

Serveur Académique Lausannois SERVAL [serval.unil.ch](http://serval.unil.ch)

## Author Manuscript

### Faculty of Biology and Medicine Publication

**This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.**

Published in final edited form as:

**Title:** The C-terminal domain of Plasmodium falciparum merozoite surface protein 3 self-assembles into alpha-helical coiled coil tetramer.

**Authors:** Gondeau C, Corradin G, Heitz F, Le Peuch C, Balbo A, Schuck P, Kajava AV

**Journal:** Molecular and biochemical parasitology

**Year:** 2009 Jun

**Volume:** 165

**Issue:** 2

**Pages:** 153-61

**DOI:** 10.1016/j.molbiopara.2009.01.015

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

Published in final edited form as:

*Mol Biochem Parasitol*. 2009 June ; 165(2): 153–161. doi:10.1016/j.molbiopara.2009.01.015.

## The C-terminal domain of *Plasmodium falciparum* merozoite surface protein 3 self-assembles into $\alpha$ -helical coiled coil tetramer

Claire Gondeau<sup>1</sup>, Giampietro Corradin<sup>2</sup>, Frédéric Heitz<sup>1</sup>, Christian Le Peuch<sup>1</sup>, Andrea Balbo<sup>3</sup>, Peter Schuck<sup>3</sup>, and Andrey V. Kajava<sup>1,\*</sup>

<sup>1</sup> Centre de Recherches de Biochimie Macromoléculaire, CNRS UMR-5237, University of Montpellier 1 and 2, Montpellier, France <sup>2</sup> Institute of Biochemistry, University of Lausanne, Epalinges, Switzerland <sup>3</sup> Dynamics of Macromolecular Assembly Section, Laboratory of Bioengineering and Physical Science, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, Maryland 20892, U.S.A

### Abstract

Proteins located on the surface of the pathogenic malaria parasite *Plasmodium falciparum* are objects of intensive studies due to their important role in the invasion of human cells and the accessibility to host antibodies thus making these proteins attractive vaccine candidates. One of these proteins, merozoite surface protein 3 (MSP3) represents a leading component among vaccine candidates; however, little is known about its structure and function. Our biophysical studies suggest that the 40 residue C-terminal domain of MSP3 protein self-assembles into a four-stranded  $\alpha$ -helical coiled coil structure where  $\alpha$ -helices are packed “side-by-side”. A bioinformatics analysis provides an extended list of known and putative proteins from different species of *Plasmodium* which have such MSP3-like C-terminal domains. This finding allowed us to extend some conclusions of our studies to a larger group of the malaria surface proteins. Possible structural and functional roles of these highly conserved oligomerization domains in the intact merozoite surface proteins are discussed.

### Keywords

$\alpha$ -helical coiled coil; oligomerization; protein structure; *Plasmodium falciparum*.

### 1. Introduction

Malaria is a major human disease that accounts for almost 2 million deaths annually. The merozoite surface of *Plasmodium falciparum* is comprised of proteins that are important for the invasion of human red cells. In addition, the merozoite surface proteins (MSP) are considered to be among the best candidate antigens for inclusion in an anti-malarial vaccine [1,2]. A number of these proteins have been identified [3]. MSP-1, 2, 4, 5, 8, and 10 proteins are linked to the membrane of the parasite via a glycosylphosphatidylinositol (GPI) membrane anchor [4–6]. These proteins, except MSP-2, have one or two epidermal growth factor (EGF)-

\*Corresponding author at: Centre de Recherches de Biochimie Macromoléculaire, CNRS UMR-5237, 1919 Route de Mende, 34293 Montpellier, Cedex 5, France. Tel: +33 4 67 61 3364; Fax: +33 4 67 521559; E-mail: E-mail: andrey.kajava@crbm.cnrs.fr.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

like domains at the carboxyl terminus. Other MSPs such as MSP3, MSP6, MSP-7 and MSP-9 are soluble, hydrophilic and are, in part, associated with the merozoite surface [7–11]. Although many proteins are known to play an important role in merozoite invasion, their three-dimensional structure and specific functions remains unknown. Therefore, a complete understanding of the structural characteristics and molecular interactions that are the basis of the invasion process is critical, not only for improving our knowledge about the basic biology of the malaria parasite, but also for the development of vaccine and the other intervention strategies to counter the disease.

In this work, we focused on MSP3 protein which is a promising candidate antigen for anti-malaria vaccines [2,12,13]. Previous studies suggested that MSP3 protein is located on the merozoite surface but does not contain a transmembrane domain or GPI anchor consensus sequence and so presumably is attached to the merozoite surface as a result of protein-protein interactions [9,13,14].

MSP3 has sequence similarity within the first ~50 residues that contain putative signal peptide with several other surface malaria proteins (Fig. 1). The precursor form of MSP3 may undergo cleavage in its N-terminal region during schizogony to generate a mature processed form [15]. The central regions of MSP3 and the other homologous proteins are different. MSP3 contains three blocks of alanine-rich heptad repeats that are predicted to form an intramolecular coiled coil structure [11], while, for example, MSP6 has an aperiodic hydrophilic sequence. In the next region toward the C-terminus, some proteins, including MSP3, share a motif (ILGWDFGGG-[AV]-P) followed by an acidic region. Finally, they exhibit a C-terminal ~40 residue domain with the highest degree of sequence similarity [9]. The sequence analysis of this C-terminal domain yielded ambiguous conclusions; initially, it was suggested that this C-terminal region contains a leucine-zipper like motif [14]. However, other authors did not identify this motif in either MSP3 or MSP6 and predicted that MSP6 has two amphipathic  $\alpha$ -helices separated by a loop thus forming intramolecular contacts with each other [9]. Previous biophysical study of the full length MSP3 and four smaller peptides [16] suggested that all tested samples have a large portion of  $\alpha$ -helical and random coil conformations. The cross-linking and analytical ultracentrifugation experiments suggested that the full length MSP3 forms elongated dimers and tetramers. The study also proposed a crucial role of the last 55 C-terminal residues in oligomerization of MSP3.

In the present study, by applying bioinformatics approaches, we have extended the list of proteins having the MSP3-like C-terminal domain from 4 to 10 in *P. falciparum* and 7 in the other species of *Plasmodium*. The sequence alignment of these proteins allowed us to demarcate more precisely the boundaries of the conserved C-terminal domain. On the other hand, the remaining parts of these malaria proteins have a high sequence diversity. This suggests that the conserved C-terminal region alone may represent a structural domain that is self-sufficient for folding into the functional structure. Furthermore, the major central regions of all analyzed proteins are well separated from the C-terminal ones by long (80–100 residues) unstructured acid regions. Such separation is typical for the domains that can self-assemble independently. In this work, we tested whether the C-terminal fragment of MSP3 can fold independently into a stable 3D structure and analyzed its structural properties.

## 2. Materials and Methods

### 2.1 Bioinformatics sequence analysis

Initial selection of malaria proteins containing the homologous C-terminal regions has been made by using a PSI-BLAST reiterative analysis [17]. Subsequently, a software package *pftools*, incorporating programs for sensitive generalized sequence profiling [18] was used to detect new proteins containing such C-terminal domains. The probability that the matches are

a product of chance alone was calculated by analyzing the score distribution obtained from a profile search against a regionally randomized version of the protein database, assuming an extreme value distribution [19].

## 2.2 Peptide synthesis

The peptides were synthesized using solid-phase Fmoc chemistry (Applied Biosystem 431A). Briefly, peptides were prepared on a *p*-alkoxybenzylalcohol resin (Wang resin) after which one equivalent of Fmoc amino acid derivatives was added. Following cleavage from the resin with Barany solution, the crude peptide was purified by RP-HPLC (C18 preparative column; H<sub>2</sub>O 0.1% TFA/ acetonitrile 0.1% TFA from 90/10 to 20/80 over 70 minutes with a flow rate of 10 mL/min). Purity (>90%) was determined by analytical C18 HPLC and mass spectroscopy (MALDI-TOF) (see Supplemental Fig. S1). The molecular weights (calculated from the sequences) of the MSP3 and MSP6 peptides are 5989 Da and 5960 Da, respectively.

## 2.3 Size exclusion chromatography

Size exclusion chromatography of MSP3 and MSP6 peptides was carried out at 4°C on a Sephacryl S-100 gel filtration column (16 mm × 60 cm) interfaced with an ÄKTA purifier 10 system (GE Healthcare Life Science). The effluent was monitored by measuring absorbance at 230 nm. Each sample was prepared by dissolving the peptide in a 100 mM sodium phosphate buffer (pH 6.1 or 7.3) to a final concentration of 0.5 mg/ml and incubated at 4°C overnight. Each sample (1 ml) was injected onto the column which was equilibrated with the sample buffer and elution was performed with a flow rate of 0.5 ml/min. The column was calibrated with the standard globular proteins albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa) and ribonuclease A (13.7 kDa) (all from GE Healthcare Life Science).

