



# Identification of a protein unique to the genus *Plasmodium* that contains a WD40 repeat domain and extensive low-complexity sequence

Gladys T. Cortés<sup>1</sup> · Martha Margarita Gonzalez Beltran<sup>2</sup> · Claudio J. Gómez-Alegría<sup>3</sup> · Mark F. Wiser<sup>4</sup>

Received: 1 March 2021 / Accepted: 11 May 2021 / Published online: 18 June 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

Proteins containing WD40 domains play important roles in the formation of multiprotein complexes. Little is known about WD40 proteins in the malaria parasite. This report contains the initial description of a WD40 protein that is unique to the genus *Plasmodium* and possibly closely related genera. The N-terminal portion of this protein consists of seven WD40 repeats that are highly conserved in all *Plasmodium* species. Following the N-terminal region is a central region that is conserved within the major *Plasmodium* clades, such as parasites of great apes, monkeys, rodents, and birds, but partially conserved across all *Plasmodium* species. This central region contains extensive low-complexity sequence and is predicted to have a disordered structure. Proteins with disordered structure generally function in molecular interactions. The C-terminal region is semi-conserved across all *Plasmodium* species and has no notable features. This WD40 repeat protein likely functions in some aspect of parasite biology that is unique to *Plasmodium* and this uniqueness makes the protein a possible target for therapeutic intervention.

**Keywords** WD40 domain · *Plasmodium* · Low-complexity sequence

---

Handling Editor: Julia Walochnik

## Highlights

- Few proteins containing WD40 domains from the malaria parasite have been characterized
- *PlasWD40-1* is a WD40 domain-containing protein only found in *Plasmodium* species
- The WD40 domain of *PlasWD40-1* is highly conserved across all *Plasmodium* species
- A large portion of *PlasWD40-1* is composed of low-complexity sequence
- *PlasWD40-1* likely plays a role related to the unique biology of *Plasmodium*

---

✉ Mark F. Wiser  
wiser@tulane.edu

<sup>1</sup> Departamento de Salud Pública, Facultad de Medicina, Grupo Biología Celular, Universidad Nacional de Colombia, Bogotá, Colombia

<sup>2</sup> Ex alumna de la Maestría en Ciencias–Bioquímica, Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Grupo UNIMOL, Bogotá, Colombia

## Introduction

Characterizing genes and proteins of the malaria parasite will lead to a better understanding of the biology of the parasite and provide information about possible therapeutic targets. Monoclonal antibodies (Mabs) were previously generated against proteins possibly located in an ER-like compartment (Cortés et al. 2003). In this previous study, we demonstrated that one of these antibodies, called Mab7, recognized a 68 kDa antigen in *P. falciparum* and other *Plasmodium* species. Fluorescence microscopy localized the 68 kDa antigen to a subcellular compartment that overlapped with the ER but exhibited

<sup>3</sup> Departamento de Farmacia, Facultad de Ciencias, Universidad Nacional de Colombia, Grupo UNIMOL, Bogotá, Colombia

<sup>4</sup> Department of Tropical Medicine, Tulane University School of Public Health and Tropical Medicine, 1440 Canal Street, Suite 2301, New Orleans, LA 70112-2824, USA

some distinctions from the ER. Subsequently, this compartment was named the *Plasmodium* export compartment (Wiser 2007). In addition, by mass spectrometry analysis of the immunopurified protein, we recently demonstrated that the antigen of Mab7 corresponds to the *P. falciparum* HSP70-2 protein (Cortés et al. 2020). The present work arises as an attempt to confirm the protein identity via the screening of recombinant DNA expression libraries with Mab7. In this study, we show that Mab7 identifies recombinant DNA clones that encode a larger-than-expected protein which is conserved in all *Plasmodium* species. This conserved gene proved not to be of the 68 kDa protein, but nonetheless, the cloned gene did prove to be interesting in that it is not found in any other species except *Plasmodium* and likely the closely related *Hepaticystis*. Another notable feature of this protein is a WD40 domain that is quite conserved across all *Plasmodium* species.

WD40 domains are found in a large number of eukaryotic proteins and participate in a wide range of functions including cell division, RNA transcription and processing, vesicular trafficking, signal transduction, cytoskeleton assembly, and the ubiquitin proteasome system (Stirnemann et al. 2010). In addition, WD40 domain-containing proteins play important roles in the assembly of dynamic multi-protein complexes and often function in cellular interaction networks. Despite the important roles WD40 proteins play in cellular physiology, little work has been done on WD40-domain proteins from the malaria parasite. Chahar et al. (2015) carried out an extensive in silico analysis of WD40-domain proteins in *Plasmodium falciparum* and identified 80 putative WD40-domain proteins. Orthologs from other organisms were identified for most of the *P. falciparum* WD40-domain proteins. However, fifteen of the WD40-domain proteins appear unique to the genus *Plasmodium*. One of these unique WD40-domain proteins has been demonstrated to be involved in the formation of adhesion protein complexes during the blood stage of the parasite (von Bohl et al. 2015). Similarly, not much is known about WD40-domain proteins in other apicomplexans. *Toxoplasma* coronin, an actin-binding protein, has been crystallized (Kallio and Kursula 2014) and a myosin from *Gregarina polymorpha* contains a WD40 domain (Heintzelman and Mateer 2008). Both of these WD40 proteins may be involved in gliding motility, which is a process unique to the Apicomplexa. Clearly, more work on the WD40 proteins of Apicomplexa and the malaria parasite is needed. In that WD40 domain-containing proteins are poorly characterized in *Plasmodium* and Apicomplexa and WD40 proteins are potential drug targets (Song et al. 2017), this unique *Plasmodium* protein was further characterized.

## Methods

### Gene cloning and sequencing

Recombinant cDNA expression libraries were obtained from Nirbhay Kumar and have been previously described (Zhang et al. 1999). Briefly, these libraries were constructed from *P. falciparum* (NF54) enriched for gametocytes. RNA was converted into cDNA and ligated into lambda ZAPII bacteriophage. For screening, the recombinant bacteriophage was mixed with *Escherichia coli* XL-1 Blue in the top agarose, plated on bacteriological agar plates, and incubated at 37 °C to form bacterial lawns. When plaques first became visible, the agar/agarose plates were overlaid with polyvinylidene difluoride (PVDF) membranes saturated with 10 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) and incubated for an additional 14–16 h at 37 °C. A total of  $5.3 \times 10^5$  plaques were screened.

The PDVF membranes from the plaque lifts were subjected to immunoblotting using monoclonal antibody-7 (Mab7). Mab7 was generated using a membrane fraction from *P. falciparum*-infected erythrocytes and recognizes a 68 kDa parasite protein (Cortés et al. 2003). Positive plaques were detected with an enzyme-conjugated anti-mouse IgG and developed as previously described (Cortés et al. 2003). Positive clones were purified by plaque purification and the insert DNA fragments were isolated by in vivo excision using M13 helper phage (Short et al. 1988). The purified recombinant phagemids were sent to the Instituto de Genética de la Universidad Nacional de Colombia for DNA sequencing with fluorescent dye terminators.

### Sequence analysis

The sequence of the cloned insert cDNA was used to carry out BLAST searches (<https://blast.ncbi.nlm.nih.gov>) of the available sequences at National Center for Biotechnology Information. The most recent search was carried out in January 2021. Orthologs of this gene were identified through additional BLAST searches and through PlasmoDB (Aurrecochea et al. 2009). Multiple sequence alignments were carried out using CLUSTALW 2.1 (Larkin et al. 2007) and pairwise alignments using the SIM alignment tool (Huang and Miller 1991). The original alignment containing all species was refined using alignments carried out within a single parasite clade as guides. PROSITE (Sigrist et al. 2013) and SPARCLE (Marchler-Bauer et al. 2017) were used to identify domains and functional sites within the sequences. Low-complexity

sequence was analyzed according to Wootton (1994). Transmembrane helices were predicted using TMHMM (Sonnhammer et al. 1998). WD40 domains were determined using the WD40-repeat protein structure predictor (WDSP) tool (Ma et al. 2019).

### Protein structure modeling

Protein modeling was carried out with Protein Homology/analogy Recognition Engine version 2.0 (Phyre<sup>2</sup>) (Kelley et al. 2015), SWISS-MODEL (Waterhouse et al. 2018), SPARKS-X (Yang et al. 2011), or QUARK (Xu and Zhang 2012). Structures were visualized with EzMol (Reynolds et al. 2018) or PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC).

