., Paguio, M. & Roepe, P. D. The antimalarial drug resistance protein *Plasmodium falciparum* chloroquine resistance transporter binds chloroquine. *Biochemistry* **43**, 8290–8296 (2004).
163. Nessler, S. *et al.* Evidence for activation of endogenous transporters in *Xenopus laevis* oocytes expressing the *Plasmodium falciparum* chloroquine resistance transporter, PfCRT. *J. Biol. Chem.* **279**, 39438–39446 (2004).
164. Maughan, S. C. *et al.* Plant homologs of the *Plasmodium falciparum* chloroquine-resistance transporter, PfCRT, are required for glutathione homeostasis and stress responses. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2331–2336 (2010).
165. Patzewitz, E.-M. *et al.* Glutathione transport: a new role for PfCRT in chloroquine resistance. *Antioxid. Redox Signal.* **19**, 683–695 (2013).
166. Bakouh, N. *et al.* Iron is a substrate of the *Plasmodium falciparum* chloroquine resistance transporter PfCRT in *Xenopus* oocytes. *J. Biol. Chem.* **292**, 16109–16121 (2017).
167. Pulcini, S. *et al.* Mutations in the *Plasmodium falciparum* chloroquine resistance transporter, PfCRT, enlarge the parasite's food vacuole and alter drug sensitivities. *Sci Rep* **5**, 14552 (2015).
168. Lewis, I. A. *et al.* Metabolic QTL analysis links chloroquine resistance in *Plasmodium falciparum* to impaired hemoglobin catabolism. *PLoS Genet.* **10**, e1004085 (2014).
169. Gabryszewski, S. J. *et al.* Evolution of Fitness Cost-Neutral Mutant PfCRT Conferring P. *falciparum* 4-Aminoquinoline Drug Resistance Is Accompanied by Altered Parasite Metabolism and Digestive Vacuole Physiology. *PLOS Pathogens* **12**, e1005976 (2016).
170. Martin, R. E. *et al.* Chloroquine transport via the malaria parasite's chloroquine resistance transporter. *Science* **325**, 1680–1682 (2009).

171. Lehane, A. M., Hayward, R., Saliba, K. J. & Kirk, K. A verapamil-sensitive chloroquine-associated H⁺ leak from the digestive vacuole in chloroquine-resistant malaria parasites. *J. Cell. Sci.* **121**, 1624–1632 (2008).
172. Lehane, A. M. & Kirk, K. Chloroquine Resistance-Confering Mutations in pfcr1 Give Rise to a Chloroquine-Associated H⁺ Leak from the Malaria Parasite's Digestive Vacuole. *Antimicrob Agents Chemother* **52**, 4374–4380 (2008).
173. Naudé, B., Brzostowski, J. A., Kimmel, A. R. & Wellems, T. E. Dictyostelium discoideum expresses a malaria chloroquine resistance mechanism upon transfection with mutant, but not wild-type, Plasmodium falciparum transporter PfCRT. *J. Biol. Chem.* **280**, 25596–25603 (2005).
174. Durrand, V. *et al.* Variations in the sequence and expression of the Plasmodium falciparum chloroquine resistance transporter (Pfcr1) and their relationship to chloroquine resistance in vitro. *Mol. Biochem. Parasitol.* **136**, 273–285 (2004).
175. Dhingra, S. K. *et al.* Global Spread of Mutant PfCRT and Its Pleiotropic Impact on Plasmodium falciparum Multidrug Resistance and Fitness. *mBio* **10**, e02731-18 (2019).
176. Chen, N. *et al.* pfcr1 Allelic types with two novel amino acid mutations in chloroquine-resistant Plasmodium falciparum isolates from the Philippines. *Antimicrob. Agents Chemother.* **47**, 3500–3505 (2003).
177. Martin, S. K., Oduola, A. M. & Milhous, W. K. Reversal of chloroquine resistance in Plasmodium falciparum by verapamil. *Science* **235**, 899–901 (1987).
178. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, A. F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in Plasmodium falciparum. *Nature* **403**, 906–909 (2000).
179. Hayward, R., Saliba, K. J. & Kirk, K. pfmdr1 mutations associated with chloroquine resistance incur a fitness cost in Plasmodium falciparum. *Mol. Microbiol.* **55**, 1285–1295 (2005).
180. Picot, S. *et al.* A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria. *Malar. J.* **8**, 89 (2009).
181. Laufer, M. K. *et al.* Return of Chloroquine Antimalarial Efficacy in Malawi. *New England Journal of Medicine* **355**, 1959–1966 (2006).
182. Lu, F. *et al.* Return of chloroquine sensitivity to Africa? Surveillance of African Plasmodium falciparum chloroquine resistance through malaria imported to China. *Parasit Vectors* **10**, (2017).
183. Mwanza, S. *et al.* The return of chloroquine-susceptible Plasmodium falciparum malaria in Zambia. *Malar J* **15**, (2016).
184. Manske, M. *et al.* Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
185. Amato, R. *et al.* Genetic markers associated with dihydroartemisinin-piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study. *Lancet Infect Dis* **17**, 164–173 (2017).
186. Spring, M. D. *et al.* Dihydroartemisinin-piperaquine failure associated with a triple mutant including kelch13 C580Y in Cambodia: an observational cohort study. *Lancet Infect Dis* **15**, 683–691 (2015).
187. Agrawal, S. *et al.* Association of a Novel Mutation in the Plasmodium falciparum Chloroquine Resistance Transporter With Decreased Piperaquine Sensitivity. *J. Infect. Dis.* **216**, 468–476 (2017).
188. Zhang, J. & Yang, J.-R. Determinants of the rate of protein sequence evolution. *Nat. Rev. Genet.* **16**, 409–420 (2015).

189. Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109–121 (2016).
190. Kimura, M. & Ohta, T. On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2848–2852 (1974).
191. Huang, Y.-F. & Golding, G. B. FuncPatch: a web server for the fast Bayesian inference of conserved functional patches in protein 3D structures. *Bioinformatics* **31**, 523–531 (2015).
192. Huang, Y.-F. & Golding, G. B. Phylogenetic Gaussian Process Model for the Inference of Functionally Important Regions in Protein Tertiary Structures. *PLOS Computational Biology* **10**, e1003429 (2014).
193. Coppée, R., Jeffares, D. C., Miteva, M. A., Sabbagh, A. & Clain, J. Comparative structural and evolutionary analyses predict functional sites in the artemisinin resistance malaria protein K13. *Sci Rep* **9**, 10675 (2019).
194. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
195. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
196. Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449 (2000).
197. Pond, S. L. K. & Frost, S. D. W. A simple hierarchical approach to modeling distributions of substitution rates. *Mol. Biol. Evol.* **22**, 223–234 (2005).
198. Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998).
199. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**, 568–573 (1998).
200. Vuong, Q. H. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* **57**, 307–333 (1989).
201. Yang, Z. & Swanson, W. J. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* **19**, 49–57 (2002).
202. Rodrigue, N., Lartillot, N., Bryant, D. & Philippe, H. Site interdependence attributed to tertiary structure in amino acid sequence evolution. *Gene* **347**, 207–217 (2005).
203. Peach, R. J. *et al.* Both extracellular immunoglobulin-like domains of CD80 contain residues critical for binding T cell surface receptors CTLA-4 and CD28. *J. Biol. Chem.* **270**, 21181–21187 (1995).
204. Deng, H., Jia, Y. & Zhang, Y. Protein structure prediction. *Int J Mod Phys B* **32**, (2018).
205. Krieger, E., Nabuurs, S. B. & Vriend, G. Homology modeling. *Methods Biochem Anal* **44**, 509–523 (2003).
206. Lee, J., Wu, S. & Zhang, Y. Ab Initio Protein Structure Prediction. in *From Protein Structure to Function with Bioinformatics* (ed. Rigden, D. J.) 3–25 (Springer Netherlands, 2009). doi:10.1007/978-1-4020-9058-5_1
207. Wodak, S. J. & Rooman, M. J. Generating and testing protein folds. *Current Opinion in Structural Biology* **3**, 247–259 (1993).
208. Kocher, J. P., Rooman, M. J. & Wodak, S. J. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J. Mol. Biol.* **235**, 1598–1613 (1994).
209. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858 (2015).
210. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacol. Rev.* **66**, 334–395 (2014).

211. Chothia, C. Structural invariants in protein folding. *Nature* **254**, 304–308 (1975).
212. Chang, J.-M., Di Tommaso, P., Taly, J.-F. & Notredame, C. Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinformatics* **13 Suppl 4**, S1 (2012).
213. Webb, B. & Sali, A. Protein Structure Modeling with MODELLER. *Methods Mol. Biol.* **1654**, 39–54 (2017).
214. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114–122 (2009).
215. Feig, M. Computational protein structure refinement: Almost there, yet still so far to go. *Wiley Interdiscip Rev Comput Mol Sci* **7**, (2017).
216. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).
217. Studer, G., Biasini, M. & Schwede, T. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics* **30**, i505–511 (2014).
218. Perutz, M. F., Kendrew, J. C. & Watson, H. C. Structure and function of haemoglobin: II. Some relations between polypeptide chain configuration and amino acid sequence. *Journal of Molecular Biology* **13**, 669–678 (1965).
219. Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J. & Wilke, C. O. Maximum allowed solvent accessibilities of residues in proteins. *PLoS ONE* **8**, e80635 (2013).
220. Goldman, N., Thorne, J. L. & Jones, D. T. Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics* **149**, 445–458 (1998).
221. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
222. Bloom, J. D., Drummond, D. A., Arnold, F. H. & Wilke, C. O. Structural determinants of the rate of protein evolution in yeast. *Mol. Biol. Evol.* **23**, 1751–1761 (2006).
223. Marcos, M. L. & Echave, J. Too packed to change: side-chain packing and site-specific substitution rates in protein evolution. *PeerJ* **3**, e911 (2015).
224. Sydykova, D. K., Jack, B. R., Spielman, S. J. & Wilke, C. O. Measuring evolutionary rates of proteins in a structural context. *F1000Res* **6**, 1845 (2017).
225. Nevin Gerek, Z., Kumar, S. & Banu Ozkan, S. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol Appl* **6**, 423–433 (2013).
226. Shahmoradi, A. *et al.* Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design. *J. Mol. Evol.* **79**, 130–142 (2014).
227. Liu, Y. & Bahar, I. Sequence evolution correlates with structural dynamics. *Mol. Biol. Evol.* **29**, 2253–2263 (2012).
228. Weiner, P. K., Langridge, R., Blaney, J. M., Schaefer, R. & Kollman, P. A. Electrostatic potential molecular surfaces. *PNAS* **79**, 3754–3758 (1982).
229. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
230. Gunsteren, W. F. van & Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angewandte Chemie International Edition in English* **29**, 992–1023 (1990).
231. Guvench, O. & MacKerell, A. D. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* **443**, 63–88 (2008).

232. Cheng, I.-C., Chen, Y.-J., Ku, C.-W., Huang, Y.-W. & Yang, C.-N. Structural and Dynamic Characterization of Mutated Keap1 for Varied Affinity toward Nrf2: A Molecular Dynamics Simulation Study. *J Chem Inf Model* **55**, 2178–2186 (2015).
233. Lu, M.-C. *et al.* A systematic molecular dynamics approach to the study of peptide Keap1-Nrf2 protein-protein interaction inhibitors and its application to p62 peptides. *Mol Biosyst* **12**, 1378–1387 (2016).
234. Wang, L., Jiang, C., Cai, R., Chen, X.-Z. & Peng, J.-B. Unveiling the Distinct Mechanisms by which Disease-Causing Mutations in the Kelch Domain of KLHL3 Disrupt the Interaction with the Acidic Motif of WNK4 through Molecular Dynamics Simulation. *Biochemistry* **58**, 2105–2115 (2019).
235. Lindahl, E. R. Molecular dynamics simulations. *Methods Mol. Biol.* **443**, 3–23 (2008).
236. González, M. A. Force fields and molecular dynamics simulations. *JDN* **12**, 169–200 (2011).
237. Rozmanov, D. & Kusalik, P. G. Robust rotational-velocity-Verlet integration methods. *Phys Rev E Stat Nonlin Soft Matter Phys* **81**, 056706 (2010).
238. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
239. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182–7190 (1981).
240. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
241. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
242. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* **7**, 146–157 (2011).
243. Ferreira, L. G., Dos Santos, R. N., Oliva, G. & Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **20**, 13384–13421 (2015).
244. Goodsell, D. S. & Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **8**, 195–202 (1990).
245. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461 (2010).
246. Louche, A., Salcedo, S. P. & Bigot, S. Protein-Protein Interactions: Pull-Down Assays. *Methods Mol. Biol.* **1615**, 247–255 (2017).
247. Lee, C. Coimmunoprecipitation assay. *Methods Mol. Biol.* **362**, 401–406 (2007).
248. Kimple, M. E., Brill, A. L. & Pasker, R. L. Overview of Affinity Tags for Protein Purification. *Curr Protoc Protein Sci* **73**, Unit-9.9 (2013).
249. Davies, T. G. *et al.* Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J. Med. Chem.* **59**, 3991–4006 (2016).
250. Fukutomi, T., Takagi, K., Mizushima, T., Ohuchi, N. & Yamamoto, M. Kinetic, thermodynamic, and structural characterizations of the association between Nrf2-DLGex degon and Keap1. *Mol. Cell. Biol.* **34**, 832–846 (2014).
251. Zor, T. & Selinger, Z. Linearization of the Bradford protein assay increases its sensitivity: theoretical and experimental studies. *Anal. Biochem.* **236**, 302–308 (1996).
252. Warrenfeltz, S. *et al.* EuPathDB: The Eukaryotic Pathogen Genomics Database Resource. *Methods Mol. Biol.* **1757**, 69–113 (2018).

253. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* (2017). doi:10.1093/bib/bbx108
254. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
255. Lefort, V., Longueville, J.-E. & Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* **34**, 2422–2424 (2017).
256. Jeffares, D. C., Tomiczek, B., Sojo, V. & dos Reis, M. A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. *Methods Mol. Biol.* **1201**, 65–90 (2015).
257. Grigoryan, G. & Keating, A. E. Structural specificity in coiled-coil interactions. *Curr Opin Struct Biol* **18**, 477–483 (2008).
258. Truebestein, L. & Leonard, T. A. Coiled-coils: The long and short of it. *Bioessays* **38**, 903–916 (2016).
259. Liu, R. & Hu, J. DNABind: a hybrid algorithm for structure-based prediction of DNA-binding residues by combining machine learning- and template-based approaches. *Proteins* **81**, 1885–1899 (2013).
260. Nguyen, H. C., Yang, H., Fribourgh, J. L., Wolfe, L. S. & Xiong, Y. Insights into Cullin-RING E3 ubiquitin ligase recruitment: structure of the VHL-EloBC-Cul2 complex. *Structure* **23**, 441–449 (2015).
261. Pinkas, D. M. *et al.* Structural complexity in the KCTD family of Cullin3-dependent E3 ubiquitin ligases. *Biochem. J.* **474**, 3747–3761 (2017).
262. Schumacher, F.-R., Sorrell, F. J., Alessi, D. R., Bullock, A. N. & Kurz, T. Structural and biochemical characterization of the KLHL3–WNK kinase interaction important in blood pressure regulation. *Biochemical Journal* **460**, 237–246 (2014).
263. Angers, S. *et al.* The KLHL12–Cullin-3 ubiquitin ligase negatively regulates the Wnt– β -catenin pathway by targeting Dishevelled for degradation. *Nature Cell Biology* **8**, 348–357 (2006).
264. Jhamb, K. & Sahoo, D. K. Production of soluble recombinant proteins in Escherichia coli: effects of process conditions and chaperone co-expression on cell growth and production of xylanase. *Bioresour. Technol.* **123**, 135–143 (2012).
265. Frangioni, J. V. & Neel, B. G. Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* **210**, 179–187 (1993).
266. Harper, S. & Speicher, D. W. Purification of proteins fused to glutathione S-transferase. *Methods Mol Biol* **681**, 259–280 (2011).
267. Massiah, M. A., Wright, K. M. & Du, H. Obtaining Soluble Folded Proteins from Inclusion Bodies Using Sarkosyl, Triton X-100, and CHAPS: Application to LB and M9 Minimal Media. *Curr Protoc Protein Sci* **84**, 6.13.1–6.13.24 (2016).
268. Park, D.-W. *et al.* Improved recovery of active GST-fusion proteins from insoluble aggregates: solubilization and purification conditions using PKM2 and HtrA2 as model proteins. *BMB Rep* **44**, 279–284 (2011).
269. Thain, A., Gaston, K., Jenkins, O. & Clarke, A. R. A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends Genet.* **12**, 209–210 (1996).
270. Cosconati, S. *et al.* Virtual Screening with AutoDock: Theory and Practice. *Expert Opin Drug Discov* **5**, 597–607 (2010).
271. Lee, Y. *et al.* Structure of the triose-phosphate/phosphate translocator reveals the basis of substrate specificity. *Nat Plants* **3**, 825–832 (2017).
272. Parker, J. L. & Newstead, S. Structural basis of nucleotide sugar transport across the Golgi membrane. *Nature* **551**, 521–524 (2017).

273. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350 (2016).
274. Ogden, K. K. *et al.* Molecular Mechanism of Disease-Associated Mutations in the Pre-M1 Helix of NMDA Receptors and Potential Rescue Pharmacology. *PLoS Genet* **13**, (2017).
275. Philips, J. & Herskowitz, I. Identification of Kel1p, a kelch domain-containing protein involved in cell fusion and morphology in *Saccharomyces cerevisiae*. *J. Cell Biol.* **143**, 375–389 (1998).
276. De Smaele, E. *et al.* Identification and characterization of KCASH2 and KCASH3, 2 novel Cullin3 adaptors suppressing histone deacetylase and Hedgehog activity in medulloblastoma. *Neoplasia* **13**, 374–385 (2011).
277. Correale, S. *et al.* Molecular organization of the cullin E3 ligase adaptor KCTD11. *Biochimie* **93**, 715–724 (2011).
278. pubmeddev & al, D. X., et. The interaction of KCTD1 with transcription factor AP-2alpha inhibits its transactivation. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/19115315>. (Accessed: 5th September 2019)
279. Zarelli, V. E. & Dawid, I. B. Inhibition of neural crest formation by Kctd15 involves regulation of transcription factor AP-2. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2870 (2013).
280. Eckert, D., Buhl, S., Weber, S., Jäger, R. & Schorle, H. The AP-2 family of transcription factors. *Genome Biology* **6**, 246 (2005).
281. Hara, T. *et al.* Novel Kelch-like Protein, KLEIP, Is Involved in Actin Assembly at Cell-Cell Contact Sites of Madin-Darby Canine Kidney Cells. *Molecular Biology of the Cell* **15**, 1172 (2004).
282. Ge, W., Hu, H., Ding, K., Sun, L. & Zheng, S. Protein interaction analysis of ST14 domains and their point and deletion mutants. *J. Biol. Chem.* **281**, 7406–7412 (2006).
283. Aguilar-Martinez, E. *et al.* Screen for multi-SUMO-binding proteins reveals a multi-SIM-binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4854-4863 (2015).
284. Rohman M, Harrison-Lavoie KJ. Separation of copurifying GroEL from glutathione-S-transferase fusion proteins. *Protein Expr. Purif.* **20**, 45-7 (2000).
285. Khan, H. *et al.* Molecular effects of cancer-associated somatic mutations on the structural and target recognition properties of Keap1. *Biochem. J.* **467**, 141–151 (2015).
286. Hast, B. E. *et al.* Cancer-derived mutations in KEAP1 impair NRF2 degradation but not ubiquitination. *Cancer Res* **74**, 808–817 (2014).
287. Padmanabhan, B. *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol. Cell* **21**, 689–700 (2006).
288. de Laurent, Z. R. *et al.* Polymorphisms in the K13 Gene in *Plasmodium falciparum* from Different Malaria Transmission Areas of Kenya. *Am. J. Trop. Med. Hyg.* **98**, 1360–1366 (2018).
289. Siddiqui, G., Srivastava, A., Russell, A. S. & Creek, D. J. Multi-omics Based Identification of Specific Biochemical Changes Associated With PfKelch13-Mutant Artemisinin-Resistant *Plasmodium falciparum*. *J. Infect. Dis.* **215**, 1435–1444 (2017).
290. Boizot, A. *et al.* The instability of the BTB-KELCH protein Gigaxonin causes Giant Axonal Neuropathy and constitutes a new penetrant and specific diagnostic test. *Acta Neuropathol Commun* **2**, 47 (2014).
291. He, Y. *et al.* Artemisinin resistance-associated markers in *Plasmodium falciparum* parasites from the China-Myanmar border: predicted structural stability of K13 propeller variants detected in a low-prevalence area. *PLoS One* **14**, (2019).
292. Halder, K., Bhattacharjee, S. & Safeukui, I. Drug resistance in *Plasmodium*. *Nat. Rev. Microbiol.* **16**, 156–170 (2018).

293. van der Reest, J., Lilla, S., Zheng, L., Zanivan, S. & Gottlieb, E. Proteome-wide analysis of cysteine oxidation reveals metabolic sensitivity to redox stress. *Nat Commun* **9**, (2018).
294. Tsuchiya, H. *et al.* Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature* **534**, 417–420 (2016).
295. Sanchez, C. P. *et al.* Phosphomimetic substitution at Ser-33 of the chloroquine resistance transporter PfCRT reconstitutes drug responses in *Plasmodium falciparum*. *J. Biol. Chem.* (2019). doi:10.1074/jbc.RA119.009464
296. Kuhn, Y. *et al.* Trafficking of the phosphoprotein PfCRT to the digestive vacuolar membrane in *Plasmodium falciparum*. *Traffic* **11**, 236–249 (2010).
297. Lekostaj, J. K., Natarajan, J. K., Paguio, M. F., Wolf, C. & Roepe, P. D. Photoaffinity labeling of the *Plasmodium falciparum* chloroquine resistance transporter with a novel perfluorophenylazido chloroquine. *Biochemistry* **47**, 10394–10406 (2008).
298. Shashidharan, A., Singh, R., Bhasker, S. & Mohankumar, C. Physicochemical characterization and functional site analysis of lactoferrin gene of Vechur cow. *Bioinformation* **6**, 275–278 (2011).
299. Pelleau, S. *et al.* Adaptive evolution of malaria parasites in French Guiana: Reversal of chloroquine resistance by acquisition of a mutation in pfcr. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11672–11677 (2015).
300. Bellanca, S. *et al.* Multiple drugs compete for transport via the *Plasmodium falciparum* chloroquine resistance transporter at distinct but interdependent sites. *J. Biol. Chem.* **289**, 36336–36351 (2014).
301. Hoelder, S., Clarke, P. A. & Workman, P. Discovery of small molecule cancer drugs: Successes, challenges and opportunities. *Mol Oncol* **6**, 155–176 (2012).
302. Li, H. *et al.* Structure- and function-based design of *Plasmodium*-selective proteasome inhibitors. *Nature* **530**, 233–236 (2016).
303. Stokes, B. H. *et al.* Covalent *Plasmodium falciparum*-selective proteasome inhibitors exhibit a low propensity for generating resistance in vitro and synergize with multiple antimalarial agents. *PLoS Pathog.* **15**, e1007722 (2019).
304. Kirkman, L. A. *et al.* Antimalarial proteasome inhibitor reveals collateral sensitivity from intersubunit interactions and fitness cost of resistance. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6863–E6870 (2018).

Annexes

Publications annexes

En 2018, Nicaise Tuikue Ndam et Bernard Tornyigah (UMR 261) ont mené une étude épidémiologique axée sur le génotypage des principaux marqueurs de résistance aux médicaments incluant le gène *pfk13* parmi des femmes enceintes présentant une infection à *P. falciparum*, au Ghana. Notre rôle fut plutôt anecdotique, avec pour objectif d'ériger un dictionnaire recensant l'ensemble des mutations non-synonymes localisées sur le domaine KREP afin d'établir ou non la présence de mutations non-rapportés jusqu'alors. Également, un second objectif fut d'étudier leur localisation sur la structure, à la lumière des observations rapportées dans le projet majeur de thèse sur Pfk13.

- **Polymorphism of genes associated with sulfadoxine-pyrimethamine and artemisinin-based combination treatments among pregnant women in Ghana.**

Tornyigah B, Houzé P, **Coppée R**, Kusi KA, Bright A, Quakyi I, Coleman N, Mama A, Clain J, Deloron P, Bosompem KM, Tahar R, Ofori MF, Ndam NT.

(En cours de soumission).

Durant cette même année, une étude clinique menée par Nicolas Argy (membre de l'UMR 261 et du laboratoire de parasitologie-mycologie à l'hôpital Bichat-Claude Bernard) sur l'atovaquone, un antipaludique également utilisé dans le traitement prophylactique de la pneumocystose, a mis en évidence une nouvelle mutation (A144V) localisée sur le cytochrome *b* de l'espèce infectieuse, *Pneumocystis jirovicii* et associée à un échec au traitement chez des patients. Une modélisation structurale du complexe *bc1* de la chaîne respiratoire mitochondriale (sous-unité du cytochrome *b*) a été générée par homologie afin de mieux comprendre l'effet de cette mutation sur la structure et son rôle associé à la résistance.

- ***Pneumocystis* cytochrome b mutants associated with atovaquone prophylaxis failure as the cause of *Pneumocystis* infection outbreak among heart transplant recipients.**

Argy N, Le Gal S, **Coppée R**, Song Z, Vindrios W, Massias L, Kao WC, Hunte C, Yazdanpanah Y, Lucet JC, Houzé S, Clain J, Nevez G.

Clinical Infectious Diseases 67:913-919. doi: <https://doi.org/10.1093/cid/ciy154>

Enfin, un dernier travail mené par Gwladys Bertin et Claire Kamaliddin a porté sur la variabilité des protéines *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1), impliquées dans la physiopathologie du paludisme grave. Pour cela, une approche multi-omique a été conduite pour déchiffrer l'expression de PfEMP1 au niveau du patient selon différents contextes cliniques. Des séquençages du génome entier et de l'ARN de *P. falciparum* ont été réalisés. L'analyse d'expression différentielle des données de séquençage ARN a révélé une expression

accrue de transcrits de plusieurs sous-domaines des PfEMP1 au sein d'échantillons de patients atteints d'un paludisme grave, par rapport aux échantillons relatifs à des patients ayant contracté un paludisme simple. L'approche a permis d'attribuer des séquences PfEMP1 à chaque échantillon et d'identifier de nouveaux peptides associés aux protéines PfEMP1 par spectrométrie de masse. Dans ce travail, nous nous sommes principalement axés sur la composition en sous-domaines des PfEMP1s et sur la diversité génétique relative à ces domaines.