## 2.4 Circular dichroism (CD) spectroscopy

CD spectra were recorded on a JASCO J-810 spectrometer (JASCO corporation, Tokyo, Japan) equipped with a temperature controller and a 0.1 cm path length cuvette. The measurements were taken at 20°C and each wave scan was performed in 0.5 nm steps between wavelengths from 260 to 185 nm with an integration time of 0.5 s at each wavelength. Peptides were dissolved in a 0.1M sodium phosphate buffer (pH 7.3 or pH 6.1).

## 2.5 Analytical ultracentrifugation

Sedimentation experiments were carried out with ProteomeLab XL-I analytical ultracentrifuges (Beckman-Coulter, Palo Alto, CA). The MSP3 samples were dialyzed overnight at 4 °C into 100 mM sodium phosphate, 100 mM NaCl pH 7.3 buffer. 3 mm Epon centerpieces were loaded with 90 µl of sample at concentrations of 1 and 2 mg/ml, and 12 mm Epon centerpieces were loaded with 360 µl of sample at concentrations of 0.2 and 0.5 mg/ml. Sample cells were prepared in duplicate and sedimented side-by-side in two centrifuges, one at the rotor speed of 42,000 rpm and one at the rotor speed of 60,000 rpm, respectively, both at a temperature of 20 °C. The evolution of refractive index profiles was recorded with the interference optical detection system. In both runs, sedimentation velocity (SV) analysis was conducted on data from the first 8 hours of sedimentation. Due to the populations of small molecular weight species, the same rotor speeds were also suitable for sedimentation equilibrium (SE) studies, which were conducted simply as a continuation of the same experiments without stopping the runs. After 13 days of taking scans in time-intervals of 6 h, the lack of detectable change in concentration gradients indicated that equilibrium was attained. This unusually long equilibration time was due to the long solution column needed to provide sufficient data for the initial SV analysis. The relatively high rotor speed was required in order to resolve the small molecular weight species from the oligomers of interest, avoiding the pitfall of measuring only a single weight-average molecular weight that would have significantly

underestimated the oligomeric state. The consistency of the global analysis of SV and SE for both runs at the two rotor speeds demonstrates the absence of significant degradation of the peptides (Table 2). The SE profiles were baseline corrected using the TI noise calculated from the initial SV analysis [20]. (This has the advantage of avoiding disassembling the cells and restarting the run, which at the high gravitational fields may lead to mechanical alterations and potential changes in the water blanks.)

Initial SV data analysis was performed using the c(s) method in SEDFIT [21]. A more refined analysis was performed by global modeling of the SE profiles at both rotor speeds and all concentrations [22], and analogously for the SV data [23], in the software SEDPHAT using the hybrid discrete/continuous model. The partial specific volume was calculated with the program SEDNTERP (kindly provided by Dr. J. Philo)[24]. Theoretical predictions of the translational hydrodynamic properties of model structures were made with the program HYDROPRO (kindly provided by Dr. Garcia de la Torre), using a radius of the atomic elements of 3.2 Å, corresponding to a hydration layer thickness of 1.5 Å [25].

### 3. Results and discussion

#### 3.1 The C-terminal coiled coil motif is common to several malaria surface proteins with otherwise different sequences

Previously, it has been described that four *P. falciparum* proteins MSP3, MSP6, MSP11 (or H103) and H101 have a similar N-terminal apolar region resembling the signal sequence followed by a common motif with the NLRng consensus sequence (Fig. 1) [9,26]. The central regions of these proteins are different, while at their C-terminal region they again share a motif with a consensus sequence gwEfgGGap followed by an acidic region and an  $\alpha$ -helical coiled coil region [26]. We were particularly interested in the C-terminal ~40 residue region that has the highest degree of sequence similarity among these proteins and alone may self-assemble into a functional structure. In an attempt to detect new proteins containing such a C-terminal domain we used a sensitive generalized sequence profiler, a Hidden-Markov-Model-based profile method implemented in the *pftools* software package [18]. Alignment of the C-terminal sequences of *P. falciparum* MSP3, MSP6, MSP11 and H101 proteins was used to construct an initial sequence profile. The best new matches with E-values < 0.001 found by this profile in the sequence databases were included in the next profile and this procedure was iterated several times. As a result, our analysis revealed that not only four previously mentioned *P. falciparum* merozoite surface proteins [26] but also six additional proteins of *P. falciparum*, namely, liver stage antigen-1 (LSA-1), glutamate-rich protein (GLURP) and as yet uncharacterized proteins MSPDBL1, MSPDBL2 having Duffy binding domains and M712, M566 (as named in [26]) have a common ~40 residue long C-terminal domain (Fig. 1). In most cases, the C-terminal domain is preceded by a glutamate-rich acidic stretch. In addition, these proteins also have a common N-terminal hydrophobic motif (a putative signal sequence) accompanied by an NLR signature. The experimentally characterized proteins are located in the parasitophorous vacuole or on the merozoite surface [26] and this allows us to suggest that the remaining uncharacterized proteins may also be located there. While the N- and C-terminal regions are similar across these surface proteins, the major central parts are strongly variable in their amino acid sequence and length (Fig. 1). Remarkably, the genes for all these proteins are located on *P. falciparum* chromosome 10, one after the other and in the same 5' to 3' orientation (see Supplemental Fig S2). We also found this C-terminal domain in several proteins of the other Plasmodium species (*P. vivax*, *P. chabaudi chabaudi*, *P. yoelii* and *P. berghei*) (Fig. 1b).

The analysis of the sequence alignment of the C-terminal domain (Fig. 1b) confirms the conservation of the  $\alpha$ -helical coiled coil pattern. It is known, that the  $\alpha$ -helical coiled coil structures share a seven residue repeat (**abcdefg**)<sub>n</sub> containing hydrophobic residues at **a** and

**d** positions and polar residues generally elsewhere. Indeed, most of the **a** and **d** positions of the consensus C-terminal sequence are large apolar residues. The exception is one **a** position in the middle of the region that is frequently occupied by polar residues. The C-terminal sequences of these malaria proteins were compared with sequence profiles which were constructed based on the known  $\alpha$ -helical coiled coil structures [27]. Their profile scores were above the cut-off levels that were chosen by tests performed against sequences of the known coiled coils. This result suggests that all the C-terminal domains form  $\alpha$ -helical coiled coil structures.

Thus, the bioinformatics analysis suggested that the ~40 residue C-terminal domain is a common element for a larger group of surface proteins. This finding added importance to the structural studies of the C-terminal peptides of MSP3 because the results of this study may be generalized for several other surface proteins. The analysis also confirmed that the C-terminal regions of all selected proteins have the  $\alpha$ -helical coiled coil sequence pattern. At the same time, we could not unambiguously predict whether the C-terminal helices interact with each other or each of them is broken into two helices which form intramolecular contacts.

### 3.2 MSP3 peptide forms $\alpha$ -helical structure

Circular dichroism (CD) spectra of the chemically synthesized peptides MSP3 (residues 302–354, here and elsewhere the numbering from Pf10\_0345 <http://plasmodb.org>) exhibits double minima at 208 and 222 nm and a maximum near 190 nm at pH 7.3 and 22°C (Fig. 2A). These spectra are characteristic of the  $\alpha$ -helical secondary structure with about 90%  $\alpha$ -helical content as calculated using the CDSSTR program [28]. Previous study of a longer C-terminal fragment of MSP3 showed that its low molecular weight oligomer comprises 31%  $\alpha$ -helices and 29% of random coil [16].

The CD spectra of the MSP6 peptide (residues 320–371) are similar to the ones of the MSP3 peptide, in that they exhibit double minima at 208 and 222 nm and a maximum near 190 nm, characteristic of the  $\alpha$ -helical secondary structure at pH 7.3 and 22°C (Fig. 2A). However, it was noticed that MSP6 peptide is less soluble and stable than the MSP3 peptide in these conditions. This property has hampered out further biophysical analysis of the MSP6 peptide.

The CD spectra of MSP3 peptide change slightly with time (Fig. 2A). Although the protein appears to remain  $\alpha$ -helical after overnight incubation at 4°C, the  $\Theta_{222}/\Theta_{208}$  ratio increases from 0.94 to 0.97 (Fig. 2A). The value of the  $\Theta_{222}/\Theta_{208}$  ratio also increases with change of pH, from 0.94 at pH 7.3 to 1.06 at pH 6.1. At pH 6.1, the MSP3 peptide reaches this  $\Theta_{222}/\Theta_{208}$  ratio almost immediately after dissolution (Fig. 2B). It is known, that a  $\Theta_{222}/\Theta_{208}$  ratio in excess of the unity is characteristic of  $\alpha$ -helical coiled coil structure [29]. The structural basis of this phenomenon may be the fact that  $\alpha$ -helices within coiled coil structures have a regular distortion and differ from usually straight  $\alpha$ -helices of globular proteins or individual  $\alpha$ -helices. A frequently used approach to provide additional evidence for the presence of  $\alpha$ -helical coiled coil structures is the CD measurement of the peptide under the same conditions but with the addition of 50% trifluoroethanol (TFE). TFE destroys coiled coil structures by decreasing the inter-helical interactions of the apolar residues, and stabilizes individual  $\alpha$ -helices by reinforcing the intra-helical H-bonds. The CD spectra of the  $\alpha$ -helical coiled coils reveal this structural change in a decrease of the  $\Theta_{222}/\Theta_{208}$  ratio below unity [29]. Our experiments show that MSP3 peptide also shows the characteristic  $\Theta_{222}/\Theta_{208}$  ratio shift from 1.06 to 0.9 at pH 6.1 upon addition of 50 % TFE (Fig. 2B) which provides additional support for the  $\alpha$ -helical coiled coil structure of MSP3.