### Epitope prediction

The sequence of the antigen-binding region of Mab7 was determined so that epitope prediction studies could be carried out. Frozen hybridoma cells were supplied to GenScript ProBio (Piscataway, NJ) and they carried out RNA extraction, and amplification and sequencing of the regions corresponding to the variable domains of the light and heavy chains. The translated sequence of the variable domain was then used to generate a three-dimensional (3D) model of the antigen-binding site of Mab7 with Lymphocyte Receptor Automated Modelling (LYRA) (Klausen et al. 2015). This Mab7 model was used to search for potential epitopes in 3-dimensional models of HSP70-2 from *Plasmodium* and models of the translation of the cloned DNA fragment using EpiPred (Krawczyk et al. 2014).

Three-dimensional models of HSP70-2 from *P. falciparum* (GenBank accession number CAD51861.1) and *P. berghei* (Swiss-Protein accession number Q25642) were generated with Phyre<sup>2</sup> (Kelley et al. 2015) using human BIP (PDB ID 5E84) as the template. SPARKS-X (Yang et al. 2011) was used to search for proteins with possible structural similarities to the protein sequence of the cloned DNA fragment. A model generated using the template from the outer surface protein A of *Borrelia burgdorferi* (PDB ID 2FKJ) was used to predict the epitope of Mab7. Other templates generating similar epitopes (data not shown) include endoglycosidase S from *Streptococcus pyogenes* (PDB ID 6EN3), vacuolar ATPase of *Saccharomyces cerevisiae* (PDB ID 6O7V), and spike protein of the type VI secretion system from *Francisella tularensis* (PDB ID 6U9E).

### Gene identifiers

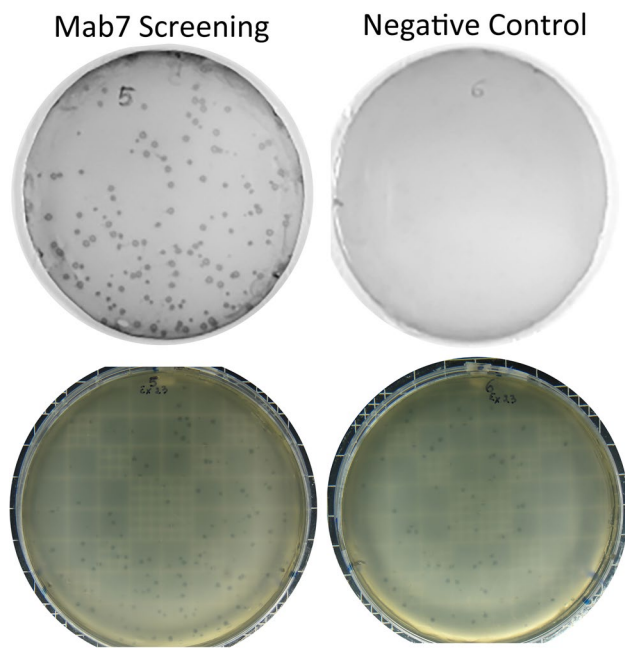
The gene ID number of the protein described in this study is PF3D7\_1228800 (*P. falciparum* 3D7). Orthologs of this gene include AK88\_02334 (*P. fragile* strain Nilgiri),

C922\_03474 (*P. inui* San Antonio 1), PADL01\_1229400 (*P. adleri* G01), PBANKA\_1443500 (*P. berghei* ANKA), PBILCG01\_1228800 (*P. billcollinsi* G01), PBLACG01\_1228200 (*P. blacklocki* G01), PCHAS\_1445700 (*P. chabaudi chabaudi*), PCOAH\_00053960 (*P. coatneyi* Hackeri), PCYB\_145450 (*P. cynomolgi* strain B), PcyM\_1450400 (*P. cynomolgi* strain M), PF7G8\_120034300 (*P. falciparum* 7G8), PfCD01\_120033700 (*P. falciparum* CD01), PfDd2\_120033800 (*P. falciparum* Dd2), PfGA01\_120033500 (*P. falciparum* GA01), PfGB4\_120034200 (*P. falciparum* GB4), PfGN01\_120034800 (*P. falciparum* GN01), PfHB3\_120033600 (*P. falciparum* HB3), PfIT\_120033900 (*P. falciparum* IT), PfKE01\_120033800 (*P. falciparum* KE01), PfML01\_120034200 (*P. falciparum* ML01), PfSD01\_120033700 (*P. falciparum* SD01), PfsN01\_120034400 (*P. falciparum* SN01), PFTG01\_120033700 (*P. falciparum* TG01), PGABG01\_1227600 (*P. gaboni* strain G01), PGAL8A\_00129900 (*P. gallinaceum* 8A), PGSY75\_1228800 (*P. gaboni* strain SY75), PKNH\_1448200 (*P. knowlesi* strain H), PKNOH\_S140266300 (*P. knowlesi* strain Malayan Strain Pk1 A), PmUG01\_14063100 (*P. malariae* UG01), PocGH01\_14054500 (*P. ovale curtisi* GH01), PPRFG01\_1226000 (*P. praefalciparum* strain G01), PRELSG\_1444400 (*P. relictum* SGS1-like), PRG01\_1232100 (*P. reichenowi* G01), PVL\_000196200 (*P. vivax-like* PvI01), PVP01\_1447000 (*P. vivax* P01), PVX\_124070 (*P. vivax* Sal-1), PY01983 (*P. yoelii yoelii* 17XNL), PY17X\_1446000 (*P. yoelii yoelii* 17X), PYYM\_1447600 (*P. yoelii yoelii* YM), YYE\_00892 (*P. vinckei vinckei* strain vinckei), and YYG\_00464 (*P. vinckei petteri* strain CR).

Gene identifiers for proteins possibly interacting with the WD40 repeat protein include PF3D7\_0416400 (histone acetyltransferase), PF3D7\_0612100 (eIF3 subunit L), PF3D7\_0702100 (PHISTb family protein), PF3D7\_0802100 (APIAP2), PF3D7\_0823300 (histone acetyl transferase GCN5), PF3D7\_0935500 (*Plasmodium* export protein, PFI1715W), PF3D7\_1246200 (actin 1), and PF3D7\_1360700 (SUMO E3 ligase).

## Results and discussion

In an attempt to identify the 68 kDa antigen of *P. falciparum*, a cDNA expression library was screened with Mab7. From a total of  $5.3 \times 10^5$  plaques analyzed, two positive clones were ultimately identified. Figure 1 shows reactivity of Mab7 with the plaques from one of these clones, and this reactivity is clearly a strong positive signal. No positive plaques were detected in the negative control that did not contain Mab7. The phagemids from both clones (named clones 5



**Fig. 1** Expression of a recombinant protein recognized by Mab7. Clone 6 derived from the screening of the recombinant cDNA library was subjected to three rounds of plaque purification. Shown are plaque lifts (upper) analyzed by immunoblotting with Mab7 (left) or a no antibody control (right) from the third plaque purification. Below the plaque lifts are pictures of the agar/agarose plates from which the plaque lifts were generated. All plaques on the agarose plate exhibited a strong positive signal in the lifts analyzed by Mab7. No plaques exhibited a positive signal in the no antibody negative control. This demonstrates that Mab7 recognizes a protein expressed by the cloned DNA fragment (GenBank Accession Number MK618679) and the recombinant DNA clone is purified

and 6) were released and sequenced, and both clones had identical 404 bp inserts. This sequence was submitted to GenBank and has the accession number MK618679. The translation of the insert DNA is on pages 32–33 of the supplemental data. Searching GenBank with the MK618679 sequence identified a *P. falciparum* gene on chromosome 12 designated as putative WD40 repeat-containing protein (gene ID = PF3D7\_1228800). This gene contains a single exon consisting of 9630 nucleotides that encodes a protein of 3209 amino acids with a predicted molecular mass of 380 kDa. This protein was previously identified as part of a survey of WD40-repeat proteins in *Plasmodium*, but not

further characterized (Chahar et al. 2015). We propose to call this protein *Plasmodium* WD40 repeat-containing protein-1 (*PlasWD40-1*) until further information about this protein is available.