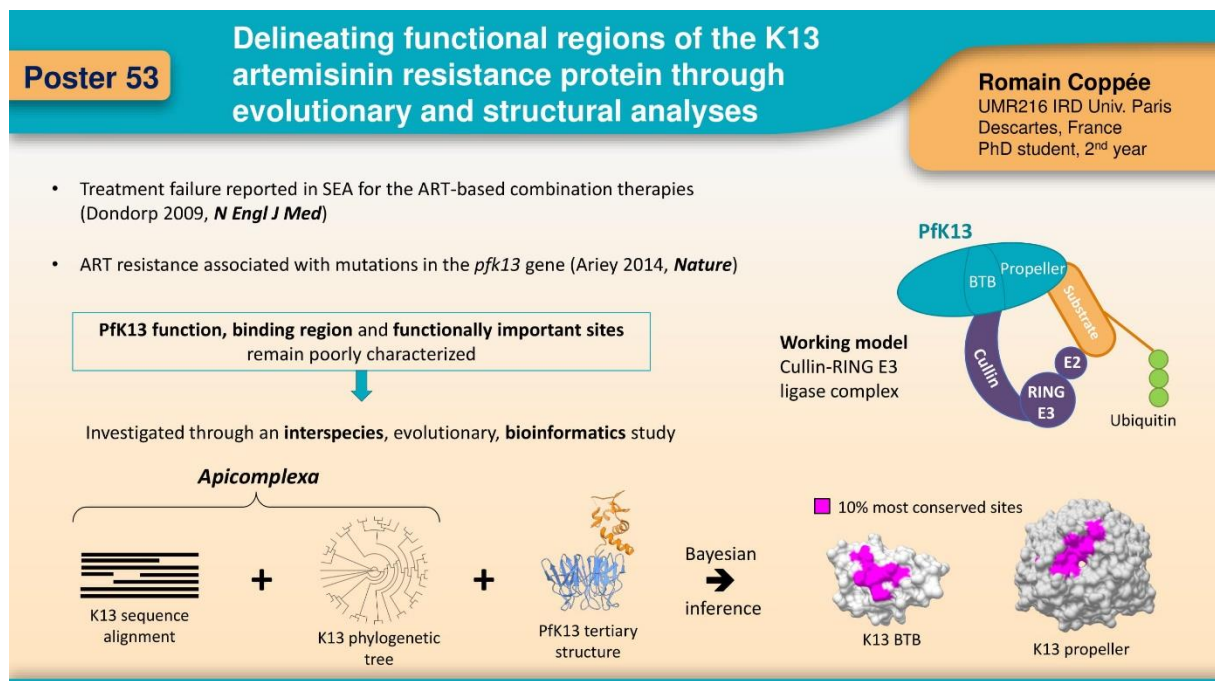
- **From genomic to LC-MS/MS evidence: Analysis of PfEMP1 in Benin malaria cases**

Kamaliddin C, Rombaut D, Guillochon E, Royo J, Ezinmegnon S, Agbota G, Huguet S, Guemouri S, Peirera C, **Coppée R**, Broussard C, Alao JM, Aubouy A, Guillonnet F, Deloron P, Bertin GI.

PLoS One 14:e0218012. doi: <https://doi.org/10.1371/journal.pone.0218012>

Poster présenté en congrès

Le poster (page suivante) a été présenté au congrès BioMalPar XIV en mai 2018 (Heidelberg, Allemagne), sous le titre « Delineating functional regions of the K13 artemisinin resistance protein through evolutionary and structural analyses ». Ce poster a été accompagné d'un flash-talk (ci-dessous).





Delineating functional regions of the K13 artemisinin resistance protein through evolutionary and structural analyses

BioMaIPar XIV – 22-25 May, 2018 – EMBL, Heidelberg, Germany

Romain Coppée¹, Audrey Sabbagh¹, Jérôme Clain¹

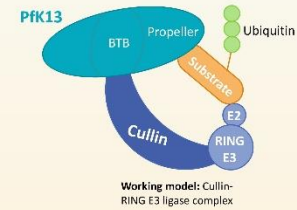
¹. UMR 216 MERIT, IRD, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

BACKGROUND

- Efficacy of first-line artemisinin (ART)-based combination therapies in Southeast Asia¹. ART resistance is associated with mutations in the *Plasmodium falciparum* **k13** (*pfk13*) gene².
- PFK13 immunoprecipitated with phosphatidylinositol-3-kinase (PI3K), which ubiquitination $\searrow$ in *pfk13* C580Y mutants³. PFK13 may exhibit **regulatory functions** consistent with the initial report of PFK13 sharing homology with Kelch-containing proteins involved in ubiquitin-based protein degradation.
- *pfk13* encodes a 726 amino acid protein (PFK13) possessing a **BTB/POZ** domain and a C-terminal **propeller** domain composed of repeated Kelch motifs² $\rightarrow$ likely falls within the **BTB-Kelch structural subgroup of proteins**⁴, but exhibits specific features.



- Generally, BTB/POZ mediates **oligomerization** architectures and contributes to **Cullin recruitment**, while the propeller domain serves as the substrate receptor^{5,6}.
- Little is known about the precise function and putative binding properties of PFK13. Nearly all *pfk13* mutations associated with ART resistance cluster in the propeller domain, suggesting a **critical role** for the propeller domain in PFK13 function, probably through directly binding a substrate protein further targeted for degradation². However, the **PFK13 propeller binding region** and **functionally important sites** remain uncharacterized.

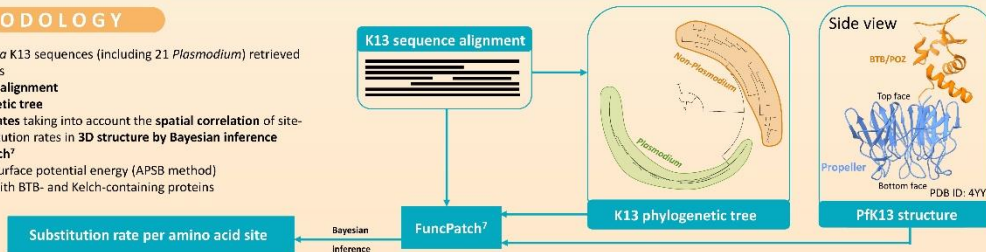


HYPOTHESIS

We hypothesized that functionally important sites of K13 are under **stronger purifying selection** and could be identified by the analysis of **k13 molecular evolution across Apicomplexa species** – a phylum of parasites including *Plasmodium* sp.

METHODOLOGY

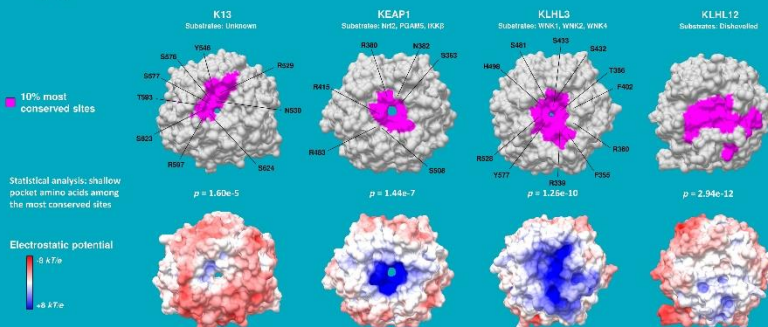
- 43 Apicomplexa K13 sequences (including 21 *Plasmodium*) retrieved from databases
- K13 **sequence alignment**
- K13 **phylogenetic tree**
- **Substitution rates** taking into account the **spatial correlation** of site-specific substitution rates in **3D structure** by **Bayesian inference** using **FuncPatch**⁷
- **Electrostatic surface potential energy** (APSB method)
- **Comparison** with BTB- and Kelch-containing proteins



RESULTS

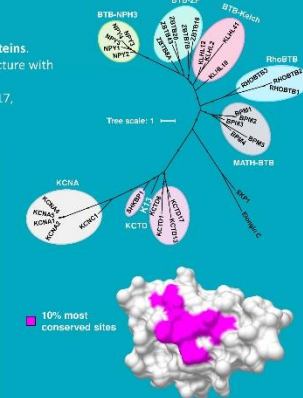
K13 PROPELLER

- Patch of slowly evolving sites:
 - located at the surface and underlying layer of a **shallow pocket**
 - overlaps with a region of the shallow pocket harboring an **electropositive surface potential energy**
 - contains **arginine and serine residues**, strictly conserved in *Apicomplexa* which are known to mediate protein-protein interactions in other propeller domains^{8,9}.



K13 BTB/POZ

- Appears **shortened** compared to the one of BTB-Kelch proteins.
- K13 BTB/POZ shows **high similarities** in sequence and structure with the one of KCTD proteins.
- Patch of slowly evolving sites similar to SHKBP1 and KCTD17, including one site involved in BTB-Cullin interactions.



REFERENCES

1. Noedl H. et al. (2008) *N. Engl. J. Med.* 359, 2619–2620
2. Ariey F. et al (2014) *Nature* 505, 50–55
3. Mbenge A. et al (2015) *Nature* 520, 683–687
4. Dhanao BS. et al. (2013) *Hum Genomics*, 7–13
5. Canning P. et al. (2013) *J. Biol. Chem.* 288, 7803–7814
6. Pinkas DM. et al (2017) *Biochem. J.* 474, 3747–3761
7. Huang YF. et al. (2015) *Bioinformatics* 31, 523–31
8. Shibata et al. (2013) *PNAS* 110, 7838–7843

CONTACT

UMR 216 MERIT "Mère et Enfant Face aux Infections Tropicales"
Université Paris Descartes, Faculté de Pharmacie, 4 Avenue de l'Observatoire,
75006 Paris, France
+33 6 26 46 40 69
romain.coppee@parisdescartes.fr



Acknowledgment: We thank YE. Huang and GB. Golding for advices in FuncPatch use. Special thanks to M. Mitrev and G. Mercereau-Pujillon for helpful discussions. We thank the Ecole Doctorale MTCT for PhD grant funding.



Printed by Université Paris Descartes

Matériel supplémentaire – Article 1

	Apicomplexa-specific N-terminal region													
	1	10	20	30	40	50	60	70	80	90				
Pfal	MEGEKVKTRKANSISNFSMTYDRESGGNGNS	DDKSGSSSEENDSNFSMNLTS	SDKNEKTENN	SFLNNSSYG	NVKDSLLES	IDMSVLD	SNFDG	SKKDFLPS						
Prei	MEGEKAKTRKANSISNFSMTYDRESGGNGNS	DDKSGSSSEENDSNFSMNLTS	SDKNEKTENN	SFLNNSSYG	NVKDSLLES	IDMSVLD	SNFDG	SKKDFLPS						
Pgab	MEGEKAK..ANSISNFSMTYDRESGGNGNS	DDKSGSSSEENDSNFSMNLTS	SDKNEKTENN	SFLNNSSYG	ANMKD	SLLES	IDMSVLD	SNFDG	SKKDFLAS					
Pviv	MEGEKIK..SNSISNFSVTYEREGGANGNS	DDKSVSSSEENESNFSMNLTS	SDKNEKTENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pcyn	MEGEKIK..TNSISNFSVTYEREGGANGNS	DDKSVSSSEENESNFSMNLTS	SDKNEKTENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pkno	MEDEKIK..SNSISNFSVTYEREGGANGNS	DDKSVSSSEENESNFSMNLTS	SDKNEKTENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pfra	MEGEKIK..SNSISNFSVTYEREGGANGNS	DDKSVSSSEENESNFSMNLTS	SDKNEKTENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pcoa	MEGEKIK..SNSISNFSVTYEREGGANGNS	DDKSVSSSEENESNFSMNLTS	SDKNEKTENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pinu	MEGEKIK..SNSISNFSVTYEREGGANGNS	DDKSVSSSEENESNFSMNLTS	SDKNEKTENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pgon	MEGEKIK..SNSVSNFSITTYEREGGANGDS	DDKSLRSSDNESNFSMNLTS	SDKNEKAENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLGS						
Pmal	MEGEKIK..SNSISNFSVTYDRESGVNCS	DDRSSESSSEENESNFSMNLTS	SDKNEKTENN	SFALNNSS	SFVN	MKDSLLES	IDLSVLD	SNFDG	TKKDFLPS					
Pwal	SDEKSMSSSDENESSF	MNLTS	SDKNEKIENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS						
Pcur	MKSNSISNFSVTYEREGGNGNS	DEKSMSSSDENESSF	MNLTS	SDKNEKIENN	SFLNNSSFAN	MKDSLLES	IDLSVLD	SNFDG	SKKDFLPS					
Pvin	MEDDKIK..SNSISNFSVTYEREGGNGNS	EERDISSEENESNFSMNLTS	GDKNKIE	DN	SFVN	IKDSLLES	IDLSVLD	SNFDG	SKNEFLPN					
Pcha	MEDDKIK..SNSISNFSVTYEREGGNGNS	EERDISSEENESNFSMNLTS	GDKNKIE	DN	SFVN	MKDSLLES	IDLSVLD	SNFDG	SKNEFLPN					
Pada	MEDDKIK..SNSISNFSVTYEREGGNGNS	EERDISSEENESNFSMNLTS	GDKNKIE	DN	SFVN	MKDSLLES	IDLSVLD	SNFDG	SKNEFLPN					
Ppet	MEDDKIK..SNSISNFSVTYEREGGNGNS	EERDISSEENESNFSMNLTS	GDKNKIE	DN	SFVN	MKDSLLES	IDLSVLD	SNFDG	SKNEFLPN					
Pyoe	MEDDKIK..SNSISNFSVTYEREGGNGNS	EERDMSSDEENESNFSMNLTS	GDKNKIE	DN	SFVN	MKDSLLES	IDLSVLD	SNFDG	SKNEFLPN					
Pber	MEDDKIK..SNSISNFSVTYEREGGNGNS	EERDMSSDEENESNFSMNLTS	GDKNKIE	DN	SFVN	IKDSLLES	IDLSVLD	SNFDG	SKNEFLPN					
Pgal	MEDCDKVK..SSSISNFSVTYDRESGVNTNS	EKLSSESEENESNFSMNLTS	INKNEKTENN	SFLNNSS	SFNVKDSLLES	IDLSVLD	SNLDT	KKGFPS						
Prcl	MDCDKIK..SNSISNFSVTYDRESVANMNS	EKLSSESEENESNFSMNLTS	SKNEKTENN	SFLDNSSFEN	IKDSLLES	IDLSVLD	SNLDT	KKGFPS						
Tgon	MNSTGNASPS	GLLRGG	ADDALAY	PPGTS	SAGALGSS	SRSPME	AGFCPP	APGPAVG	TGVGGLLYS	ATASPEFLCSST	AGGCP	PA		
Hham	MNSTSHASAS	GLLRGG	ADDALAY	PLGTS	SAGALGSS	SRSPME	AGFCPP	APGPAVG	TGVGGLLYS	ATASPEFLCSST	AGGCP	PA		
Ncan	MNPTGRV	STSGAGSVPR	RGSDADD	IVYVPT	PPAL	PCRS	TL	PALGA	VGGVTAF	PSAAPEFLY	GAAGARG	L	AGG	
Ppau	MLVTGSS	GLGVGRGMA	ALPGAGSG	HAAAGSG	EQPHGSG	MTTL	LRSSSSSG	NEGCGGAR	RTASS	SAGGLLNS	GEP	PPSSS	ALSSMLHQS	QQ
Eace	MDPY	ASSSGG	PLNAGV	ALQ	SAHST	ISSHSG	AP	TGGP	HDRG	GAP	GGPL	SYGL	PS	SGG
Efal	MEPY	SSANGG	PLGLGG	GILHTS	INTQ	ANRNS	GGPL	AYGL	PNYP	SGS	APP	AGLL	PP	
Bmic														
Tpar														
Tann														
Tori														
Tequ														
Bbov														
Bbig														
Ebru														
Emax														
Cmur														
Cand														
Chom														
Cpar														
Cmel														
Cubi														
Cbai														

	Apicomplexa-specific N-terminal region														
	100	110	120	130	140	150	160	170							
Pfal	NLSRTFNNMCKDNIGNKYLKLLNKKKDTITNENNNIN	HNNNNNNL	TANNITNN	LINNNMMN	SGSIMNT	NKKENFLD	AA	NL							
Prei	NLSRTFNNMCKDNIGNKYLKLLNKKKDTITNENNNIN	HNNNN	LTANNITNN	LINNNMMN	SGSIMNT	NKKENFLD	AA	NL							
Pgab	NLSRTFNNMCKDNIGNKYLKLLNKKKDTITNENNNIN	HNNNN	NLTANNITNN	LINNNMMN	SGSIMNT	NKKENFLD	AA	NL							
Pviv	NLSKNFNNLSKENLGNKYLNKLLNKS	DSGIFMCKSKKDMN	L	MENN	LGSNNL	PVKSGN	KKKEGFMD	SS	TP	I					
Pcyn	NLSKNFNNLSKENLGNKYLNKLLNKS	DSGIFMCKSKKDMN	L	MENN	LGSNNL	PVKSGN	KKKEGFMD	SS	TP	I					
Pkno	NLSKNFNNLSKENLGNKYLNKLLNKS	DSGLFMSKKNKDMN	L	MDNN	MGSNNL	PVKSGN	KKKEGFMD	SS	TP	I					
Pfra	NLSKNLNNLSKENLGNKYLNKLLNKS	DSGIFMCKSKKDMN	L	MENN	LGSNNL	PVKSGN	KKKEGFMD	SS	TP	I					
Pcoa	NLSKNFNNLSKENLGNKYLNKLLNKNDAVFM	PKGKDMN	L	MENN	MGSNNL	PVKSGN	KKKEGFMD	SS	TP	I					
Pinu	NLSKNFNNLSKENLGNKYLNKLLNKS	DSGLFTSKGKDMN	L	MENN	LGSNNP	PINTGN	KKKEGFMD	SS	TP	I					
Pgon	NLSKNFNNLSKEHLSSKYLNKLLNKS	DSGIFMSKKNKDMH	L	MENN	VGSNNP	PVKN	NNKKKEGFMD	SS	TP	I					
Pmal	NFSKNFNNLSKENISNKYLKLLNKS	DSGIFMCKSKKDMN	L	TDAS	NNVNI	SVKN	NNTKKEIFMD	DA	TA	L					
Pwal	NFSKNFNNLSKENISNKYLKLLNKS	DSGLFMSKKNKDLN	L	LTDV	GNLNNNSGNSH	GNGANV	PLRNN	NRKDSFMD	SG	QNPLN					
Pcur	NFSKNFNNLSKENISNKYLKLLNKS	DSGLFMSKKNKDN	L	LTDV	GNLNNNSGNSH	GNGANV	PLRNN	NRKDSFMD	SG	QNPLN					
Pvin	NFSKNLNNLT	KDNIGNKYLKLLNKNDAVFM	PKGKDNNSIDL	NALNVGNNN	IGNNNIV	VSNDGG	NNKMHV	KIGNNNING	SGT	GAP	NKKEIFMD	SS	GA	SI	
Pcha	NFSKNLNNLT	KDNIGNKYLKLLNKNDAVFM	PKGKDNNSIDL	NALNVGNNN	IGNNNIV	VSNDGG	NNKMHV	KIGNNNING	SGT	GAP	NKKEIFMD	SS	GA	SI	
Pada	NFSKNLNNLT	KDNIGNKYLKLLNKNDAVFM	PKGKDNNSIDL	NALNVGNNN	IGNNNIV	VSNDGG	NNKMHV	KIGNNNING	SGT	GAP	NKKEIFMD	SS	GA	SI	
Ppet	NFSKNLNNLT	KDNIGNKYLKLLNKNDAVFM	PKGKDNNSIDL	NALNVGNNN	IGNNNIV	VSNDGG	NNKMHV	KIGNNNING	SGT	GAP	NKKEIFMD	SS	GA	SI	
Pyoe	NFSKNLNNLT	KDNIGNKYLKLLNKNDAVFM	PKGKDNNSIDL	NALNVGNNN	IGNNNIV	VSNDGG	NNKMHV	KIGNNNING	SGT	GAP	NKKEIFMD	SS	GA	SI	
Pber	NFSKNLNNLT	KDNIGNKYLKLLNKNDAVFM	PKGKDNNSIDL	NALNVGNNN	IGNNNIV	VSNDGG	NNKMHV	KIGNNNING	SGT	GAP	NKKEIFMD	SS	GA	SI	
Pgal	NLSKNLNNLSKENISNKYLKLLNKS	DSGLFMSKKNKDN	L	LTDNN	KILNN	NNKKEGFMD	SS	TP	SNH	I					
Prcl	NLSKNFNNLSKENISNKYLKLLNKS	DSGLFMSKKNKDN	L	LTDNN	KILNN	NNKKEGFMD	SS	TP	SNH	I					
Tgon	ACAGFLSS	RVLANG	GLSRLSS	RPFD	ASCAT	LPRL	SG	APT	GDG	AGALH	AL	HAL	HAP	LA	EKR
Hham	ACAGFLSS	RVLANG	GLSRLSS	RPFD	ASCAT	LPRL	SG	APT	GDG	AGALH	AL	HAL	HAP	LA	EKR
Ncan	TGVSYLS	GGGERGFANG	RVTRPLP	SETSVLA	ASHRP	LAS	DPE	LA	SH	K	FL	TP	LA	A	P
Sneu	RHFLAV	AGGATAGVNG	SSSGM	VLLG	ARGGA	ATE	QQ	LVA	AS	FAM	LR	DA	DEA	AV	AAAA
Eace	YGGSA	ADT	SLHAS	MLRK	ETQ	SRT	TD	LP	SV	IP	AG	TT	SG	TY	T
Efal	SGISS	ADS	TVHAS	VGR	KRP	ACT	TD	AL	GK	AP	DL	SG	L	GS	GP
Bmic															
Tpar															
Tann															
Tori															
Tequ															
Bbov															
Bbig															
Ebru															
Emax															
Cmur															
Cand															
Chom															
Cpar															
Cmel															
Cubi															
Cbai															

Apicomplexa-specific N-terminal region										Coiled-coil containing domain									
	180	190	200	210	220	230	240	250	260	270									
Pfal	.INDDSG.LNNLKKFSTVNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Prai	.INDDSG.LNNLKKFSTVNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pgab	.INDDSG.LNNLKKFSTVNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pviv	.NANEDNA.MNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pcyn	.NANEDNA.MNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pkno	.NANEDNA.MNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pfra	.NANEDNA.MNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pcoa	.NANEDNA.MNNLKKFSTVNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pinu	.NANEDNA.MNNLKKFSTVNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pgon	.NANEDNA.MNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pmal	.NANEDNA.MNNLKKFSTVNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pwal	.TNEENNA.LNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pcun	.VNEENNA.LNNLKKYGNANN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pvin	.NMDDNSTMHNARKYKSTNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pcha	.NMDDNSTMHNARKYKSTNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pada	.NMDDNSTMHNARKYKSTNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Ppet	.NMDDNSTMHNARKYKSTNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Ppye	.NMDDNSTMHNARKYKSTNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pber	.NMDDNSTMHNARKYKSTNN.VNDTYEKKIIETELSDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Pgal	.NINEDST.LNTLKKYTN.NNNLNDIYKKIIENELNDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Prel	.NEDST.LNTLKKYTN.NNNLNDIYKKIIENELNDAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Tgon	.EFGPAEFLDQGGGEERMGRAASDDTDEAPAFASAE	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Iham	.EFGPAEFLDQGGGEERMGRAASDDTDEAPAFASAE	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Ncan	.DANPQ..FNGDGSVDESPL...ANGD.EDASAGAFASAE	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Sneu	.AVRRRRQGGGGGGGGKEEEDDEEDREDESGSGFSTAG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Eace	GAPPLSVVTPETLEAPGGGCGSLG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Efal	.CPTAGLALGLPGGSLGVPLTVVTPETLEKGGGCGSLG	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Bmic	.YEFHPTSGNIDKSID..SFDINVLRCANEEIFVPFGST	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Tpar	.TNEFKGLTSLDPTDPLSLADIVPLKIS..NGPDQDEVGSS	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Tann	.TNEFKGLTSLDPTDPLSLADIVPLKIS..NGPDQDEVGSS	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Tori	.TNEFKGLTSLDPTDPLSLADIVPLKIS..NGPDQDEVGSS	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Tequ	.KGR...LDIANRADLPISDGLVPVKLSTSTTPTETINV	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Bbiv	.AAIHGVSRRDQDSTYSLSDVAQLRLLS.ERDELQNEAD	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Bbig	.ASLTAGASARDVGSIGLSGLDAPLRPLPP.ERVDLTGST	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Ebru	DLRRTFVAGLWHTAEALKRERNDLVRRKKE	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Emax	MVDFVSWLWHTAEALKRERNDLVRRKKE	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Cmur	MNVIKDESITNGVTPKTAIAKSVMS	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Cand	MNVIKDESITNGVTPKTAIAKSVMS	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Chom	MDVTRDITTESGLVNDLLISGVLLDDDDGLDTLIK	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Cpar	MDVTRDITTESGLVNDLLISGVLLDDDDGLDTLIK	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Cmel	DVTDRITTESGLVNDLLISGVLLDDDDGLDTLIK	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Cubi	MDNAIKDNTIESTVCRNGSFINGVLLDDDDGLDTLIK	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															
Cbai	DILLSD...PACVLDSDISNLSLIC	DFENMVLDLR	TFINWLKKTQNNFIREKDLFKDKKKELEMERVRLYKEL	ENRKNIEEOK															

Coiled-coil containing domain																BTB domain			
	280	290	300	310	320	330	340	350	360	370									
Pfal	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Prai	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pgab	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pviv	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pcyn	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pkno	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pfra	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pcoa	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pinu	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pgon	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pmal	LHDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pwal	LYDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pcur	LYDERKKLDIDISNGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pvin	LQDERKKLDIDISNGYKQIKKEEERKRRFDDERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pcha	LQDERKKLDIDISNGYKQIKKEEERKRRFDDERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pada	LQDERKKLDIDISNGYKQIKKEEERKRRFDDERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Ppet	LQDERKKLDIDISNGYKQIKKEEERKRRFDDERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Ppye	LQDERKKLDIDISNGYKQIKKEEERKRRFDDERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pber	LQDERKKLDIDISNGYKQIKKEEERKRRFDDERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGAIFESRHTLTQ																
Pgal	LLEERKKLDIDIANGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGVVPEASRHTLTQ																
Prel	LLEERKKLDIDIANGYKQIKKEEERKRRFDEERLRFLOE	DKIKLVLYLEKEKYFYQYKNNFENDKKIVDANIA	ETMIDINVGGVVPEASRHTLTQ																
Tgon	IEEERKKAAQAEATQIKQI	QVEDARKKLNARGFEEKEAHHRRRTLLEKEFRFQAEALDAEKKRIVDANIA	ETMIDINVGGVVPEASRHTLTQ																
Iham	IEEERKKAAQAEATQIKQI	QVEDARKKLNARGFEEKEAHHRRRTLLEKEFRFQAEALDAEKKRIVDANIA	ETMIDINVGGVVPEASRHTLTQ																
Ncan	IEEERKKAAQAEATQIKQI	QVEDARKKLNARGFEEKEAHHRRRTLLEKEFRFQAEALDAEKKRIVDNT	ETMIDINVGGVVPEASRHTLTQ																
Sneu	IEEERKKAAQAEASQIKQI	QVEDARKKLNARMFEEKEAHHRRRTLLEGEFRFQAEADGLEAKKRIVDNT	ETMIDINVGGVVPEASRHTLTQ																
Eace	IEEERAAAHAELEQQQLKQIE	EDNARGKLQKQKQFLQEDTARRQVLEERFRFQAEVLAQQRQVDA	VAEETVVDINVGGVVPEASRHTLTQ																
Efal	IEEERVAQAELQEQMQQLKQVQIE	EDNARAKLQVQRQKQEEENQRRQAVLDREFRFQAEVLAQERQVLDGSS	ETMIDINVGGVVPEASRHTLTQ																
Bmic	LAEEERKKQNAETASQLQKQIE	EDDARRVFEERAKMEKEEELHKKRRMMAEAKLLQKQIKLFEAEKKKIIDNA	IAGETMIDINVGGVVPEASRHTLTQ																
Tpar	LAEDRRRQNAETASQLKQIE	EDDARRRLEDWCKLEHREDFVRKYMOMAEAKLLQKQIKLLEAEKKRVVDC	IAGETMIDINVGGVVPEASRHTLTQ																
Tann	LAEDRRRQNAETASQLKQIE	EDDARRRLEDWCKLEHREDFVRKYMOMAEAKLLQKQIKLLEAEKKRVVDC	IAGETMIDINVGGVVPEASRHTLTQ																
Tori	LAEDRRRQNAETASQLKQIE	EDDARRRLEDWCKLEHREDFVRKYMOMAEAKLLQKQIKLLEAEKKRVVDC	IAGETMIDINVGGVVPEASRHTLTQ																
Tequ	LAEDRRRQNAETASQLKQVQIE	EDDARRRIEEDLCKLEHREDFVRKYMOMLETKFRQVLEAEKKRVVDTK	IAGETMIDINVGGVVPEASRHTLTQ																
Bbiv	LAEDRRRRQAEVASQLKQVQIE	EDDARRRIEERQRKLDHREDFVFNMYOMLESAKLRQVLEAEQDRQVLDGK	ETMIDINVGGVVPEASRHTLTQ																
Bbig	LSDEERRRRQAEVASQLKHWQIE	EDDARRRIEERQRKLDHREDFVFNMYOMLESAKLRQVLEAEQDRQVLDGK	ETMIDINVGGVVPEASRHTLTQ																
Ebru	IEEERAAAHAELEQQQLKQIE	EDNARGKLQKQKLFQEDENIRRRQVMDRESFQAEVLAQERQVDA	VAEETVVDINVGGVVPEASRHTLTQ																
Emax	IEEERAAAHAELEQQQLKQIE	EDNARGKLQKQKLFQEDENIRRRQVMDRESFQAEVLAQERQVDA	VAEETVVDINVGGVVPEASRHTLTQ																
Cmur	IEEERAAAHAELEQQQLKQIE	EDNARGKLQKQKLFQEDENIRRRQVMDRESFQAEVLAQERQVDA	VAEETVVDINVGGVVPEASRHTLTQ																
Cand	IQMLEKKAEADILQSFIAKLTQIE	QEQMQLNEKEEDNFQEQMANYMMEYDLDQKAKILEDKTLVDTIA	EQGVQVNVGGIPEASRHTLTQ																
Chom	MCKLNTLTTEETETIAKLLIKIE	EDNNAKKLLNKEEYFKKQOELANHWOMIENKLNRLELEIKDKKITEINF	GAETVEINVGGTINPEASRILTE																
Cpar	MCKLNTLTTEETETIAKLLIKIE	EDNNAKKLLNKEEYFKKQOELANHWOMIENKLNRLELEIKDKKITEINF	GAETVEINVGGTINPEASRILTE																
Cmel	VHKLNTLTTEETETIAKLLIKIE	EDNNAKKLLNKEEYFKKQOELANHWOMIENKLNRLELEIKDKKITEINF	GAETVEINVGGTINPEASRILTE																
Cubi	IQKLLNTLTTEETETIAKLLIKIE	EDNNAKKLLNKEEYFKKQOELNTNYMOMIENKLNRLELEIKDKKITEINF	GAETVEINVGGTINPEASRILTE																
Cbai	IKRLNLTLENEVIGFGRKLIKIE	QEQMNYMLLEKEEELFNRRQALANYMOMIERIKDEKLEIKDKKITEINF	GSKGTVEINVGGTINPEASRILTE																

[illegible]

Sequence logo for the Propeller domain. The x-axis represents residue positions from 470 to 540. The y-axis lists 30 different protein sequences. Each sequence is represented by a row of letters (A, C, G, T) indicating the amino acid at each position. A red box highlights the region from position 470 to 500, and another red box highlights the region from position 530 to 540. The logo shows that the sequence is highly conserved in the 470-500 region and less conserved in the 530-540 region.