### 3.3 Size exclusion chromatography reveals two discrete molecular species

To investigate the ability of MSP3 peptide to form oligomers, we performed size exclusion chromatography experiments (Fig. 3). The peptide was dissolved in 0.1M sodium phosphate, pH 7.3 and 6.1, and incubated overnight at 4 °C before injection. In both cases, the sample eluted as a double peak with apparent molecular masses of about 35 kDa and 25 kDa when compared to the standard globular proteins ovalbumin (43 kDa) and chymotrypsinogen (25 kDa). The first peak with elution volume 48 mL, which may correspond to the high molecular weight (HMW) oligomer, represents about 70 % of the MSP3 sample (estimated by peak integration) while the second peak with elution volume 60 mL, which may correspond to low molecular weight (LMW) oligomers represents about 30 %. The symmetry of the peaks suggests two discrete molecular species under these conditions. Apparent molecular masses of both peaks (35 kDa and 25 kDa) are much larger than the calculated monomeric mass of 6 kDa. This suggests that MSP3 either forms large oligomers (hexamer and tetramer) and/or has a highly non-globular shape that contributes to increased surface area and frictional properties of the peptides. To determine the oligomerization state of MSP3 species independent of shape considerations we performed analytical ultracentrifugation studies (next section).

To investigate the individual conformational properties of these two molecular species, we recorded their CD spectrum immediately after the separation. The measurements show that both oligomers have  $\alpha$ -helical conformation (Fig. 4). Although, in some conditions, the  $\Theta_{222}/\Theta_{208}$  ratios of the MSP3 oligomer species are below unity, their increases relative to the ratios obtained in 50% TFE propose transformation of the oligomeric structures towards the  $\alpha$ -helical coiled coil. This ratio of the HMW species is larger than the LMW one (Table 1) in two different conditions. This suggests that the HMW oligomer has a more characteristic  $\alpha$ -helical coiled coil structure than the LMW one.

The thermal stability of both MSP3 species was monitored by CD spectroscopy in 0.1M sodium phosphate at pH 7.3. The temperature dependence of the molar ellipticity at 222 nm is reported in Fig. 4. The secondary structure of both species denaturates as the temperature is increased. A sigmoidal trace was recorded for the species that eluted in the first peak (elution volume of 48 ml), reflecting a structural transition at 36 °C (Fig. 4A). It worth mentioning that this thermal stability is modest compared with many previously characterized coiled coils [30,31]. The molar ellipticity at 222 nm of the second lower apparent molecular weight MSP3 species increases gradually with temperature without a noticeable transition (Fig. 4B).

### 3.4 Analytical ultracentrifugation reveals the presence of dimers and tetramers

In order to examine in more detail the oligomeric state of the MSP3 peptide (residues 311–354), sedimentation equilibrium (SE) and sedimentation velocity (SV) experiments were conducted in 100 mM sodium phosphate buffer at pH 7.3. As shown in Fig. 5A, the  $c(s)$  sedimentation coefficient distributions exhibit one main peak. The peak sedimentation coefficients (short  $s$ -values) are identical within 0.01 S and without a concentration-dependent trend, indicating the presence of a main oligomeric species that is quite stable ( $k_{\text{off}} < 10^{-5} \text{ sec}^{-1}$ ) on the time-scale of the SV experiment. At the lowest concentration, a second peak at a lower  $s$ -value emerged, suggesting partial dissociation at concentrations below 0.5 mg/ml.

Also apparent from the  $c(s)$  analysis (as well as from visual inspection of the SV signal near the meniscus that continually decreases with time (see Fig. 5B), and the shallow exponential contribution in the SE data (see Fig. 5D) is the presence of a signal contribution from a very small species characteristic of buffer salts (suggesting that the dialysis with the very low MW cutoff membrane was not completely successful). However, this signal is well-separated from that of MSP3 in both SE and SV and does not complicate the analysis, and was accounted for throughout (with a best-fit estimate buoyant molar mass from the SE analysis of ~ 90 Da).

With the focus on the determination of the identity of the largest MSP3 species, we globally modeled all available SE data, and separately all available SV data from the different rotor speeds and loading concentrations, with a model incorporating a large species with unknown molecular weight and dimeric species. An excellent fit to the data was achieved with the molecular weight estimates of the large oligomer of 19.2 and 20.2 kDa, respectively, from SE and SV analysis, which is close to the theoretical value of 19.9 kDa for the MSP3 tetramer (Fig. 5 and Supplemental Figs S3 and S4). The best-fit value of the sedimentation coefficient for the largest species was  $s_{20,w} = 1.82$  S, a value too high to be accounted for by even the most compact hypothetical trimeric structures (see below). The results of fits with several alternate models are listed in Table 2. Models accounting for largest species to be trimers only can be clearly rejected, and also models with the monomer as the only dissociation product fit the data distinctly worse (although some monomer may be present in addition to the dimer). However, based on our SE and SV data shown in Table 2, we cannot exclude the co-existence of trimers as minority species with alternate fold.

The presence of two peaks during SEC but only one major peak in analytical ultracentrifugation appears to be in conflict. However, the loading concentration on SEC was lower and being further diluted on the column, shifting the dimer-tetramer equilibrium more towards the dimer. Further, for self-association processes with very slow kinetics such as displayed by MSP3, it is notoriously difficult to establish that the sample is in chemical equilibrium at the start of the experiment [32,33], which may contribute to quantitative differences in the populations of the dimer and tetramer.

Our results differ from the previous observation (reported by two of us, A.B. and P.S. in [16]) on the ability of the longer C-terminal region (57 residues) to form both tetramers and hexamers but agree with the oligomeric states previously observed for the full length protein [16].

### 3.5 Structural models of the MSP3 oligomers

The analytical ultracentrifugation data may provide reliable constraints for support or refute of the 3D atomic models of the analyzed samples. The principal requirement for this is a possibility to build a limited number of plausible molecular models for the analysed protein. It was impossible for the full length MSP3 protein and its fragments [16] due to their large size and presence of long regions with random coil conformation. On the other hand, the MSP3 C-terminal domain studied in this work is smaller and  $\alpha$ -helical, and this allows us to construct atomic models and compare their hydrodynamic properties with the data from sedimentation velocity. The models were built by using the crystal structures of GCN4 leucine zipper mutants [34,35] as initial templates and applying the molecular modeling program Insight II [36].

The main constraints for the modeling were: (1) the predominant structural state of MSP3 is a tetramer and (2) the peptides have  $\alpha$ -helical coiled coil structure. The first tested model represented two coiled coil dimers of MSP3(311–354) which interact with each other “end-to-end” [16] (Fig. 6A). The predicted s-value for such a tetramer is 1.52 S, far below the experimentally observed value of 1.82 S. (The difference in the predicted and observed Stokes radius is  $\sim 5$  Å, far too large to be accounted for by uncertainties in the hydration layer thickness of the model [25].) Therefore, this model was rejected. The second model, a four-stranded  $\alpha$ -helical coiled coil structure, is more compact and exhibits a lower translational frictional coefficient, which would lead to a theoretical s-value of 1.75 S. The deviation from the experimental value is 0.07 S (or 3.5%), which is within the typical range of difference between hydrodynamic prediction from structures and experimental values [25].

The experimental constraints do not allow discrimination between two possible  $\alpha$ -helical coiled coil tetramers: (i) a structure with four-fold axis of symmetry and parallel and in-register arrangement of the  $\alpha$ -helices and (ii) a structure with an antiparallel orientation of two coiled



coil dimers forming side-by-side contacts (Fig. 6B, C). The full length protein has been observed to be highly extended [16], which may render the antiparallel orientation more favorable. Thus, our results agree with the oligomeric states previously observed for the full length protein and lead to the conclusion that this oligomerization occurs via C-terminal domains. At the same time, our data provide more details about possible structures of the C-terminal oligomers.

It is worth mentioning that the coiled coil motif of MSP3-like peptides is interrupted in the middle by a polar residue at position **a**. With the assumption that the  $\alpha$ -helix is not broken at this place, the polar side chains will interact with each other in the coiled coil structure, as was already observed in several known coiled coil structures [37]. This was an additional constraint that we used to model the MSP3 oligomers, especially when we searched for the exact positioning of two coiled coil dimers in the side-by-side tetramer (Fig. 6B). The interruption of the coiled coil motif may also be an indication of helix-turn-helix structure. However, this alternative was not considered, because all plausible models formed from such helix-turn-helix resulted in sphere-like shapes which do not fit the sedimentation velocity data.