*PlasWD40-1* is substantially larger than the 68 kDa protein recognized by Mab7. Furthermore, through mass spectrometry analysis, the 68 kDa protein has been recently identified as the ER-resident heat-shock protein-70 from *P. falciparum* referred to as *PfHSP70-2* (Cortés et al. 2020). Thus, the detection of clones 5 and 6 by Mab7 is likely due to cross-reaction between the protein expressed by the cloned cDNA fragment and Mab7. In other words, Mab7 not only recognizes *PfHSP70-2*, but also recognizes a protein expressed by the cloned DNA fragment (MK618679).

To investigate a possible common epitope shared by both proteins, the amino acid sequences of these two proteins were compared. A pair-wise alignment identified a 38 amino acid fragment with 29% sequence identity (Fig. 2). The sequence of Mab7 was then used to generate a 3-dimensional model of the antigen-binding site of Mab7. This antigen-binding site model was then used to search for conformational epitopes in 3-dimensional models of *PfHSP70-2* and MK618679. A 12 amino acid long epitope was identified in *PfHSP70-2* (Fig. 2, highlighted in orange). As expected, since Mab7 reacts equally well with HSP70-2 from *P. falciparum* and *P. berghei* (Cortés et al. 2003), this same epitope was also identified in the *P. berghei* ortholog *PbHSP70-2* (data not shown). Two potential epitopes of six and seven amino acids in length (highlighted in yellow), which overlap with the predicted *PfHSP70-2* epitope, were identified in MK618679. Similar epitopes were identified in three other models generated from the sequence of the cloned insert fragment (data not shown). The overlapping potential epitopes identified in these modeling studies may explain the reactivity of Mab7 with both *PfHSP70-2* and the MK618679 insert.

### Expression of *PlasWD40-1*

The cloning of *PlasWD40-1* from a cDNA library prepared from parasite mRNA indicates that the gene is expressed in blood stage parasites. This is corroborated by several transcriptome studies. Messenger RNA levels of the gene peak during the late trophozoite stage (Llinás et al. 2006;

|                  |     |                     |                        |          |
|------------------|-----|---------------------|------------------------|----------|
| <i>PfHSP70-2</i> | 536 | KMINDAEKFADEDKNL    | REKVEAKNNLDNYIQSMKATVE | 575      |
| MK618679         | 2   | QIIDDTEHFNNVEEKEKEK | NKGSYFMKNYNVSKKLNNE    | 39       |
|                  |     | * * * *             | **                     | ** * * * |

**Fig. 2** Predicted epitopes of *PfHSP70-2* and cloned sequence. A pairwise alignment between *PfHSP70-2* (Gene ID PF3D7\_0917900) and the translation of the cloned DNA fragment (Genebank Accession MK618679) was carried out. Shown is the highest scoring over-

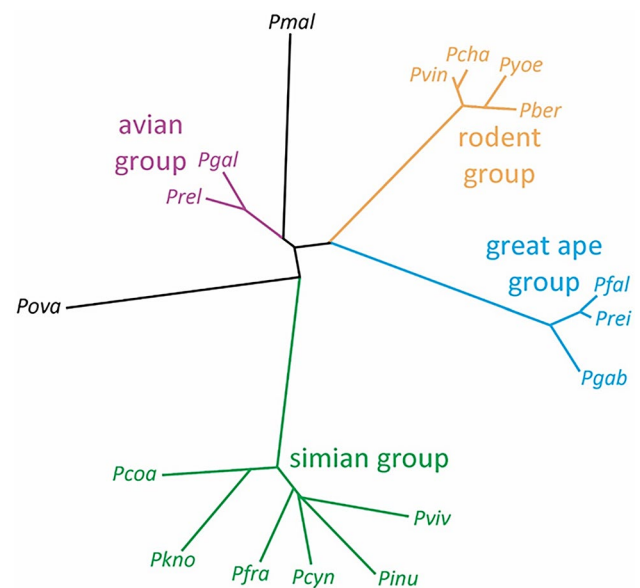
lap between the two sequences consisting of 38 residues with 29% identity (denoted with asterisks \*). Predicted epitopes recognized by Mab7 are highlighted in orange and yellow

Rovira-Graells et al. 2012). The gene is also expressed in gametocytes (López-Barragán et al. 2011) and sexually committed parasites (Pelle et al. 2015). Transcripts are also detected in the ookinete stage (López-Barragán et al. 2011), as well as oocyst and sporozoite stages (Zanghi et al. 2018). Similar expression profiles are observed in the ortholog from rodent malaria parasites (Otto et al. 2014), *P. knowlesi* (Lapp et al. 2015), and *P. vivax* (Zhu et al. 2016). During the blood stage, the transcript is also associated with polysomes (Bunnik et al. 2013) implying translation as well as transcription of the gene. Furthermore, proteomic analyses using mass spectrophotometry have detected peptides of PlasWD40-1 in blood stages as well as sporozoites (<https://plasmodb.org>). These results indicate that the gene is expressed in all life cycle stages of the malaria parasite. Insertional mutagenesis by the *piggyBAC* transposon suggests that the gene is dispensable during the in vitro blood stage culture (Zhang et al. 2018). However, the fact that orthologs of the gene have been found in all *Plasmodium* species does argue that the gene is essential for some aspect of the parasite's life cycle.

### Orthologues in other species

The amino acid sequence of PlasWD40-1 was used to search GenBank and orthologs were identified in all other *Plasmodium* species. However, no homologous proteins were identified in non-*Plasmodium* species, except for a partial sequence of 363 amino acids from *Hepaticystis* (Gene Bank Accession Number VWU53150.1). *Hepaticystis* is generally placed as a sister group to mammalian *Plasmodium* species (Martinsen et al. 2008; Galen et al. 2018) and thus is often considered to be within the *Plasmodium* clade. Complete genomes are available for many other Apicomplexa, such as *Toxoplasma*, *Cryptosporidium*, *Theileria*, and *Babesia*, and no homologues were found in these species. PlasWD40-1 belongs to ortholog group OG6\_158520 (Chen et al. 2006). Sequences from seven non-*Plasmodium* species are also included in OG6\_158520. However, this inclusion of non-*Plasmodium* species in the ortholog group is likely an artefact since there are only short regions of homology between these proteins and PlasWD40-1 (data not shown). Therefore, the PlasWD40 gene appears to be unique to the genus *Plasmodium*. However, the Haemosporida are not widely sampled in terms of genomics and it is possible that this gene may be present in other members of the Haemosporida.

The amino acid sequences from 17 *Plasmodium* species were aligned (Supplemental Data) and this alignment was used to generate an unrooted phylogenetic tree (Fig. 3). The phylogenetic analysis revealed four major branches designated as great ape, simian, avian, and rodent. *P. malariae* and *P. ovale* sequences did not form a strong association with any of these groups and branched near the base. This phylogenetic tree is in strong agreement with what is known



**Fig. 3** Phylogenetic relationships of the WD40 repeat-containing protein among *Plasmodium* species. Shown is an unrooted neighbor-joining tree derived from the Clustal 2.1 alignment of the WD40 repeat-containing protein from *P. falciparum* (Pfal, CZT99447.1), *P. reichenowi* (Prel, CDO65564.1), *P. gaboni* (Pgab, SOV16446.1), *P. vivax* (Pviv, EDL47649.1), *P. cynomolgi* (Pcyn, GAB69117.1), *P. fragile* (Pfra, KJP88059.1), *P. coatneyi* (Pcoa, ANQ11029.1), *P. inui* (Pinu, EUD66004.1), *P. knowlesi* (Pkno, CAQ42237.1), *P. berghei* (Pber, CDS52152.1), *P. chabaudi* (Pcha, CDR16910.1), *P. vinckei* (Pvin, EUD74513.1), *P. yoelii* (Pyoe, ETB61791.1), *P. ovale* (Pova, SCQ17100.1), *P. malariae* (Pmal, SCP03571.1), *P. relictum* (Prel, CRH02710.1), and *P. gallinaceum* (Pgal, CRG93592.1). Four major groups are observed: great ape (blue), simian (green), avian (purple), and rodent (orange). *P. malariae* and *P. ovale* do not show a strong affinity for any of the four groups

about the evolution of the genus *Plasmodium* and indeed others have shown trees with similar topologies and these four major groups (Loy et al. 2017). For example, *P. falciparum*, *P. reichenowi*, and *P. gaboni* have long been grouped together in the subgenus *Laverania* and their evolutionary relationships are well characterized (Sundararaman et al. 2016). Similarly, *P. vivax* forms a clade with malaria parasites found in macaques of Asia (Cornejo and Escalante 2006), the rodent parasites are a monophyletic group (Ramiro et al. 2012), and the avian parasites split from the mammalian species (Böhme et al. 2018). In addition, the PlasWD40-1 gene is located on chromosome 12 of the great ape species and chromosome 14 of the other species. The synteny of chromosome 12 from the great ape species and chromosome 14 from the other species has been previously noted (Tachibana et al. 2012).