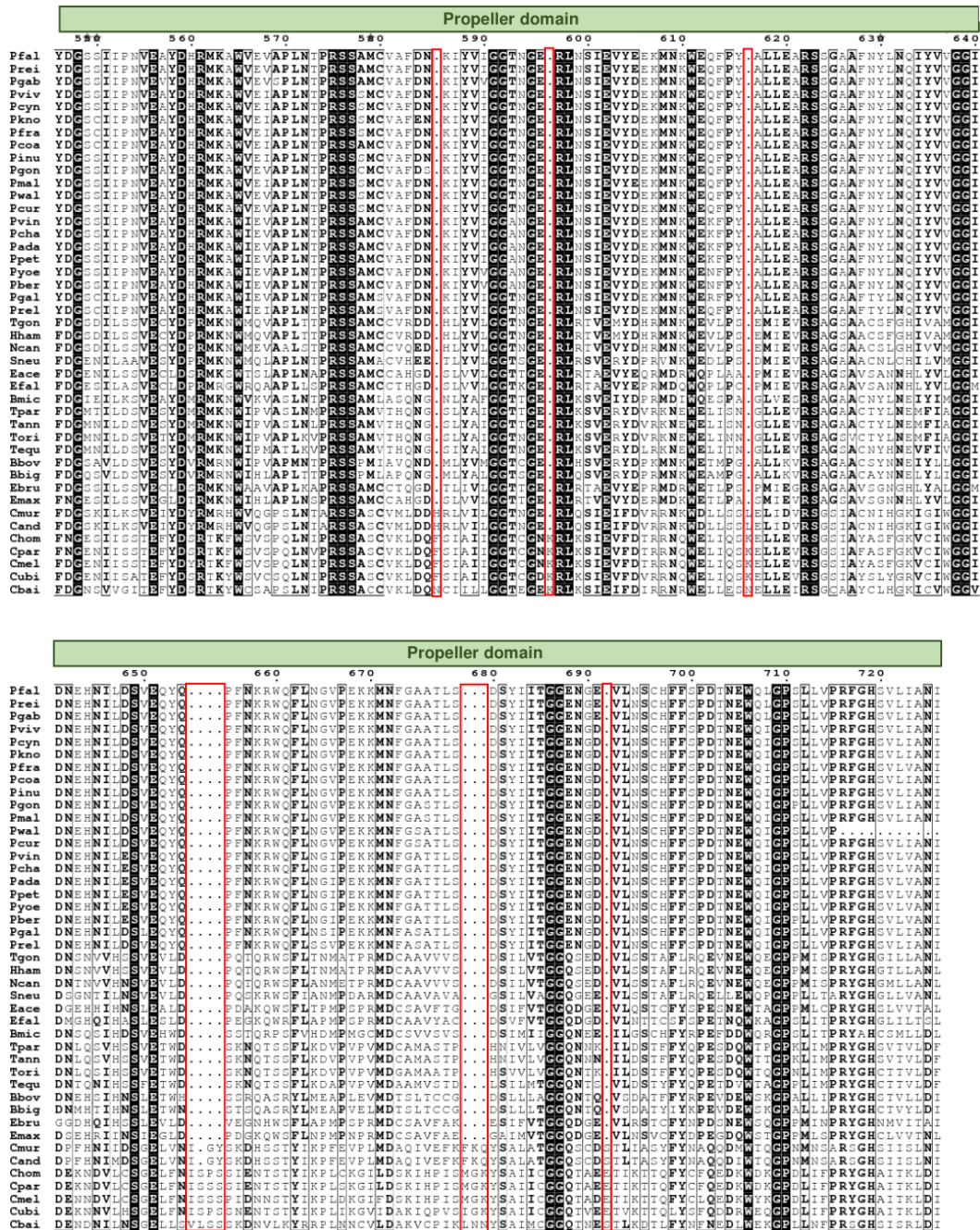


Figure S1. K13 multiple amino acid sequence alignment. The K13 multiple sequence alignment was performed using Mafft version 7. Strictly conserved positions are written in white letters and shaded in black, and highly conserved positions are boxed with black lines and written in white, bold letters (the level of conservation is based on BLOSUM62 scoring matrix). A four-letter code is used for the name of species (full species names are provided in Supplementary Table S1). The regions of the alignment with highly divergent positions, as well as positions containing gaps in at least 30% of all sequences, were manually removed (red boxes). The multiple alignment was displayed with the ESPrnt 3.0 server. The annotated domains of K13 (Coiled-coil containing, BTB and propeller (or KREP)) are shown above the K13 multiple sequence alignment.

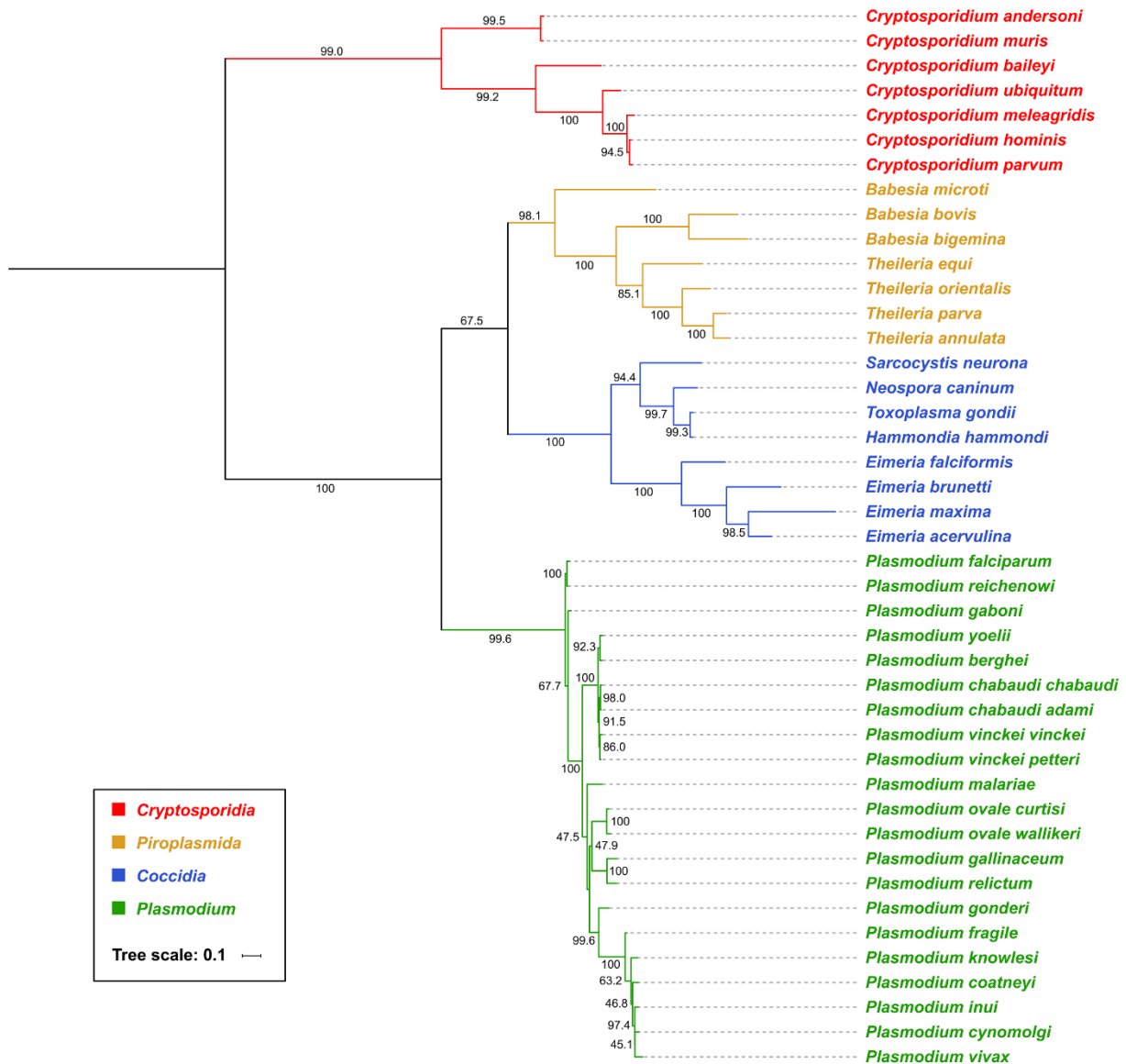


Figure S2. Maximum-likelihood phylogenetic tree of *k13* orthologous protein-coding sequences. All bootstrap values are shown as percentages at nodes using the approximate likelihood-ratio aLRT SH-like method. Colors are associated with the different *genera* or *phyla*: *Plasmodium*, green; *Cryptosporidia*, red; *Coccidia*, blue; and *Piroplasmida*, orange. *Cryptosporidia* was used as outgroup to root the tree. *Hematozoa* should include both *Plasmodium* and *Piroplasmida* species.

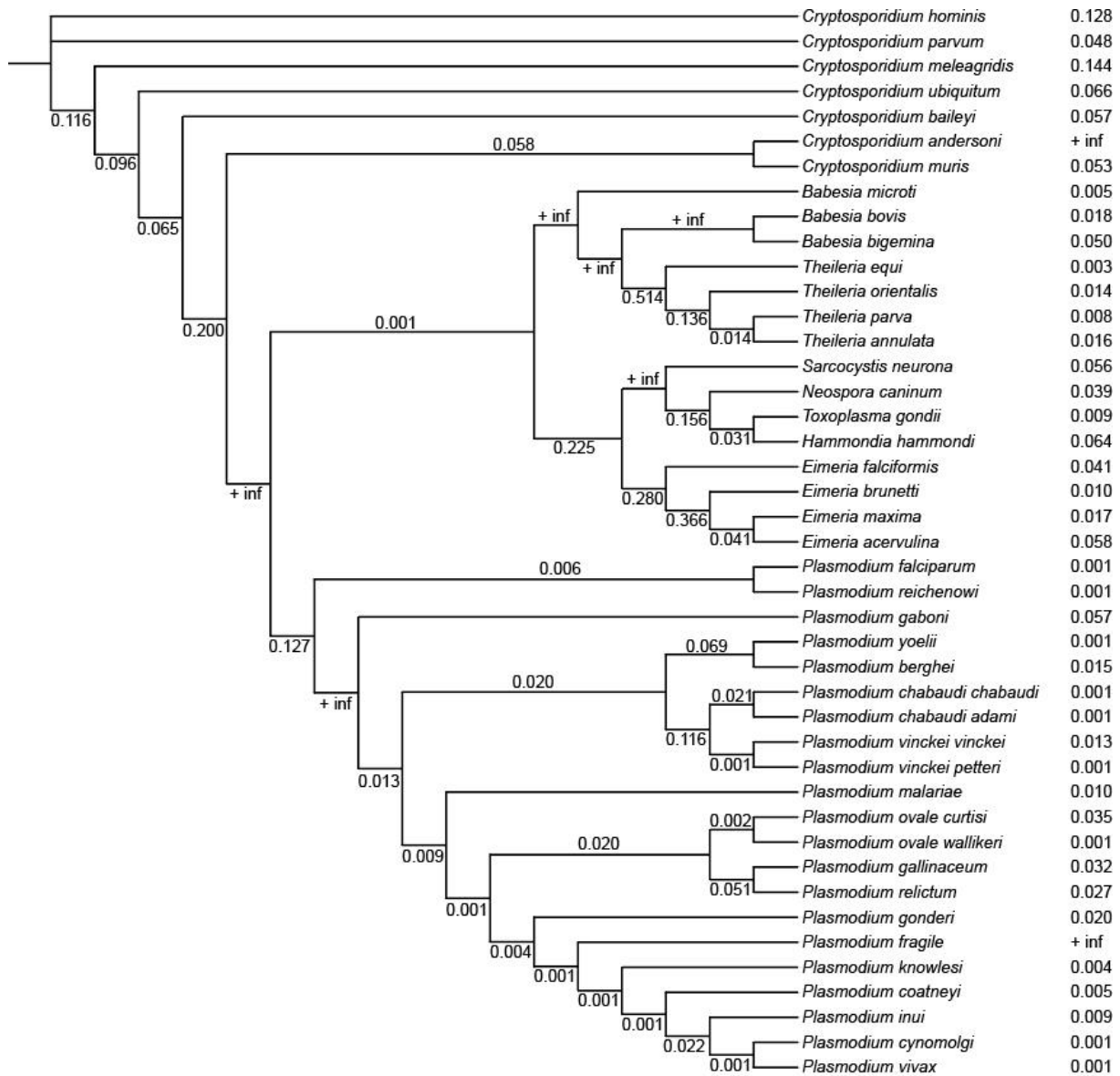


Figure S3. Branch-specific ω values estimated with PAML under the free-ratio (FR) model for the *k13* gene. ω estimates with “+ inf” values for which the estimation of d_s was equal to 0.0 (*i.e.* no silent substitutions were observed) are designated as extremely low. They are therefore not indicative of a positive selection episode. Branch lengths were ignored for ease of representation.

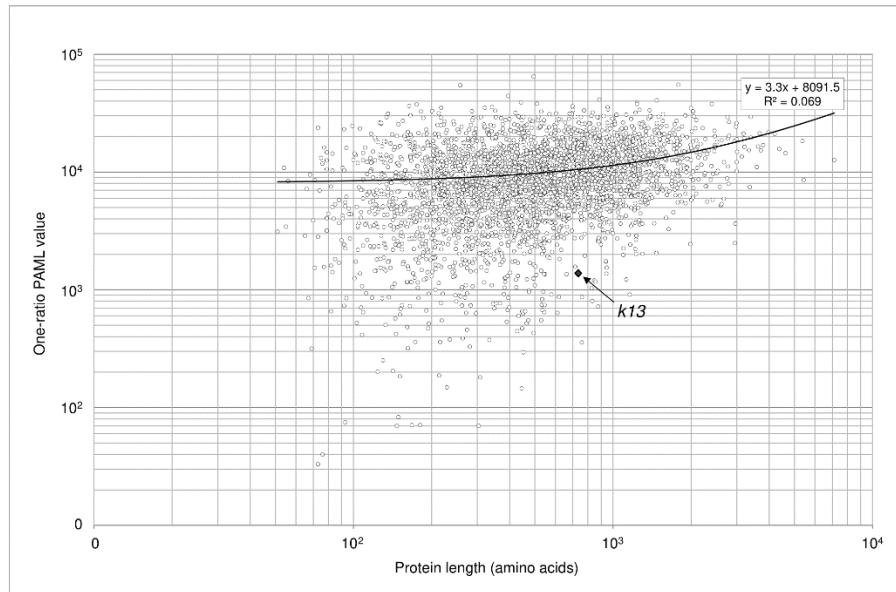


Figure S4. Positive correlation between protein length and one-ratio PAML value for the *Plasmodium* protein-coding genes. The one-ratio PAML values were multiplied by 10^5 and then plotted using a logarithmic scale. K13 is shown as a full-black diamond.

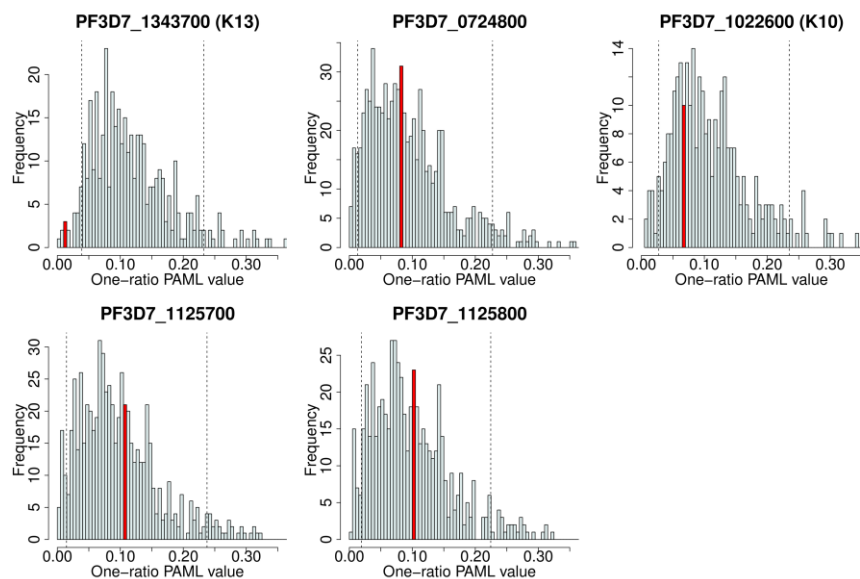


Figure S5. Conservation level of *k13* and four other KREP-containing protein-coding sequences compared to all *Plasmodium* genes. The conservation level (d_N/d_S) was estimated for 3,256 orthologous genes among six *Plasmodium* species (*P. falciparum*, *P. berghei*, *P. chabaudi*, *P. vivax*, *P. yoelii* and *P. knowlesi*) under the one-ratio PAML model. Histograms show the distribution of one-ratio PAML values for protein-coding genes whose length is comprised in an interval of ± 100 amino acids centered on the length of each KREP-containing protein (sized-rank) which corrects for a correlation between protein length and one-ratio PAML value. PlasmoDB accession numbers of the five KREP-containing proteins investigated are provided above each histogram. Red bars indicate the position of each of these KREP-containing proteins in the d_N/d_S distribution, with vertical dashed lines showing the five percent cutoff of the most and less conserved protein-coding genes.

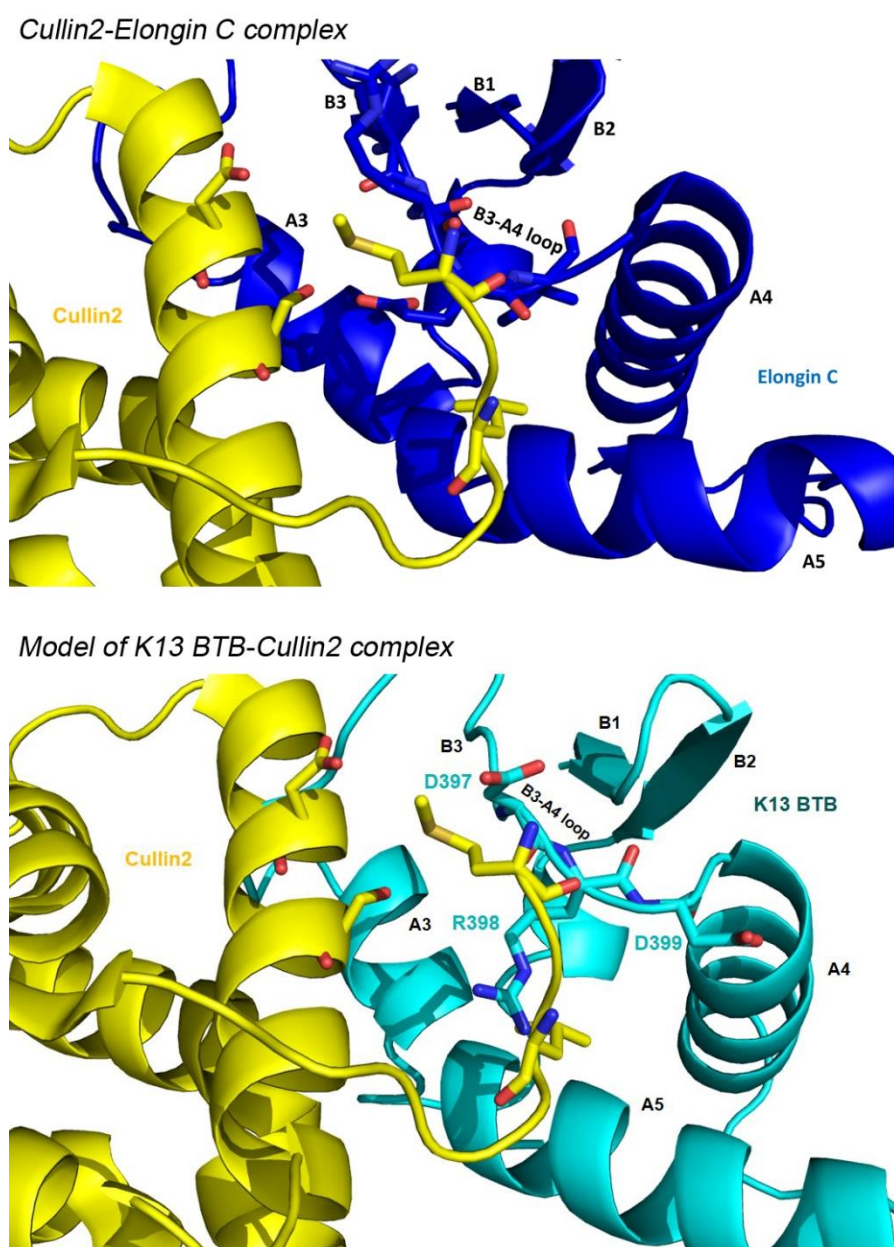


Figure S6. Model of the K13 BTB-Cullin2 interaction using the X-ray structure of the Elongin C-Cullin2 complex as reference. K13 BTB, Elongin C and Cullin2 are shown in cyan, blue and yellow, respectively. The B3-A4 loop in Elongin C contains amino acids that are reported to make contacts with Cullin2 (*upper structure*). The amino acid sites forming the B3-A4 loop of K13 BTB (positions 397-399 using the PfK13 sequence numbering) are strictly conserved during *Apicomplexa* evolution. The structural alignment of K13 BTB with Elongin C showed that amino acid sites 397 and 398 are oriented towards the Cullin protein with contacts distance ≤ 3.0 Å (*lower structure*). The structures were visualized with PyMOL (Schrödinger, LLC).

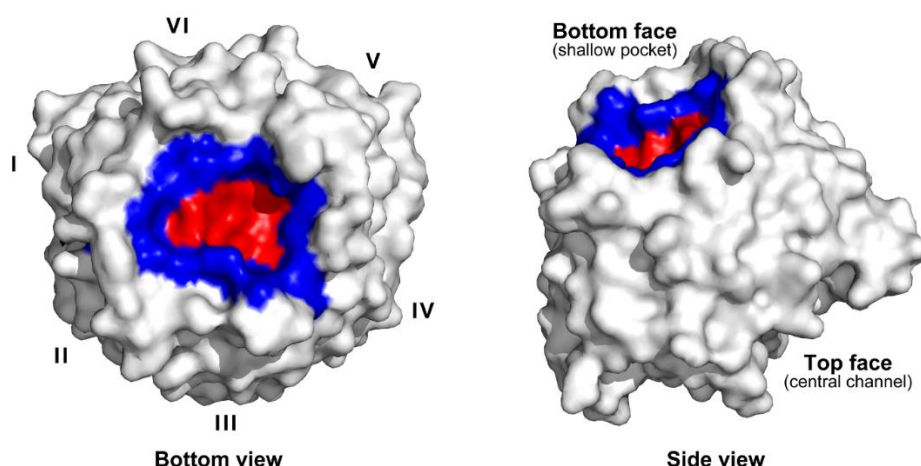


Figure S7. Structural delineation of the shallow pocket in the K13 KREP domain. No conventional definition exists for the shallow pocket delineation in KREP folds. Consequently, we defined it as the amino acids forming the surface plan of the pocket plus those protruding out of the plan. We used the Pfk13 BTB-KREP structure solved by X-ray diffraction at a resolution of 1.81 Å (PDB code 4YY8, chain A). The surface plan of the K13 shallow pocket included the positions 482, 530, 576, 577, 623, 624 and 672 ($n = 7$), shown in red. The amino acid positions protruding out of the plan included the positions 451, 456, 498, 529, 546, 551, 593, 597, 640, 671, 688 and 717 ($n = 12$), shown in blue. The structure is shown as surface and visualized with PyMOL (Schrödinger, LLC).