In order to provide further independent evidence for the large oligomeric species to be the tetramer, we have constructed a model for the most compact trimeric form (Fig. S5). It has a quasi-spherical contour and would have the least hydrodynamic friction and the highest possible sedimentation coefficient of 1.62 S. This is still significantly below the experimentally measured value of  $s_{20,w} = 1.82$  S. (Only at a radius of the atomic elements of  $\sim 1.5$  Å would the theoretical s-value match the experimental one, which is less than half the commonly observed value and even smaller than the average van der Waals radius of 1.8 Å, leaving essentially no room for hydration.) This supports the molecular weight result from SV boundary modeling and SE that the largest species and main oligomer is the tetramer.

#### 4. Conclusions

Our biophysical studies suggest that the 40 residue C-terminal domain of MSP3 protein from *P. falciparum* self-assembles into a four-stranded  $\alpha$ -helical coiled coil structure with  $\alpha$ -helices that are packed “side-by-side”. The other homologous MSP6 peptide also is able to form  $\alpha$ -helical oligomers, though biophysical characterization of this peptide was hampered by its general tendency to aggregate. On the other hand, in this work, by applying bioinformatics approaches, we extended the list of proteins having this MSP3-like domain from 4 to 10 in *P. falciparum* and 7 in the other species of *Plasmodium*. The literature search provided us with additional structural data related to a third protein from this extended list of proteins having the MSP3-like domain, the glutamate-rich protein GLURP from *P. faciparum*. This large protein (1233 residues), otherwise called glutamate-rich tandem repeat protein, exists in an oligomeric form as revealed by gel filtration chromatography and non-denaturing gel electrophoresis [38]. The discrepancy between the estimated molecular sizes determined from these experiments suggested that the protein is a long rod-shaped protein. In non-denaturing electrophoresis, most of GLURP migrates as an apparent monomer and an oligomer of three to four subunits [38]. It is important to mention that full length MSP3 also forms two types of oligomers: dimer and tetramer [16]. Thus, all these data suggest that the C-terminal domains of the other identified surface proteins of *P. falciparum* and the other malaria species may also serve as a homo-oligomerization domain. Except the C-terminal domain, the sequence of the mature part of these proteins is quite different, rich in hydrophilic unfolded regions and with dispersed common short motifs (Fig. 1A). A most probable architecture of such molecules could be a spider-like oligomer with a core formed by C-terminal domains and semi-rigid arms protruding from the core (Fig. 6D). Previously, it was shown that molecules with such architecture can have a high avidity for membrane receptors due to their multivalency [39]. This suggests that the oligomerization of these malaria surface proteins by the C-terminal

domain can impart multivalency to these molecules, which may be necessary for effective binding. These proteins may interact with the merozoite surface proteins, such as, for example, cysteine protease ABRA for MSP3 [40] and the MSP1 and MSP7 complex for MSP6 [9]. It is also noteworthy to mention that the C-terminal fragments of MSP3 and MSP6 are targets of a parasite growth inhibitory antibody response [12,41]. Although this model is highly speculative, it is in agreement with the observed importance of both N- and C-termini in the association of MSP3 on the parasite surface [15,40] because, the C-terminal region is critical for multivalency and the N-terminal region may be involved in binding with the surface proteins. On the other hand, MSP3 and GLURP also bind the erythrocyte membrane [42,43]. Our spider-like model can explain how it can switch from interaction with the merozoite to the erythrocytes surface (Fig. 6D). Recently, it was shown that several MSP3 peptides including the first half of the C-terminal domain may interact specifically with erythrocytes and inhibit merozoite invasion [43]. This suggests that the C-terminal domain itself may have a binding site on the surface of erythrocytes.

## Acknowledgements

This work was supported in part by the Intramural Research Program of the National Institutes of Health, NIBIB.

## List of abbreviations

<b>MSP</b>	merozoite surface protein
<b>CD</b>	circular dichroism
<b>TFE</b>	trifluoroethanol
<b>SE</b>	sedimentation equilibrium
<b>SV</b>	sedimentation velocity
<b>LMW</b>	low molecular weight
<b>HMW</b>	high molecular weight

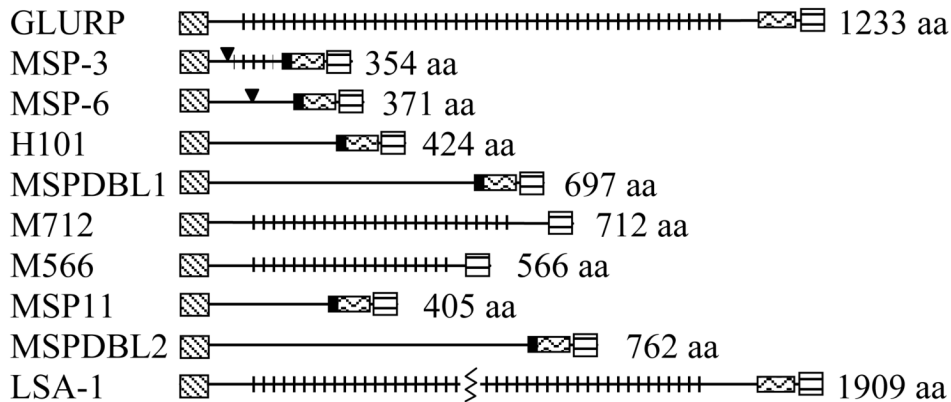
## References

1. Richie TL, Saul A. Progress and challenges for malaria vaccines. *Nature* 2002;415:694–701. [PubMed: 11832958]
2. Meraldi V, Nebie I, Tiono AB, Diallo D, Sanogo E, Theisen M, Druilhe P, Corradin G, Moret R, Sirima BS. Natural antibody response to *Plasmodium falciparum* Exp-1, MSP-3 and GLURP long synthetic peptides and association with protection. *Parasite Immunol* 2004;26:265–72. [PubMed: 15541030]
3. Gaur D, Mayer DC, Miller LH. Parasite ligand-host receptor interactions during invasion of erythrocytes by *Plasmodium* merozoites. *Int J Parasitol* 2004;34:1413–29. [PubMed: 15582519]
4. Gerold P, Schofield L, Blackman MJ, Holder AA, Schwarz RT. Structural analysis of the glycosyl-phosphatidylinositol membrane anchor of the merozoite surface proteins-1 and -2 of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1996;75:131–43. [PubMed: 8992312]

5. Marshall VM, Silva A, Foley M, Cranmer S, Wang L, McColl DJ, Kemp DJ, Coppel RL. A second merozoite surface protein (MSP-4) of *Plasmodium falciparum* that contains an epidermal growth factor-like domain. *Infect Immun* 1997;65:4460–7. [PubMed: 9353020]
6. Black CG, Wang L, Wu T, Coppel RL. Apical location of a novel EGF-like domain-containing protein of *Plasmodium falciparum*. *Mol Biochem Parasitol* 2003;127:59–68. [PubMed: 12615336]
7. Stahl HD, Bianco AE, Crewther PE, Anders RF, Kyne AP, Coppel RL, Mitchell GF, Kemp DJ, Brown GV. Sorting large numbers of clones expressing *Plasmodium falciparum* antigens in *Escherichia coli* by differential antibody screening. *Mol Biol Med* 1986;3:351–68. [PubMed: 3534513]
8. Pachebat JA, Ling IT, Grainger M, Trucco C, Howell S, Fernandez-Reyes D, Gunaratne R, Holder AA. The 22 kDa component of the protein complex on the surface of *Plasmodium falciparum* merozoites is derived from a larger precursor, merozoite surface protein 7. *Mol Biochem Parasitol* 2001;117:83–9. [PubMed: 11551634]
9. Trucco C, Fernandez-Reyes D, Howell S, Stafford WH, Scott-Finnigan TJ, Grainger M, Ogun SA, Taylor WR, Holder AA. The merozoite surface protein 6 gene codes for a 36 kDa protein associated with the *Plasmodium falciparum* merozoite surface protein-1 complex. *Mol Biochem Parasitol* 2001;112:91–101. [PubMed: 11166390]
10. Weber JL, Lyon JA, Wolff RH, Hall T, Lowell GH, Chulay JD. Primary structure of a *Plasmodium falciparum* malaria antigen located at the merozoite surface and within the parasitophorous vacuole. *J Biol Chem* 1988;263:11421–5. [PubMed: 3042768]
11. McColl DJ, Anders RF. Conservation of structural motifs and antigenic diversity in the *Plasmodium falciparum* merozoite surface protein-3 (MSP-3). *Mol Biochem Parasitol* 1997;90:21–31. [PubMed: 9497029]
12. Singh S, Soe S, Mejia JP, Roussillon C, Theisen M, Corradin G, Druilhe P. Identification of a conserved region of *Plasmodium falciparum* MSP3 targeted by biologically active antibodies to improve vaccine design. *J Infect Dis* 2004;190:1010–8. [PubMed: 15295710]
13. Oeuvray C, Bouharoun-Tayoun H, Gras-Masse H, Bottius E, Kaidoh T, Aikawa M, Filgueira MC, Tartar A, Druilhe P. Merozoite surface protein-3: a malaria protein inducing antibodies that promote *Plasmodium falciparum* killing by cooperation with blood monocytes. *Blood* 1994;84:1594–602. [PubMed: 8068948]
14. McColl DJ, Silva A, Foley M, Kun JF, Favaloro JM, Thompson JK, Marshall VM, Coppel RL, Kemp DJ, Anders RF. Molecular variation in a novel polymorphic antigen associated with *Plasmodium falciparum* merozoites. *Mol Biochem Parasitol* 1994;68:53–67. [PubMed: 7891748]
15. Pearce JA, Hodder AN, Anders RF. The alanine-rich heptad repeats are intact in the processed form of *Plasmodium falciparum* MSP3. *Exp Parasitol* 2004;108:186–9. [PubMed: 15582517]
16. Burgess BR, Schuck P, Garboczi DN. Dissection of merozoite surface protein 3, a representative of a family of *Plasmodium falciparum* surface proteins, reveals an oligomeric and highly elongated molecule. *J Biol Chem* 2005;280:37236–45. [PubMed: 16135515]
17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402. [PubMed: 9254694]
18. Bucher P, Karplus K, Moeri N, Hofmann K. A flexible motif search technique based on generalized profiles. *Comput Chem* 1996;20:3–23. [PubMed: 8867839]
19. Hofmann K, Bucher P. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem Sci* 1995;20:347–9. [PubMed: 7482699]
20. Schuck P. Sedimentation equilibrium analysis of interference optical data by systematic noise decomposition. *Anal Biochem* 1999;272:199–208. [PubMed: 10415089]
21. Schuck P. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys J* 2000;78:1606–1619. [PubMed: 10692345]
22. Vistica J, Dam J, Balbo A, Yikilmaz E, Mariuzza RA, Rouault TA, Schuck P. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal Biochem* 2004;326:234–256. [PubMed: 15003564]
23. Schuck P. On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal Biochem* 2003;320:104–124. [PubMed: 12895474]

24. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. Computer-aided interpretation of analytical sedimentation data for proteins. *Analytical Ultracentrifugation in Biochemistry and Polymer Science* 1992;90–125.
25. Garcia de la Torre J, Huertas ML, Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J* 2000;78:719–730. [PubMed: 10653785]
26. Pearce JA, Mills K, Triglia T, Cowman AF, Anders RF. Characterisation of two novel proteins from the asexual stage of *Plasmodium falciparum*, H101 and H103. *Mol Biochem Parasitol* 2005;139:141–51. [PubMed: 15664649]
27. Villard V, Agak GW, Frank G, Jafarshad A, Servis C, Nebie I, Sirima SB, Felger I, Arevalo-Herrera M, Herrera S, Heitz F, Backer V, Druilhe P, Kajava AV, Corradin G. Rapid identification of malaria vaccine candidates based on alpha-helical coiled coil protein motif. *PLoS ONE* 2007;2:e645. [PubMed: 17653272]
28. Johnson WC. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* 1999;35:307–12. [PubMed: 10328265]
29. Lau SY, Taneja AK, Hodges RS. Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J Biol Chem* 1984;259:13253–61. [PubMed: 6490655]
30. Lu SM, Hodges RS. Defining the minimum size of a hydrophobic cluster in two-stranded alpha-helical coiled-coils: effects on protein stability. *Protein Sci* 2004;13:714–26. [PubMed: 14978309]
31. Privalov PL, Potekhin SA. Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol* 1986;131:4–51. [PubMed: 3773768]
32. Yikilmaz E, Rouault TA, Schuck P. Self-association and ligand-induced conformational changes of iron regulatory proteins 1 and 2. *Biochemistry* 2005;44:8470–8. [PubMed: 15938636]
33. Brown PH, Balbo A, Schuck P. Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation. *Curr Protoc Immunol*. 2008Chapter 18:Unit 18 15
34. Harbury PB, Zhang T, Kim PS, Alber T. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* 1993;262:1401–7. [PubMed: 8248779]
35. Deng Y, Zheng Q, Liu J, Cheng CS, Kallenbach NR, Lu M. Self-assembly of coiled-coil tetramers in the 1.40 Å structure of a leucine-zipper mutant. *Protein Sci* 2007;16:323–8. [PubMed: 17189475]
36. Dayring HE, Tramonato A, Sprang SR, Fletterick RJ. Interactive program for visualization and modelling of proteins, nucleic acids and small molecules. *J Mol Graphics* 1986;4:82–87.
37. Akey DL, Malashkevich VN, Kim PS. Buried polar residues in coiled-coil interfaces. *Biochemistry* 2001;40:6352–60. [PubMed: 11371197]
38. Giraldo LE, Grab DJ, Wisner MF. Molecular characterization of a *Plasmodium chabaudi* erythrocyte membrane-associated protein with glutamate-rich tandem repeats. *J Eukaryot Microbiol* 1998;45:528–34. [PubMed: 9783454]
39. Terskikh AV, Le Doussal JM, Cramer R, Fisch I, Mach JP, Kajava AV. Peptabody™: a new type of high avidity binding protein. *Proc Natl Acad Sci U S A* 1997;94:1663–8. [PubMed: 9050835]
40. Mills KE, Pearce JA, Crabb BS, Cowman AF. Truncation of merozoite surface protein 3 disrupts its trafficking and that of acidic-basic repeat protein to the surface of *Plasmodium falciparum* merozoites. *Mol Microbiol* 2002;43:1401–11. [PubMed: 11952894]
41. Singh S, Soe S, Roussillon C, Corradin G, Druilhe P. *Plasmodium falciparum* merozoite surface protein 6 displays multiple targets for naturally occurring antibodies that mediate monocyte-dependent parasite killing. *Infect Immun* 2005;73:1235–8. [PubMed: 15664972]
42. Wisner MF, Sartorelli AC, Patton CL. Association of *Plasmodium berghei* proteins with the host erythrocyte membrane: binding to inside-out vesicles. *Mol Biochem Parasitol* 1990;38:121–34. [PubMed: 2181301]
43. Rodriguez LE, Curtidor H, Ocampo M, Garcia J, Puentes A, Valbuena J, Vera R, Lopez R, Patarroyo ME. Identifying *Plasmodium falciparum* merozoite surface antigen 3 (MSP3) protein peptides that bind specifically to erythrocytes and inhibit merozoite invasion. *Protein Sci* 2005;14:1778–86. [PubMed: 15987906]

(a)



▼ - cleavage site, ▨ - hydrophobic and NLR region, +++ - tandem repeats,  
 ■ - gly-containing motif, ☞ - acidic region, ☐ - coiled coil

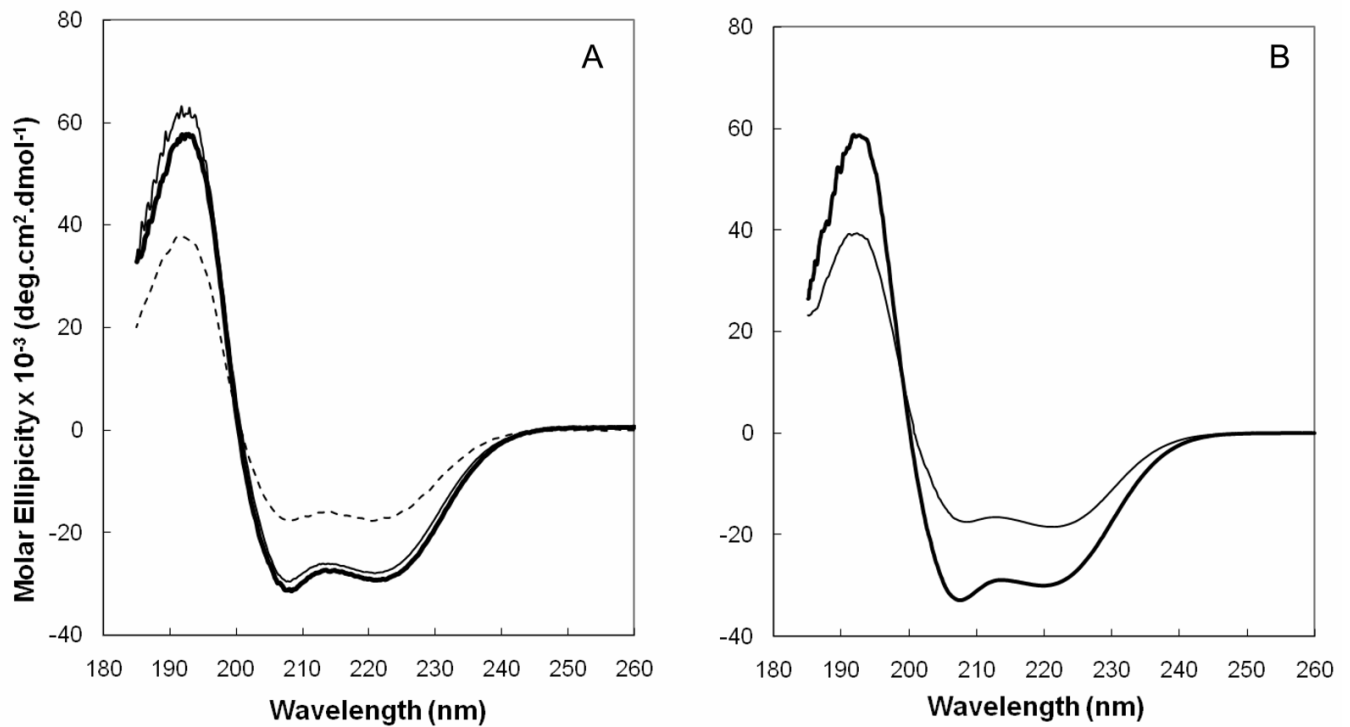
(b)