Complete sequences of PlasWD40-1 are available from PlasmoDB and other databases from 13 isolates of *P. falciparum*. These *P. falciparum* sequences are essentially identical with the differences between isolates being found

almost exclusively in the number of tandem repeats in low-complexity regions (data not shown). One exception is a 33 amino acid insert found in some of the isolates. This insert corresponds to a putative intron in some isolates, but not other isolates. The 33 extra amino acids are part of a conserved region in the great ape group (Supplemental Data, pages 26–27), and therefore, we feel the designation of this putative intron is in error. The need for the intron to maintain the reading frame is due to a variation in the number of adenosines in a poly-A tract, which is a common sequencing error.

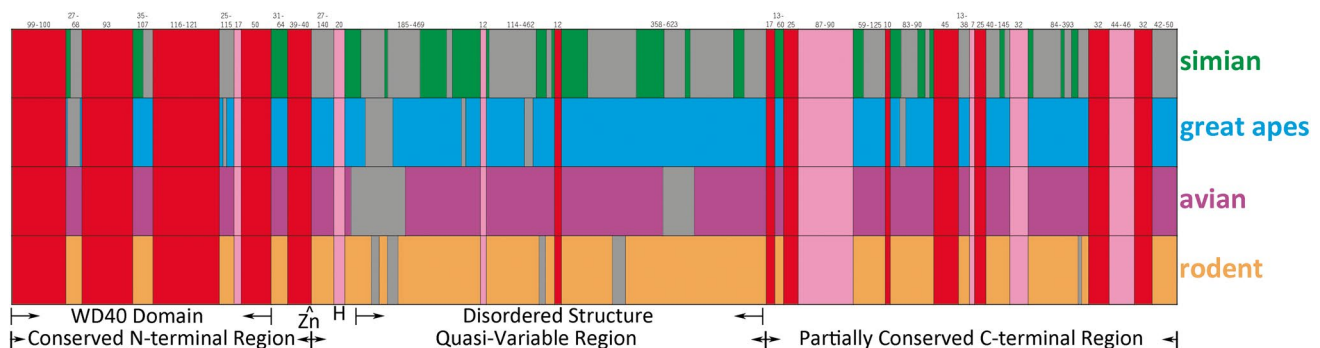
Sequence alignments of *Plas*WD40-1 from species within each of the four *Plasmodium* clades were also carried out. These clade-specific alignments were then used to refine the original alignment with all 17 species. The combined alignments revealed three major regions of *Plas*WD40-1 based upon sequence homology and structural features (Fig. 4). These three regions are (1) a conserved N-terminal region containing the WD40 domain, (2) a quasi-conserved central region containing a large amount of low-complexity sequence, and (3) a partially conserved C-terminal region.

### WD40 repeat domain

Seven WD40 repeats are found at the beginning of *Plas*WD40-1 (Supplemental Data, pages 2–7). These repeats exhibit the characteristic features of a WD40 repeat (Wang et al. 2013). For example, the repeats are approximately 40 amino acids and form four anti-parallel  $\beta$ -strands, and many of the repeats have the iconic tryptophan (W) and aspartate

(D) that is part of the namesake WD40. In addition, a conserved tetrad of aspartate (D), histidine (H), serine (S), and tryptophan (W) is present in most of the repeats. The hydrogen bonds formed by this DHSW tetrad stabilize the  $\beta$ -sheet. Furthermore, most WD40 domains are composed of seven WD40 repeats and are often described as folding into seven-bladed propeller-like structures with each blade composed of the four-stranded anti-parallel  $\beta$ -sheet (Jain and Pandey 2018). Each propeller blade is formed from three  $\beta$ -strands of the repeat and a  $\beta$ -strand of the adjacent WD40 repeat. The last blade is then formed from the remaining  $\beta$ -strand of the first repeat and the three remaining  $\beta$ -strands of the last repeat. This closes the ring and provides stability to the WD40 domain.

These WD40 repeats were readily identified with the WDSP tool (Wang et al. 2013). The average score for the WD40 domain from all 17 species was 72.1, which is in the high confidence range. The individual repeats range in size from 39 to 49 amino acids and exhibit WDSP scores greater than 70 for six of the seven repeats (Table 1). Note that scores greater than 62 are considered high confidence, scores between 44 and 62 are medium confidence, and scores less than 44 are low confidence (Ma et al. 2019). The individual repeats exhibit a high level of sequence homology between *Plasmodium* species with 51–78% sequence identity among the 17 species. If conservative substitutions are considered, the identical plus similar residues range from 83 to 98%. Within a clade, the sequence identity approaches 100% for the WD40 repeats. This amount of homology is rather remarkable since WD40 sequence motifs generally do not



**Fig. 4** Structure of the WD40 repeat-containing protein. A schematic representation of the WD40 repeat protein based on the alignment (Supplemental Data). Blocks in red are highly conserved in all species; blocks in pink are partially conserved in all species; blocks in green are conserved among species in the simian group; blocks in blue are conserved among species in the great ape group; blocks in purple are conserved among species in the avian group; blocks in orange are conserved among species in the rodent group; and blocks in gray are not conserved. The protein is divided into three major regions: a conserved N-terminal region, a central quasi-variable region, and a partially conserved C-terminal region. The conserved N-terminal region is composed primarily of a WD40 domain consist-

ing of seven WD40 repeats. A zinc-protease motif (Zn) is also found at the end of the N-terminal conserved region in most species. In the quasi-conserved region, there is little sequence conservation across all species. However, there is substantial sequence conservation among the species within a group. There are three small regions of conserved sequence in this quasi-variable region including a 20-residue region of primarily hydrophobic amino acids (H) and two 12 amino acid regions of either partially or highly conserved sequence. A large portion of this quasi-variable region is a large continuous block of predicted disordered structure in all species. The C-terminal region is interspersed with conserved, partially conserved, and conserved within the group sequences

**Table 1** WDSP scores and homology across the seven WD40 repeats

| Repeat | Size | All species (17) |          |           | Simian (6) |          | Great ape (3) |          | Rodent (4) |          | Avian (2) |          |
|--------|------|------------------|----------|-----------|------------|----------|---------------|----------|------------|----------|-----------|----------|
|        |      | Score            | Identity | + Similar | Score      | Identity | Score         | Identity | Score      | Identity | Score     | Identity |
| WD40-1 | 41   | 77.4             | 32 (78)  | 38 (93)   | 77.5       | 36 (88)  | 77.3          | 40 (98)  | 77.5       | 39 (95)  | 77.5      | 40 (98)  |
| WD40-2 | 42   | 80.2             | 28 (67)  | 35 (83)   | 79.3       | 39 (93)  | 79.0          | 41 (98)  | 79.3       | 40 (95)  | 87.5      | 41 (98)  |
| WD40-3 | 40   | 83.2             | 24 (60)  | 36 (90)   | 85.5       | 36 (90)  | 80.3          | 38 (95)  | 82.0       | 40 (100) | 85.0      | 40 (100) |
| WD40-4 | 39   | 77.9             | 23 (59)  | 35 (90)   | 80.8       | 33 (85)  | 76.7          | 36 (92)  | 77.8       | 35 (90)  | 75.0      | 37 (95)  |
| WD40-5 | 46   | 44.1             | 29 (63)  | 45 (98)   | 45.2       | 33 (72)  | 43.0          | 46 (100) | 43.0       | 44 (96)  | 42.5      | 46 (100) |
| WD40-6 | 49   | 74.1             | 25 (51)  | 46 (94)   | 72.8       | 39 (80)  | 73.7          | 47 (96)  | 63.0       | 45 (92)  | 76.0      | 49 (100) |
| WD40-7 | 41   | 70.8             | 28 (68)  | 38 (93)   | 67         | 37 (90)  | 71.7          | 40 (98)  | 73.8       | 40 (98)  | 74.0      | 41 (100) |
| Domain | -    | 72.1             | -        | -         | 73.3       | -        | 71.3          | -        | 71.0       | -        | 73.6      | -        |

Shown are the number of amino acids (size), the scores generated by the WDSP tool, the number of identical residues (%), and the number of identical + similar residues (%) for each of the seven WD40 repeats. The data for each of the four clades are also shown and the number in parentheses is the number of species included in the analysis. Average WDSP scores for the entire WD40 domain are also shown for all 17 species and within the four major groups

exhibit high levels of sequence identity (Stirnemann et al. 2010). Considering avian and mammalian *Plasmodium* species may have diverged between 10 million (Böhme et al. 2018) and 64 million (Silva et al. 2015) years ago, the high level of homology implies strong selective forces to maintain this level of sequence conservation. In addition, the partial sequence of *PlasWD40-1* from *Hepaticystis* exhibits a high level of sequence identity with WD40 repeats 3–6 (Supplemental Data, pages 3–6).