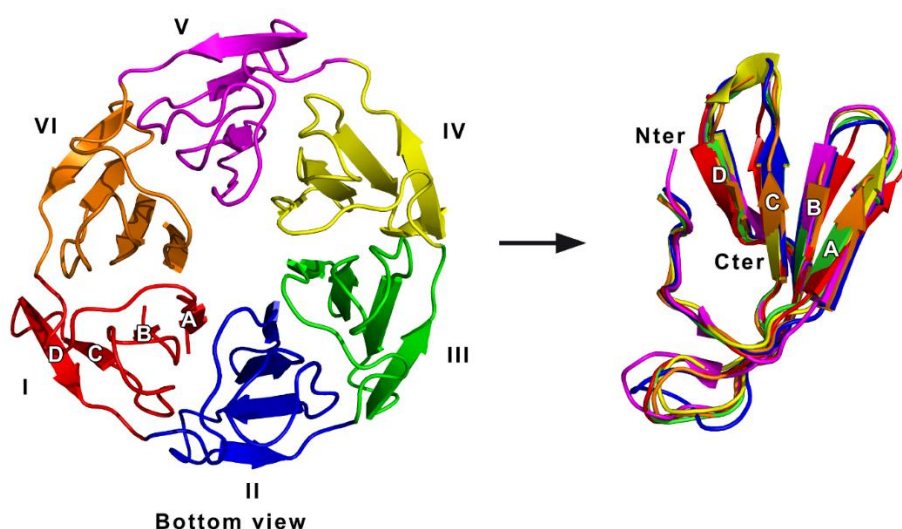


Figure S8. Structure-based alignment of the six blade repeats of Pfk13 KREP. We used the Pfk13 BTB-KREP structure solved by X-ray diffraction at a resolution of 1.81 Å (PDB code 4YY8, chain A). The Pfk13 KREP structure is shown as cartoon. The six blades are colored differently: red, blade I; blue, blade II; green, blade III; yellow, blade IV; magenta, blade V; orange, blade VI. The four strands within each blade are labelled A to D, from the innermost to the outermost. The KREP structure was divided into six PDB files, each one containing the atomic coordinates of one blade. Then, the six blade structures were aligned by minimizing the root-mean-square deviation (RMSD) of atomic positions using the *align* function in PyMOL (Schrödinger, LLC) (*right*). This structure-based alignment of the K13 KREP blades was then used to produce the amino acid sequence alignment of the six blades shown in Figure 4a. The blade length ranged from 46 (blade VI) to 49 (blade V) amino acids, the pairwise blade sequence conservation level ranged from 15.2% (blades III/VI) to 33.3% (blades IV/V), and the pairwise blade structure comparison indicated a RMSD ranging from 0.394 Å² (blades I/III) to 0.669 Å² (blades I/IV).

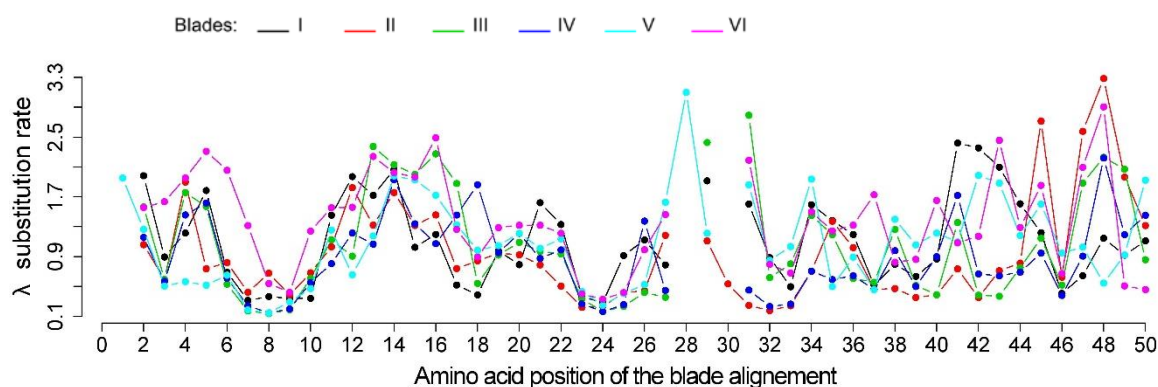
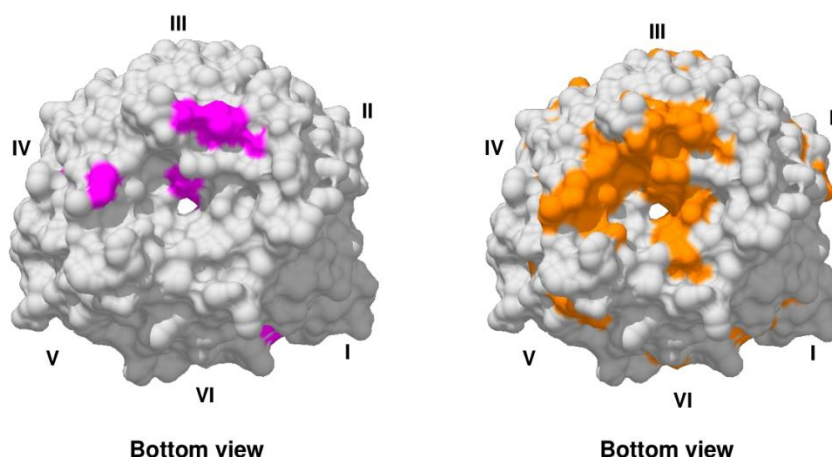


Figure S9. Conservation level of the amino acid sites for each blade along the structure-based amino acid alignment of the six blades. Conservation level of amino acids (λ substitution rate) was estimated with the FuncPatch server. Each blade is associated with a specific color. Because the structure-based amino acid alignment of the six blades contained some gaps, there are missing values.

Most conserved propeller sites (based on ω estimates)

10% cutoff
25% cutoff



Protein	Conservation level	Shallow pocket positions	Remaining positions	<i>p</i> value
K13	High ^{10%}	1	27	0.704
	Others	18	238	
	High ^{25%}	10	61	6.9×10^{-3}
	Others	9	204	

Figure S10. Conservation level of the amino acid sites forming the shallow pocket of K13 KREP, estimated using PAML data. Conservation level of amino acid sites (ω substitution rate) was estimated with the PAML program which does not take into account the protein tertiary structure. The 10% (left structure) and 25% (right structure) most conserved sites are mapped on the three-dimensional structure of PfK13 KREP in magenta and orange, respectively. Only one position of the shallow pocket belonged to the 10% of the most conserved positions (position 530 using the PfK13 sequence numbering). This result is consistent with initial reports by FuncPatch' authors. Without information provided by the

tertiary structure, the most conserved positions of K13 KREP inferred by PAML corresponded to the signature amino acids of KREP proteins: the conserved arginine/lysine (R/K) in DA loops, the diglycine doublet (GG) after B strands, and the tryptophan (W) in C strands. However, when we extended the cutoff to the 25% most conserved sites, the shallow pocket was statistically enriched in conserved positions (10/19 positions: 482, 529, 546, 551, 576, 593, 597, 671, 672 and 688). Contingency tables were subjected to chi-squared or Fisher's exact tests.

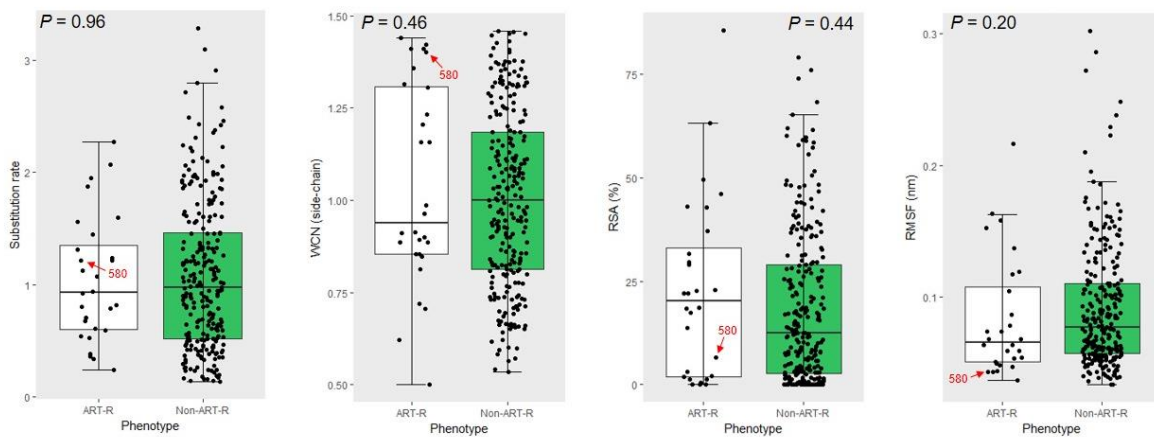


Figure S11. Absence of significant difference in evolutionary- and structure-based parameters between K13 KREP positions associated or not with an ART-R mutation. Box plots showing the distribution of, from *left to right* respectively: the site-specific substitution rates λ inferred from FuncPatch analysis, the relative solvent accessibility (RSA) values, the side-chain weighted contact number (WCN_{sc}) values and the root-mean-square fluctuation (RMSF) values. λ , RSA, WCN_{sc} , and RMSF values were calculated using the PfK13 BTB-KREP structure solved by X-ray diffraction at a resolution of 1.81 Å (PDB code 4YY8, chain A). K13 KREP positions were categorized as either associated with a reported ART-R mutation (*ART-R group*, white) or not (*non-ART-R group*, green). All values are provided in the Supplementary Dataset S2. The position 580 – associated with the predominant *pfk13* C580Y allele in SEA – is indicated by the red arrow. Box boundaries represent the first and third quartiles and the length of whiskers correspond to 1.5 times the interquartile range. The difference between groups was evaluated by non-parametric Mann-Whitney U test.

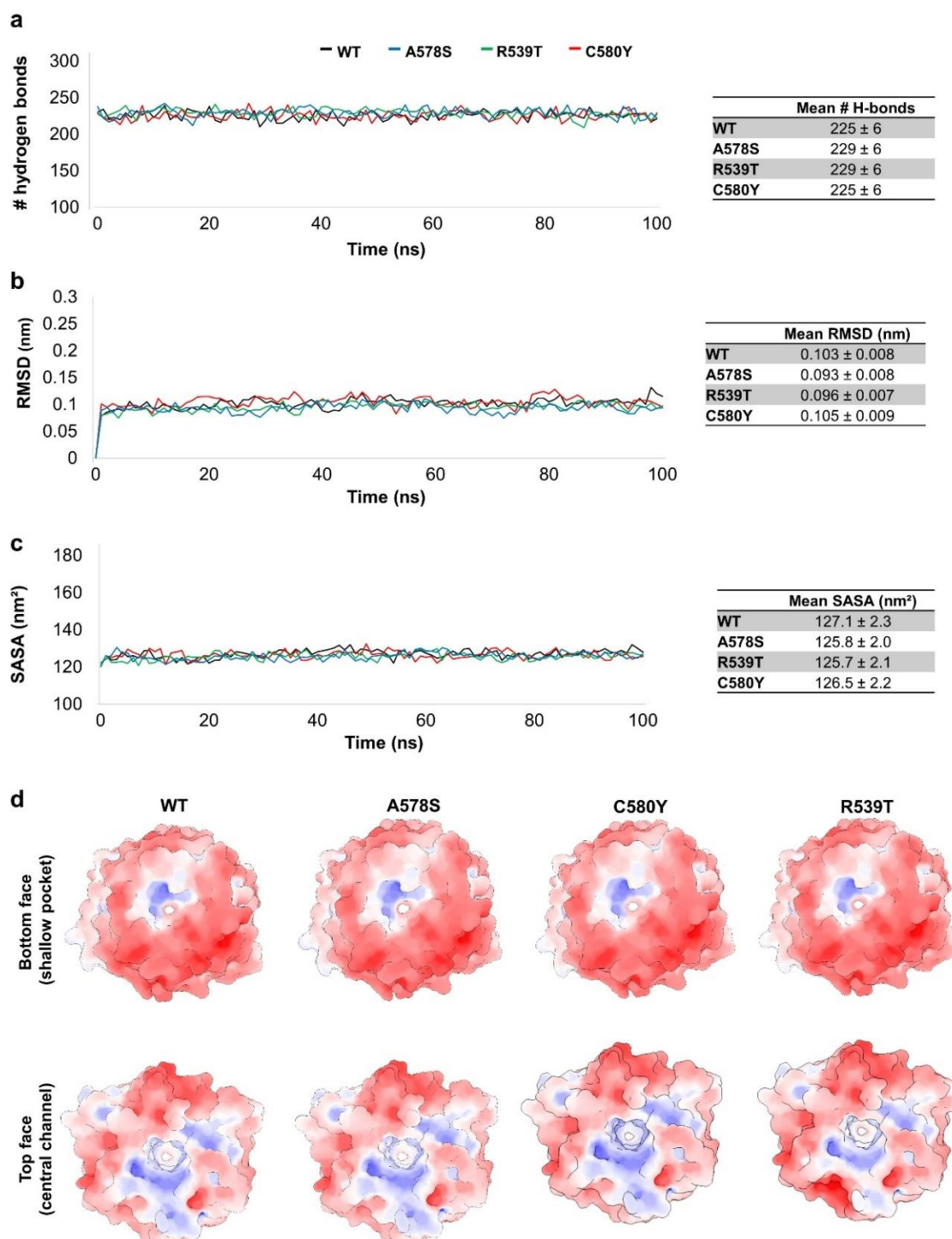


Figure S12. Global impacts of the A578S, C580Y and R539T mutations on the PfK13 KREP structures, as assessed by molecular dynamics simulations. Molecular dynamics simulations were carried out on the KREP structure using GROMACS during 100 ns at a temperature of 300 K in an all-atom system. The first five nanoseconds (ns) correspond to the equilibration phase. Results were saved at each ns considering all-atom KREP structures. C580Y and R539T are two common mutations conferring ART-R in Southeast Asian parasites. A578S mutation is commonly found in African parasites but does not confer ART-R. **(a)** Root-mean-square deviation (RMSD) values of wild type and mutant PfK13 KREP structures. Wild type and mutant structures were very stable and followed similar trajectories, with a mean RMSD close to 0.1 nm. **(b)** Evolution of hydrogen bond number during trajectories. A hydrogen bond was counted when the donor-acceptor distance is ≤ 3.5 Å and donor-hydrogen-acceptor

angle $\leq 30^\circ$. Wild type and mutant structures showed similar average number of hydrogen bonds. **(c)** Solvent accessible surface area (SASA) values of wild type and mutant Pfk13 KREP structures. Wild type and mutant structures showed similar SASA, suggesting no drastic changes in the KREP fold. **(d)** Electrostatic surface potential of wild type and mutant Pfk13 KREP structures, estimated with the APBS method. Electrostatic potential values are in units of kT/e at 298 K, on a scale of $-8 kT/e$ (red) to $+8 kT/e$ (blue). White color indicates a neutral potential. The electrostatic potential is shown for the bottom face (shallow pocket) and for the top face (central channel). Wild type and mutant structures showed similar electrostatic surface potentials at the end of the simulations.

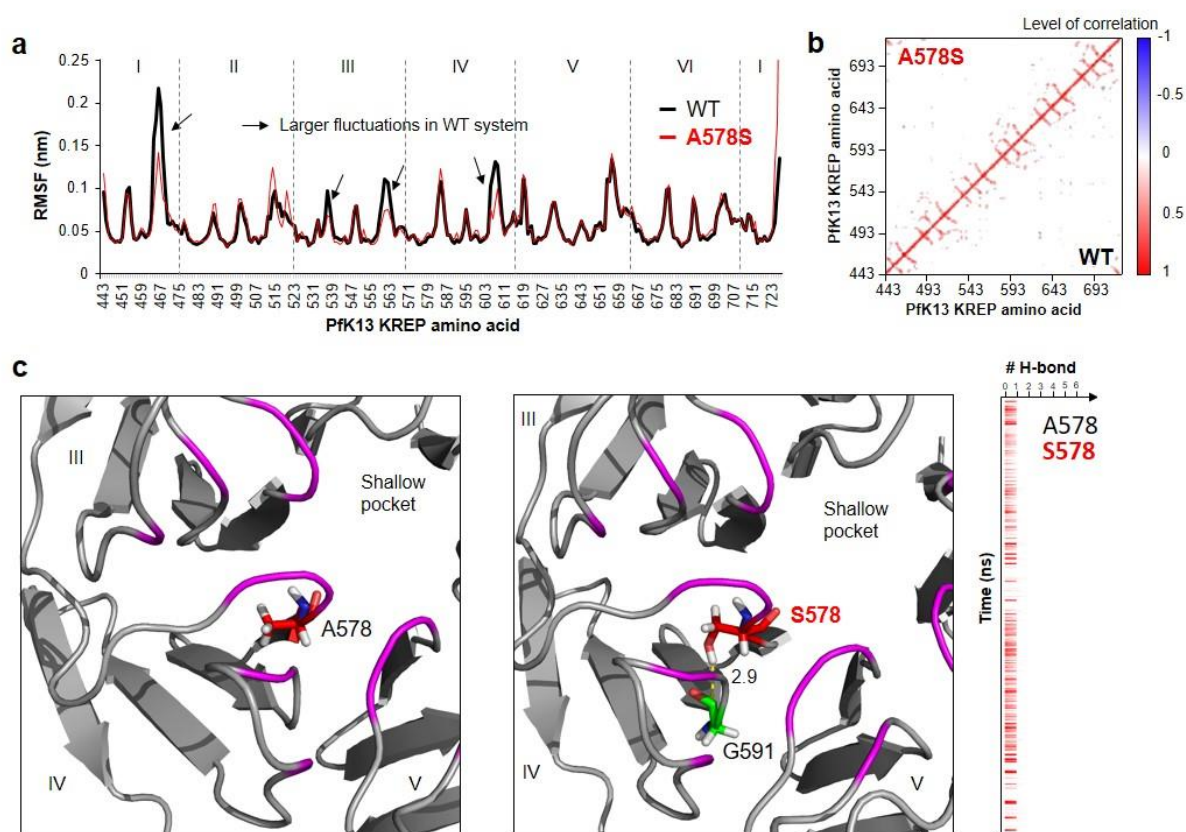


Figure S13. Local impacts of the A578S mutation on the Pfk13 KREP structure, as assessed by molecular dynamics simulations. Molecular dynamics simulations on wild type and A578S mutant Pfk13 KREP structure were carried out using GROMACS during 100 ns at a temperature of 300 K in an all-atom system. The first five nanoseconds (ns) correspond to the equilibration phase. **(a)** Root-mean-square fluctuation (RMSF) values of wild type and A578S mutant Pfk13 KREP structures. RMSF per position was calculated on Ca atoms (excluding the first five ns, corresponding to the equilibration phase). **(b)** Dynamical cross-correlation maps (DCCMs) of wild type and A578S mutant Pfk13 KREP structures. Wild type and mutant systems are shown in *bottom right* and *top left*, respectively. Positive or negative correlation for a pair of residue implies that the two residues move in the same (red) or opposition (blue) directions, respectively. No major difference in correlated and anti-correlated motions of KREP residues was found between the A578S ART-S and the wild type systems. Maps were generated using Bio3D in R. **(c)** Local impacts of A578S ART-S mutation on Pfk13 KREP structure. Blade number (I to VI) and location of the shallow pocket are indicated. We observed the presence of one recurrent hydrogen bond (G591, blade IV), which was not found in the wild type system.

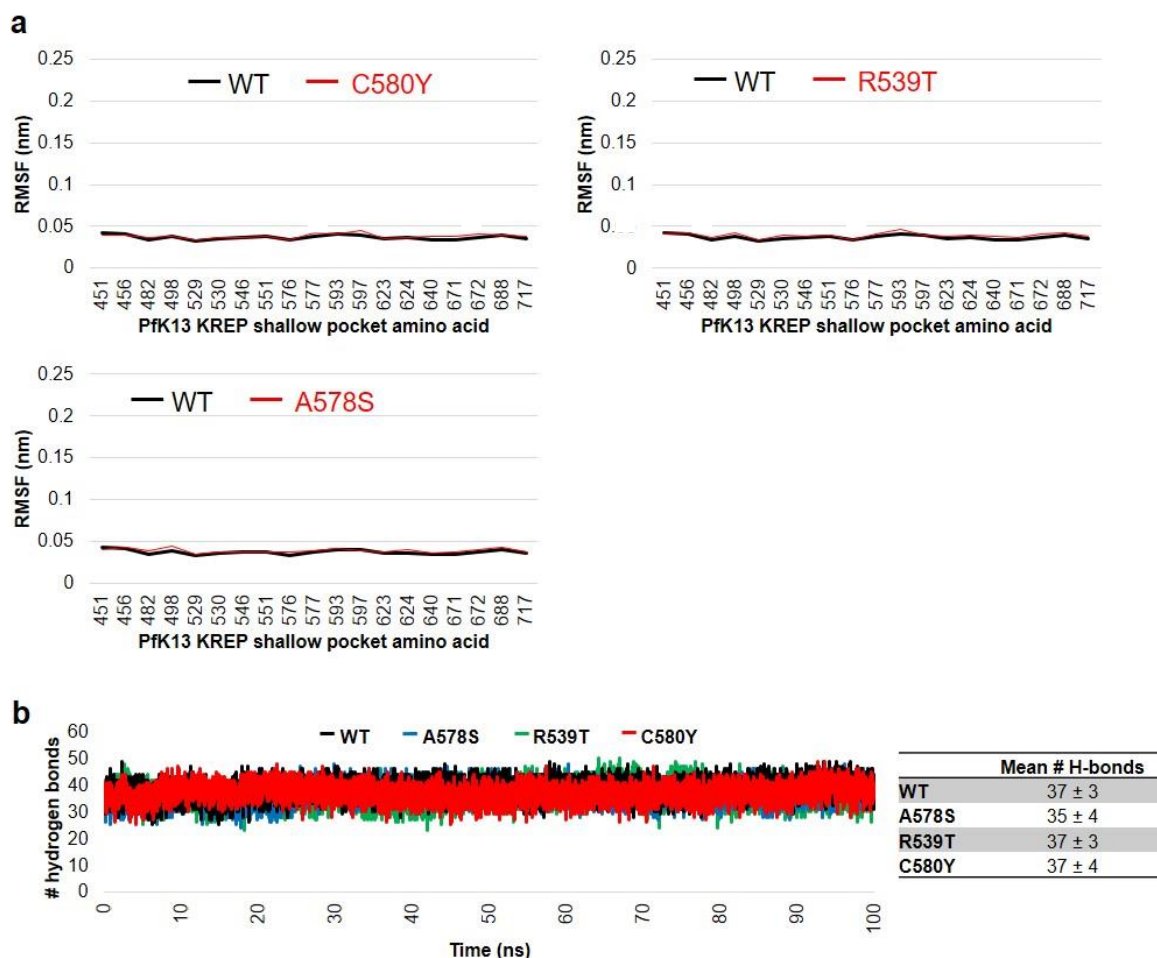


Figure S14. Lack of difference in the number of H-bonds and amino acid fluctuations (RMSFs) between WT and mutant systems for amino acids forming the shallow pocket. Molecular dynamics simulations were carried out on the KREP structure using GROMACS during 100 ns at a temperature of 300 K in an all-atom system. The first ns (0 to 5 ns) correspond to an equilibration phase. **(a)** WT-mutant comparison of the root-mean-square fluctuations (RMSFs) associated to residues forming the shallow pocket. RMSF per position was calculating on α atoms (excluding the first five ns, corresponding to the equilibration phase). WT and mutants are shown in black and red colors. **(b)** Evolution of hydrogen bond number for amino acids forming the shallow pocket during trajectories. A hydrogen bond was counted when the donor-acceptor distance is ≤ 3.5 Å and donor-hydrogen-acceptor angle $\leq 30^\circ$. Wild type and mutant structures showed similar average number of hydrogen bonds. No crucial hydrogen bonds were lost in any of the simulations.

Table S1. List of K13 orthologous sequences used in this study.

Accession Number ^a	Species	Abbreviation ^b	Description	Source	Length ^c	Hosts ^d
PF3D7_1343700	<i>Plasmodium falciparum</i>	Pfal	Kelch protein kelch-13	PlasmoDB	726	Humans
XP_012764805.1	<i>Plasmodium reichenowi</i>	Prei	Kelch protein, putative	NCBI	724	Monkeys
XP_018640212.1	<i>Plasmodium gaboni</i>	Pgab	Putative kelch protein K13	NCBI	725	Monkeys
XP_001614215.1	<i>Plasmodium vivax</i>	Pviv	Kelch domain-containing protein	NCBI	712	Humans / Monkeys
XP_008815744.1	<i>Plasmodium inui</i>	Pinu	Hypothetical protein C922_01923	NCBI	712	Monkeys
XP_002259918.1	<i>Plasmodium knowlesi</i>	Pkno	Kelch-motif containing protein	NCBI	712	Humans / Monkeys
XP_012337473.1	<i>Plasmodium fragile</i>	Pfra	Hypothetical protein AK88_04440	NCBI	712	Monkeys
PCYB_122000	<i>Plasmodium cynomolgi</i>	Pcyn	Kelch domain-containing protein	PlasmoDB	713	Monkeys
XP_008622491.1	<i>Plasmodium vinckei vinckei</i>	Pvin	Hypothetical protein YYE_00585	NCBI	738	Rodents
EUD71747.1	<i>Plasmodium vinckei petteri</i>	Ppet	Hypothetical protein YYG_03164	NCBI	738	Rodents
CDR16038.1	<i>Plasmodium chabaudi chabaudi</i>	Pcha	Kelch protein kelch-13, putative	NCBI	738	Rodents
SCM11204.1	<i>Plasmodium chabaudi adami</i>	Pada	Kelch protein K13, putative	NCBI	738	Rodents
XP_730901.1	<i>Plasmodium yoelii</i>	Pyoe	Hypothetical protein	NCBI	738	Rodents
CDS51002.1	<i>Plasmodium berghei</i>	Pber	Kelch protein kelch-13, putative	NCBI	738	Rodents
PCOAH_00041530	<i>Plasmodium coatneyi</i>	Pcoa	Uncharacterized protein	PlasmoDB	712	Monkeys
PGAL8A_00254800	<i>Plasmodium gallinaceum</i>	Pgal	Kelch protein K13, putative	PlasmoDB	706	Birds
PRELSG_1254200	<i>Plasmodium relictum</i>	Prel	Kelch protein K13, putative	PlasmoDB	704	Birds
GAW82138.1	<i>Plasmodium gonderi</i>	Pgon	Kelch protein K13	NCBI	712	Monkeys
PmUG01_12021200	<i>Plasmodium malariae</i>	Pmal	Kelch protein K13, putative	PlasmoDB	714	Humans
ANG83523.1	<i>Plasmodium ovale wallikeri</i>	Pwal	Kelch protein K13, partial	NCBI	687	Humans
PocGH01_12019400	<i>Plasmodium ovale curtisi</i>	Pcur	Kelch protein K13, putative	PlasmoDB	725	Humans
ESS31969.1	<i>Toxoplasma gondii</i>	Tgon	Kelch repeat and K ⁺ channel tetramerisation domain containing protein	NCBI	818	—
XP_008882882.1	<i>Hammondia hammondi</i>	Hham	Kelch repeat and K ⁺ channel tetramerisation domain containing protein	NCBI	820	—
XP_003882792.1	<i>Neospora caninum</i>	Ncan	Kelch motif domain-containing protein	NCBI	804	—
XP_013247290.1	<i>Eimeria acervulina</i>	Eace	Kelch motif domain-containing protein	NCBI	654	—
EfaB_MINUS_11414	<i>Eimeria falciformis</i>	Efal	Unspecified product	ToxoDB	662	—
CDJ54128.1	<i>Eimeria brunetti</i>	Ebru	Kelch motif domain-containing protein, putative	NCBI	506	—
Ebh_scaff_2461				GeneDB		
Ebh_scaff_22						
XP_013334533.1	<i>Eimeria maxima</i>	Emax	Kelch motif domain-containing protein, putative	NCBI	509	—
Ema_scaff_872				GeneDB		

Accession Number ^a	Species	Abbreviation ^b	Description ^c	Source	Length ^d	Hosts ^e
XP_012648586.1	<i>Babesia microti</i>	Bmic	Uncharacterized protein	NCBI	557	—
XP_766008.1	<i>Theileria parva</i>	Tpar	Hypothetical protein	NCBI	576	—
XP_004829715.1	<i>Theileria equi</i>	Tequ	Kelch repeat domain containing protein	NCBI	556	—
XP_009689263.1	<i>Theileria orientalis</i>	Tori	Uncharacterized protein TOT_010000427	NCBI	603	—
XP_001609009.1	<i>Babesia bovis</i>	Bbov	Kelch repeat and K+ channel tetramerisation domain containing protein	NCBI	569	—
tann.chr01	<i>Theileria annulata</i>	Tann	Protein with kelch domain	GeneDB	576	—
XP_012767816.1	<i>Babesia bigemina</i>	Bbig	Kelch repeat and K+ channel tetramerisation domain containing protein, putative	NCBI	634	—
XP_002141433.1	<i>Cryptosporidium muris</i>	Cmur	Kelch motif family protein	NCBI	561	—
XP_667484.1	<i>Cryptosporidium hominis</i>	Chom	Hypothetical protein	NCBI	579	—
CmeUKMEL1_04065	<i>Cryptosporidium meleagridis</i>	Cmel	Kelch motif family protein	CryptoDB	580	—
cubi_02935	<i>Cryptosporidium ubiquitum</i>	Cubi	KELCH domain-containing protein	CryptoDB	580	—
Cbai_ctg7180000000685	<i>Cryptosporidium baileyi</i>	Cbai	—	CryptoDB	566	—
cand_030750	<i>Cryptosporidium andersoni</i>	Cand	Kelch motif family protein	CryptoDB	561	—
XP_628432.1	<i>Cryptosporidium parvum</i>	Cpar	POZ+kelch domain protein with kelch repeats at the C-terminus	NCBI	580	—
SN3_01300125	<i>Sarcocystis neurona</i>	Sneu	Kelch protein	ToxoDB	707	—

^a Identifier of the protein sequence retrieved from sequence databases.