GLURP	KKSSFIT <sup>Y</sup> ISTK-KKKV <sup>S</sup> QTTVSVMINAYDGV <sup>I</sup> QV <sup>S</sup> TIKGI <sup>A</sup> KDIVIE <sup>F</sup> QNI	1233
MSP-3	QNL <sup>I</sup> ISK <sup>N</sup> QNNNEK <sup>N</sup> VKEA <sup>E</sup> ESIMK <sup>T</sup> LAG <sup>L</sup> IKGN <sup>O</sup> IDS <sup>T</sup> LKDI <sup>V</sup> EELSK <sup>Y</sup> EKN <sup>H</sup>	355
MSP-6	QNL <sup>I</sup> ISK <sup>N</sup> K <sup>K</sup> KND-E <sup>T</sup> KK <sup>T</sup> AENT <sup>V</sup> KT <sup>L</sup> VGL <sup>E</sup> NEK <sup>N</sup> E <sup>I</sup> DS <sup>T</sup> INN <sup>L</sup> VQ <sup>E</sup> MI <sup>H</sup> LE <sup>S</sup> NN	371
H101	YSL <sup>I</sup> IY <sup>K</sup> NYK <sup>D</sup> ND-KSE <sup>K</sup> TAQ <sup>T</sup> LIT <sup>A</sup> LIS <sup>L</sup> ING <sup>K</sup> NEL <sup>D</sup> AT <sup>I</sup> RR <sup>L</sup> K <sup>H</sup> R <sup>F</sup> ME <sup>F</sup> FT <sup>Y</sup> N	424
MSPDBL1	YR <sup>I</sup> LS <sup>V</sup> SYK <sup>D</sup> NN-E <sup>V</sup> KN <sup>V</sup> AE <sup>S</sup> IV <sup>K</sup> KL <sup>F</sup> SL <sup>E</sup> ND <sup>N</sup> N <sup>L</sup> ET <sup>I</sup> FK <sup>G</sup> L <sup>T</sup> ED <sup>M</sup> TD <sup>L</sup> EQ <sup>K</sup>	697
M712	HD <sup>Y</sup> D <sup>Y</sup> NY <sup>F</sup> END <sup>N</sup> YN <sup>I</sup> Q <sup>N</sup> V <sup>K</sup> D <sup>N</sup> L <sup>V</sup> KK <sup>V</sup> ND <sup>F</sup> ME <sup>S</sup> DN <sup>L</sup> V <sup>N</sup> TF <sup>K</sup> GI <sup>A</sup> GG <sup>V</sup> TSE <sup>F</sup> GY	712
M566	ND <sup>A</sup> SS <sup>E</sup> E <sup>I</sup> K <sup>D</sup> SS-DE <sup>K</sup> ESH <sup>E</sup> EL <sup>F</sup> K <sup>V</sup> F <sup>L</sup> E <sup>L</sup> INK <sup>N</sup> D <sup>L</sup> V <sup>K</sup> EN <sup>L</sup> KK <sup>I</sup> T <sup>N</sup> N <sup>L</sup> N <sup>E</sup> M <sup>H</sup> L <sup>S</sup> T <sup>L</sup> Y <sup>P</sup>	566
MSP-11	QK <sup>L</sup> IS <sup>N</sup> E <sup>Y</sup> K <sup>K</sup> T--E <sup>E</sup> KK <sup>S</sup> LE <sup>D</sup> H <sup>V</sup> N <sup>L</sup> L <sup>F</sup> N <sup>F</sup> L <sup>Q</sup> T <sup>N</sup> N <sup>O</sup> L <sup>D</sup> PS <sup>L</sup> K <sup>D</sup> LE <sup>N</sup> EL <sup>T</sup> FE <sup>T</sup> N <sup>Y</sup>	405
MSPDBL2	RNL <sup>I</sup> ISK <sup>N</sup> Y <sup>K</sup> NY <sup>N</sup> -E <sup>I</sup> DK <sup>N</sup> V <sup>H</sup> TL <sup>V</sup> NS <sup>I</sup> I <sup>S</sup> L <sup>L</sup> E <sup>E</sup> GN <sup>S</sup> DS <sup>T</sup> LN <sup>S</sup> LS <sup>K</sup> D <sup>I</sup> T <sup>N</sup> LE <sup>K</sup> N	762
LSA-1	YDE <sup>H</sup> IK <sup>K</sup> Y <sup>K</sup> KN <sup>V</sup> K-Q <sup>V</sup> N <sup>K</sup> E <sup>K</sup> E <sup>K</sup> FI <sup>K</sup> SL <sup>F</sup> HI <sup>F</sup> FD <sup>G</sup> NE <sup>T</sup> L <sup>Q</sup> I <sup>V</sup> DEL <sup>S</sup> E <sup>D</sup> IT <sup>K</sup> Y <sup>E</sup> M <sup>K</sup> L	1909
PV110965	ND <sup>S</sup> GI <sup>G</sup> -----S <sup>V</sup> TT <sup>E</sup> T <sup>E</sup> N <sup>L</sup> V <sup>D</sup> DL <sup>M</sup> N <sup>L</sup> LN <sup>S</sup> DN <sup>G</sup> V <sup>D</sup> Q <sup>S</sup> L <sup>K</sup> V <sup>L</sup> A <sup>Q</sup> DM <sup>A</sup> Q <sup>Y</sup> EL <sup>N</sup> H <sup>L</sup> NA <sup>E</sup> SE <sup>V</sup> A	441
PV097690	Q <sup>A</sup> LL <sup>E</sup> D <sup>N</sup> Y <sup>K</sup> N <sup>F</sup> T-DE <sup>K</sup> KK <sup>A</sup> E <sup>D</sup> L <sup>T</sup> K <sup>N</sup> I <sup>N</sup> T <sup>I</sup> D <sup>G</sup> AG <sup>V</sup> I <sup>E</sup> T <sup>L</sup> K <sup>D</sup> FA <sup>K</sup> DM <sup>N</sup> Q <sup>E</sup> IL <sup>N</sup> M	1127
PV097695	DK <sup>Q</sup> FG <sup>D</sup> E <sup>F</sup> DAY <sup>N</sup> -D <sup>I</sup> KK <sup>V</sup> T <sup>E</sup> AL <sup>V</sup> KT <sup>M</sup> TS <sup>L</sup> V <sup>S</sup> ED <sup>P</sup> S <sup>V</sup> GD <sup>I</sup> T <sup>N</sup> E <sup>F</sup> LS <sup>D</sup> M <sup>N</sup> Q <sup>L</sup> EL <sup>S</sup> W	899
PC104352	Q <sup>D</sup> K <sup>I</sup> N <sup>K</sup> -----D <sup>S</sup> KK <sup>E</sup> A <sup>K</sup> E <sup>I</sup> M <sup>K</sup> T <sup>L</sup> V <sup>T</sup> M <sup>F</sup> DD <sup>A</sup> D <sup>V</sup> I <sup>D</sup> E <sup>P</sup> V <sup>G</sup> D <sup>F</sup> S <sup>D</sup> DIS <sup>Q</sup> FE <sup>I</sup> I	307
PC101322	KK <sup>L</sup> ISE <sup>Y</sup> Y <sup>Q</sup> N <sup>M</sup> K-N <sup>I</sup> K <sup>N</sup> K <sup>I</sup> K <sup>D</sup> Y <sup>V</sup> KS <sup>A</sup> MD <sup>L</sup> ID <sup>G</sup> N <sup>N</sup> G <sup>I</sup> SK <sup>A</sup> FS <sup>N</sup> LS <sup>E</sup> D <sup>I</sup> S <sup>N</sup> E <sup>F</sup> L <sup>K</sup> L	351
PY01017	ND <sup>P</sup> I <sup>N</sup> A <sup>E</sup> SR <sup>N</sup> V <sup>V</sup> -E <sup>S</sup> K <sup>Q</sup> E <sup>A</sup> K <sup>E</sup> I <sup>M</sup> K <sup>T</sup> L <sup>V</sup> T <sup>M</sup> F <sup>D</sup> D <sup>V</sup> D <sup>V</sup> F <sup>D</sup> E <sup>P</sup> V <sup>F</sup> D <sup>F</sup> S <sup>D</sup> DIS <sup>Q</sup> FE <sup>I</sup> I	316
PB105994	KK <sup>L</sup> ISE <sup>Y</sup> Y <sup>A</sup> N <sup>M</sup> K-N <sup>I</sup> K <sup>N</sup> K <sup>I</sup> E <sup>D</sup> Y <sup>V</sup> KS <sup>M</sup> M <sup>N</sup> L <sup>I</sup> D <sup>S</sup> NY <sup>G</sup> I <sup>S</sup> K <sup>L</sup> LS <sup>N</sup> FS <sup>E</sup> N <sup>V</sup> S <sup>N</sup> E <sup>F</sup> Y <sup>N</sup> I	576
consensus	li yk eVkk ae IVktLV LIId n Id tik L dMs FF	
	<u>d a d a d a d a d a d</u>	