Divergent sequences are found between WD40 repeats two and three (Supplemental Data, pages 2–3), repeats four and five (Supplemental Data, page 4), and repeats six and seven (Supplemental Data, page 6). These inter-WD40-repeat sequences are largely conserved within the four clades, but not conserved across the four *Plasmodium* clades. Similarly, the *Hepaticystis* sequence between repeats four and five and repeats six and seven is distinct from the *Plasmodium* sequences. These inserted sequences between the WD40 repeats tend to be of low complexity and blocks of tandem repeats are sometimes observed. In addition, *P. vivax* has a glycine-rich insert of 16 residues between WD40 repeats five and six (Supplemental Data, page 5). The inserts between the WD40 repeats would be exposed on the bottom face of the WD40 domain (Wang et al. 2013). The significance of these inserts within the WD40 domain is not clear.

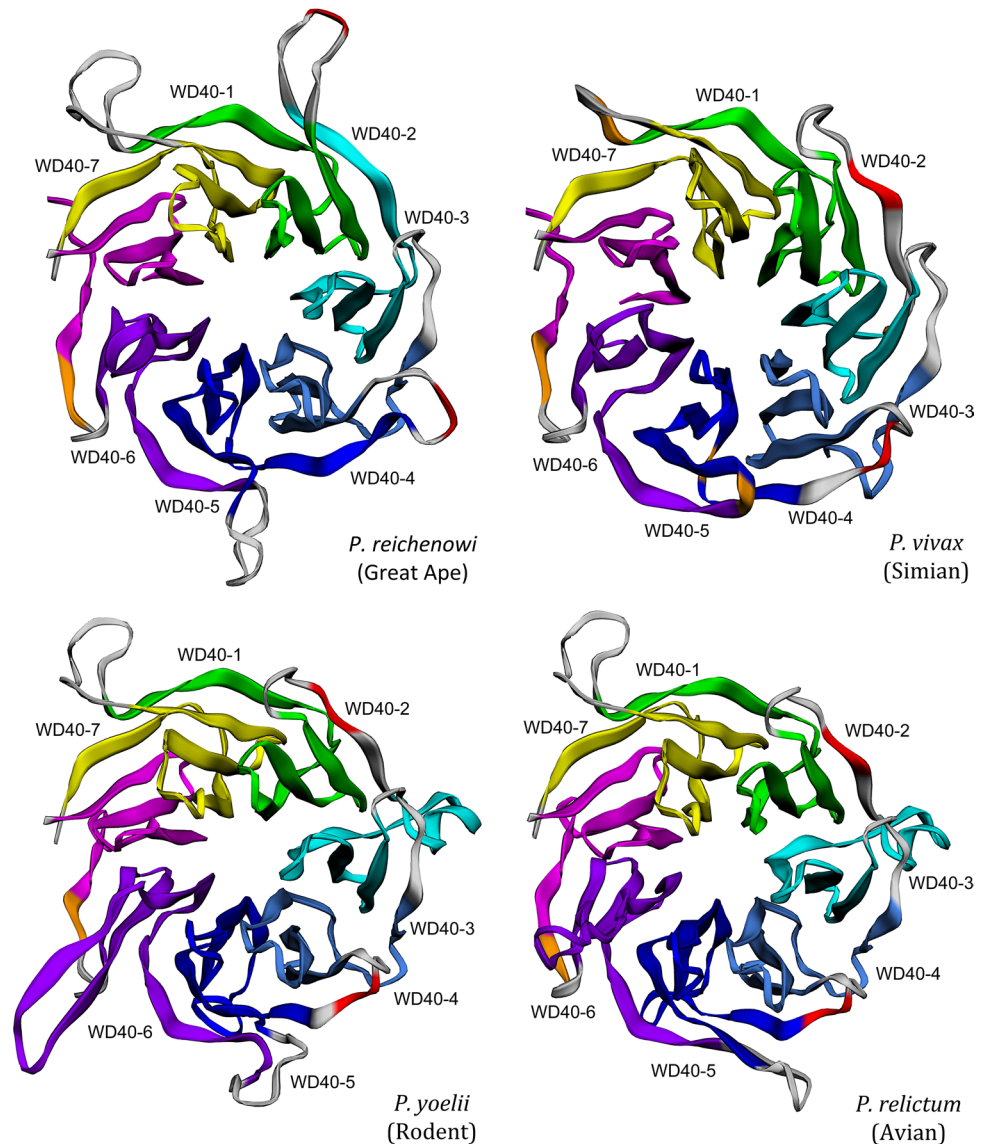
The WD40 domains of the *PlasWD40-1* sequences were subjected to 3D modeling with Phyre<sup>2</sup> and SWISS-MODEL. Both 3D modeling programs readily identified the WD40 domain of *PlasWD40-1* and a wide variety of WD40 domain-containing proteins were identified as templates. Many of the resulting models exhibited the canonical seven-bladed propeller (Jain and Pandey 2018) with each of the blades being formed from three  $\beta$ -strands of one WD40 repeat and the N-terminal  $\beta$ -strand of the adjacent WD40 repeat (Fig. 5). Therefore, the modeling further supports the

validity of this WD40 domain. Since the primary function of the WD40 domain is to bind other proteins, the high level of conservation of the WD40 domain in both sequence and structure implies that *PlasWD40-1* binds to other protein(s) that may also be conserved in the genus *Plasmodium*.

### Zinc protease motif

A search for known motifs and domains revealed a zinc protease motif at the end of the conserved N-terminal region in most species (Fig. 4). Most WD40-domain proteins also possess additional domains with other functional activities (Jain and Pandey 2018). This zinc protease motif is found within a highly conserved stretch of 39 amino acids that is separated from the WD40 domain by a small region of sequence that is conserved within clades (Supplemental Data, page 8). Zinc protease motifs are defined by a core element of HExxH and more stringently defined by [GSTALIVN]-{PCHR}-{KND}-H-E-[LIVMFYW]-{DEHRKP}-H-{EKPC}-[LIVMFYWGSPQ] where residues enclosed by [] are allowed and residues enclosed by {} are not allowed (Rawlings and Barrett 1995). The two histidines (H) bind the zinc atom and the glutamate (E) provides the nucleophile to form the active site. *P. coatneyi* has a conservative replacement of the active site glutamate (E) in the core element (HDxxH vs HExxH). This substitution of glutamate (E) for aspartate (D) would likely have minimal effect on the presumptive protease activity. However, members of the great ape clade (subgenus *Laverania*) have PDxxH for the core element, as well as an asparagine-rich insert immediately before the core element. Proline is never found within the zinc protease motif, possibly because it would disrupt the  $\alpha$ -helix that includes the HExxH motif. The proline and poly-asparagine insertion raises questions about the validity

**Fig. 5** Modeling of the WD40 domain from *Plas*WD40-1. Shown are representative examples of models generated with Phyre<sup>2</sup> using the receptor for activated C-kinase 1 (PDB ID 3DM0) as template from each of the four *Plasmodium* clades (as indicated). The colors of the seven WD40 repeats are yellow (WD40-1), green (WD40-2), turquoise (WD40-3), light blue (WD40-4), dark blue (WD40-5), purple (WD40-6), and maroon (WD40-7). Red indicates positions where the inter-WD40-repeat sequences were manually removed before modeling, and orange indicates positions where sequence was removed by the modeling program. Remaining sequence is in light gray



and universality of this putative zinc protease motif, especially in the species from the great ape clade.

