Published in final edited form as:

J Mol Biol. 2008 May 23; 379(1): 105–121. doi:10.1016/j.jmb.2008.03.039.

Solution conformation, backbone dynamics and lipid interactions of the intrinsically unstructured malaria surface protein MSP2

Xuecheng Zhang^{1,2}, Matthew A. Perugini^{3,4}, Shenggen Yao¹, Christopher G. Adda², Vincent J. Murphy², Andrew Low¹, Robin F. Anders², and Raymond S. Norton^{1,*}

¹ The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville 3050, Australia

² Department of Biochemistry, La Trobe University, Bundoora, 3086, Australia

³ Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, 3010, Australia

⁴ Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville, 3010, Australia

Abstract

Merozoite surface protein 2 (MSP2), one of the most abundant proteins on the surface of the merozoite stage of *Plasmodium falciparum*, is a potential component of a malaria vaccine, having shown some efficacy in a clinical trial in Papua New Guinea. MSP2 is a GPI-anchored protein consisting of conserved N- and C-terminal domains and a variable central region. Previous studies have shown that it is an intrinsically unstructured protein with a high propensity for fibril formation, in which the conserved N-terminal domain plays a key role. Secondary structure predictions suggest that MSP2 contains long stretches of random coil with very little α -helix or β -strand. Circular dichroism spectroscopy confirms this prediction under physiological conditions (pH 7.4) and in more acidic solutions (pH 6.2 and 3.4). Pulsed field gradient NMR diffusion measurements showed that MSP2 under physiological conditions has a large effective hydrodynamic radius consistent with an intrinsic pre-molten globule state, as defined by Uversky. This was supported by sedimentation velocity studies in the analytical ultracentrifuge. NMR resonance assignments have been obtained for FC27 MSP2, allowing the residual secondary structure and backbone dynamics to be defined. There is some motional restriction in the conserved C-terminal region in the vicinity of an intramolecular disulfide bond. Two other regions show motional restrictions, both of which display helical structure propensities. One of these helical regions is within the conserved N-terminal domain, which adopts essentially the same

* To whom correspondence should be addressed. ray.norton@wehi.edu.au. FAX: +61 3 9345 2686. Phone +61 3 9345 2306.

Accession numbers

Resonance assignments have been deposited in BioMagResBank36 with accession number 15590.

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.

conformation in full-length MSP2 as in corresponding peptide fragments. We see no evidence of long-range interactions in the full-length protein. MSP2 associates with lipid micelles, but predominantly through the N-terminal region rather than the C-terminus, which is GPI-anchored to the membrane in the parasite.

Keywords

intrinsically unstructured protein; merozoite surface protein 2; NMR; backbone dynamics; lipid interactions

Introduction

Malaria is one of the most important infectious diseases in the world resulting in half a billion clinical episodes and as many as two million deaths annually.^{1–3} Merozoite surface protein 2 (MSP2), one of the most abundant proteins on the merozoite surface of *Plasmodium falciparum*, has been shown to be a promising vaccine target.^{4,5} MSP2 is a glycosyl-phosphatidyl-inositol (GPI) anchored protein with conserved N- and C-terminal domains and a variable central region (Figure 1). The patterns of diversity in this central region have led to all MSP2 alleles being classified into two families, FC27 and 3D7 (Figure 1).^{6–11} Recent studies have shown that full-length MSP2 is an intrinsically unstructured protein (IUP) and prone to form amyloid-like fibrils, in which the N-terminal region (MSP2_{1–25}) plays a key role.^{12–14} In addition, there is evidence that MSP2 exists in an oligomeric form on the merozoite surface.¹²

Intrinsically unstructured proteins are widespread,^{15–17} and the proteome of the malaria parasite is predicted to contain a remarkably high proportion of this class of proteins.¹⁸ Some IUPs are involved in important regulatory functions and may only fold into ordered structures upon recognition of their biological partners, while others are prone to form amyloid fibrils which are associated with neurodegenerative conditions and systemic or localized amyloidoses.¹⁹ Recently, some IUPs were found to form functional amyloid-like fibrils on bacterial and fungal surfaces where they may facilitate attachment to host cells or other surfaces.^{19,20}