**Figure 1. A family of malaria surface proteins containing common C-terminal domain**

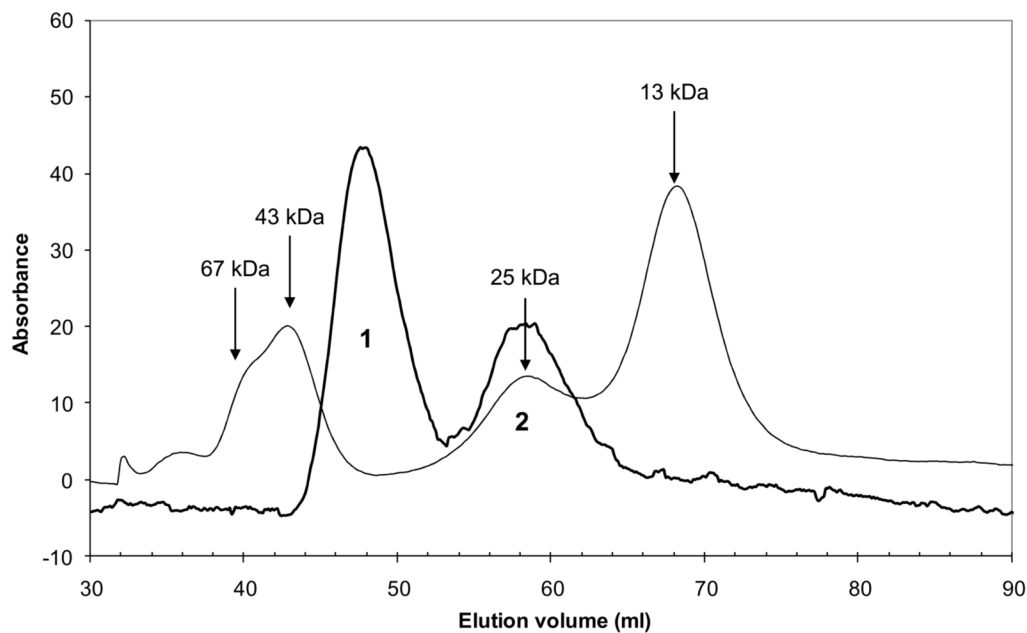
(a) A schematic diagram of the domain organization of *P. falciparum* proteins that contain a common C-terminal coiled coil domain. (b) Sequence alignment of the C-terminal domain of malaria proteins. Bold uppercase letters of the consensus indicate conserved apolar residues. The upper part of the alignment contains *P. falciparum* proteins: GLURP, glutamate rich protein (PF10\_0344), MSP3 (PF10\_0345), MSP6 (PF10\_0346), H101 (PF10\_0347), MSPDBL1 (PF10\_0348), M712 (PF10\_0350), M566 (PF10\_0351), MSP11 (PF10\_0352), MSPDBL2 (PF10\_0355), LSA-1, liver stage antigen-1 (PF10\_356). The lower part of the alignment contains *P. vivax* Sal-1 proteins: putative MSP3 (PV110965), putative MSP (PV097690), putative MSP3a (PV097695). *P. chabaudi* hyp. proteins PC104352 and

PC101322. *P. yoelii* latency associated antigen (PY01017). *P. berghei* hyp. protein PB105994. The conserved 40-residue region of the alignment with the (**abcdefg**)<sub>n</sub> heptad repeats is underlined and their **a** and **d** positions are shown.



**Figure 2. MSP3 and MSP6 adopt an  $\alpha$ -helical structure in a 100 mM sodium phosphate buffer at pH 7.3**

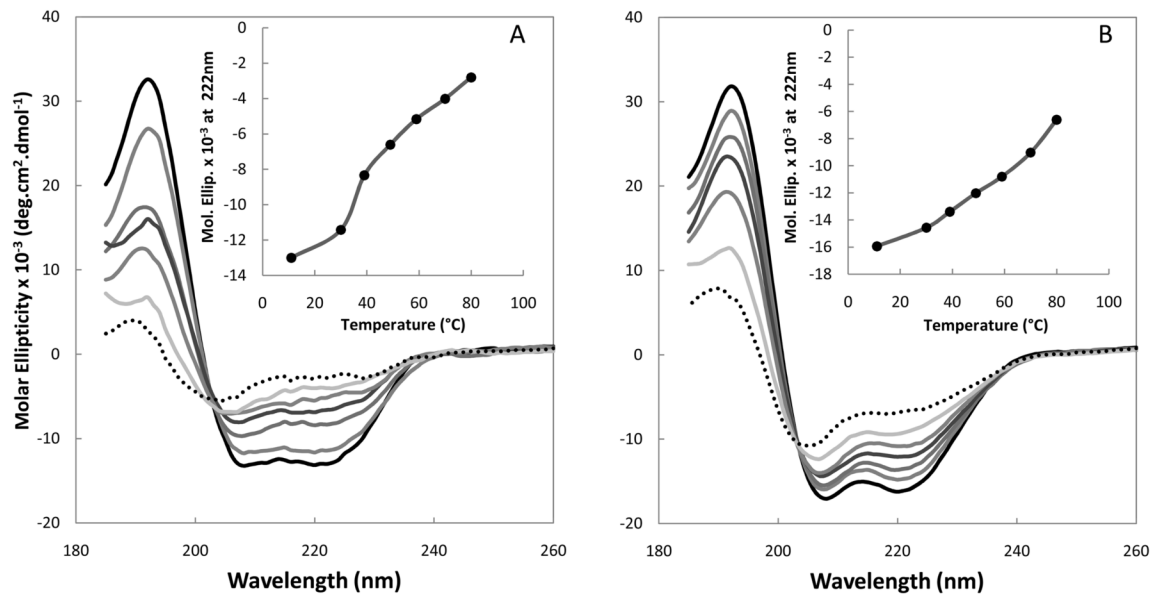
A) Circular dichroism spectra of peptide MSP3 (thin line, 45  $\mu$ M) and MSP6 (bold line, 25  $\mu$ M) were recorded immediately after solubilisation at 22  $^{\circ}$ C. A CD spectrum of the same solution of MSP3 was also recorded after incubation overnight at 4  $^{\circ}$ C (dashed line). B) Effect of trifluoroethanol (TFE). Circular dichroism spectra of peptide MSP3 were recorded in 100 mM sodium phosphate buffer at pH 6.1 at 8  $^{\circ}$ C (thin line, 49  $\mu$ M) and in the same solution + 50 % TFE (bold line, 25  $\mu$ M).



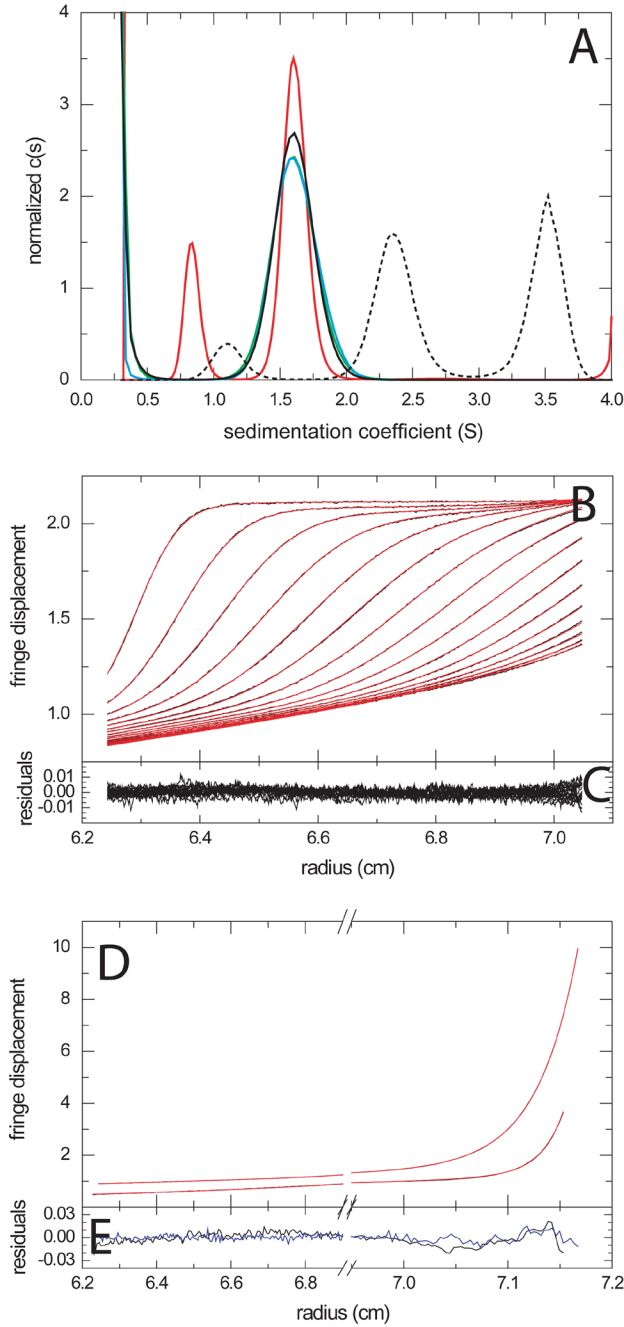
**Figure 3. Determination of apparent molecular weight by calibrated size exclusion chromatography**

The measurements were carried out on a Sephacryl S-100 size-exclusion column at 4°C. The elution profile of MSP3 peptide in a 100 mM sodium phosphate buffer (pH 6.1) is monitored by measuring absorbance at 230 nm (bold line). The protein standards used [albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa) and ribonuclease A (13.7 kDa)] were eluted in the same conditions (absorbance at 280 nm) (thin line).





