The protozoan parasite *Plasmodium* causes malaria, with hundreds of millions of cases recorded annually. Protection against malaria infection can be conferred by antibodies against merozoite surface protein (MSP)-1, making it an attractive vaccine candidate. Here we present the structure of the C-terminal domains of MSP-1 (known as MSP-119) from *Plasmodium knowlesi*. The structure reveals two tightly packed epidermal growth factor-like domains oriented head to tail. In domain 1, the molecule displays a histidine binding site formed primarily by a highly conserved tryptophan. The protein carries a pronounced overall negative charge primarily due to the large number of acidic groups in domain 2. To map protein binding surfaces on MSP-119, we have analyzed the crystal contacts in five different crystal environments, revealing that domain 1 is highly preferred in protein-protein interactions. A comparison of MSP-119 structures from *P. knowlesi*, *P. cynomolgi*, and *P. falciparum* shows that, although the overall protein folds are similar, the molecules show significant differences in charge distribution. We propose the histidine binding site in domain 1 as a target for inhibitors of protein binding to MSP-1, which might prevent invasion of the merozoite into red blood cells.

A necessary step in the life cycle of the malaria parasite is its entry into the red blood cell of a mammalian host. This process involves many surface proteins specifically expressed at the merozoite stage of the parasite, including merozoite surface protein (MSP)-1 (also known as merozoite surface antigen-1 or MSA-1). MSP-1, a required component of the merozoite invasion machinery, associates noncovalently with other parasite proteins, including the C-terminal portion of MSP-6 (1). MSP-1 is synthesized as a 180–225-kDa polypeptide capable of binding sialic acid (2, 3). As a requirement for merozoite entry into a red cell, MSP-1 undergoes two processing steps, the first at merozoite release from an infected cell and the second during invasion of a red cell (4) (Fig. 1a). In the first processing step, the precursor polypeptide is cleaved into four chains (with apparent molecular masses of 83, 30, 38, and 42 kDa) held together by noncovalent contacts (4–6). In the second processing step, the 42-kDa segment (MSP-142) is cleaved into two fragments with apparent molecular masses of 33 kDa (MSP-133) and 19 kDa (MSP-119 or MSP-1inv) (4). MSP-133 sheds from the surface, whereas MSP-119 remains anchored to the merozoite membrane by a glycosylphosphatidylinositol tail attached to the C-terminal residue (7–10).

Because high antibody titer against the C-terminal region of MSP-1 protects against severe malaria, this molecule represents an attractive vaccine candidate (11–13). Antibodies against MSP-119 correlate with clinical immunity (14, 15); however, recombinant MSP-119 is poorly immunogenic, presumably due to poor processing of the highly disulfide-bonded protein (16, 17). Antibodies to the C-terminal region of MSP-1 fall into two categories: protective antibodies that inhibit merozoite invasion of red cells, and blocking antibodies that enhance invasion by interfering with the binding of protective antibodies. Protective antibodies include those that recognize the native conformation of MSP-119 (7, 8, 20), those that block proteolytic processing of MSP-142 (22), and certain antibodies to the 83-kDa N-terminal fragment of intact MSP-1 (22). Blocking antibodies bind to the first domain of MSP-119, near the N terminus introduced by the processing of MSP-142 (22, 23).

The extremely stable MSP-119 contains two consecutive epidermal growth factor (EGF)-like domains, a small disulfide-rich fold also found in diverse proteins including Factor Xa (24), E-selectin (25), cyclooxygenase-2 (26), and integrin αvβ3 (27). In MSP-1, disulfide bonds are integral to the correct folding of the molecule (20). Although the full-length MSP-1 is highly polymorphic, the C-terminal EGF-like domains in MSP-119 are much less so, and they represent a conserved segment of a variable molecule. MSP-119 shows high sequence identity across *Plasmodium* species: *P. knowlesi* MSP-119 shares 82% sequence identity with *P. cynomolgi* and 51% sequence identity with *P. falciparum*.