Understanding the structure and dynamics of the unstructured state of IUPs can provide valuable insight into the conformational space available to these proteins and ultimately to their function.^{21,22} This class of proteins cannot be described adequately by a single conformation, but rather by an ensemble of rapidly interconverting conformations.²³ Indeed, IUPs are unsuitable for study by X-ray crystallography because of their disordered nature, but high-resolution NMR can provide detailed insights into their structure and dynamics, and studies of many such proteins have been reported.^{24–30} Experimental challenges associated with NMR studies of disordered proteins include extensive peak overlap, particularly for proton resonances, and their tendency to aggregate in solution. On the other hand, the substantial intrinsic dispersion of ¹⁵N and ¹³O chemical shifts, the fact that chemical shifts tend to be close to random coil values, and the typically sharp peaks in spectra of disordered proteins all favour resonance assignments, which are crucial for detailed NMR studies of proteins.²⁴

The aims of this study were to characterize the structure and stability of recombinant MSP2 in aqueous solution by employing circular dichroism (CD) spectroscopy, analytical ultracentrifugation and NMR spectroscopy. Our results show that MSP2 is largely unfolded, but with some nascent structures present under all the conditions investigated. Residue-specific NMR studies on the FC27 isoform of MSP2 defined three motionally restricted regions in the protein. We further identified a key role of the N-terminal helical region in the interaction of FC27 MSP2 with lipid micelles. The results provide a clear picture of the solution properties of monomeric FC27 MSP2, and provide a basis for understanding the nature of MSP2 on the merozoite surface.

Results

Sequence analysis

Previous studies have shown that MSP2 is an intrinsically unstructured protein (IUP).¹² The sequence characteristics of intrinsically unstructured proteins include the presence of numerous uncompensated charged groups (often negative) and a low content of hydrophobic residues. This allows the estimation of the 'boundary' mean hydrophobicity value $\langle H \rangle_b$, below which a polypeptide chain with a given mean net charge $\langle R \rangle$ will most probably be unfolded: $\langle H \rangle_b = (\langle R \rangle + 1.151)/2.785$ (Figure S1, Supplementary Material). Both the 3D7 and FC27 isoforms of MSP2 possess much lower mean hydrophobicity values than the boundary value predicted from their mean net charges (Figure S1, Supplementary Material), indicating that they are likely to be unstructured under physiological conditions.

The secondary structures of the 3D7-6H and FC27-6H MSP2 isoforms were predicted from their amino acid sequences. The GOR IV31 output (Figure S2, Supplementary Material) shows that both isoforms are predicted to be largely unstructured, with extensive regions of random coil and minimal structured regions comprised of α -helix and β -strand/sheet. Similar outputs were obtained with the nnPredict algorithm (data not shown).³²

CD spectroscopy

CD spectra of purified recombinant 3D7-6H and FC27-6H isoforms of MSP2 were recorded at pH 3.4, 6.2 and 7.4 for freshly prepared samples and also following 7 days incubation at room temperature (Figure 2). The spectra for 3D7-6H under all conditions examined are remarkably similar, showing a single minimum at approximately 198 nm, indicative of a largely unstructured protein (Figure 2a). However, there is also a broad but weak negative ellipticity between 215 and 235 nm, indicative of a small proportion of residual α -helical and/or β -strand structure (Figure 2a). Nonlinear least squares analyses confirm this assertion, yielding excellent fits ($\text{rmsd} < 0.17$; fits not shown) to approximately 5–6% α -helix, 1–2% β -strand, 1–2% turn and 90–92 % disordered structure. These data indicate that the 3D7-6H isoform remains equivalently disordered over the pH range 3.4 to 7.4 even after incubation for 7 days at room temperature. These results were all obtained in buffers containing 150 mM NaCl, but in the absence of salt the spectra were essentially identical (data not shown).

The CD spectra obtained for freshly prepared FC27-6H at pH 3.4, 6.2 and 7.4 showed a single minimum at approximately 198 nm, indicating that this isoform is also largely

unstructured in solution (Figure 2b). However, following 7 days incubation at room temperature at acidic pH in the presence of 150 mM NaCl, the CD spectrum of FC27-6H changed significantly, with a broader and more negative minimum observed between 205 and 220 nm, suggesting that the protein had become more structured in solution. This is confirmed by nonlinear least squares regression analyses, which indicate that the FC27-6H isoform adopts approximately 4-fold greater α -helical structure and more than 10-fold greater β -structure following 7 days incubation at pH 3.4 (Figure 2c). By contrast, no significant changes were observed in the CD spectra of FC27-6H following 7 days incubation at pH 6.2 and pH 7.4 (Figure 2b).