^b Four-letter species code used in Supplementary Figure S1.

^c Number of amino acids.

^d Vertebrate host(s) are only provided for *Plasmodium* species.

Table S2. Results of PAML analyses for the *k13* gene.

Model	<i>L</i>	Parameters	κ	Positively selected codons
FR (free-ratio)	-29,511.0	ω can vary between branches	1.19	NA
M0 (one-ratio)	-29,739.4	$\omega = 0.022$	1.39	NA
M1a (neutral)	-29,727.8	$\omega_0 = 0.021, p_0 = 0.994$ $\omega_1 = 1.000, p_1 = 0.006$	1.41	NA
M2a (selection)	-29,727.8	$\omega_0 = 0.021, p_0 = 0.994$ $\omega_1 = 1.000, p_1 = 0.002$ $\omega_2 = 1.000, p_2 = 0.004$	1.41	None detected
M3 (discrete) ($k = 5$)	-29,216.3	$\omega_0 = 0.0002, p_0 = 0.105$ $\omega_1 = 0.0065, p_1 = 0.245$ $\omega_2 = 0.0202, p_2 = 0.420$ $\omega_3 = 0.0491, p_3 = 0.210$ $\omega_4 = 0.1100, p_4 = 0.020$	1.43	None detected
M7 (beta)	-29,225.2	$\beta(0.98, 40.98)$ $\omega \leq 1.00, p = 1.000$	1.43	NA
M8 (beta& ω)	-29,225.2	$\beta(0.98, 40.98)$ $\omega \leq 1.00, p_0 = 0.999$	1.43	None detected

Note – ω , d_N/d_S ratio; p_n , proportion of codon sites in the site class ω_n ; L , log-likelihood value; κ , transition/transversion rate ratio; $\beta(p, q)$, shape parameters for the β -distribution of ω . Candidate codon sites for positive selection are identified using the Bayes empirical Bayes (BEB) inference for models M2a and M8, and the Naïve empirical Bayes (NEB) inference for model M3.

Table S3. Distribution of the conserved KREP shallow pocket positions for K13 and some BTB-Kelch proteins.

Protein	Conservation level ^a	Shallow pocket positions ^b	Remaining positions	<i>p</i> value
K13 _{KREP}	High	9	20	1.60×10^{-5}
	Others	10	245	
KEAP1 _{KREP}	High	12	17	1.44×10^{-7}
	Others	12	244	
KLHL2 _{KREP}	High	13	16	1.26×10^{-10}
	Others	6	251	
KLHL3 _{KREP}	High	12	17	1.36×10^{-9}
	Others	6	251	
KLHL12 _{KREP}	High	14	15	2.94×10^{-12}
	Others	5	255	

^a The KREP positions were categorized into two groups: the *high* conservation level group contained the 10% most conserved positions according to the λ substitution rates, the *others* group included the KREP positions with lower levels of conservation.

^b The shallow pocket of K13 KREP was defined as the positions forming the surface plan of the pocket and protruding out of the plan. A view and delineation of the PfK13 KREP shallow pocket is provided in Supplementary Figure S7. The same structural delineation of the KREP shallow pocket was applied to the four other BTB-Kelch KREP domains. Contingency tables were subjected to chi-squared test.

Table S4. Conservation level of the protein sites associated with artemisinin resistance or artemisinin sensitive mutations.

AA ^a position	Ref. AA ^a (<i>P. fal</i>)	PfK13 mutant ^b	Amino acid variation across <i>Apicomplexa</i> species		λ -based rank n, (%) ^c
			<i>Plasmodium</i> (n = 21)	Non- <i>Plasmodium</i> (n = 22)	
446	F	F446I	F (21)	F (9); Y (7); L (6)	110 (38.73)
449	G	G449A	G (21)	G (22)	32 (11.27)
452	D	D452E	D (21)	N (14); S (4); D (4)	164 (57.75)
458	N	N458Y	N (20); S (1)	N (6); D (6); R (5); K (2); S (1); G (1); P (1)	222 (78.17)
469	C	C469Y	C (21)	C (10); G (4); V (3); R (2); K (1); Q (1); N (1)	185 (65.14)
476	M	M476I	M (21)	M (21); L (1)	81 (28.52)
479	K	K479I	K (21)	E (13); P (4); A (3); S (2)	117 (41.20)
481	A	A481V	A (21)	A (13); M (7); I (1); T (1)	96 (33.80)
493	Y	Y493H	Y (21)	Y (10); A (5); F (4); C (3)	136 (47.89)
515	R	R515K	R (21)	R (12); I (4); L (3); T (3)	103 (36.27)
522	S	S522C	S (21)	A (13); P (5); K (2); S (1); Q (1)	194 (68.31)
527	P	P527H	P (21)	P (13); G (5); A (4)	71 (25.00)
537	N	N537I; N537D	N (21)	D (10); G (5); E (2); N (2); S (1); A (1); R (1)	271 (95.42)
538	G	G538V	G (21)	G (10); N (3); D (3); E (2); S (2); A (1); K (1)	242 (85.21)
539	R	R539T	R (21)	R (22)	74 (26.06)
543	I	I543T	I (21)	I (10); V (10); A (2)	137 (48.24)
553	P	P553L	P (21)	S (10); D (6); K (3); A (2); G (1)	209 (73.59)
561	R	R561H	R (21)	R (22)	36 (12.68)
568	V	V568G	V (13); I (8)	V (9); L (4); S (3); A (2); G (2); M (1); C (1)	263 (92.61)
574	P	P574L	P (21)	P (20); A (2)	85 (29.93)
575	R	R575K	R (21)	R (22)	15 (5.28)
#578	A	A578S	A (12); S (8); C (1)	A (20); P (2)	76 (26.76)
579	M	M579I	M (21)	M (15); S (6); C (1)	114 (40.14)
580	C	C580Y	C (19); S (2)	C (14); V (4); L (2); I (1); A (1)	183 (64.44)
584	D	D584V	D (20); E (1)	D (9); N (7); G (4); E (2)	158 (55.63)
667	P	P667T	P (21)	P (12); E (3); A (2); C (1); S (1); I (1); D (1); N (1)	252 (88.73)
673	F	F673I	F (21)	C (11); A (5); S (3); T (2); G (1)	186 (65.49)
675	A	A675V	A (17); S (4)	A (8); I (6); M (4); L (2); V (2)	219 (77.11)
719	H	H719N	H (20); ? (1)	H (22)	28 (9.86)

^a AA: amino acid position. Ref. AA indicates the wild type residue in PfK13.

^b Validated and candidate PfK13 artemisinin resistance mutations (according to WHO) are shown in bold and normal letters, respectively; A578S (indicated with the # symbol) does not confer resistance to artemisinin.

^c The rank attributed to positions was based on the site-specific substitution rates λ (Supplementary Dataset S1) estimated for the 284 studied Kelch-repeat propeller (KREP) positions using FuncPatch. The lower the rank, the more conserved was the position.

Table S5. List of BTB-containing sequences used for the different BTB-containing protein families.

Accession Number ^a	Protein	Protein family ^b	Species	Length	Description ^c	Extended BTB ^d
Q8H1D3	NPY1	BTB-NPH3	<i>Arabidopsis thaliana</i>	571	BTB/POZ domain-containing protein NPY1	yes (C-ter only)
O80970	NPY2	BTB-NPH3	<i>Arabidopsis thaliana</i>	634	BTB/POZ domain-containing protein NPY2	yes (C-ter only)
Q9FN09	NPY3	BTB-NPH3	<i>Arabidopsis thaliana</i>	579	BTB/POZ domain-containing protein NPY3	yes (C-ter only)
O64814	NPY4	BTB-NPH3	<i>Arabidopsis thaliana</i>	481	BTB/POZ domain-containing protein NPY4	yes (C-ter only)
Q0WL52	NPY5	BTB-NPH3	<i>Arabidopsis thaliana</i>	580	BTB/POZ domain-containing protein NPY5	yes (C-ter only)
O15156	ZBTB7B	BTB-ZF	<i>Homo sapiens</i>	539	Zinc finger and BTB domain-containing protein 7B	yes (C-ter only)
Q96BR9	ZBTB8A	BTB-ZF	<i>Homo sapiens</i>	441	Zinc finger and BTB domain-containing protein 8A	yes (C-ter only)
Q05516	ZBTB16	BTB-ZF	<i>Homo sapiens</i>	673	Zinc finger and BTB domain-containing protein 16	yes (C-ter only)
Q9HC78	ZBTB20	BTB-ZF	<i>Homo sapiens</i>	741	Zinc finger and BTB domain-containing protein 20	yes (C-ter only)
O43298	ZBTB43	BTB-ZF	<i>Homo sapiens</i>	467	Zinc finger and BTB domain-containing protein 43	yes (C-ter only)
O95198	KLHL2	BTB-Kelch	<i>Homo sapiens</i>	593	Kelch-like protein 2	yes
Q53G59	KLHL12	BTB-Kelch	<i>Homo sapiens</i>	568	Kelch-like protein 12	yes
Q14145	KLHL19	BTB-Kelch	<i>Homo sapiens</i>	624	Kelch-like ECH-associated protein 1	yes
O60662	KLHL41	BTB-Kelch	<i>Homo sapiens</i>	606	Kelch-like protein 41	yes
O94844	RHOBTB1	RhoBTB	<i>Homo sapiens</i>	696	Rho-related BTB domain-containing protein 1	yes
Q9BYZ6	RHOBTB2	RhoBTB	<i>Homo sapiens</i>	727	Rho-related BTB domain-containing protein 2	yes
O94955	RHOBTB3	RhoBTB	<i>Homo sapiens</i>	611	Rho-related BTB domain-containing protein 3	yes
Q8L765	BPM1	MATH-BTB	<i>Arabidopsis thaliana</i>	407	BTB/POZ and MATH domain-containing protein 1	yes
Q9M8J9	BPM2	MATH-BTB	<i>Arabidopsis thaliana</i>	406	BTB/POZ and MATH domain-containing protein 2	yes
O22286	BPM3	MATH-BTB	<i>Arabidopsis thaliana</i>	408	BTB/POZ and MATH domain-containing protein 3	yes
Q9SRV1	BPM4	MATH-BTB	<i>Arabidopsis thaliana</i>	465	BTB/POZ and MATH domain-containing protein 4	yes
Q1EBV6	BPM5	MATH-BTB	<i>Arabidopsis thaliana</i>	410	BTB/POZ and MATH domain-containing protein 5	yes
Q8TBC3	SHKBP1	KCTD-like	<i>Homo sapiens</i>	707	SH3KBP1-binding protein 1	no
Q719H9	KCTD1	KCTD	<i>Homo sapiens</i>	257	BTB/POZ domain-containing protein KCTD1	no
Q8NC69	KCTD6	KCTD	<i>Homo sapiens</i>	237	BTB/POZ domain-containing protein KCTD6	no
Q8WZ19	KCTD13	KCTD	<i>Homo sapiens</i>	329	BTB/POZ domain-containing adapter for CUL3-mediated RhoA degradation protein 1	no
Q8N5Z5	KCTD17	KCTD	<i>Homo sapiens</i>	321	BTB/POZ domain-containing protein KCTD17	no
Q09470	KCNA1	KCNA	<i>Homo sapiens</i>	495	Potassium voltage-gated channel subfamily A member 1	no
P16389	KCNA2	KCNA	<i>Homo sapiens</i>	499	Potassium voltage-gated channel subfamily A member 2	no
P22001	KCNA3	KCNA	<i>Homo sapiens</i>	575	Potassium voltage-gated channel subfamily A member 3	no
P22459	KCNA4	KCNA	<i>Homo sapiens</i>	653	Potassium voltage-gated channel subfamily A member 4	no
P22460	KCNA5	KCNA	<i>Homo sapiens</i>	613	Potassium voltage-gated channel subfamily A member 5	no
P63208	SKP1	—	<i>Homo sapiens</i>	163	S-phase kinase-associated protein 1	yes (N-ter only)
Q15369	Elongin C	—	<i>Homo sapiens</i>	112	Elongin-C	no

^a Identifier of the protein sequence retrieved from the UniProtKB database.

^b BTB-containing proteins are classified into different families or subfamilies on the basis of the protein domain composition.

^c Functional annotation of the protein, provided by UniprotKB database.

^d BTB domains from BTB-ZF, BTB-NPH3, BTB-Kelch, RhoBTB and MATH-BTB proteins contain additional N- and/or C-terminal structural elements.

Table S6. Full list of species and GenBank entries used for SHKBP1, KCTD17, KEAP1, KLHL2, KLHL3 and KLHL12 proteins.

Species	GenBank accession numbers					
	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Acanthaster planci</i>	—	—	—	XP_022079940.1	—	—
<i>Acanthisitta chloris</i>	—	XP_009082248.1	—	—	—	—
<i>Acanthochromis polyacanthus</i>	—	—	XP_022046307.1	XP_022074174.1	XP_022075435.1	XP_022043757.1
<i>Acinonyx jubatus</i>	—	—	XP_014933811.1	—	XP_014919947.1	—
<i>Acropora digitifera</i>	—	—	—	XP_015764133.1	—	—
<i>Ailuropoda melanoleuca</i>	—	—	XP_002921417.1	XP_011233125.2	XP_011227565.2	XP_002925475.1
<i>Alligator mississippiensis</i>	—	XP_019345267.1	—	XP_019336971.1	KYO20331.1	—
<i>Alligator sinensis</i>	—	XP_014376482.1	XP_006023504.1	XP_014374102.1	XP_014376891.1	XP_006038484.1
<i>Amazona aestiva</i>	—	KQK76492.1	—	—	—	—
<i>Amphiprion ocellaris</i>	XP_023125107.1	—	XP_023128544.1	XP_023135194.1	XP_023143999.1	XP_023131332.1
<i>Anas platyrhynchos</i>	—	EOB03950.1	—	—	XP_005018740.1	—
<i>Anolis carolinensis</i>	—	XP_008108870.1	XP_003216447.1	XP_008116333.1	XP_008103098.1	XP_003220513.1
<i>Anser cygnoides domesticus</i>	—	XP_013036383.1	—	XP_013030701.1	XP_013031336.1	—
<i>Antrostomus carolinensis</i>	—	XP_010160965.1	—	—	KFZ58705.1	—
<i>Aotus nancymae</i>	—	XP_012331822.1	XP_012291394.1	XP_021530947.1	XP_012290872.1	—
<i>Aplysia californica</i>	—	—	—	XP_005099881.1	—	—
<i>Aptenodytes forsteri</i>	—	XP_009288225.2	—	XP_009270590.1	KFM12065.1	XP_009280168.1
<i>Apteryx australis mantelli</i>	—	XP_013802726.1	—	XP_013797640.1	—	—
<i>Aquila chrysaetos canadensis</i>	—	XP_011592211.1	—	XP_011572555.1	XP_011569540.1	—
<i>Astyanax mexicanus</i>	XP_022535160.1	—	XP_007231370.1	XP_022537500.1	XP_007246470.2	XP_022526735.1
<i>Austrofundulus limnaeus</i>	XP_013889136.1	—	XP_013860381.1	XP_013875667.1	—	XP_013871669.1
<i>Balaenoptera acutorostrata scammoni</i>	XP_007166868.1	XP_007165722.1	—	XP_007176776.1	—	—
<i>Balearica regulorum gibbericeps</i>	—	KFO11200.1	—	XP_010297540.1	—	—
<i>Bison bison bison</i>	XP_010847541.1	XP_010858924.1	XP_010826664.1	—	XP_010842774.1	—
<i>Boleophthalmus pectinirostris</i>	XP_020794892.1	—	XP_020794190.1	XP_020788618.1	XP_020779108.1	XP_020785201.1
<i>Bos indicus</i>	XP_019833810.1	XP_019816894.1	—	—	XP_019820522.1	—
<i>Bos mutus</i>	ELR60093.1	XP_014333583.1	ELR51175.1	XP_005898408.1	—	—
<i>Bos taurus</i>	NP_001076076.1	—	NP_001094612.1	XP_005217484.1	XP_612749.5	—
<i>Branchiostoma belcheri</i>	—	—	—	XP_019635414.1	—	—
<i>Bubalus bubalis</i>	XP_006050123.1	XP_006057052.1	XP_006068412.1	XP_006073262.1	—	XP_006078512.1
<i>Buceros rhinoceros silvestris</i>	—	XP_010143072.1	—	—	—	—
<i>Calidris pugnax</i>	—	XP_014819558.1	—	—	—	—
<i>Callithrix jacchus</i>	XP_008986284.1	XP_017819005.1	—	XP_002745302.1	—	XP_008983687.1
<i>Callorhinchus milii</i>	—	—	XP_007907175.1	XP_007904738.1	—	XP_007899790.1
<i>Calypte anna</i>	—	XP_008494604.1	—	XP_008492043.1	KFO97435.1	—
<i>Camelus bactrianus</i>	XP_010945566.1	XP_010958659.1	—	—	—	XP_010954536.1
<i>Camelus dromedarius</i>	—	XP_010993082.1	—	—	—	—

Species	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Camelus ferus</i>	XP_006177980.1	—	XP_006179430.1	XP_006190874.1	XP_006180366.1	XP_006173510.2
<i>Canis lupus familiaris</i>	XP_533665.2	—	XP_533917.2	—	XP_005626632.1	XP_537119.2
<i>Capra hircus</i>	XP_017917921.1	XP_017903788.1	—	—	—	XP_013825912.1
<i>Cariama cristata</i>	—	XP_009708538.1	—	XP_009707350.1	XP_009707735.1	—
<i>Carlito syrichta</i>	—	—	XP_008049111.2	XP_021568132.1	—	—
<i>Castor canadensis</i>	XP_020022544.1	XP_020008051.1	XP_020029886.1	—	—	—
<i>Cavia porcellus</i>	XP_003462052.1	—	—	XP_003479693.2	XP_003473223.2	XP_003474875.1
<i>Cebus capucinus imitator</i>	XP_017358686.1	XP_017364405.1	—	—	XP_017356455.1	—
<i>Ceratotherium simum simum</i>	XP_004441608.1	XP_014636452.1	XP_004442687.1	XP_014645362.1	XP_004420216.1	XP_014649616.1
<i>Cercocebus atys</i>	XP_011942145.1	XP_011893527.1	XP_011949164.1	—	—	XP_011893112.1
<i>Chaetura pelagica</i>	—	XP_010004144.1	—	XP_009994230.1	KFU83910.1	XP_009997930.1
<i>Charadrius vociferus</i>	—	XP_009880208.1	—	—	XP_009889406.1	—
<i>Chelonia mydas</i>	—	XP_007062699.1	—	XP_007067100.1	XP_007053651.1	XP_007057964.1
<i>Chinchilla lanigera</i>	XP_005414100.1	XP_005379244.1	XP_013360585.1	XP_013368945.1	XP_005376559.1	XP_005375289.1
<i>Chlorocebus sabaues</i>	—	XP_007973871.1	—	—	XP_008012737.1	—
<i>Chrysemys picta bellii</i>	—	XP_005302503.1	XP_005279608.1	XP_005307496.1	XP_005295945.1	XP_005311075.1
<i>Chrysocloris asiatica</i>	XP_006871474.1	XP_006865277.1	XP_006875157.1	XP_006864603.1	XP_006866164.1	XP_006834185.1
<i>Ciona intestinalis</i>	—	—	—	—	XP_009862126.1	—
<i>Clupea harengus</i>	—	—	XP_012694314.1	XP_012686788.1	XP_012689714.1	—
<i>Colius striatus</i>	—	—	—	—	XP_010202177.1	—
<i>Colobus angolensis palliatus</i>	XP_011807875.1	—	XP_011810238.1	—	—	—
<i>Columba livia</i>	—	XP_021138056.1	—	XP_005502561.1	XP_005503512.2	—
<i>Condylura cristata</i>	XP_004694187.1	XP_017593373.1	XP_004688795.1	XP_012580367.1	XP_004686812.1	XP_004685574.1
<i>Corvus brachyrhynchos</i>	—	—	—	XP_008636811.1	KFO58285.1	XP_008636237.1
<i>Corvus cornix cornix</i>	—	XP_019148111.1	—	—	—	—
<i>Coturnix japonica</i>	—	XP_015717966.1	—	XP_015716366.1	XP_015731168.1	XP_015740339.1
<i>Cricetulus griseus</i>	—	XP_003514245.1	—	—	XP_007614336.1	XP_003498896.1
<i>Crassostrea virginica</i>	—	—	—	XP_022318258.1	—	—
<i>Cricetulus griseus</i>	—	—	XP_003499782.1	XP_007638805.1	—	—
<i>Crocodylus porosus</i>	—	XP_019400405.1	—	XP_019390124.1	XP_019390681.1	—
<i>Cuculus canorus</i>	—	XP_009566732.1	—	XP_009563451.1	—	XP_009558874.1
<i>Cyanistes caeruleus</i>	—	XP_023791561.1	—	—	—	—
<i>Cynoglossus semilaevis</i>	XP_008307749.1	—	XP_008315818.2	XP_008315684.1	XP_016894992.1	XP_008317759.1
<i>Cyprinodon variegatus</i>	—	—	XP_015254374.1	XP_015239330.1	XP_015228777.1	—
<i>Cyprinus carpio</i>	—	—	—	—	—	XP_018964306.1
<i>Danio rerio</i>	XP_001920935.4	—	NP_001106948.1	XP_021329567.1	—	NP_001007329.1
<i>Dasypus novemcinctus</i>	—	—	—	—	XP_004450558.1	—
<i>Delphinapterus leucas</i>	XP_022441330.1	XP_022452419.1	—	—	—	XP_022417643.1
<i>Desmodus rotundus</i>	—	XP_024434983.1	—	—	—	—
<i>Dipodomys ordii</i>	XP_012888217.1	—	XP_012889414.1	XP_012881565.1	XP_012871485.1	—
<i>Echinops telfairi</i>	XP_004710557.1	XP_012859580.1	XP_004716860.1	XP_004717436.1	—	—
<i>Egretta garzetta</i>	—	XP_009646023.1	—	XP_009634645.1	—	—

Species	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Elephantulus edwardii</i>	XP_006900961.1	XP_006890212.1	XP_006898831.1	XP_006896716.1	XP_006899557.1	—
<i>Enhydra lutris kenyonii</i>	XP_022380965.1	—	XP_022382690.1	XP_022370359.1	XP_022348825.1	XP_022364617.1
<i>Eptesicus fuscus</i>	XP_008137806.1	XP_008142891.1	XP_008156282.1	XP_008142537.1	XP_008140142.1	—
<i>Equus asinus</i>	XP_014715764.1	XP_014706708.1	XP_014721012.1	—	XP_014683945.1	—
<i>Equus caballus</i>	XP_003362363.2	XP_023487576.1	XP_001492331.3	XP_005607800.1	XP_001504377.2	XP_001495985.1
<i>Equus przewalskii</i>	XP_008532834.1	—	—	—	XP_008512037.1	XP_008524709.1
<i>Erinaceus europaeus</i>	XP_007536779.1	XP_007519412.1	XP_007526023.1	—	XP_007517633.1	XP_007522166.1
<i>Esox lucius</i>	XP_010890266.1	—	XP_010871089.1	—	XP_019901531.1	XP_010888908.1
<i>Exaiptasia pallida</i>	—	—	—	XP_020894482.1	—	—
<i>Falco cherrug</i>	—	—	—	—	XP_005440701.2	XP_005447275.1
<i>Falco peregrinus</i>	—	—	—	XP_005230556.1	XP_005242799.2	—
<i>Felis catus</i>	XP_003997840.1	XP_023113233.1	XP_023101067.1	—	XP_019688114.1	XP_019677561.1
<i>Ficedula albicollis</i>	—	XP_005039846.1	—	—	XP_016157084.1	XP_005059410.1
<i>Fukomys damarensis</i>	XP_010612769.1	—	—	XP_010628165.1	XP_010612413.1	KFO24396.1
<i>Fulmarus glacialis</i>	—	XP_009572356.1	—	—	—	—
<i>Fundulus heteroclitus</i>	XP_021164550.1	—	XP_012705937.1	XP_012725245.1	XP_021168459.1	XP_012730968.1
<i>Galeopterus variegatus</i>	—	—	XP_008561317.1	—	XP_008573655.1	—
<i>Gallus gallus</i>	—	XP_004937809.2	—	NP_001264749.1	XP_015149440.1	XP_015154220.1
<i>Gavia stellata</i>	—	XP_009813946.1	—	—	KFV59300.1	—
<i>Gavialis gangeticus</i>	—	XP_019370816.1	—	XP_019377290.1	XP_019361778.1	—
<i>Gekko japonicus</i>	—	XP_015281963.1	—	—	—	—
<i>Geospiza fortis</i>	—	XP_005421877.1	—	XP_005415620.1	XP_005418758.1	XP_005426852.1
<i>Gorilla gorilla gorilla</i>	—	XP_018874316.1	—	—	—	—
<i>Haliaeetus leucocephalus</i>	—	XP_010583711.1	—	—	XP_010577185.1	—
<i>Haliaeetus albicilla</i>	—	—	—	XP_009928870.1	—	—
<i>Haplochromis burtoni</i>	—	—	XP_014196664.1	—	—	—
<i>Heterocephalus glaber</i>	XP_004874267.1	XP_012927957.1	EHB12931.1	XP_004848282.1	XP_004836251.1	EHB03738.1
<i>Hippocampus comes</i>	—	—	XP_019717662.1	XP_019717542.1	XP_019739597.1	XP_019744013.1
<i>Hipposideros armiger</i>	XP_019485272.1	XP_019520201.1	XP_019514927.1	—	XP_019517118.1	XP_019493677.1
<i>Homo sapiens</i>	XP_006723537.1	NP_001269613.1	XP_011526754.1	NP_009177.3	NP_059111.2	NP_067646.1
<i>Ictalurus punctatus</i>	XP_017322144.1	XP_017339989.1	XP_017337726.1	XP_017316882.1	XP_017329595.1	XP_017324276.1
<i>Ictidomys tridecemlineatus</i>	XP_005336496.1	XP_005322340.1	XP_005336185.1	XP_013215786.1	XP_005333265.1	XP_005330387.1
<i>Jaculus jaculus</i>	—	XP_012804799.1	XP_004667509.1	XP_004657220.1	XP_004665149.1	—
<i>Kryptolebias marmoratus</i>	—	—	XP_017285965.1	XP_017261728.1	XP_017273010.1	XP_017294510.1
<i>Labrus bergylta</i>	XP_020514016.1	—	XP_020503327.1	XP_020502841.1	—	XP_020493361.1
<i>Larimichthys crocea</i>	XP_019133021.1	—	XP_019129802.1	XP_010747705.1	XP_010741230.1	XP_019108919.1
<i>Lates calcarifer</i>	—	—	XP_018520553.1	XP_018546403.1	XP_018549229.1	XP_018542865.1
<i>Latimeria chalumnae</i>	—	—	XP_005994373.1	—	XP_005991519.1	XP_005988446.1
<i>Lepidothrix coronata</i>	—	—	—	XP_017661316.1	XP_017683358.1	XP_017676878.1
<i>Lepisosteus oculatus</i>	XP_006627705.1	—	—	—	—	XP_015197756.1
<i>Leptonychotes weddellii</i>	—	XP_006751188.1	XP_006741120.1	—	—	—
<i>Leptosomus discolor</i>	—	—	—	XP_009950311.1	KFQ10383.1	KFP98732.1