To better understand the process of invasion of the *Plasmodium* merozoite into a red blood cell, we have solved the structure of the C-terminal domains of MSP-1 from *P. knowlesi* by x-ray crystallography. We have refined the structure to an R-factor of 23.4% and R-free of 26.4% to 2.4 Å resolution (Table I). Our structure reveals features of MSP-1 novel to *P. knowlesi* and provides new insight into MSP-1 from other *Plasmodium* species. With four copies in the crystallographic asymmetric unit, we observe MSP-1 in four different environments, allowing us to infer novel properties of the molecule.
Expression and Purification—DNA encoding the C-terminal 92 amino acids of the of P. knowlesi (Malayan H strain) MSP-1 (GenBank™ accession code AAC24615) was inserted into the vector YEpRPEU-3 and expressed in the Saccharomyces cerevisiae BankTM accession code AAG24615) was inserted into the vector P. knowlesi

**MATERIALS AND METHODS**

Crystallization and X-ray Data Collection—Purified MSP-1p protein was concentrated to 10 mg/ml for crystallization trials. Crystals were grown via vapor diffusion with a solution containing 30% polyethylene glycol 6000, 100 mM HEPES, pH 7.0. Crystals were transferred into the same solution supplemented with 20% glycerol before cooling to 100 K in a nitrogen stream. X-ray data were collected on a Rigaku RU-200 rotating anode generator equipped with nickel mirrors and a RAXIS 4 detector (Molecular Structure Corp., The Woodlands, TX). 180° of diffraction data were processed using DENZO (37), SCALEPACK (37), and TRUNCATE (38) to a resolution limit of 2.4 Å.

Phasing and Refinement—Molecular replacement was performed with the AMoRe package (38) using coordinates from the C-terminal domains of MSP-1 from P. cynomolgi (29). The P. cynomolgi model was rotated and translated against the 8–4 Å P. knowlesi diffraction amplitudes, using correlation coefficients to rank solutions. Four successive objects were placed in the crystallographic asymmetric unit, whereas insertion of a fifth object failed to improve the correlation coefficient. Inspection of the packing showed no steric clashes in a unit cell with 49% solvent content. Rigid body refinement in the programs AMoRe (38) and Crystallography and NMR System (CNS) (39) was followed by model building in the program O (40). Residue numbering of the P. knowlesi structure corresponds to the mature secreted protein and begins at the first residue of the cleaved a mating factor secretory signal. Refinement protocols in CNS included conjugate gradient minimization, simulated annealing, and temperature factor refinement. Models were built into simulated annealing composite omit maps calculated in CNS. Tight 4-fold noncrystallographic symmetry restraints (300 kcal/mol-Å2) were imposed on all atoms in the early stages of refinement and later relaxed for atoms that differ among the four copies in the asymmetric unit. Refinement steps were accepted only if they reduced the R-free (of a test set comprised of 280 reflections, 5% of the total, selected using resolution shells). The R-work and R-free are 23.4% and 26.4%, respectively, using all reflections to 2.4 Å.

**Calculations**—Least squares superpositions were performed using the program LSQMAN (41) with a distance cutoff of 3.8 Å, and coordinate transformations were applied using the program MOLEMAN2 (41). Structural comparisons with the NMR structure of P. falciparum MSP-1p (28) use the most representative member of the ensemble. Electrostatic calculations include only the native MSP-1p sequence, without the secretory signal and purification epitope. Molecular figures were prepared using the programs MOLSCRIPT (42), BOBSCRIPT (43), and GRASP (44).

**RESULTS**

Overall Description of the Structure—The C-terminal region of P. knowlesi MSP-1 contains two successive EGF-like domains packed together tightly and related by a rotation of ~170° (Fig. 1b). The molecule is roughly square-shaped, with dimensions of about 35 Å × 35 Å × 15 Å. The angle between domains 1 and 2 leads to close approach of the N terminus of domain 1 and the C terminus of domain 2. Thus, as domain 2 is anchored in the membrane by a glycosylphosphatidylinositol linkage attached to the C-terminal residue, the N-terminal residue of domain 1 (i.e. the cleavage site in MSP-1p) also points toward the membrane, in an appropriate orientation for processing by proteases attached to the merozoite membrane (28).