Analytical ultracentrifugation

Sedimentation velocity studies of 3D7-6H and FC27-6H MSP2 were undertaken at initial concentrations of 0.5 mg/mL in solutions buffered at pH 3.4, 6.2 and 7.4. The absorbance versus radial position profiles for freshly prepared 3D7-6H at pH 3.4 are plotted at 6-min intervals in Figure 3a. The data were fitted to a single discrete species model that yielded a nonlinear least squares best fit to a species with a sedimentation coefficient of 1.60 S, diffusion coefficient of 4.92×10^{-7} cm²/s and molecular mass of 24.7 kDa (Table 1). These data suggest that 3D7-6H exists primarily as a monomer in aqueous solution. This was confirmed by sedimentation equilibrium analysis (Table 1). Furthermore, the axial ratio (a/b) of 17.2, calculated from the sedimentation coefficient and monomeric molecular mass assuming a prolate ellipsoidal shape (Table 1), indicates that the 3D7-6H monomer is significantly elongated in solution. Similar data were also obtained for freshly prepared 3D7-6H at pH 6.2 and 7.4 (Table 1) and for 3D7-6H incubated at room temperature for 24, 48, 72 and 168 h in solutions buffered at pH 3.4, 6.2 and 7.4 (data not shown). These data demonstrate that the 3D7-6H isoform is a stable monomer in solution even when incubated at pH 3.4, 6.2 or 7.4 for up to 7 days.

FC27-6H also exists primarily as a stable monomer when freshly prepared in solutions buffered at pH 6.2 and pH 7.4 (Table 1). However, in solution buffered at pH 3.4, FC27-6H readily forms large aggregates. This is demonstrated in Figure 3b, where the initial absorbance versus radial scan collected at 40,000 rpm shows evidence of sample heterogeneity, with the presence of a broad, rapidly-sedimenting boundary corresponding to large FC27-6H aggregates, as well as a steep, slow-moving boundary corresponding to monomeric FC27-6H. When these data are fitted to a non-interacting mixture of 4 species, the nonlinear least squares best fit yields the parameters reported in Table 2, suggesting that several aggregates are present at $t = 0$ ranging from approximately 100 S to > 330 S. However, following 24 h incubation at room temperature, aggregates of FC27-6H are no longer observed, even when radial scans were collected at 10,000 rpm (data not shown). This indicates that the aggregates of FC27-6H are extremely large (> 500 S) and have therefore sedimented to the bottom of the cell during the acceleration phase of the experiment. Nevertheless, the signal due to monomeric FC27-6H is still detected, thus providing an indirect means of assessing the degree of aggregation. This is illustrated in Figure 3c, which shows the initial radial absorbance scan for FC27-6H following 0, 24, 48, 72 and 168 h incubation at room temperature and pH 3.4. These data demonstrate that the proportion of monomeric FC27-6H decreases exponentially as a function of time (Figure

3d), whilst the proportion of FC27-6H aggregate is presumably increasing exponentially. When FC27-6H is incubated at pH 6.2 and pH 7.4, significantly more monomer is evident in the absorbance versus time profiles (Figure 3d), suggesting that more neutral pH stabilizes the monomer. By contrast, the proportion of monomeric 3D7-6H MSP2 remains relatively constant as a function of time, even when incubated at pH 3.4 (Figure 3d). This indicates that FC27-6H, but not 3D7-6H, was susceptible to aggregation in acidic solution.

NMR translation diffusion measurements

One-dimensional ^1H NMR spectra of MSP2 under both physiological and acidic conditions show typical characteristics of an unstructured protein, with limited chemical shift dispersion and substantial peak overlap (Figures S3-S6, Supplementary Material). The effective hydrodynamic radius can be calculated from measurements of the diffusion coefficient by means of the Stokes-Einstein equation. However, this approach has the disadvantages that it requires knowledge of the viscosity of the solution and relies on interpretation of absolute values of diffusion coefficients, which are subject to errors from inaccuracies in measuring the exact shape and strength of the gradient pulse and from spin-echo instability. To overcome these problems, dioxane was added to the samples as a reference molecule. The hydrodynamic radius of the protein could then be determined from $R_h^{\text{prot}} = R_h^{\text{ref}}(D_{\text{ref}}/D_{\text{prot}})^{1/3}$ where R_h^{prot} is the protein hydrodynamic radius and R_h^{ref} is the radius of the reference molecule, D_{prot} and D_{ref} represent the measured diffusion coefficients of the protein and reference molecule (Figure S7 in the Supplementary Material shows an example of $\ln(I/I_0)$ versus $\gamma^2 g^2 \delta^2 (-\delta/3)$ plot, the slope of which equals $-D$, for FC27 MSP2 in PBS).