Species	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Limosa lapponica baueri</i>	—	—	—	—	—	PKU41125.1
<i>Lingula anatina</i>	—	—	—	—	XP_013388754.1	—
<i>Lipotes vexillifer</i>	XP_007458702.1	XP_007454129.1	—	XP_007471591.1	XP_007461459.1	XP_007456221.1
<i>Lonchura striata domestica</i>	—	XP_021409774.1	—	XP_021405931.1	XP_021407958.1	—
<i>Loxodonta africana</i>	—	XP_023415713.1	XP_003413141.1	XP_010593899.1	XP_010589155.1	XP_003410281.1
<i>Macaca fascicularis</i>	XP_005589339.1	XP_005567444.1	—	—	—	—
<i>Macaca mulatta</i>	XP_014979475.1	NP_001252736.1	—	XP_001100501.1	NP_001252766.1	—
<i>Macaca nemestrina</i>	XP_011763031.1	XP_011710770.2	—	XP_011714137.1	XP_011714779.1	—
<i>Manacus vitellinus</i>	—	—	—	XP_008931371.1	KFW78569.1	XP_008927962.2
<i>Mandrillus leucophaeus</i>	XP_011839136.1	—	XP_011849866.1	XP_011823166.1	XP_011824557.1	—
<i>Manis javanica</i>	XP_017528133.1	XP_017508523.1	XP_017495620.1	XP_017511714.1	XP_017513581.1	XP_017514597.1
<i>Marmota marmota marmota</i>	XP_015352327.1	—	—	XP_015341732.1	—	—
<i>Maylandia zebra</i>	XP_004565061.1	—	XP_004567922.1	XP_004549335.1	XP_004553661.1	XP_004546484.1
<i>Meleagris gallopavo</i>	—	XP_019470588.1	—	XP_010707769.1	—	—
<i>Melopsittacus undulatus</i>	—	—	—	XP_005145559.1	—	—
<i>Meriones unguiculatus</i>	XP_021506303.1	XP_021492399.1	XP_021504873.1	XP_021493522.1	—	—
<i>Merops nubicus</i>	—	—	—	—	KFQ31548.1	—
<i>Mesitornis unicolor</i>	—	XP_010188342.1	—	—	—	—
<i>Mesocricetus auratus</i>	XP_005086624.1	XP_005067012.1	XP_005078609.1	XP_005075242.1	XP_021087667.1	—
<i>Microcebus murinus</i>	XP_012630007.1	XP_012642587.1	XP_012646096.1	XP_012626034.1	XP_020137732.1	XP_012599823.1
<i>Microtus ochrogaster</i>	XP_005371646.1	XP_005354301.1	XP_005346939.1	XP_005367805.1	—	XP_005348437.1
<i>Miniopterus natalensis</i>	XP_016064707.1	—	XP_016063423.1	XP_016053241.1	XP_016059235.1	—
<i>Mizuhopecten yessoensis</i>	—	—	—	XP_021358880.1	—	—
<i>Monodelphis domestica</i>	XP_007492042.1	XP_016281346.1	XP_007488618.1	XP_007496215.1	XP_007473357.1	—
<i>Monopterus albus</i>	—	—	XP_020452724.1	XP_020476449.1	XP_020463137.1	XP_020473490.1
<i>Mus caroli</i>	XP_021021679.1	XP_021038727.1	XP_021027535.1	—	—	—
<i>Mus musculus</i>	XP_006539730.2	XP_006521523.2	NP_057888.1	XP_017168500.1	XP_006517071.1	NP_001298065.1
<i>Mus pahari</i>	XP_021049670.1	XP_021072585.1	XP_021062655.1	—	XP_021071267.1	XP_021053254.1
<i>Mustela putorius furo</i>	XP_004776173.1	XP_012907038.1	XP_004748492.1	XP_004779602.2	XP_012907923.1	XP_012913608.1
<i>Myotis brandtii</i>	XP_005868856.1	—	XP_005859245.1	XP_014390466.1	XP_005881064.1	XP_005886426.1
<i>Myotis davidii</i>	XP_006759262.1	—	ELK28701.1	XP_015424193.1	—	XP_006757531.1
<i>Myotis lucifugus</i>	—	XP_006107324.2	XP_006104108.1	XP_014311670.1	—	—
<i>Nannospalax galili</i>	XP_008852225.1	XP_008849546.1	XP_008848262.1	XP_008829715.1	XP_008830034.1	XP_008828554.1
<i>Nanorana parkeri</i>	—	—	XP_018431113.1	—	—	XP_018412396.1
<i>Neolamprologus brichardi</i>	XP_006801280.1	—	XP_006787448.1	—	—	XP_006794161.1
<i>Neomonachus schauinslandi</i>	XP_021556625.1	—	—	—	—	—
<i>Neophocaena asiaeorientalis asiaeorientalis</i>	XP_024624432.1	XP_024617351.1	—	—	—	—
<i>Nestor notabilis</i>	—	XP_010017253.1	—	—	—	—
<i>Nipponia nippon</i>	—	XP_009473450.1	—	XP_009473047.1	KFQ93934.1	—
<i>Nomascus leucogenys</i>	XP_012352070.1	XP_003264762.1	—	XP_003257978.1	XP_003266467.1	XP_003264617.1
<i>Nothobranchius furzeri</i>	XP_015806312.1	—	—	XP_015797306.1	—	XP_015824029.1
<i>Notothernia coriiceps</i>	—	—	XP_010766987.1	XP_010789689.1	—	—

Species	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Numida meleagris</i>	—	XP_021241564.1	—	XP_021250584.1	—	XP_021232836.1
<i>Ochotona princeps</i>	XP_004598066.1	XP_004589821.2	XP_004600089.2	XP_004591480.1	XP_004586491.1	XP_004578811.1
<i>Octodon degus</i>	XP_004648039.1	—	XP_004632901.1	XP_004630664.1	XP_004632291.1	XP_004625419.1
<i>Octopus bimaculoides</i>	—	—	—	XP_014770021.1	—	—
<i>Odobenus rosmarus divergens</i>	XP_004396404.1	—	XP_004398925.1	XP_004411320.2	XP_004397909.1	XP_004392834.2
<i>Odocoileus virginianus texanus</i>	XP_020759499.1	—	XP_020758035.1	—	XP_020748682.1	XP_020759424.1
<i>Oncorhynchus kisutch</i>	XP_020363280.1	—	—	—	XP_020357787.1	XP_020315264.1
<i>Oncorhynchus mykiss</i>	XP_021443142.1	—	XP_021412413.1	—	XP_021439271.1	XP_021421846.1
<i>Oncorhynchus tshawytscha</i>	XP_024300321.1	—	—	—	—	—
<i>Ophiophagus hannah</i>	—	—	ETE73917.1	—	—	ETE69140.1
<i>Orbicella faveolata</i>	—	—	—	XP_020601220.1	—	—
<i>Orcinus orca</i>	XP_004271317.1	XP_012392053.1	XP_004277407.1	XP_004284035.1	XP_004282135.1	XP_012394372.1
<i>Oreochromis niloticus</i>	XP_013129084.1	—	XP_003447974.1	XP_019215626.1	XP_003443442.1	XP_005448525.1
<i>Orycteropus afer afer</i>	XP_007941168.1	XP_007939819.1	XP_007951766.1	XP_007952393.1	XP_007937122.1	XP_007939336.1
<i>Oryctolagus cuniculus</i>	—	—	XP_002723322.1	XP_008271730.1	XP_008253281.1	XP_002717633.2
<i>Oryzias latipes</i>	XP_023817440.1	—	XP_020557948.1	XP_020559732.1	XP_004073059.3	XP_004070913.1
<i>Oryzias melastigma</i>	XP_024133077.1	—	—	—	—	—
<i>Otolemur garnettii</i>	XP_003802151.1	XP_003783288.1	XP_003798009.1	—	XP_003782213.1	—
<i>Ovis aries</i>	—	—	—	XP_011952679.1	XP_004008868.1	XP_012043275.1
<i>Ovis aries musimon</i>	XP_011986482.1	XP_012004229.1	XP_011991418.1	—	XP_011971358.1	—
<i>Pan paniscus</i>	XP_003812529.1	XP_003821599.2	XP_003809635.1	—	XP_003829282.1	XP_014198142.1
<i>Pan troglodytes</i>	XP_016791502.1	XP_001160085.2	NP_001266890.1	XP_001150117.3	XP_016810041.1	—
<i>Panthera pardus</i>	XP_019287597.1	XP_019302361.1	—	—	XP_019279815.1	—
<i>Panthera tigris altaica</i>	—	—	XP_007079390.1	—	XP_007077902.1	—
<i>Pantholops hodgsonii</i>	XP_005963560.1	XP_005957223.1	—	—	—	—
<i>Papio anubis</i>	XP_017807946.1	XP_003905544.2	XP_003914932.1	—	—	—
<i>Paralichthys olivaceus</i>	—	—	XP_019939843.1	XP_019956734.1	XP_019941392.1	XP_019960243.1
<i>Paramormyrops kingsleyae</i>	XP_023668026.1	—	—	—	—	—
<i>Parus major</i>	—	XP_015483734.1	—	XP_015480275.1	—	XP_015506363.1
<i>Patagioenas fasciata monilis</i>	—	—	—	—	OPJ88162.1	OPJ82394.1
<i>Pelecanus crispus</i>	—	XP_009486864.1	—	—	—	—
<i>Pelodiscus sinensis</i>	—	XP_014434404.1	—	—	—	XP_006135792.1
<i>Peromyscus maniculatus bairdii</i>	XP_006995894.1	XP_006979168.2	XP_006987077.1	XP_015861554.1	XP_006976813.2	XP_006982523.1
<i>Phascolarctos cinereus</i>	—	XP_020838182.1	XP_020850692.1	—	—	—
<i>Phalacrocorax carbo</i>	—	—	—	KFW84676.1	—	—
<i>Phascolarctos cinereus</i>	—	—	—	XP_020831993.1	—	—
<i>Phoenicopterus ruber ruber</i>	—	KFQ81133.1	—	—	—	—
<i>Physeter catodon</i>	XP_007128153.1	XP_007111830.1	XP_007107317.1	—	XP_007112795.1	XP_007121223.1
<i>Picoides pubescens</i>	—	XP_009900889.1	—	XP_009902865.1	KFV61276.1	—
<i>Piliocolobus tephrosceles</i>	XP_023084903.1	XP_023077918.1	—	—	—	—
<i>Pleuragramma antarctica</i>	—	—	SCW25703.1	—	—	—
<i>Poecilia formosa</i>	XP_007570740.1	—	XP_007551655.1	XP_007549688.1	XP_007577950.1	XP_007553643.1

Species	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Poecilia latipinna</i>	XP_014915223.1	—	XP_014907996.1	XP_014915653.1	XP_014902577.1	—
<i>Poecilia mexicana</i>	—	—	XP_014837107.1	XP_014829877.1	XP_014860420.1	XP_014840751.1
<i>Poecilia reticulata</i>	XP_008400256.1	—	XP_008413531.1	XP_008400816.1	—	XP_008412643.1
<i>Pogona vitticeps</i>	—	XP_020643409.1	XP_020654350.1	XP_020657605.1	XP_020638337.1	XP_020638148.1
<i>Pongo abelii</i>	XP_024092291.1	XP_003779591.1	NP_001126406.1	XP_002815314.1	NP_001127192.1	—
<i>Propithecus coquereli</i>	XP_012512503.1	XP_012493505.1	XP_012496759.1	—	XP_012501789.1	—
<i>Protobothrops mucrosquamatus</i>	—	XP_015687420.1	—	XP_015676122.1	—	XP_015671067.1
<i>Pseudopodoces humilis</i>	—	XP_005527632.2	—	XP_005517662.1	XP_005523584.1	XP_005530239.1
<i>Pterocles gutturalis</i>	—	—	—	XP_010076229.1	—	—
<i>Pteropus alecto</i>	XP_006905479.1	XP_006914536.1	—	—	—	XP_006908363.1
<i>Pteropus vampyrus</i>	XP_011370513.1	XP_011385740.1	XP_011378571.1	XP_011362014.1	XP_011360672.1	—
<i>Pundamilia nyererei</i>	—	—	XP_005752286.1	—	XP_005726524.1	XP_005731643.1
<i>Pygocentrus nattereri</i>	XP_017537184.1	—	XP_017548442.1	XP_017572936.1	XP_017545599.1	XP_017548954.1
<i>Python bivittatus</i>	—	XP_007421076.1	XP_007426630.1	—	—	XP_007426111.1
<i>Rattus norvegicus</i>	NP_001258009.1	XP_008763932.2	NP_476493.2	XP_008770267.1	—	XP_006249905.1
<i>Rhinolophus sinicus</i>	XP_019606931.1	XP_019575200.1	XP_019600540.1	XP_019600193.1	XP_019595782.1	XP_019606448.1
<i>Rhinopithecus roxellana</i>	XP_010379143.1	XP_010387491.1	XP_010372068.1	XP_010362880.1	XP_010386363.1	—
<i>Rousettus aegyptiacus</i>	—	XP_016017310.1	XP_016009006.1	XP_016010073.1	XP_016016102.1	—
<i>Saccoglossus kowalevskii</i>	—	—	—	XP_002730487.1	—	—
<i>Saimiri boliviensis boliviensis</i>	XP_010329404.1	XP_003932948.2	XP_010347939.1	XP_010330383.1	XP_003934008.1	XP_010347190.1
<i>Salmo salar</i>	XP_014017560.1	XP_014034374.1	XP_013986100.1	—	XP_014068166.1	XP_014002172.1
<i>Salvelinus alpinus</i>	XP_023838330.1	—	—	—	—	—
<i>Sarcophilus harrisii</i>	—	—	XP_003760514.1	—	—	XP_003767642.1
<i>Scleropages formosus</i>	XP_018598136.1	—	KPP71934.1	XP_018591491.1	XP_018589121.1	XP_018597687.1
<i>Serinus canaria</i>	—	—	—	—	XP_009089645.1	XP_009095577.1
<i>Seriola dumerili</i>	XP_022594116.1	—	XP_022622580.1	XP_022593877.1	XP_022615920.1	XP_022618872.1
<i>Seriola lalandi dorsalis</i>	XP_023256419.1	—	XP_023274978.1	XP_023265271.1	XP_023253862.1	—
<i>Sinocyclocheilus anshuiensis</i>	—	—	XP_016349766.1	XP_016299079.1	XP_016310893.1	XP_016330970.1
<i>Sinocyclocheilus grahami</i>	—	—	XP_016109850.1	XP_016136233.1	XP_016145128.1	XP_016094876.1
<i>Sinocyclocheilus rhinoceros</i>	XP_016369534.1	—	XP_016364948.1	XP_016414377.1	XP_016409147.1	XP_016372389.1
<i>Sorex araneus</i>	XP_004620791.1	XP_004610551.2	XP_004616619.1	XP_004610419.1	XP_012788506.1	XP_004610091.1
<i>Stegastes partitus</i>	—	—	XP_008275180.1	XP_008291203.1	XP_008303524.1	XP_008275958.1
<i>Strongylocentrotus purpuratus</i>	—	—	—	—	XP_011677555.1	—
<i>Struthio camelus australis</i>	—	XP_009682597.1	—	XP_009675470.1	—	—
<i>Sturnus vulgaris</i>	—	—	—	—	XP_014738920.1	—
<i>Stylophora pistillata</i>	—	—	—	XP_022807913.1	—	—
<i>Sus scrofa</i>	XP_020950056.1	XP_020947187.1	NP_001108143.1	XP_005666707.1	XP_020940448.1	XP_003130659.3
<i>Taeniopygia guttata</i>	—	XP_012424696.1	—	XP_002197028.2	XP_002187183.3	—
<i>Takifugu rubripes</i>	XP_011615817.1	—	XP_003972643.1	XP_003972526.2	XP_011609237.1	XP_003963586.1
<i>Tauraco erythrolophus</i>	—	—	—	KFV09010.1	—	—
<i>Thamnophis sirtalis</i>	—	—	—	—	XP_013925240.1	—
<i>Tinamus guttatus</i>	—	XP_010222429.1	—	—	—	—

Species	SHKBP1 (n = 124)	KCTD17 (n = 139)	KEAP1 (n = 135)	KLHL2 (n = 162)	KLHL3 (n = 158)	KLHL12 (n = 129)
<i>Trichechus manatus latirostris</i>	XP_004390909.1	XP_004373880.1	XP_004378408.1	XP_004390697.1	XP_004384746.1	XP_004375335.1
<i>Tupaia chinensis</i>	XP_006140529.1	XP_014446395.1	XP_006161600.1	—	—	XP_006159222.1
<i>Tursiops truncatus</i>	—	XP_019774047.1	—	—	—	—
<i>Tyto alba</i>	—	—	—	—	KFV43663.1	—
<i>Ursus maritimus</i>	—	—	—	XP_008683826.1	—	—
<i>Vicugna pacos</i>	XP_006215068.1	XP_015097025.1	XP_006206340.1	XP_006213249.1	XP_006204669.1	XP_006213306.2
<i>Xenopus laevis</i>	XP_018084584.1	—	XP_018111773.1	XP_018097193.1	—	XP_018104893.1
<i>Xenopus tropicalis</i>	XP_002938848.1	—	NP_001008024.1	XP_002934798.2	—	—
<i>Xiphophorus maculatus</i>	XP_014327689.1	—	XP_023189591.1	XP_005809438.1	XP_023184792.1	XP_023187727.1
<i>Zonotrichia albicollis</i>	—	XP_014123680.1	—	XP_005486355.1	—	—

Note – The list of species and GenBank entries contained only protein sequences: i) having an unambiguous description (*i.e.* including the name of the queried KCTD or BTB-Kelch protein); ii) that aligned with $\geq 80\%$ sequence coverage and had $\geq 60\%$ sequence identity with the query sequence; iii) that contained all the annotated domains (using the domain annotation automatically generated by the Uniprot Knowledgebase); and iv) without any gap in positions of the conserved domains. Proteins that did not pass these filtering steps were discarded.

Materials and Methods S1. Description of PAML models and interpretations. The heterogeneity of ω among lineages of the *k13* phylogenetic tree (branch models) was tested by comparing the free-ratio (FR) model, which assumes as many ω parameters as the number of branches in the tree, to the one-ratio (M0) model which supposes only one ω value for all branches.^{1,2}

M1a allows codon sites to fall into two site classes, either with $\omega < 1$ (purifying selection) or $\omega = 1$ (neutral evolution), whereas model M2a extends model M1a with a further site class as $\omega > 1$ (positive selection). Model M3 includes a discrete distribution of independent ω with k classes of sites ($k = [3, 4, 5]$ in this study), with ω values and corresponding proportions estimated from the dataset. Model M7 assumed a β -distribution of ten ω ratios limited to the interval $[0, 1]$ with two shape parameters p and q , whereas model M8 adds an additional site class with ω possibly > 1 as M2a does. The heterogeneity of ω across codon sites was tested by comparison of models M0-M3, while comparison of paired models M1a-M2a and M7-M8 allowed to detect positive selection.¹

Model comparisons were made using likelihood ratio tests (LRTs).³ For each of the LRTs, twice the log-likelihood difference between alternative and null models ($2\Delta\ell$) was compared to critical values from a chi-squared distribution with degrees of freedom equal to the difference in the number of estimated parameters between both models.⁴ Candidate sites for positive selection were pinpointed using the Bayes empirical Bayes (BEB) inference which calculates the posterior probability that each codon site falls into a site class affected by positive selection (in models M2a and M8), as described by Yang and colleagues.⁵ For model M3, in which no BEB approach is implemented yet, the Naïve empirical Bayes (NEB) approach was used to identify those sites evolving under positive selection.

Three codon substitution models were used and compared for all models: F1x4 and F3x4, which assume equal nucleotide frequencies and individual codon frequencies at all codon positions, respectively, and the parameter-rich model F61, which estimates codon frequencies separately for each codon.^{6,7} Since the three codon substitution models yielded similar results ($p < 0.01$ in each pairwise comparison, Spearman's rank correlation), we only presented those obtained with the most widely used F3x4 codon model. The analyses were run multiple times with different ω starting values to check the consistency of the results.

For PAML model M3 with k site classes of ω ratios, the posterior mean of ω value at each codon site was calculated as the average of the ω ratios across the k ω site classes weighted by their posterior probabilities.⁸

Materials and Methods S2. Setting up of KCTD and BTB-Kelch datasets. Each *Homo sapiens* KCTD and BTB-Kelch protein sequence was successively submitted as query sequence for a blastp search⁹ (BLOSUM62 scoring matrix, max target sequences fixed at 1,000) against the NCBI non-redundant protein database to retrieve orthologous sequences from a large amount of species. The output lists were then filtered according to specific criteria so as to keep only protein sequences having an unambiguous description (*i.e.* a description that includes the name of the queried KCTD or BTB-Kelch protein), and that aligned with $\geq 80\%$ sequence coverage and had $\geq 60\%$ sequence identity with the query sequence. The multiple protein sequence alignment of each set of orthologous sequences was then generated using MAFFT version 7¹⁰ (E-INS-I strategy with BLOSUM62 scoring matrix, gap opening penalty 2.0 and offset 0.1). A second filtering step was performed to remove incomplete or miss-annotated sequences, *i.e.* the sequences that did not contain all the annotated domains (using the domain annotation automatically generated by the Uniprot Knowledgebase) and/or that included a gapped position located in one of the annotated domains. The final multiple protein sequence alignments included: i) 124 sequences $\times$ 103 aligned positions for SHKBP1; ii) 139 sequences $\times$ 102 aligned positions for KCTD17; iii) 135 sequences $\times$ 285 aligned positions for KEAP1; iv) 162 sequences $\times$ 286 aligned positions for KLHL2; v) 158 sequences $\times$ 286 aligned positions for KLHL3; and vi) 129 sequences $\times$ 289 aligned positions for KLHL12. The full list of orthologous sequences used for each mammalian KCTD and BTB-Kelch protein is provided in Table S6.

Materials and Methods S3. Molecular dynamics simulations set up. At startup, the K13 KREP crystallographic structure was retrieved from the PDB repository, PDB ID: 4ZGC.¹¹ One monomer (chain B) and the BTB domain of the two monomers were excluded for molecular dynamics simulations. In order to evaluate the influence of ART-R mutations on the PfK13 KREP fold, three models were generated: (i) K13 KREP harboring the most common C580Y mutation in natural populations; (ii) K13 KREP harboring the R539T mutation which confers the highest level of ART-R; and (iii) K13 KREP harboring the A578S mutation which serves as control since the mutation do not confer ART-R. Single mutations were introduced using the *swapaa* function of UCSF Chimera by substituting the residue with the most probable rotameric conformation.¹² All missing atoms were then added using Swiss PDB Viewer.¹³ Generated models were checked for quality using MolProbity¹⁴ which showed no outliers and $> 98\%$ of residues (including *in silico* introduced mutations) in favored regions. Molecular dynamics simulations were carried out using the GROMACS package, v. 5.0.7.¹⁵ with the improved side-chain torsion potentials force field Amber99ss-ILDN for amino acid interaction.¹⁶ Protein systems were immersed in a dodecahedron box of TIP3P water molecules¹⁷ preserving at least 13 Å of separation between the solute and the edges of the box. The Particle Mesh Ewald (PME)¹⁸ approach was employed with van der Waals and Coulomb non-bonded interactions truncated at 10 Å. Bond lengths were constrained using the LINCS algorithm¹⁹ that allowed a 2 fs time step in all simulations. The ionization state of residues was set to be consistent with neutral pH, and counter-ions Na^+ were then added by randomly replacing water molecules to ensure the overall charge neutrality of the system. The whole systems consisted of $\sim 35\,000$ atoms. To release conflicting contacts, solvated systems were subjected to energy minimization using the steepest descent algorithm over 5 000 steps until the maximum force $< 1000.0 \text{ kJ/mol}^{-1}/\text{nm}^{-1}$. Before MD productions, each solvated system was subjected to two-step equilibration. In the first step, systems were equilibrated for 100 ps in the NVT ensemble at 300 K with the V-rescale temperature coupling.²⁰ The equilibrated systems from the NVT ensemble were then treated to constant pressure (NPT) ensemble for 100 ps using the Parrinello-Rahman barostat²¹ under an isothermal-isobaric pressure of 1.0 bar. Position restraints were applied to all atoms during equilibration steps to avoid configuration changes. MD productions were run for 100 ns in the absence of any restraints. During MD productions, the V-rescale thermostat coupled with the Parrinello-Rahman barostat were used to maintain the temperature and pressure at 300 K and 1.0 bar, respectively. The trajectories were stored at every 10 ps. Each wild type and mutant system was subjected to three molecular dynamics replicates with the same parameters and algorithms but different velocities. MD trajectories were then analyzed using inbuilt GROMACS tools. The first 5 ns of trajectories (which referred as equilibration time), were removed *prior* to analyses.

References (related to supplementary materials and methods).