Each EGF-like domain in P. knowlesi MSP-1 contains a segment of random coil followed by four antiparallel β strands (Fig. 1b). Typically, EGF-like domains contain six cysteines in three disulfide bonds with a connectivity of 1–3, 2–4, and 5–6. The MSP-1 EGF-like domains follow this paradigm, except that
in domain 1, valine and tryptophan replace the middle (2–4) disulfide. The valine and tryptophan pack into the domain 1 hydrophobic core, in the same volume as the disulfide bond in canonical EGF-like domains.

The crystals of *P. knowlesi* MSP-119 contain four molecules in the asymmetric unit, revealing the protein in four different environments. The four molecules are highly similar, with r.m.s. deviations of 0.4–0.6 Å for 77–81 Ca atoms, with differences primarily at the two termini and residues 69–74 on the loop between β strands 1 and 2 in domain 2 (Fig. 1c). This loop shows differences across the four molecules in the crystallographic asymmetric unit and also within one molecule (as reflected by high B factors). The flexibility in this loop and in the termini corresponds to the mobility seen in the NMR structure of the *P. falciparum* MSP-119 (28).

In *P. knowlesi* MSP-119, the extensive interface between the two domains buries 26 residues and 1090 Å² of surface area. This produces a rigid interface with little variation in the interdomain angle among the four copies in the asymmetric unit (Fig. 1c). The residues in the interface are among the most conserved in the MSP-119 protein: 23 of 26 residues are identical between *P. knowlesi* and *P. cynomolgi*, with three conservative substitutions; and 14 of 26 residues are identical between *P. falciparum* and *P. knowlesi* (28). We have observations of MSP-119 in five different crystal environments. We counted the number of times each residue participates in crystal contacts and mapped that number onto the surface of the protein (Fig. 3a). The *P. knowlesi* and *P. cynomolgi* crystal contacts correlate: of the 9 residues involved in crystal contacts in all four *P. knowlesi* copies, 7 make contacts in the *P. cynomolgi* crystals as well. The seven residues (Ile14, Asp15, Leu28, Trp34, Glu52, Ala50, and Ser51) participate in crystal contacts in all five crystal packing environments. All of these except Ser51 are conserved between *P. knowlesi* and *P. cynomolgi*, suggesting protein binding sites conserved across species. All of these except Leu28 differ between *P. knowlesi* and *P. falciparum*, resulting in distinctly shaped MSP-119 surfaces across these species.

**Histidine Binding Site**—In all four copies of MSP-119 in the crystallographic asymmetric unit, electron density appears in a shallow depression on the surface (Fig. 2a). In each copy, the density traces back to a histidine from the hexahistidine tag of a symmetry-related molecule. For each of the four bound histidines, the plane of the imidazole ring stacks roughly parallel to the plane of the indole ring of Trp54, and the carboxylic acid of Glu52 pairs with the bound (and presumably protonated) histidine. In the histidine-binding pocket, Trp54 forms the floor, and Glu52 and Arg55 form the walls (Fig. 2b). The hydrophobic tryptophan and the bound histidine appear in the middle of a region of negative electrostatic potential (Fig. 2c). Although the conformation of the C-terminal hexahistidine tail differs in the four copies in the asymmetric unit, one histidine from the tail fills a histidine-binding pocket of a symmetry-related molecule. In the four copies in the asymmetric unit, each histidine binding site in the center of domain 1, falls in an ideal position to participate in protein-protein interactions on the surface of the merozoite. In *P. falciparum* MSP-119, domain 1 has been shown to be the binding site of monoclonal antibody G17.12 (33).