Initial modeling of the region containing the zinc protease motif did not identify any known structures that could serve as templates. Therefore, an ab initio approach was used to model this region. Generally, the xxxHExxHxx motif is found within an  $\alpha$ -helix and the two histidine residues coordinate with the zinc atom (Rawlings and Barrett 1995). Some of the predicted models had the HExxH active site adjacent to an  $\alpha$ -helix and in a region with several  $\alpha$ -helices (Fig. 6), as is common in metalloproteases. Considering that the ab initio approach is based on probabilities of folding, and since the motif is surrounded by  $\alpha$ -helices, it is reasonable to assume that the Zn-protease motif is within an  $\alpha$ -helix. In addition, the critical histidine and glutamate residues are near the zinc atom suggesting the possibility of an intact active site. Furthermore, a groove that could serve

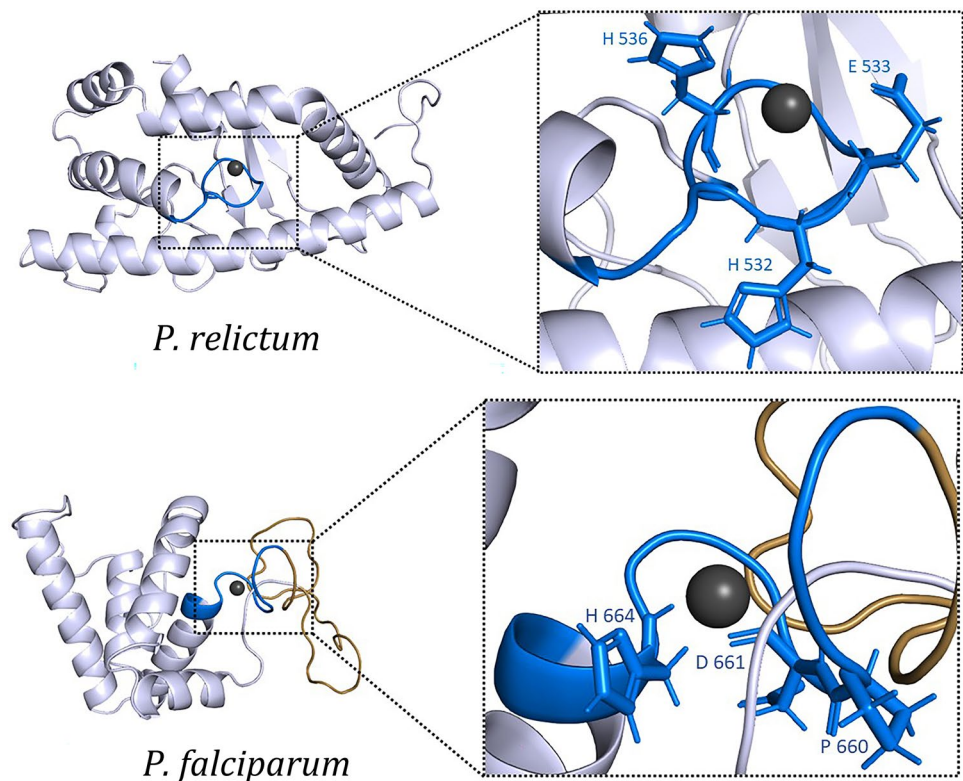
as a substrate-binding site is found near the putative active site. As expected, the *P. falciparum* model (Fig. 6) is rather disordered in the proximity of the putative active site due to the poly-asparagine region adjacent to the putative zinc-binding site and the substitution of PDxxH for the canonical HExxH. Experimental studies are needed to resolve the inconsistencies of the modeling data with the sequence data and to determine if *Plas*WD40-1 has zinc-binding or metalloprotease activity.

### Low-complexity disordered region

Following the highly conserved N-terminal region is a central region that exhibits little homology across all the species. However, within the four major clades, there is a fair amount of homology (note the amount of green, blue, purple, and orange in the central region of Fig. 4). Most of



**Fig. 6** Modeling of the putative zinc protease motif. Structures were built following the ab initio approach using the QUARK tool. The zinc atom (dark gray sphere) was inserted with PyMOL. Shown are the hypothetical structural models of the Zn-protease motif plus 100 residues on each side of the canonical xxxHExxHxx sequence (blue) from *PlasWD40-1* of *P. relictum* and *P. falciparum*. Dotted boxes show enlargements of the canonical HEXxH motif and the Zn atom. The dashed ellipse in the *P. relictum* model denotes a groove that could function as a substrate-binding site. The golden-brown regions in the *P. falciparum* model denote the poly-asparagine insert found in the great ape group



the variation in this central region is found in the simian group and a lot of this variation is due to the two major branches within the simian group (Fig. 3). There are three small regions of 12–20 residues (Supplemental Data, pages 9, 14, and 18) that are conserved across all species within this quasi-variable central region. Presumably, these three conserved regions play some role in the overall structure of *PlasWD40-1*.

The quasi-variable region is also characterized by a substantial amount of low-complexity sequence. Low-complexity sequence is characterized by a low diversity of amino acid residues, and this low diversity is often due to tandem repeats. Proteins with tandem repeats and low complexity are relatively common in *Plasmodium* with more than half of the genes containing low-complexity sequence (DePristo et al. 2006). The sequences from the great ape group contain several asparagine-rich regions, which have been noted in *P. falciparum* proteins in general (Chaudhry et al. 2018). The rodent parasites have the least amount of low-complexity sequence within the *PlasWD40-1* sequence.

Although there is little conservation across the central region of *PlasWD40-1* in terms of sequence identity, this central region is predicted to primarily exhibit a disordered structure in all 17 species. Furthermore, the central region has the signature features of disordered regions, namely low-complexity sequence and an amino acid compositional bias with a low content of bulky hydrophobic amino acids and a high proportion of polar and charged amino acids (Dyson

and Wright 2005). Tandemly repeated sequences are a common feature of disordered regions (Tompa 2003). Although intrinsically disordered protein regions lack a well-defined structure, they often carry out important functions that often involve molecular recognition or function as linkers between protein domains. Thus, even though the sequence is not highly conserved across all *Plasmodium* species, the function of the central region may be preserved.

### C-terminal region

The C-terminal region of *PlasWD40-1* exhibits some sequence conservation, but not as much as the N-terminal region. BLAST searches with the C-terminal regions did not identify any highly homologous sequences that were from non-*Plasmodium* species. Furthermore, no motifs or other characterized domains were identified except for a predicted membrane-spanning sequence near the C-terminus in most of the species (Supplemental Data, page 36). However, there are no signal sequences in *PlasWD40-1*, and this may represent a conserved hydrophobic region that is important for the folding and structure of the protein.

### Interactome

Both the WD40 domain and disordered protein sequence are likely involved in protein–protein interactions, and especially interactions involved in scaffolding and the assembly

of dynamic multi-subunit complexes. An analysis of protein–protein interactions using the yeast two-hybrid screening has been carried out in *P. falciparum* (LaCount et al. 2005), and several interactions with *PlasWD40-1* were identified (Table 2). None of the interactions corresponded to the N-terminal WD40 domain. This is not unexpected in that binding to the WD40 is highly dependent on the three-dimensional structure of the domain which may not be maintained in the yeast two-hybrid system. Most of the interactions identified by the yeast two-hybrid screening corresponded to the central disordered regions of *PlasWD40-1* and are in support of the contention this quasi-conserved region may function in protein–protein interactions. The only binding sequence outside of the central disordered region significantly overlaps with the original cDNA clone from the lambda expression library (Supplemental Data, page 33).

The most often observed interaction of *PlasWD40-1* was with itself. This could imply that the protein forms oligomers with itself. However, one of the binding partners is twelve amino acid residues of a poly-asparagine region containing eleven asparagines and one aspartate (Supplemental Data, pages 11–12). The high frequency of poly-asparagine repeats in *P. falciparum* (Chaudhry et al. 2018) raises questions about the validity of this protein–protein interaction. Similarly, the regions of *PlasWD40-1* that binds

to the GNC5 histone acetyltransferase and the L subunit of translation initiation factor 3 are also largely a poly-asparagine repeats (Supplemental Data, page 19). Likewise, the region of *PlasWD40-1* that binds to actin is primarily a low-complexity repeat that is not conserved within the great ape group (Supplemental Data, page 15).