1. Nielsen R, Yang Z. (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148: 929–936.
2. Yang Z. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 15: 568–573.
3. Vuong QH. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.
4. Yang Z, Nielsen R, Goldman N, Pedersen AM. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431–449.
5. Yang Z, Wong WSW, Nielsen R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22: 1107–1118.
6. Goldman N, Yang Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11: 725–736.
7. Muse SV, Gaut BS. (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11: 715–724.
8. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. (2007) *Mol Biol Evol* 24: 1586–1591.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410.
10. Katoh K, Standley DM. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772–780.
11. Jiang DQ, Tempel W, Loppnau P, Graslund S, He H, et al. (2015) Crystal structure analysis of Kelch protein (with disulfide bond) from *Plasmodium falciparum*. Protein Data Bank.
12. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612.
13. Guex N, Peitsch MC. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723.
14. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12–21.
15. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25.
16. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–1958.
17. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
18. Darden T, York D, Pedersen L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
19. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472
20. Bussi G, Donadio D, Parrinello M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
21. Parrinello M, Rahman A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.

Matériel supplémentaire – Article 2

```

1      10      20      30      40      50      60      70      80      90
Pfal MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Prai MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pviv MTILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...EDVPISGRKIANFLKLVYH...EIRRNISIIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pinu MRILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GDIPISGRKIANFLKLVYH...EIRRNISIIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pkno MKILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...DEAPMSRKILYFLKLVYH...EIRRNISIIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pfra MKILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GDAPISGRKILYFLKLVYH...EIRRNISIIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pber MTILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSGRSGVKAFFKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pcyn MTILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNVPIGRKILYFLKLVYH...EIRRNISIIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pvin MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GTNFRMKNKAKVFLKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pcha MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSRWGCAKRIKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pcho MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSGRSGVKAFFKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Ppet MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSRWGCAKRIKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pcoa MKILAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GDAPMGOKILYFLKLVYH...EIRRNISIIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pgab MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pcur MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...ERISLFRKILNFI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pwal MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...ERISLFRKILNFI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pada MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSRWGCAKRIKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Ppra MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pbla MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pbil MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Padi MKFAASKKN...NOKNSKKN...DERYRELNLVQEGNGSGISGGGSLGKCAHVFLKLFKEIKDNIFIYHLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pmal MKIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...HRISVIRKINFI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Pgal MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSRWGCAKRIKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH
Prel MTIVAKKKK...KEDSQITFE...DERYRELDSHAQNESEIQ...GNNSRWGCAKRIKLI...GNMKNVYVYLLSIIYLVGVCMNRIFAARLTNKGNYSEVTSETH

```

```

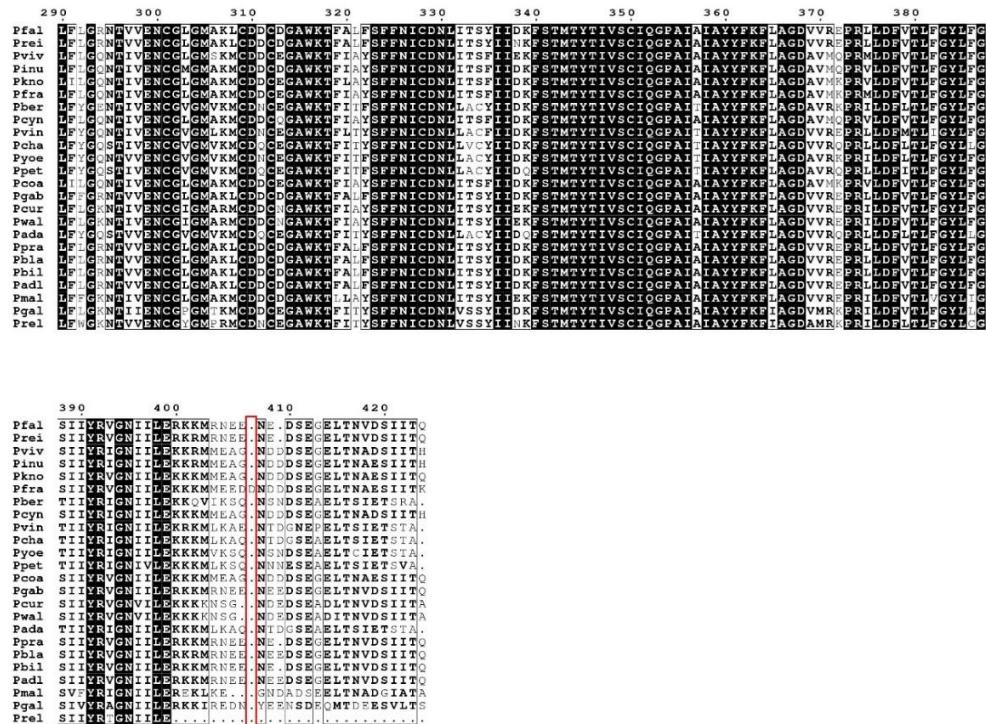
100      110      120      130      140      150      160      170      180      190
Pfal NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Prai NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pviv NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pinu NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pkno NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pfra NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pber NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pcyn NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pvin NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pcha NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pcho NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Ppet NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pcoa NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pgab NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pcur NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pwal NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pada NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Ppra NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pbla NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pbil NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Padi NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pmal NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Pgal NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV
Prel NTCIMMIF...IVYSLFGNKK...GNKSRHRSFNLQFAISMLDASVIAFICGLTRTTGNOSEFVLOLSPINNMFCFLLLYRHYHLYNVLGAIIV

```

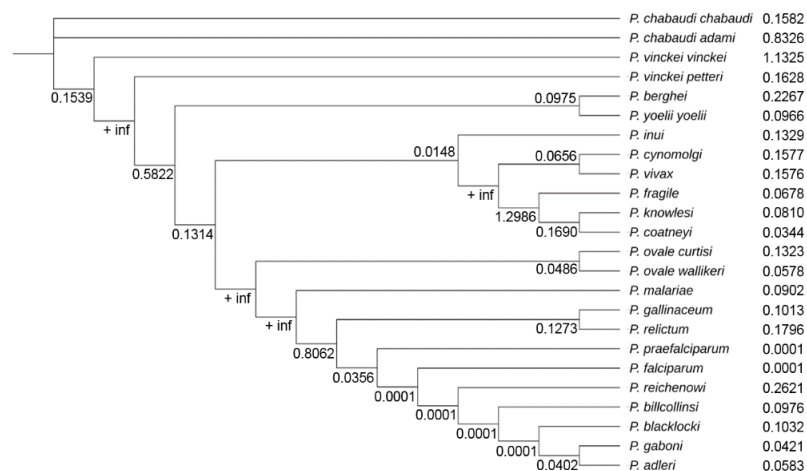
```

200      210      220      230      240      250      260      270      280
Pfal VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Prai VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pviv VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pinu VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pkno VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pfra VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pber VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pcyn VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pvin VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pcha VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pcho VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Ppet VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pcoa VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pgab VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pcur VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pwal VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pada VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Ppra VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pbla VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pbil VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Padi VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pmal VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Pgal VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC
Prel VTIADVEMKLSFENQRENSIFNLVLIISALIEVGSNMTREIVEKKYKINILRLNAMVSFFOLFSTCLILEVYTLPELQKQLHPYNEITNTNKGFFAC

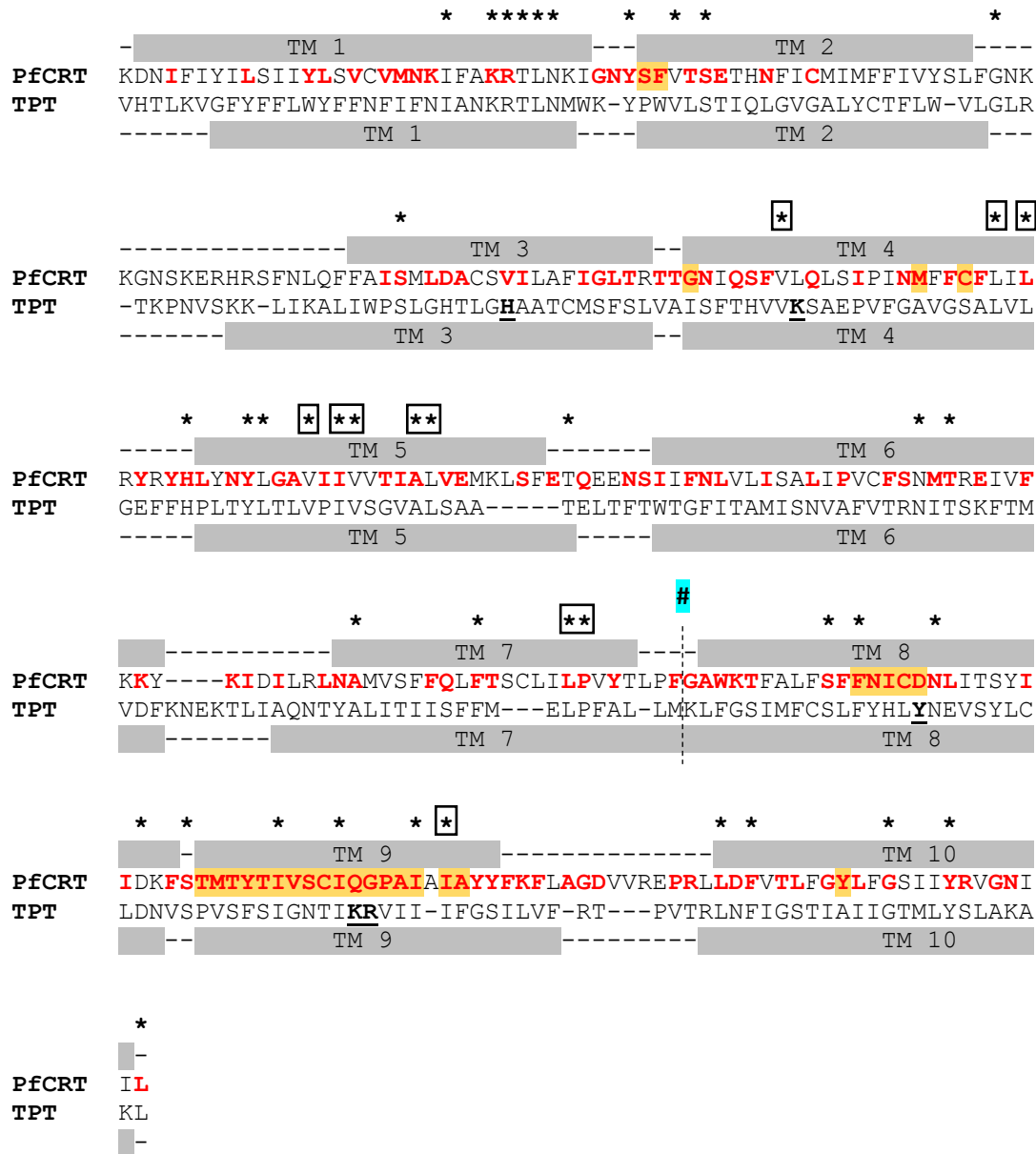
```



Supplementary Figure S1. CRT multiple amino acid sequence alignment. The CRT multiple sequence alignment was performed using Mafft version 7. Strictly conserved positions are written in white letters and shaded in black. Highly conserved positions are boxed with black lines and written in black, bolded letters. The level of conservation is based on BLOSUM62 scoring matrix. A four-letter code is used for the name of species (full species names are provided in Supplementary Table S1). The regions of the alignment with highly divergent positions, as well as positions containing gaps in at least 30% of all sequences, were manually removed in some subsequent analyses (red boxes). The multiple sequence alignment was displayed with the ESPrpt 3.0 server.



Supplementary Figure S2. Branch-specific ω values estimated with PAML under the free-ratio (FR) model for the *crt* gene. ω estimates (dN/dS) with “+ inf” values for which the estimation of d_s was equal to 0.0 (i.e. no silent substitutions observed) are designated as extremely low. Branch lengths were ignored for ease of representation. There was suggestive evidence of positive selection or neutral evolution for *P. vinckeii vinckeii* and during the speciation event of *P. fragilis* with *P. knowlesi* and *P. coatneyi*. All the remaining lineages evolved under strong purifying selection.



Supplementary Figure S3. Sequence alignment of PfCRT with TPT template used to predict the PfCRT tertiary structure at the occluded state.

X: strictly conserved CRT amino acid position among *Plasmodium* species.

X or **X**: position belonging to the top 10% conserved sites of PfCRT (FuncPatch).

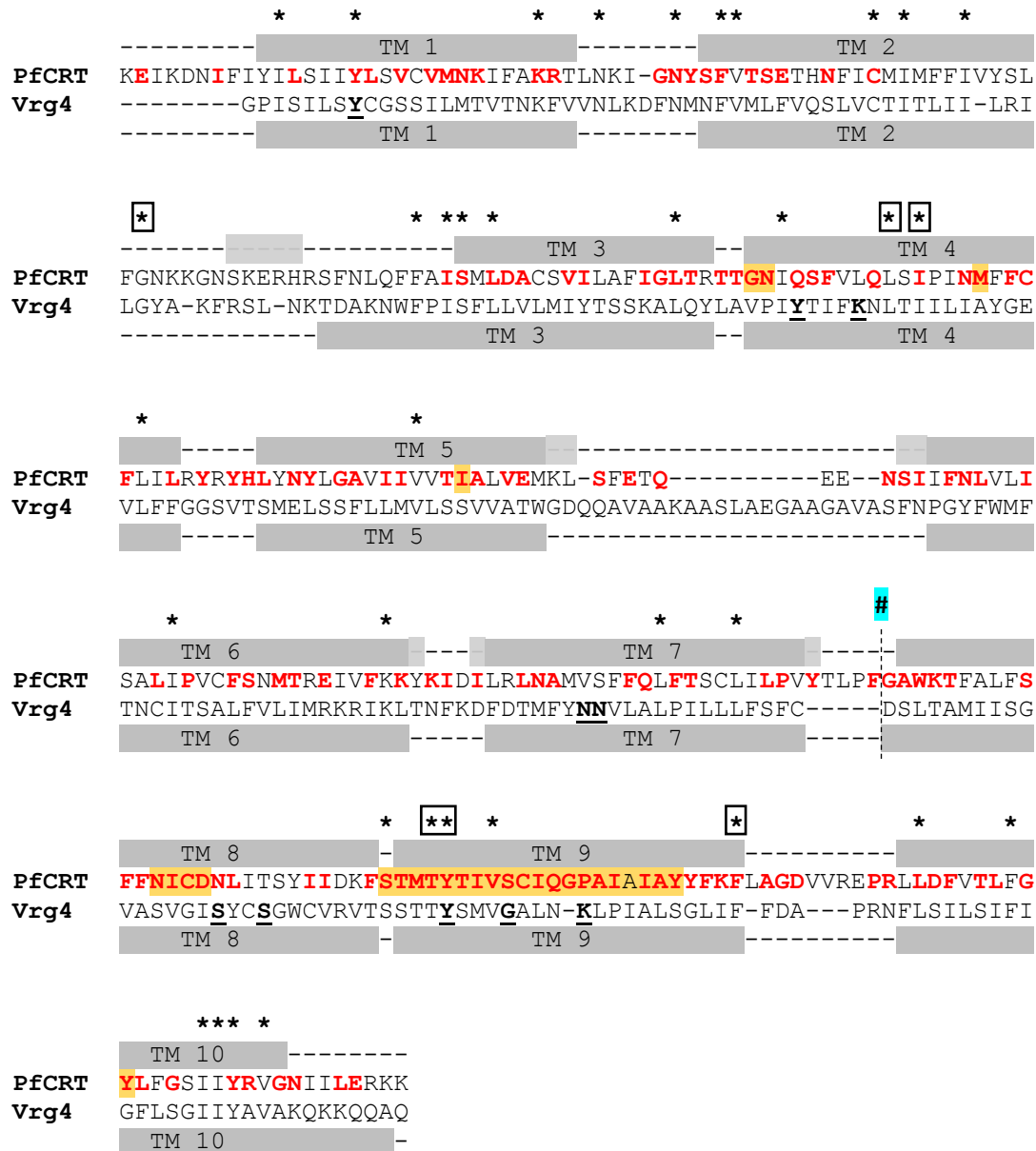
X: amino acid position from template sequence involved in substrate binding.

*****: identical amino acid between PfCRT and template sequence.

X: amino acid predicted to be important to align based on profile-profile alignment.

--: secondary structure is indicated above PfCRT and below PTP template. It is based on a homology model for PfCRT and an experimentally-determined tertiary structure for PTP (PDB file 5y79, chain A). Grey shading and dash signs indicate helices and loops respectively.

#: refers to the positions 269 to 313, corresponding to the long vacuolar loop of PfCRT, that were removed because of insufficient amino acid coverage in DMT templates.



Supplementary Figure S4. Sequence alignment of PfCRT with Vrg4 template used to predict the PfCRT tertiary structure at the inward-facing state.

X: strictly conserved CRT amino acid position among *Plasmodium* species.

X or **X**: position belonging to the top 10% most conserved sites of PfCRT (FuncPatch).

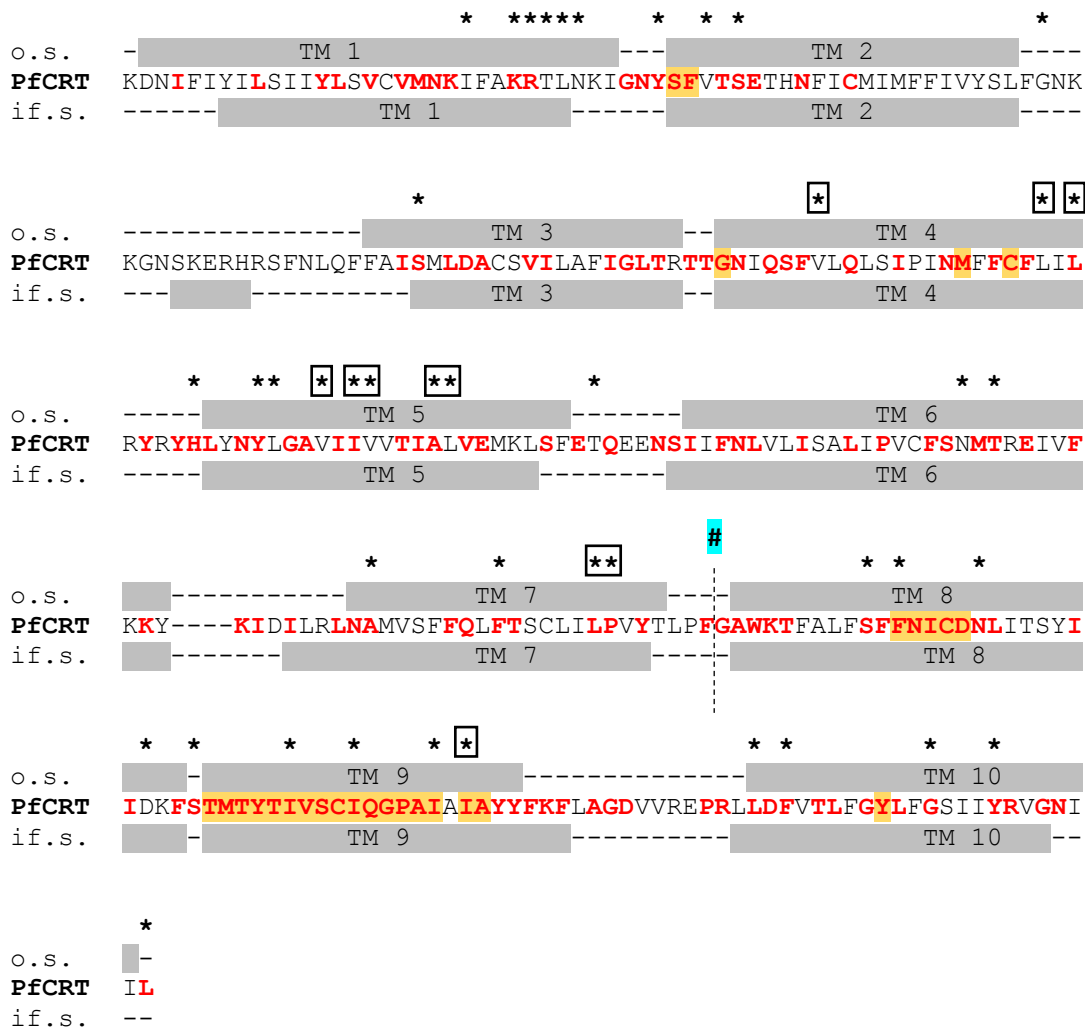
X: amino acid position from template sequence involved in substrate binding.

*****: identical amino acid between PfCRT and template sequences.

[]: amino acid predicted to be important to align based on profile-profile alignment.

--: secondary structure is indicated above PfCRT and below Vrg4 template. It is based on a homology model for PfCRT and an experimentally-determined tertiary structure for Vrg4 (PDB file 5oge, chain A). Grey shading and dash signs indicate helices and loops respectively.

#: refers to the positions 269 to 313, corresponding to the long vacuolar loop of PfCRT, that were removed because of insufficient amino acid coverage in DMT templates.



Supplementary Figure S5. Sequence alignment of the occluded and inward-facing states of PfCRT model.

o.s.: occluded state.

if.s.: inward-facing state.

X: strictly conserved CRT amino acid position among *Plasmodium* species.

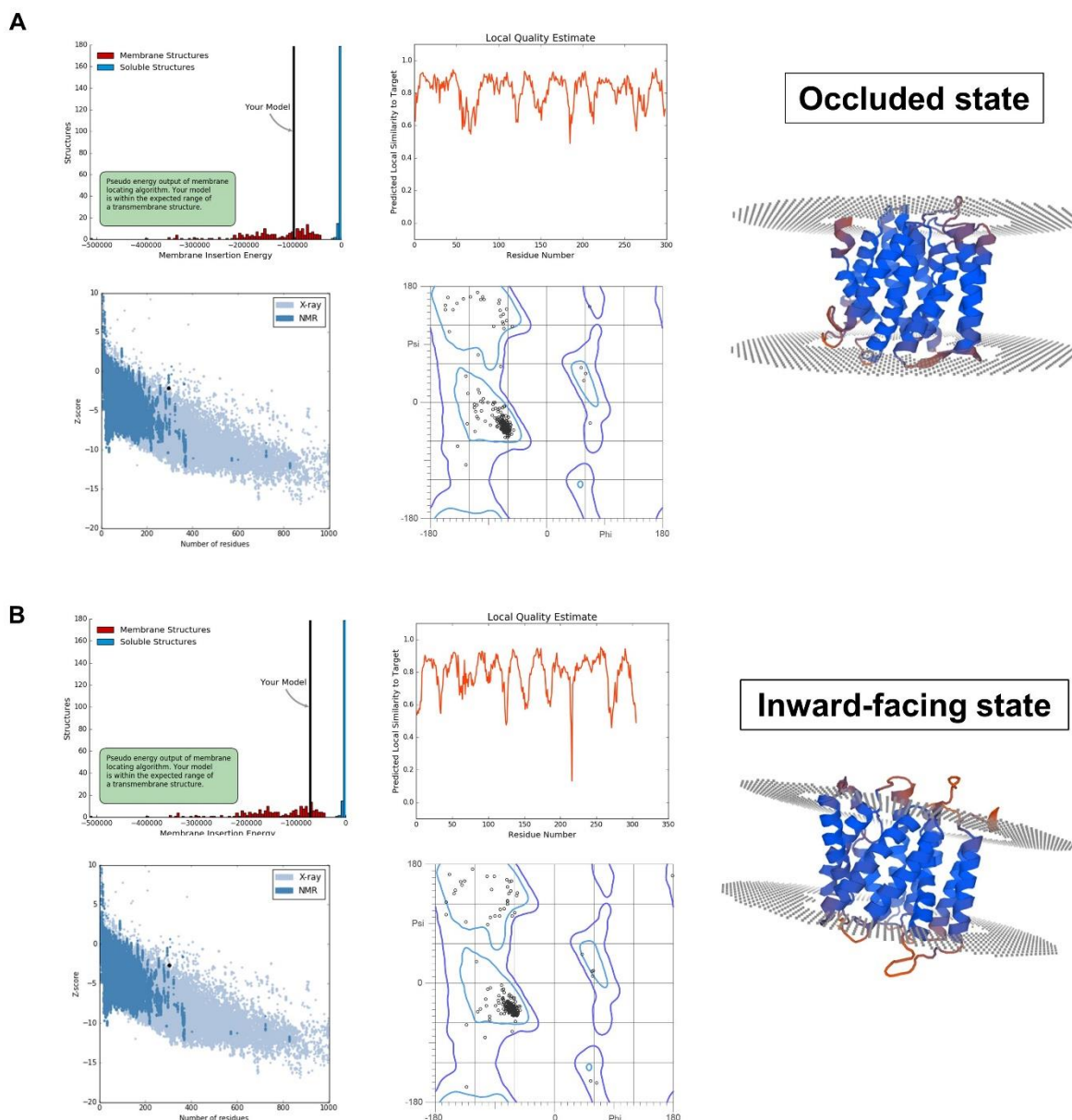
X or X: position belonging to the top 10% most conserved sites of PfCRT (FuncPatch).

*: identical amino acid between PfCRT and template sequences.

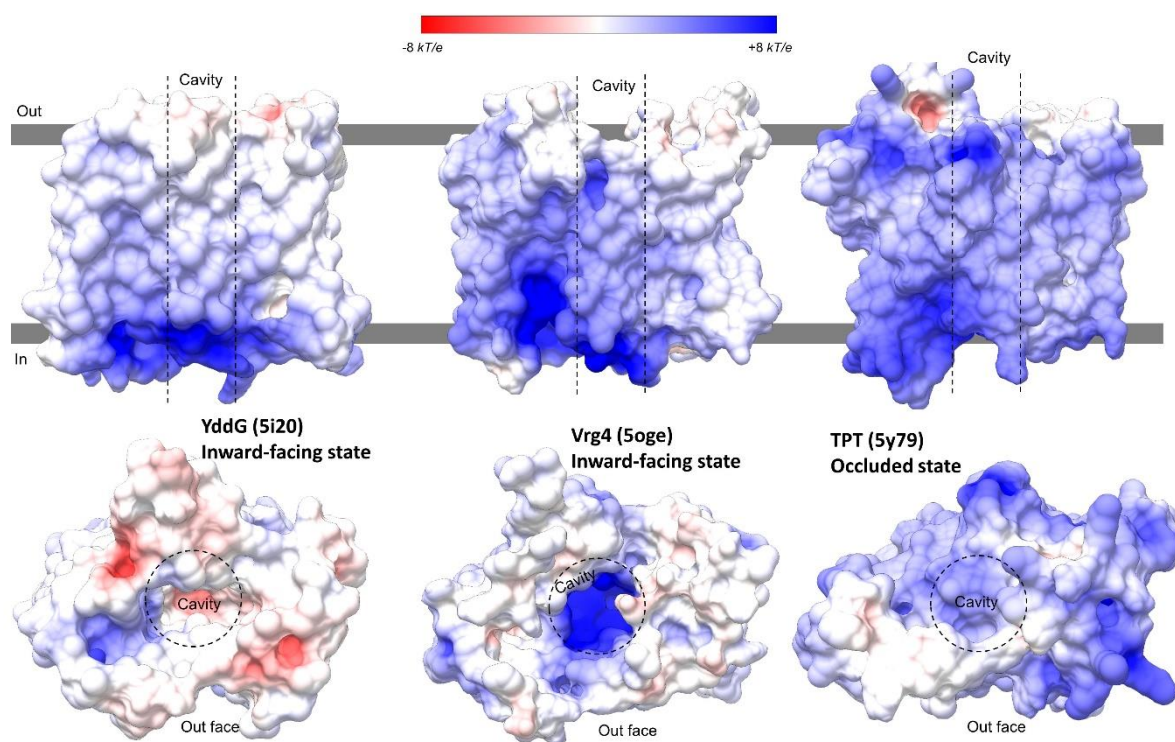
[]: amino acid predicted to be important to align based on profile(CRT)-profile(DMT) alignment.