**Comparison across Species**—The high level of sequence identity in MSP-119 across *Plasmodium* species results in highly similar three-dimensional structures of the molecule (Figs. 3b and 4a). The *P. cynomolgi* MSP-119 superimposes on the *P. knowlesi* structure with an r.m.s. deviation of 1.3 Å for 78 Ca atoms, whereas the *P. falciparum* structure superimposes on the *P. knowlesi* structure with an r.m.s. deviation of 1.7 Å for 65
atoms (Table II). A superposition of the four
P. knowlesi
MSP-119 molecules with the
P. cynomolgi
and
P. falciparum
homologues reveals that the differences occur primarily at the
termini and in the 69–74 loop, the same loop that shows var-
iability among the four molecules in the
P. knowlesi
crystal.

MSP-119 from
P. knowlesi
exhibits a striking charge distri-
bution (Fig. 4).
The molecule carries a pronounced negative
potential, with an overall charge of
−7 in the 92 residues of the
two domains. There are no buried ion pairs in
P. knowlesi
MSP-119; rather, all charged residues are surface-accessible.
The overall charge is species-specific because the equivalent
segments of MSP-1 from
P. cynomolgi
and
P. falciparum
have overall charges of −4 and −2, respectively (Fig. 4, c and d). In

Hydrophobicity maps of the three
Plasmodium
MSP-119 struc-
tures show a patch of nonpolar residues along the interface
between the two domains (Fig. 4, b, c, and d). Of the 26 residues con-
tributing to the interface, 15 are hydrophobic, and 3 are cysteines
in disulfide bonds. Because the molecule is only 15 Å thick, some
hydrophobic residues in the interdomain interface (including
Tyr25, Tyr27, Leu37, Leu38, and Phe88) remain partially solvent-
accessible. Surprisingly, the large patch of hydrophobic residues
on the back surface is among the least active surfaces in protein
binding, as judged by the crystal contact analysis (Fig. 3).

The arrangement of charged and hydrophobic residues on
MSP-119 may maintain specificity of the interactions between
MSP-119 and its binding partners such as the other MSP-1
fragments and MSP-6. Complementary changes between resi-
dues in MSP-119 and its binding partners may allow evolution
of the MSP-119 protein surface.

DISCUSSION

The region around Trp34 appears to be a site on
P. knowlesi
MSP-119 well suited for protein-protein interactions. First, all
EGF-like domain structures contain at least three pairs of
disulfide bonds, with one exception: the first domain of MSP-
Of the 40 independent EGF/laminin-like domain structures currently in the Protein Data Bank (34), only *Plasmodium* MSP-119 has lost one of its three canonical disulfide bonds. The loss of a structurally important and highly conserved disulfide bond is most likely due to a functional requirement. Second, all of the *Plasmodium* species (except *P. falciparum*) contain a tryptophan in domain 1 of MSP-119 in place of the canonical cysteine. The disulfide is replaced with a tryptophan and a smaller residue; whereas the smaller residue may become a valine, isoleucine, or threonine, the other member of the pair is always tryptophan, suggesting functional selection. Third, tryptophan is the residue most commonly found in protein-protein interfaces (30), so the replacement of a strictly conserved disulfide-forming cysteine with a tryptophan suggests the evolution of a protein binding site. Fourth, there are no instances of crystals of MSP-119 with an empty histidine-binding pocket. In all five independent crystallographic environments, a histidine or lysine is found stacked above the tryptophan ring. This pocket fills when the concentration of nearby protein becomes sufficiently high. Fifth, the crystal contact map of MSP-119 indicates that the tryptophan and nearby regions in domain 1 are very active in participating in protein-protein interactions.

The sequence differences between *P. falciparum* and the other *Plasmodium* may be due to the strong bias against tryptophan in *P. falciparum* (35). The virulent *P. falciparum* displays somewhat different disease pathology compared with the other *Plasmodium* species; perhaps the differences in MSP-119 proteins are related to the differences in disease severity. For example, *P. falciparum* invades a broader spectrum of red blood cells and reaches higher levels of parasitemia compared with other malaria species; potentially due to loss of receptor specificity in *P. falciparum*. Some of this specificity may be conferred by the MSP-1 protein during red blood cell invasion. Although *P. falciparum* lacks the tryptophan defining the floor of the binding pocket, all the other species of *Plasmodium* have this tryptophan, which generates a site for binding histidine. We propose the use of this histidine binding site as a target for the design of compounds that bind, for example, *P. vivax* MSP-1. From such a scaffold compound, new inhibitors that protect against merozoite invasion of a red blood cell might be developed.