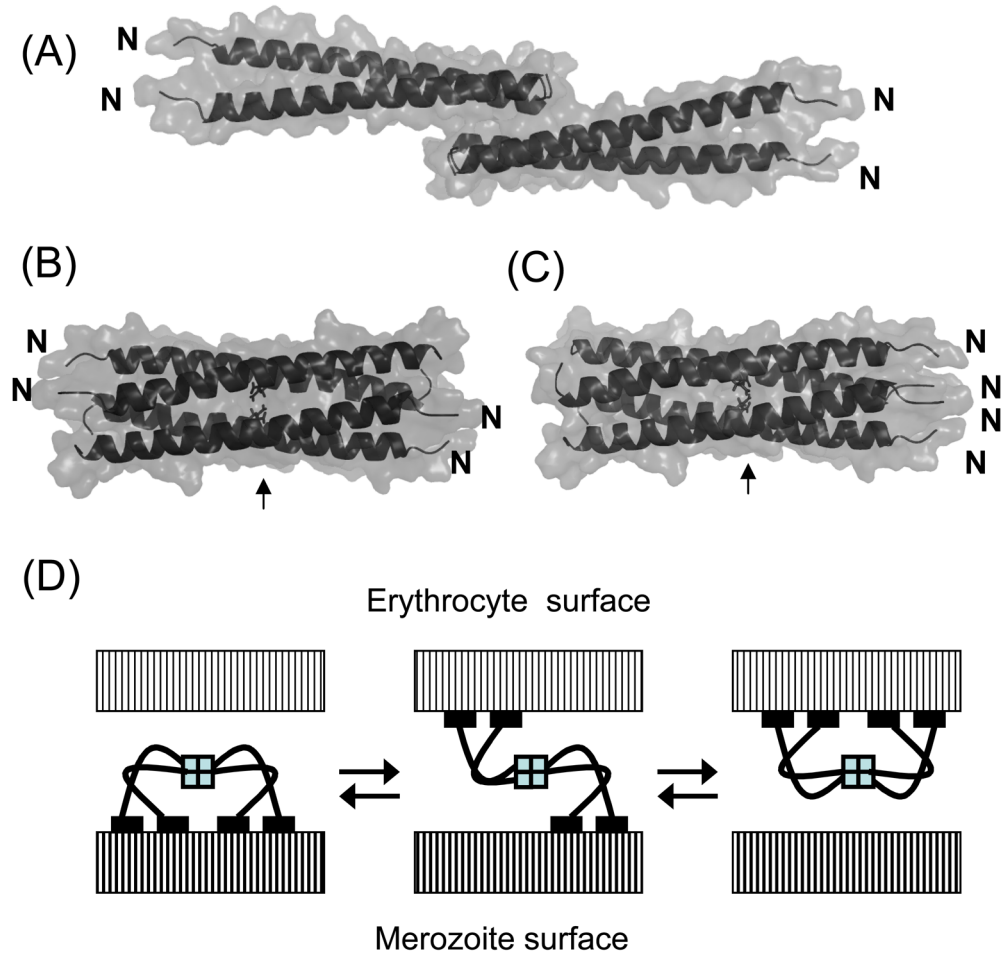
**Figure 4. Thermal denaturation of the two species isolated after size exclusion chromatography** CD spectra of each species obtained after size exclusion chromatography in 100 mM sodium phosphate buffer at pH 7.3 were recorded at increasing temperatures from 11°C (black solid line) to 80°C (dotted line). A) CD spectra of the species eluted in the first peak (elution volume of 48 ml). B) CD spectra of the species eluted in the second peak (elution volume of 60 ml). Insets show molar ellipticity of the species at 222 nm as a function of temperature.



**Figure 5. SE and SV analytical ultracentrifugation of MSP3**

(A) Calculated  $c(s)$  distributions derived from SV data recorded at 60,000 rpm, normalized by the loading signal, at concentrations of 2 (solid black line), 1 (solid blue line), 0.5 (solid green line), and 0.2 mg/ml (solid red line), respectively. The root-mean-square deviation (rmsd) of the  $c(s)$  analyses were  $\sim 0.003$  fringes throughout. Similar distributions were obtained from the analysis of the SV data at 40,000 rpm, conducted side-by-side with replicate sample cells (data not shown). Also shown is the  $c(s)$  distribution obtained for MSP6 at 2 mg/ml (black dotted line) (B) The evolution of the SV fringe profiles measured at all concentrations and rotor speeds were globally fit with a discrete species model accounting for buffer signal, a MSP3 dimer, and a MSP3 tetramer. The local rmsd to each data set was between 0.002 and 0.004

fringes. The full global fit is shown in the Supplemental Information Figure S3. As representative data from the global fit, displayed here are only the data at the highest concentration at 60,000 rpm, with every third scan shown (solid lines), and the calculated best-fit distributions (red dashed lines). (C) Residuals from the fit shown in Panel B. (D) Analogously to the SV analysis, all SE data were globally fit with a model comprised of buffer species, dimer and tetramer. The local rmsd for each data set ranged from 0.003 – 0.01 fringes. The full global fit is shown in the Supplemental Information Fig. S3. As representative data, displayed are the SE fringe profiles at 60,000 rpm for the 2 mg/ml loading concentration (black solid lines) and the best-fit model (red dashed lines). (E) Residuals of the data shown in Panel D.



**Figure 6. Structural models of the C-terminal tetramerization domain of MSP3**

(A) A model of  $\alpha$ -helical coiled coil dimers which interact with each other end-to-end and (B) side-by-side. (C) A parallel and in-register coiled coil tetramer. Arrows point to interior Asn residues.

(D) A schematic representation of a possible mechanism of merozoite adhesion to erythrocytes via MSP3. MSP3 oligomer may have several arrangements of its “arms” relative to the merozoite and erythrocyte (three of them are shown) which have a dynamic equilibrium between each other. At a given time, the “arms” of several MSP3 oligomers may bind to both merozoite and erythrocyte and link them.

**Table 1**

Circular dichroism data for MSP3 before and after size exclusion chromatography. Values of the  $\Theta_{222}/\Theta_{208}$  ratio were calculated from the CD spectra recorded at 8 °C.

$\Theta_{222}/\Theta_{208}$ ratio			
Buffer <sup>a</sup>	Before elution	After elution (elution volume)	
		Peak 1 <sup>c</sup>	Peak 2 <sup>c</sup>
A	0.97	0.98	0.94
B	1.01	1.09 <sup>b</sup>	0.95

<sup>a</sup> Buffer A: 100 mM sodium phosphate pH 7.3; Buffer B: 100 mM sodium phosphate pH 6.1

<sup>b</sup> In the presence of 50 % TFE, this value decreases to 0.91.

<sup>c</sup> Ratios were calculated from CD spectra of samples corresponding to elution volumes of 48mL (Peak 1) and 60 mL (Peak 2).

Table 2

Fit of models to the SE and SV data.

Model <sup>(a)</sup>	SE	$\chi^2$ <sup>(b)</sup>	SV	$\chi^2$ <sup>(c)</sup>
	best-fit Mw of largest species (kDa)		best-fit Mw of largest species (kDa)	s-value of largest species (S)
1 - 3	n/a	107	n/a	1.774
1 - x	19.0	2.05	17.9	1.785
<b>2 - x<sup>(c)</sup></b>	<b>20.2</b>	<b>1.61</b>	<b>19.2</b>	<b>1.818</b>
1 - 2 - x	19.7	1.02	19.2	1.816
1 - 2 - 3 - 4	n/a	0.968	n/a	1.827

<sup>(a)</sup>The model additionally includes the small Mw component accounting for the signal contribution characteristic for buffer salts detected by the interference optical system.

<sup>(b)</sup>The  $\chi^2$  value is a reduced  $\chi^2$  (based on 434,000 data points in SV and 2,113 data points in SE, respectively), operationally assuming standard deviations of the data acquisition to be uniformly 0.005 fringes for interference optical acquisition in both SV and SE. On this scale, based on F-statistics, the critical  $\chi^2$  values for a two standard deviations confidence interval is at 1.03 for SE, and 0.381 for SV, respectively

<sup>(c)</sup>This is the fit shown in Fig. 5B/D and S3/S4.