--: secondary structure of the occluded and inward-facing states are indicated above and below PfCRT sequence respectively. Grey shading and dash signs indicate helices and loops respectively.

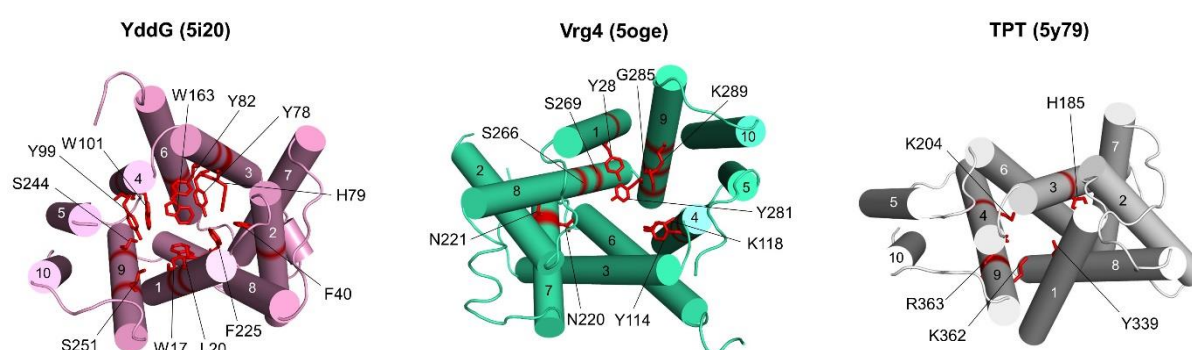
#: refers to the positions 269 to 313, corresponding to the long vacuolar loop of PfCRT, that were removed because of insufficient amino acid coverage in DMT templates.



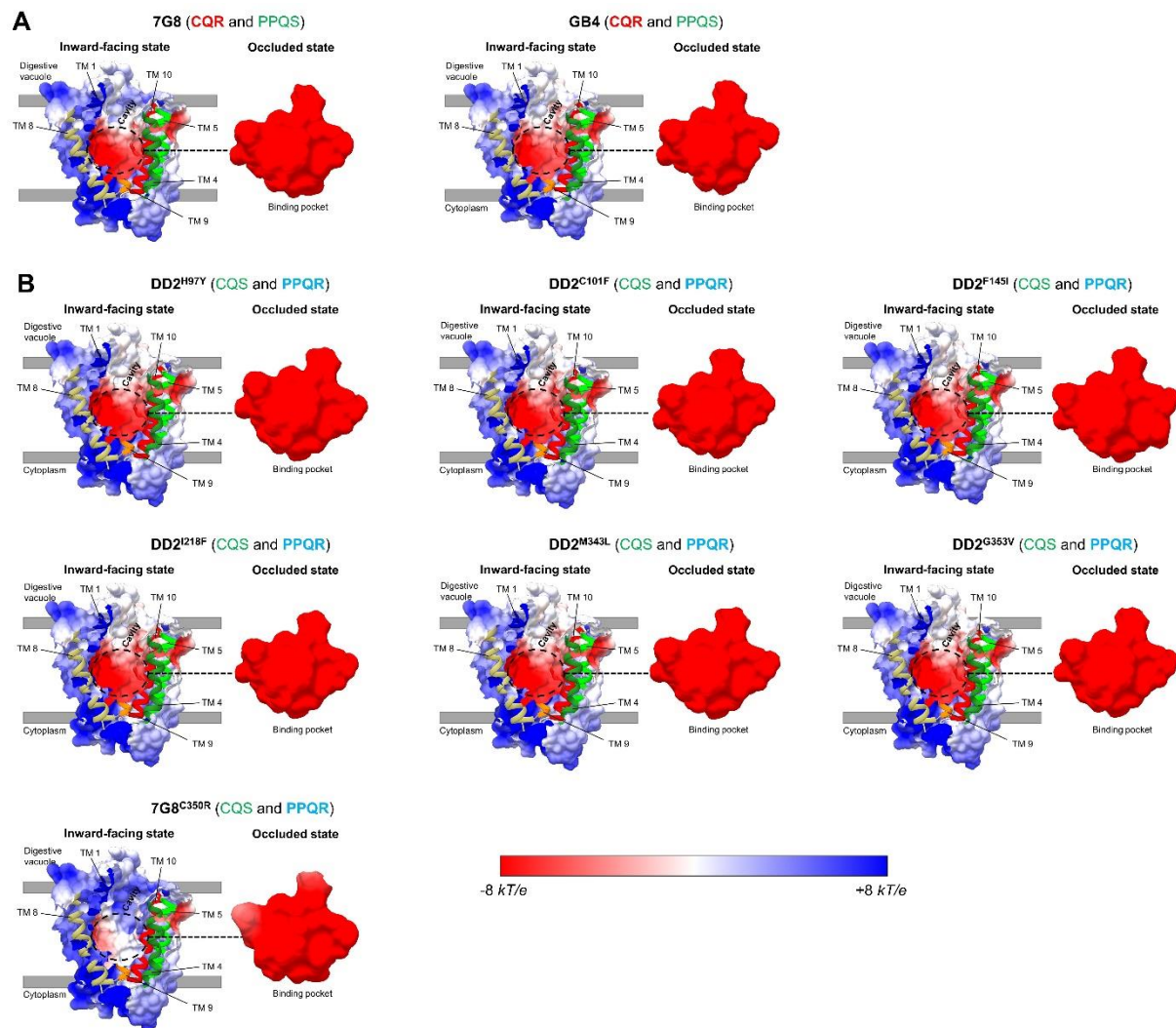
Supplementary Figure S6. Validation of the refined, energy-minimized PfCRT models (inward-facing and occluded states) obtained by homology modeling. General and transmembrane-specialized 3D quality metrics of the PfCRT models in an occluded (A) and an inward-facing (B) state. *Upper plots:* QMEANBrane analyses showed that the pseudo-energy of PfCRT models (*left histograms*) fell within the range of expected pseudo-energies in transmembrane structures. The graph (*right*) exhibited the per-residue local quality estimate for PfCRT models. As shown in the 3D structures, residues located in TMs had high quality estimates (most of them ranging from 0.8 to 1, colored in blue in the 3D structure), while the loops were associated with lower scores (colored in orange in 3D structure). *Bottom plots:* ProSA II showed the distribution of Z-scores calculated for all experimentally determined protein chains available in the Protein Data Bank (*left plot*). PfCRT models (represented as a black circle in the plots) were located in the range of structured resolved with NMR method. The Ramachandran plot from MolProbity (*right plot*) showed that most residues of PfCRT models had favorable ψ and ϕ angles, without outliers.



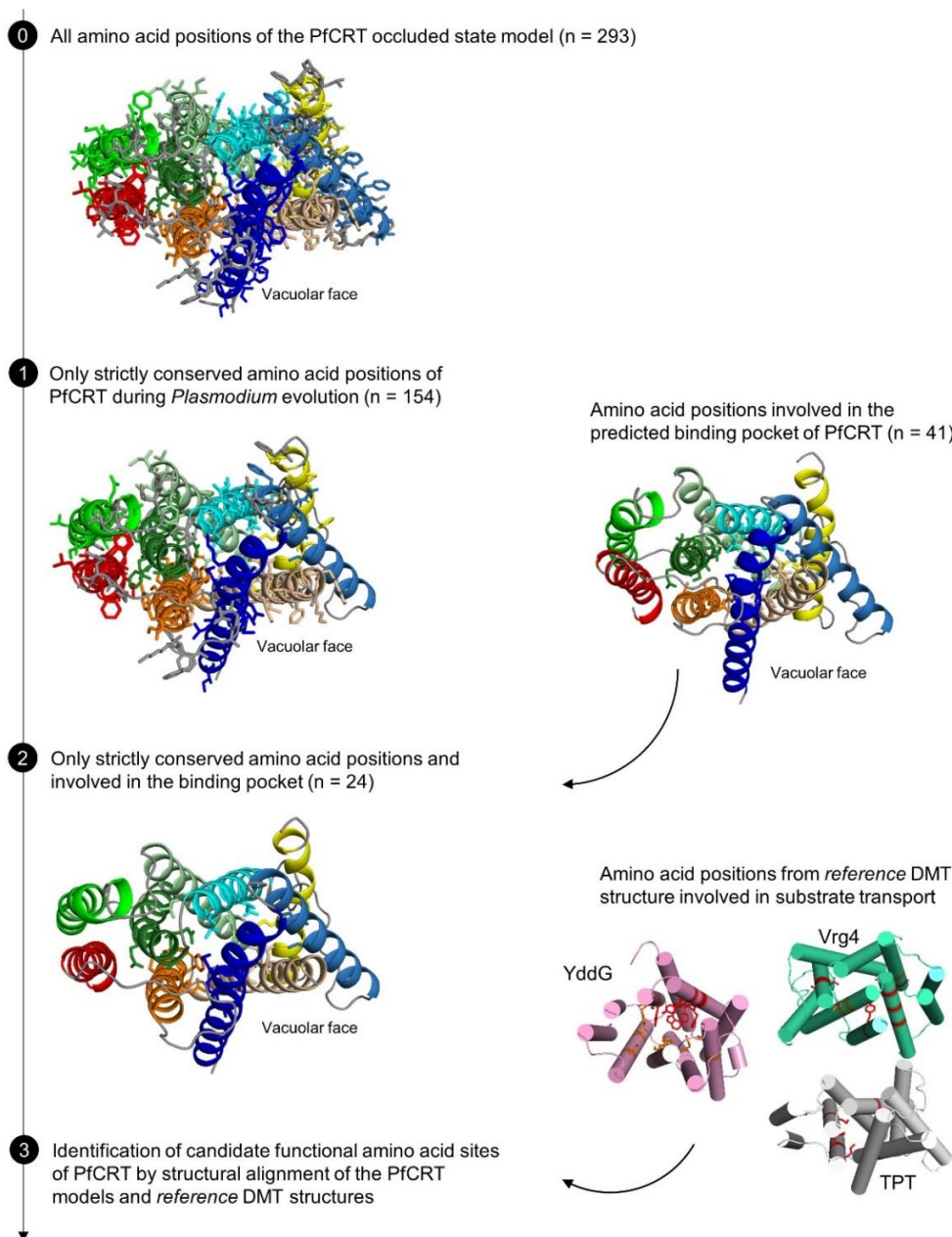
Supplementary Figure S7. Electrostatic surface potential of the different transporter proteins investigated in this study. Electrostatic surface potential of the three DMT proteins (including the two templates used for modeling PfCRT), estimated with the APBS method. Electrostatic potential values are in units of kT/e at 298 K, on a scale of $-8 kT/e$ (red) to $+8 kT/e$ (blue). The white color indicates a neutral potential. The transporters are shown as surface from the side (*top* structures) and from the extracellular face that corresponds to the vacuolar face in PfCRT (*bottom* structures). The name of the proteins and the corresponding PDB identifiers (within brackets) are indicated.



Supplementary Figure S8. Amino acid sites of the reference transporter proteins investigated in the study shown experimentally to bind physiological substrate(s). The key amino acid sites binding with the substrate(s) of each DMT transporter are shown in red color as sticks. The overall structures are shown as cartoons from the extracellular face corresponding to the PfCRT vacuolar face. The name of the proteins and corresponding PDB identifiers (within brackets) are provided above each structure.



Supplementary Figure S9. Electrostatic surface potential highlighting the cavity and binding pocket of PfCRT in a CQR or PPQR context. The electrostatic surface potential is displayed for the CQR haplotypes 7G8 and GB4 (A) and for seven other PPQR mutant structures (B). For each structure are indicated: the TMs – colored according to figure 2A –, the phenotype (chloroquine resistance, CQR; chloroquine sensitivity, CQS; piperazine resistance, PPQR; piperazine sensitivity, PPQS). The structure is shown from the side to highlight the cavity in the inward-facing state model (*left* structures), in addition to a zoom on the binding pocket from the occluded state model (*right* structures).



Supplementary Figure S10. Pipeline describing the filtering of PfCRT amino acid sites for the identification of PfCRT candidate functional sites. The PfCRT occluded state model is colored according to TMs. YddG, Vrg4 and TPT proteins are colored in pink, blue-green and grey, respectively. Functional amino acid sites of these proteins are shown in stick and colored in red. Starting with the PfCRT occluded state model covering 293 amino acid positions of the PfCRT sequence, only the strictly conserved amino acid sites at the *Plasmodium* scale were selected (n = 154). Next, the amino acid sites involved in the architecture of the predicted binding pocket (based on our model) were selected (n = 41). Among these sites, those that are strictly conserved during *Plasmodium* evolution were kept (n = 24). Finally, the PfCRT models (occluded and inward-facing states) were structurally aligned with the different *reference* DMT structures used in this study (*i.e.* YddG, Vrg4 and TPT proteins). We then focused on the subset of PfCRT amino acid sites that structurally aligned with experimentally-tested and validated amino acid sites of the reference DMT proteins.

Supplementary Table S1. List of CRT orthologous sequences used in this study.

Species	Abbreviation ^a	Description	Accession number	Source	Length (AA)
<i>Plasmodium falciparum</i>	Pfal	chloroquine resistance transporter	PF3D7_0709000	PlasmoDB	424
<i>Plasmodium reichenowi</i>	Prei	chloroquine resistance transporter	PRCDC_0707200	PlasmoDB	424
<i>Plasmodium vivax</i>	Pviv	chloroquine resistance transporter	PVX_087980	PlasmoDB	424
<i>Plasmodium inui</i>	Pinu	—	—	—	424
<i>Plasmodium knowlesi</i>	Pkno	chloroquine resistance transporter	PKNH_0107600	PlasmoDB	424
<i>Plasmodium fragile</i>	Pfra	hypothetical protein	AK88_01012	PlasmoDB	425
<i>Plasmodium berghei</i>	Pber	chloroquine resistance transporter	PBANKA_1219500	PlasmoDB	425
<i>Plasmodium cynomolgi</i>	Pcyn	chloroquine resistance transporter	PCYB_011610	PlasmoDB	424
<i>Plasmodium vinckei vinckei</i>	Pvin	chloroquine resistance transporter	YYE_01536	PlasmoDB	425
<i>Plasmodium chabaudi chabaudi</i>	Pcha	chloroquine resistance transporter	PCHAS_1220200	PlasmoDB	424
<i>Plasmodium yoelii yoelii</i>	Pyoe	chloroquine resistance transporter	PY17X_1222700	PlasmoDB	424
<i>Plasmodium vinckei petteri</i>	Ppet	chloroquine resistance transporter	YYG_02380	PlasmoDB	425
<i>Plasmodium coatneyi</i>	Pcoa	chloroquine resistance transporter	PCOAH_00000850	PlasmoDB	424
<i>Plasmodium gaboni</i>	Pgab	chloroquine resistance transporter	PGSY75_0709000	PlasmoDB	424
<i>Plasmodium malariae</i>	Pmal	chloroquine resistance transporter	PmUG01_01020700	PlasmoDB	423
<i>Plasmodium ovale curtisi</i>	Pcur	chloroquine resistance transporter	PocGH01_01016900	PlasmoDB	423
<i>Plasmodium ovale wallikeri</i>	Pwal	chloroquine resistance transporter	POVWA2_000940	PlasmoDB	423
<i>Plasmodium gallinaceum</i>	Pgal	chloroquine resistance transporter	PGAL8A_00049200	PlasmoDB	424
<i>Plasmodium relictum</i>	Prel	chloroquine resistance transporter	PRELSG_0107100	PlasmoDB	398
<i>Plasmodium chabaudi adami</i>	Pada	chloroquine resistance transporter	SCN62053.1	GenBank	424
<i>Plasmodium adleri</i>	Padl	chloroquine resistance transporter	PADL01_0707800	PlasmoDB	424
<i>Plasmodium billcollinsi</i>	Pbill	chloroquine resistance transporter	PBILCG01_0708300	PlasmoDB	423
<i>Plasmodium blacklocki</i>	Pbla	chloroquine resistance transporter	PBLACG01_0708200	PlasmoDB	425
<i>Plasmodium praefalciparum</i>	Ppra	chloroquine resistance transporter	PPRFG01_0708600	PlasmoDB	424

Note – For *P. ovale* spp. and *P. inui* sequences, we manually performed exon skipping to reconstruct the full cDNA and amino acid sequences. Only the bird-infecting *P. relictum* species seemed incomplete with an absence of about twenty amino acid positions in the C-terminal region of CRT. ^a Four-letter species code used in supplementary figure 1.

Supplementary Table S2. Results of PAML analyses for the *crt* gene.

Model	ℓ	Parameters	κ	Positively Selected Codons
free-ratio	-10,436.32	ω can vary between branches	2.31	Not allowed
M0	-10,566.42	$\omega = 0.119$	2.27	Not allowed
M1a	-10,383.20	$\omega_0 = 0.077, p_0 = 0.817$ $\omega_1 = 1.000, p_1 = 0.183$	2.72	Not allowed
M2a	-10,383.20	$\omega_0 = 0.077, p_0 = 0.817$ $\omega_1 = 1.000, p_1 = 0.056$ $\omega_2 = 1.000, p_2 = 0.127$	2.72	None detected
M3 ($k = 4$)	-10,254.80	$\omega_0 = 0.000, p_0 = 0.299$ $\omega_1 = 0.054, p_1 = 0.330$ $\omega_2 = 0.241, p_2 = 0.266$ $\omega_3 = 0.603, p_3 = 0.105$	2.41	None detected
M7	-10,255.75	$\beta(0.357, 2.047)$ $\omega \leq 1.000, p = 1.000$	2.41	Not allowed
M8	-10,255.75	$\beta(0.357, 2.047)$ $\omega \leq 1.000, p_0 = 1.000$	2.41	None detected

Note – ω , d_N/d_S ratio; p_n , proportion of codon sites in the site class ω_n ; ℓ , log-likelihood value; κ , transition/transversion rate ratio; $\beta(p, q)$, shape parameters for the β -distribution of ω . Candidate codon sites for positive selection are identified using the Bayes empirical Bayes (BEB) inference for models M2a and M8, and the Naïve empirical Bayes (NEB) inference for model M3.

Supplementary Table S3. List of the PfCRT amino acid sites involved in the architecture of the binding pocket (based on the predicted PfCRT occluded state model).

PfCRT 3 letters	PfCRT 1 letter	PfCRT position	2D Location ^a	Drug resistance-related position ^b
TYR	Y	68	TM 1	-
CYS	C	72	TM 1	Chloroquine
ASN	N	75	TM 1	Chloroquine
LYS	K	76	TM 1	Chloroquine
ALA	A	79	TM 1	-
HIS	H	97	TM 2	Piperaquine
ASN	N	98	TM 2	-
CYS	C	101	TM 2	Piperaquine
ASP	D	137	TM 3	-
SER	S	140	TM 3	-
VAL	V	141	TM 3	-
ALA	A	144	TM 3	-
PHE	F	145	TM 3	Piperaquine
LEU	L	148	TM 3	-
GLN	Q	156	TM 4	-
SER	S	157	TM 4	-
VAL	V	159	TM 4	-
LEU	L	160	TM 4	-
GLN	Q	161	TM 4	-
SER	S	163	TM 4	-
LEU	L	217	TM 6	-
ALA	A	220	TM 6	Chloroquine
LEU	L	221	TM 6	-
VAL	V	224	TM 6	-
CYS	C	225	TM 6	-
ASN	N	228	TM 6	-
ARG	R	231	TM 6	-
GLN	Q	253	TM 7	-
LEU	L	254	TM 7	-
PHE	F	322	TM 8	-
ASN	N	326	TM 8	Chloroquine
ASP	D	329	TM 8	-
ASN	N	330	TM 8	-
THR	T	333	TM 8	-
TYR	Y	345	TM 9	-
THR	T	346	TM 9	-
SER	S	349	TM 9	-
CYS	C	350	TM 9	Piperaquine
GLN	Q	352	TM 9	-
GLY	G	353	TM 9	Piperaquine
ILE	I	356	TM 9	Chloroquine

^a 2D location indicates on which TM the amino acid site is located.

^b “-” indicates no known relationship with a drug resistance phenotype.

Supplementary Table S4. PAML results for different structure-based partitionings of the *crt* gene.

Model	ℓ	r_2	κ	ω
Cytoplasm (n = 146 codon sites) / vacuole (n = 196 codon sites) partitioning based on the PfCRT inward-facing state model				
A (homogeneous model)	-7,914.23	—	2.34	0.091
B (different r_s)	-7,906.94	0.80	2.34	0.090
C (different r_s and π_s)	-7,922.95	0.70	2.35	0.092
D (different r_s , κ , and ω)	-7,897.08	0.89	$\kappa_1 = 2.58$; $\kappa_2 = 2.19$	$\omega_1 = 0.121$; $\omega_2 = 0.070$
E (different r_s , κ , ω and π_s)	-7,913.51	0.78	$\kappa_1 = 2.59$; $\kappa_2 = 2.25$	$\omega_1 = 0.124$; $\omega_2 = 0.072$
TMD-connecting loops (n = 116 codon sites) / TMDs (n = 226 codon sites) partitioning based on the PfCRT inward-facing state model				
A (homogeneous model)	-7,914.23	—	2.34	0.091
B (different r_s)	-7,885.07	0.62	2.32	0.090
C (different r_s and π_s)	-7,882.14	0.85	2.27	0.090
D (different r_s , κ , and ω)	-7,878.79	0.70	$\kappa_1 = 2.31$; $\kappa_2 = 2.35$	$\omega_1 = 0.117$; $\omega_2 = 0.076$
E (different r_s , κ , ω and π_s)	-7,876.75	0.94	$\kappa_1 = 2.08$; $\kappa_2 = 2.40$	$\omega_1 = 0.111$; $\omega_2 = 0.078$

Note – ℓ , log-likelihood value; r_s , nucleotide substitution rates (r_2 is the nucleotide substitution rate of the second partition, while r_1 is fixed at 1); π_s , codon frequencies; κ , transition/transversion rate ratio; ω , d_N/d_S rate ratio; n, number of codon sites; TM,

transmembrane helix; non-TM, loop connecting two TMs. For the cytoplasm / vacuole partitioning, r_1 corresponds to the nucleotide substitution rate of codon sites located at the cytoplasmic-half side of PfCRT; for the TM / non-TM loops partitioning, r_1 is the nucleotide substitution rate for codon sites located within non-TM loops. For the two series of partitioning tests shown here, codon sites were partitioned with the inward-facing state model of PfCRT and we discarded the poorly conserved N- and C-terminal regions, both located in the cytoplasm (this alteration was conservative). Similar results were obtained when codon sites were partitioned with the occluded state model than with the inward-facing state model (data not shown).

Supplementary Table S5. List of the 10% most conserved PfCRT amino acid sites based on the spatially correlated site-specific substitution rates (using the inward facing state model).

PfCRT 3 letters	PfCRT 1 letter	PfCRT position	2D Location ^a	Involved in ^b	
				Cavity	Binding pocket
GLY	G	153	TM 4	Yes	No
ASN	N	154	TM 4	Yes	No
MET	M	168	TM 4	No	No
ILE	I	194	TM 5	Yes	No
ASN	N	326	TM 8	Yes	Yes
ILE	I	327	TM 8	No	No
CYS	C	328	TM 8	No	No
ASP	D	329	TM 8	Yes	Yes
SER	S	341	TM 8-TM 9 loop	No	No
THR	T	342	TM 9	Yes	No
MET	M	343	TM 9	No	No
THR	T	344	TM 9	No	No
TYR	Y	345	TM 9	Yes	Yes
THR	T	346	TM 9	Yes	Yes
ILE	I	347	TM 9	No	No
VAL	V	348	TM 9	No	No
SER	S	349	TM 9	Yes	Yes
CYS	C	350	TM 9	Yes	Yes
ILE	I	351	TM 9	No	No
GLN	Q	352	TM 9	Yes	Yes
GLY	G	353	TM 9	Yes	Yes
PRO	P	354	TM 9	No	No
ALA	A	355	TM 9	No	No
ILE	I	356	TM 9	Yes	Yes
ALA	A	357	TM 9	Yes	No
ILE	I	358	TM 9	No	No
ALA	A	359	TM 9	No	No
TYR	Y	360	TM 9	Yes	No
TYR	Y	384	TM 10	Yes	No

Note – The 10% most conserved sites of PfCRT inward-facing state at the *Plasmodium* scale were identified using the FuncPatch server. The results were very similar when we used the PfCRT occluded state model.

^a 2D location indicates on which TM or loop the amino acid site is located.

^b For each amino acid site, we indicate whether it is located in the cavity (inward-facing state) and/or the binding pocket (occluded state) of PfCRT.

Supplementary Table S6. Strictly conserved PfCRT amino acid sites involved in the architecture of the binding pocket (based on the PfCRT occluded state model).

PfCRT 3 letters	PfCRT 1 letter	PfCRT position	2D Location ^a	Drug resistance-related position ^b
TYR	Y	68	TM 1	-
ASN	N	75	TM 1	Chloroquine
LYS	K	76	TM 1	Chloroquine
ASN	N	98	TM 2	-
CYS	C	101	TM 2	Piperaquine
ASP	D	137	TM 3	-
VAL	V	141	TM 3	-
LEU	L	148	TM 3	-
GLN	Q	156	TM 4	-
SER	S	157	TM 4	-
GLN	Q	161	TM 4	-
LEU	L	221	TM 6	-
ARG	R	231	TM 6	-
GLN	Q	253	TM 7	-
ASN	N	326	TM 8	Chloroquine
ASP	D	329	TM 8	-
ASN	N	330	TM 8	-
TYR	Y	345	TM 9	-
THR	T	346	TM 9	-
SER	S	349	TM 9	-
CYS	C	350	TM 9	Piperaquine
GLN	Q	352	TM 9	-
GLY	G	353	TM 9	Piperaquine
ILE	I	356	TM 9	Chloroquine

Note – The identification of the strictly conserved amino acid sites of PfCRT was based on a set of 24 different *Plasmodium* species (supplementary table 1).

^a 2D location indicates on which TM the amino acid site is located.

^b We indicate whether the position is related to a drug resistance mutation; “–” indicates no known relationship with a drug resistance phenotype. Of note, the position 76, associated with the key chloroquine resistance mutation K76T, belongs to the strictly conserved amino acid sites and is involved in the architecture of the binding pocket.

Supplementary Table S7. DMT template proteins used to generate PfCRT models by homology modeling.

	Vrg4 (PDB: 5oge) – inward-facing state	TPT protein (PDB: 5y79) – occluded state
DMT subfamily ^a	2.A.7.13 GDP-Mannose:GMP amntiporter (GMA)	2.A.7.9 Triose-phosphate Transporter (TPT)
Protein length	337	329
Sequence covering in structure	21-330	20-315
# TMD	10	10
% seq. id. with PfCRT ^b	12.7	14.5
Organism	<i>Starkeya novella</i>	<i>Galdieria sulphuraria</i>
Chain used for PfCRT modeling	E	A
Structure resolution	3.22	2.2
Method	X-ray diffraction	X-ray diffraction
Released structure date	2017-11-22	2017-10-04
Reference	Parker & Newstead, Nature, 2017	Lee <i>et al.</i> , Nat. Plants, 2017

^a Code and subfamily are based on the Transport Classification Database (TCDB).

^b The % sequence identity of the given transporter with PfCRT was provided by the Phyre2 server.