Our structure reveals a new feature of the MSP-119 molecule, a histidine-binding pocket in domain 1. Because antibodies that protect against malarial infections often bind specifically to domain 1, this new property of the molecule might be used to engineer tighter binding and thus more protective antibodies. Alternatively, this histidine-binding pocket might be used to interfere with the binding of blocking antibodies. For example, a small molecule inhibitor containing an imidazole scaffold may be designed to overlap with the footprint of blocking an-

![Fig. 4. Electrostatic potential and hydrophobicity.](image-url)

*a*, the amino acid sequences from the crystals of MSP-119 from *P. knowlesi*, *P. cynomolgi*, and *P. falciparum*. Cysteines are highlighted in yellow. *b–d*, surface representations of MSP-119 from three *Plasmodium* species are shown, colored by electrostatic potential (left) and hydrophobicity (right). Molecules are shown in two views 180° apart. Electrostatic potentials contoured from −10 kT (red) to +10 kT (blue) are plotted onto the molecular surface of MSP-119, from *P. knowlesi*, *P. cynomolgi*, and *P. falciparum*, respectively. Hydrophobicity values calculated from modified solvent transfer experiments (45) are contoured from −2 kcal/mol (most hydrophobic, in yellow) to 2 kcal/mol (hydrophilic, in green). The hydrophobic patches found on the surface tend to be from exposed portions of residues forming the interface between domains 1 and 2.
tibodies and thus disrupt the binding of blocking antibodies. The specific disruption of MSP-1\textsubscript{19} blocking antibodies without perturbation of protective antibodies would greatly enhance the immune response against the malaria parasite.

Acknowledgments—We thank T. Allison and the other members of the Structural Biology Section for discussions.

REFERENCES

1. Trucco, C., Fernandez-Reyes, D., Howell, S., Stafford, W. H., Scott-Finnigan, T. J., Grainger, M., Ogun, S. A., Taylor, W. R., and Holder, A. A. (2001) Mol. Biochem. Parasitol. 112, 91–101
2. Perkins, M. E., and Rocco, L. J. (1988) J. Immunol. 141, 3190–3196
3. Su, S., Sanadi, A. R., Ifon, E., and Davidson, E. A. (1993) J. Immunol. 151, 2309–2317
4. Holder, A. A., Blackman, M. J., Burghaus, P. A., Chappel, J. A., Ling, I. T., McCallum-Deighton, N., and Shai, S. (1992) Mem. Inst. Oswaldo Cruz 87, Suppl. 3, 57–42
5. Holder, A. A., Sandhu, J. S., Hillman, Y., Davey, L. S., Nicholls, S. C., Cooper, H., and Lockyer, M. J. (1987) Parasitology 94, 199–208
6. McBride, J. S., and Heidrich, H. G., Donachie, S., McBride, J. S., and Holder, A. A. (1990) J. Exp. Med. 172, 379–392
7. Chang, S. P., Gibson, H. L., Lee-Ng, C. T., Barr, P. J., and Hui, G. S. (1992) Mol. Biochem. Parasitol. 49, 29–33
8. Blackman, M. J., Chappel, J. A., Riley, E. M., Druilhe, P., and Holder, A. A. (1994) J. Exp. Med. 179, 307–315
9. Blackman, M. J., Ling, I. T., Nicholls, S. C., and Holder, A. A. (1991) Mol. Biochem. Parasitol. 49, 35–44
10. al-Yaman, F., Genton, B., Howell, S., Stafford, W. H., Scott-Finnigan, T. J., Grainger, M., Ogun, S. A., Taylor, W. R., and Holder, A. A. (2001) Mol. Biochem. Parasitol. 112, 91–101

Structure of P. knowlesi MSP-1\textsubscript{19}