Several of the proteins interacting with *PlasWD40-1* are involved in transcription and chromatin remodeling, known roles for some WD40 containing proteins. For example, transcription factor APIA2 binds to a region that is partially conserved in all *Plasmodium* species (Supplemental Data, page 14). However, it should be noted that this partially conserved region is rather hydrophobic and could promote promiscuous interactions. Discerning the exact interactome of *PlasWD40-1* will require more direct experimental studies. In addition, yeast two-hybrid assays in other *Plasmodium* species may prove useful in identifying common proteins that interact with *PlasWD40-1* in other *Plasmodium* species.

## Conclusions

*PlasWD40-1* is a large protein (2065–3217 amino acids) found in the genus *Plasmodium*. The protein consists of a N-terminal WD40-repeat domain that is highly conserved among all *Plasmodium* species. The N-terminal region

**Table 2** Proteins interacting with *PlasWD40-1* as detected by the yeast two-hybrid assay

| Bait              | Prey              | Residues   | Obs | Int | Description   | Function                                | Comments   |
|-------------------|-------------------|------------|-----|-----|---|---|--|
| <i>PlasWD40-1</i> | <i>PlasWD40-1</i> | 997–1027   | 5   | 9   | Conserved <i>Plasmodium</i> protein with WD40 repeats           | Unknown                                 | Residues 943–955 are part of a poly-N repeat         |
| <i>PlasWD40-1</i> | <i>PlasWD40-1</i> | 943–955    |     |     |   |   |  |
| <i>PlasWD40-1</i> | <i>PfPIAS</i>     | 1657–1732  | 3   | 1   | Small ubiquitin-like modifier (SUMO) E3 ligase                  | Protein SUMOylation                     |  |
| <i>PlasWD40-1</i> | <i>PFI1715W</i>   | 2826–2937* | 2   | 27  | <i>Plasmodium</i> exported protein of subgenus <i>Laverania</i> | Unknown                                 | The binding region overlaps with original cDNA clone |
| MAL7P1.7          | <i>PlasWD40-1</i> | 1783–1870  | 1   | 23  | <i>Plasmodium</i> exported protein, PHISTb family               | Unknown                                 |  |
| <i>PlasWD40-1</i> | HAT1              | 1657–1742  | 1   | 5   | Histone acetyltransferase-1                                     | Gene expression and chromatin structure |  |
| <i>PlasWD40-1</i> | GNC5              | 927–970    | 1   | 24  | Histone acetyltransferase, GNC5                                 | Gene expression and chromatin structure | Residues 922–953 are a poly-N repeat                 |
| APIA2             | <i>PlasWD40-1</i> | 1208–1272  | 1   | 28  | Transcription factor with AP2 domain of plants and Apicomplexa  | Gene expression                         |  |
| <i>PlasWD40-1</i> | MAL6P1            | 927–963    | 1   | 26  | Translation initiation factor 3 subunit L                       | Protein translation                     | Residues 922–953 are a poly-N repeat                 |
| Actin-1           | <i>PlasWD40-1</i> | 1352–1377  | 1   | 5   | Actin monomer of microfilaments                                 | Parasite motility and invasion          | Residues 1352–1377 are within tandem repeats         |

Shown are the proteins interacting with *PlasWD40-1* as either bait or prey. The residues refer to the amino acid sequence of *PlasWD40-1* that interact with the binding partner and are underlined in the Supplement Data file. Obs.=the number of times a particular interaction was observed. Int.=the total number of protein interactions observed for the protein interacting with *PlasWD40-1*. \*The only binding sequence not found in central disordered region

may also contain a Zn-protease site in some *Plasmodium* species. This is followed by a central domain of quasi-conserved sequence that primarily has low-complexity sequence and a disordered structure. The sequence in this central region is not highly conserved across all species, but the central region is relatively well conserved among species within each of the four *Plasmodium* clades. The C-terminal region is partially conserved across all species. In that the function of WD40 domains and disordered protein structure is primarily to bind other proteins, it is likely that *PlasWD40-1* plays some role in forming multi-protein complexes. The specific proteins that bind to *PlasWD40-1* remain to be determined.

The apparent restriction of *PlasWD40-1* to the genus *Plasmodium* implies that the function is somewhat unique to *Plasmodium*. There are certainly many unique features in the biology of *Plasmodium* that required multi-protein complexes and scaffolding. For example, *PfWLP1* is another WD40 repeat protein unique to *Plasmodium* (von Bohl et al. 2015). *PfWLP1* forms complexes with adhesion proteins in the micronemes of merozoites and in the subpellicular space in gametocytes. In addition, proteins unique to *Plasmodium* that play important roles in the parasite's biology make attractive therapeutic targets. And initial studies suggest that WD40 proteins are viable drug targets (Song et al. 2017).

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00436-021-07190-z>.

**Acknowledgements** Acknowledgments to the Instituto Nacional de Salud, Bogotá, D.C., Colombia, where the original work was initiated.

**Author contribution** Cortés, Gladys T.: Investigation, Methodology (gene cloning), Writing — Review & Editing, Conceptualization, Resources, Funding acquisition. Gonzalez Beltran, Martha Margarita: Methodology (molecular modeling). Gómez-Alegría, Claudio J.: Writing — Review & Editing, Conceptualization, Resources, Supervision. Wiser, Mark F.: Writing — Original Draft, Visualization, Formal analysis (in silico), Methodology (molecular modeling), Conceptualization, Supervision.

**Funding** This work was supported by the Universidad Nacional de Colombia, Vicerrectoría de Investigación y extensión (Strategic Research Project-Hermes 42917), Medicine Faculty (Public Health Department and Laboratorio de equipos comunes), Programa de doctorado del Instituto de Biotecnología, Departamento de Farmacia, Facultad de Ciencias and Doctorado en Medicina Tropical, Universidad de Cartagena- SUE-Caribe. This project was also funded in part by a grant from the Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología (Colciencias-project no. 110140820398).

**Data Availability** Data has been deposited in databases where appropriate and the datasets and materials are available from the corresponding author on reasonable request.

**Code availability** Not applicable.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

- Aureochochea C, Brestelli J, Brunk BP et al (2009) PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 37:539–543. <https://doi.org/10.1093/nar/gkn814>
- Böhme U, Otto TD, Cotton JA et al (2018) Complete avian malaria parasite genomes reveal features associated with lineage-specific evolution in birds and mammals. *Genome Res* 28:547–560. <https://doi.org/10.1101/gr.218123.116>
- Bunnik EM, Chung D-WD, Hamilton M et al (2013) Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. *Genome Biol* 14:R128. <https://doi.org/10.1186/gb-2013-14-11-r128>
- Chahar P, Kaushik M, Gill SS et al (2015) Genome-wide collation of the *Plasmodium falciparum* WDR protein superfamily reveals malarial parasite-specific features. *PLoS ONE* 10:e0128507. <https://doi.org/10.1371/journal.pone.0128507>
- Chaudhry SR, Lwin N, Phelan D et al (2018) Comparative analysis of low complexity regions in Plasmodia. *Sci Rep* 8:335. <https://doi.org/10.1038/s41598-017-18695-y>
- Chen F, Mackey AJ, Stoeckert CJ, Roos DS (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34:D363–D368. <https://doi.org/10.1093/nar/gkj123>
- Cornejo O, Escalante AA (2006) The origin and age of *Plasmodium vivax*. *Trends Parasitol* 22:558–563. <https://doi.org/10.1016/j.pt.2006.09.007>
- Cortés GT, Winograd E, Wiser MF (2003) Characterization of proteins localized to a subcellular compartment associated with an alternate secretory pathway of the malaria parasite. *Mol Biochem Parasitol* 129:127–135. [https://doi.org/10.1016/S0166-6851\(03\)00097-5](https://doi.org/10.1016/S0166-6851(03)00097-5)
- Cortés GT, Wiser MF, Gómez-Alegría CJ (2020) Identification of *Plasmodium falciparum* HSP70-2 as a resident of the Plasmodium export compartment. *Heliyon* 6:e04037–e04037. <https://doi.org/10.1016/j.heliyon.2020.e04037>
- DePristo MA, Zilversmit MM, Hartl DL (2006) On the abundance, amino acid composition, and evolutionary dynamics of low-complexity regions in proteins. *Gene* 378:19–30. <https://doi.org/10.1016/j.gene.2006.03.023>
- Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208. <https://doi.org/10.1038/nrm1589>
- Galen SC, Borner J, Martinsen ES et al (2018) The polyphyly of *Plasmodium*: comprehensive phylogenetic analyses of the malaria parasites (order Haemosporida) reveal widespread taxonomic conflict. *R Soc Open Sci* 5:171780. <https://doi.org/10.1098/rsos.171780>
- Heintzelman MB, Mateer MJ (2008) GpMyoF, a WD40 repeat-containing myosin associated with the myonemes of *Gregarina polymorpha*. *J Parasitol* 94:158–168. <https://doi.org/10.1645/GE-1339.1>
- Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. *Adv Appl Math* 12:337–357. [https://doi.org/10.1016/0196-8858\(91\)90017-D](https://doi.org/10.1016/0196-8858(91)90017-D)
- Jain BP, Pandey S (2018) WD40 repeat proteins: signalling scaffold with diverse functions. *Protein J* 37:391–406. <https://doi.org/10.1007/s10930-018-9785-7>
- Kallio JP, Kursula I (2014) Recombinant production, purification and crystallization of the *Toxoplasma gondii* coronin WD40

- domain. *Acta Crystallogr Sect f, Struct Biol Commun* 70:517–521. <https://doi.org/10.1107/S2053230X14005196>
- Kelley LA, Mezulis S, Yates CM et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10:845–858. <https://doi.org/10.1038/nprot.2015.053>
- Klausen MS, Anderson MV, Jespersen MC et al (2015) LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Res* 43:W349–W355. <https://doi.org/10.1093/nar/gkv535>
- Krawczyk K, Liu X, Baker T et al (2014) Improving B-cell epitope prediction and its application to global antibody-antigen docking. *Bioinformatics* 30:2288–2294. <https://doi.org/10.1093/bioinformatics/btu190>
- LaCount DJ, Vignali M, Chettier R et al (2005) A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* 438:103–107. <https://doi.org/10.1038/nature04104>
- Lapp SA, Mok S, Zhu L et al (2015) *Plasmodium knowlesi* gene expression differs in ex vivo compared to in vitro blood-stage cultures. *Malar J* 14:110. <https://doi.org/10.1186/s12936-015-0612-8>
- Larkin MA, Blackshields G, Brown NP et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Llinás M, Bozdech Z, Wong ED et al (2006) Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res* 34:1166–1173. <https://doi.org/10.1093/nar/gkj517>
- López-Barragán MJ, Lemieux J, Quiñones M et al (2011) Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* 12:1–13. <https://doi.org/10.1186/1471-2164-12-587>
- Loy DE, Liu W, Li Y et al (2017) Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. *Int J Parasitol* 47:87–97. <https://doi.org/10.1016/j.ijpara.2016.05.008>
- Ma J, An K, Zhou J-B et al (2019) WDSPPdb: an updated resource for WD40 proteins. *Bioinformatics* 35:4824–4826. <https://doi.org/10.1093/bioinformatics/btz460>
- Marchler-Bauer A, Bo Y, Han L et al (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* 45:D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- Martinsen ES, Perkins SL, Schall JJ (2008) A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Mol Phylogenet Evol* 47:261–273. <https://doi.org/10.1016/j.ympev.2007.11.012>
- Otto TD, Böhme U, Jackson AP et al (2014) A comprehensive evaluation of rodent malaria parasite genomes and gene expression. *BMC Biol* 12:86. <https://doi.org/10.1186/s12915-014-0086-0>
- Pelle KG, Oh K, Buchholz K et al (2015) Transcriptional profiling defines dynamics of parasite tissue sequestration during malaria infection. *Genome Med* 7:19. <https://doi.org/10.1186/s13073-015-0133-7>
- Ramiro RS, Reece SE, Obbard DJ (2012) Molecular evolution and phylogenetics of rodent malaria parasites. *BMC Evol Biol* 12:219. <https://doi.org/10.1186/1471-2148-12-219>
- Rawlings ND, Barrett AJ (1995) Evolutionary families of metalloproteases. *Methods Enzymol* 248:183–228. [https://doi.org/10.1016/0076-6879\(95\)48015-3](https://doi.org/10.1016/0076-6879(95)48015-3)
- Reynolds CR, Islam SA, Sternberg MJE (2018) EzMol: a web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *J Mol Biol* 430:2244–2248. <https://doi.org/10.1016/j.jmb.2018.01.013>
- Rovira-Graells N, Gupta AP, Planet E et al (2012) Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res* 22:925–938. <https://doi.org/10.1101/gr.129692.111>
- Short JM, Fernandez JM, Sorge JA, Huse WD (1988) Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res* 16:7583–7600. <https://doi.org/10.1093/nar/16.15.7583>
- Sigrist CJA, De Castro E, Cerutti L et al (2013) New and continuing developments at PROSITE. *Nucleic Acids Res* 41:D344–D347. <https://doi.org/10.1093/nar/gks1067>
- Silva JC, Egan A, Arze C et al (2015) A new method for estimating species age supports the coexistence of malaria parasites and their mammalian hosts. *Mol Biol Evol* 32:1354–1364. <https://doi.org/10.1093/molbev/msv005>
- Song R, Wang Z-D, Schapira M (2017) Disease association and druggability of WD40 repeat proteins. *J Proteome Res* 16:3766–3773. <https://doi.org/10.1021/acs.jproteome.7b00451>
- Sonnhammer EL, Von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6:175–182
- Stirnemann C, Petsalaki E, Russell R, Müller C (2010) WD40 proteins cellular networks. *Trends Biochem Sci* 35:565–574. <https://doi.org/10.1016/j.tibs.2010.04.003>
- Sundaraman SA, Plenderleith LJ, Liu W et al (2016) Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun* 7:11078. <https://doi.org/10.1038/ncomms11078>
- Tachibana SI, Sullivan SA, Kawai S et al (2012) *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat Genet* 44:1051–1055. <https://doi.org/10.1038/ng.2375>
- Tompa P (2003) Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays* 25:847–855. <https://doi.org/10.1002/bies.10324>
- von Bohl A, Kuehn A, Simon N et al (2015) A WD40-repeat protein unique to malaria parasites associates with adhesion protein complexes and is crucial for blood stage progeny. *Malar J* 14:435. <https://doi.org/10.1186/s12936-015-0967-x>
- Wang Y, Jiang F, Zhuo Z et al (2013) A method for WD40 repeat detection and secondary structure prediction. *PLoS ONE* 8:e65705. <https://doi.org/10.1371/journal.pone.0065705>
- Waterhouse A, Bertoni M, Bienert S et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W296–W303. <https://doi.org/10.1093/nar/gky427>
- Wiser MF (2007) Export and trafficking of *Plasmodium* proteins within the host erythrocyte. *Acta Biológica Colomb* 12:3–18
- Wootton JC (1994) Non-globular domains in protein sequences: automated segmentation using complexity measures. *Comput Chem* 18:269–285. [https://doi.org/10.1016/0097-8485\(94\)85023-2](https://doi.org/10.1016/0097-8485(94)85023-2)
- Xu D, Zhang Y (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins* 80:1715–1735. <https://doi.org/10.1002/prot.24065>
- Yang Y, Faraggi E, Zhao H, Zhou Y (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics* 27:2076–2082. <https://doi.org/10.1093/bioinformatics/btr350>
- Zanghì G, Vembar SS, Baumgarten S et al (2018) A specific PfEMP1 is expressed in *P. falciparum* sporozoites and plays a role in hepatocyte infection. *Cell Rep* 22:2951–2963. <https://doi.org/10.1016/j.celrep.2018.02.075>
- Zhang K, Fujioka H, Lobo CA et al (1999) Cloning and characterization of a new asparagine-rich protein in *Plasmodium*

- falciparum*. Parasitol Res 85:956–963. <https://doi.org/10.1007/s004360050666>
- Zhang M, Wang C, Otto TD et al (2018) Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. Science 360:eaap7847. <https://doi.org/10.1126/science.aap7847>
- Zhu L, Mok S, Imwong M et al (2016) New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. Sci Rep 6:20498. <https://doi.org/10.1038/srep20498>

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