Tables 3 and 4 summarize the hydrodynamic properties derived from NMR diffusion data for FC27-6H and 3D7-6H MSP2, respectively. Under physiological conditions (PBS, pH 7.0), MSP2 is much more extended than both folded and molten globular states, yet more compact than a fully unfolded coil and close to an intrinsic pre-molten globule, as described by Uversky.¹⁵ The two forms of MSP2 have hydrodynamic radii similar to other proteins of comparable size that exist as intrinsic pre-molten globules¹⁵ (Table S1, Supplementary Material). The hydrodynamic properties of MSP2 under acidic conditions (10 mM HOAc, pH 3.6) with and without salt, and at physiological pH in the absence of salt (10 mM phosphate, pH 7.25) were also measured. The results show that salt makes little difference to the hydrodynamic dimensions of MSP2 at physiological pH (Table 3). By contrast, adding salt at acidic pH transforms the more extended, intrinsic coil-like state present at this pH to an intrinsic pre-molten globule-like state akin to that found at physiological pH (Table 4).

NMR resonance assignments

In order to obtain residue-specific structure and dynamics information of MSP2 by NMR, we chose 10 mM HOAc as the solvent, in which CD and NMR spectra show that MSP2 adopted a state similar to that under physiological conditions and the protein remained monomeric for sufficient time to enable a complete set of NMR spectra to be acquired. A ^1H - ^{15}N HSQC spectrum of uniformly ^{15}N -labelled FC27 MSP2 at 25 °C and pH 3.5 is shown in Figure 4. Amide ^1H chemical shifts were dispersed over a narrow range, 7.8–8.7 ppm, as expected for a largely disordered and flexible conformation. However, the chemical

shift dispersion in the ^{15}N dimension, which depends more on residue type and context sequence than on conformation, was still substantial; as expected,^{34,35} this was also the case for ^{13}O resonances. In addition, the peaks were relatively sharp and intense because of the high mobility of the polypeptide backbone (see below). Moreover, because the protein was largely unstructured, chemical shifts were mostly close to random coil values. All these factors made resonance assignments for this protein feasible under these conditions.

Using HNCACB, CBCA(CO)NH, HNCA, HN(CO)CA, and HNCO spectra, backbone and some side chain resonances from 204 out of the 208 non-proline residues were assigned; resonances from the N-terminal Met from the vector, as well as Lys6, Lys134, and Lys207, were not observed. The cross peaks were separated into distinct regions typical of the residue random coil positions (Gly, Ser/Thr, Ala and other backbone amides), emphasizing the unfolded nature of the protein. Most cross-peaks in the ^1H - ^{15}N HSQC spectrum displayed similar intensities and there was no line broadening suggestive of intermediate conformational exchange processes. Peaks from corresponding residues of the two repeated sequences (41-72 and 73-104) in the central region, with the exception of the first five residues of each repeat region, overlap with each other and thus give rise to single intense peaks in the ^1H - ^{15}N HSQC spectrum. Of the five resolved residues, the first two (41-42 and 73-74) were assigned unambiguously, while the latter three (43-45 and 75-77) were not. For subsequent data analyses, corresponding resonances of the two repeat sequences were treated as a single peak, while overlapped resonances from other regions were excluded. The assignments are tabulated in the Supplementary Material (Table S2) and have been deposited in BioMagResBank³⁶ with accession number 15590.

Deviations of HN, H^α , C^α , and CO chemical shifts from random coil values for FC27 MSP2 in 10 mM HOAc are plotted in Figure 5. Rather than uniform deviations in the vicinity of zero, as expected for a random coil, significant and specific patterns of deviations were observed. This is consistent with CD data, which showed that some secondary structure was present in this predominantly disordered protein. Negative deviations were found for almost all H^α and HN resonances and positive deviations for C^α and CO resonances, suggesting that backbone conformations with helical structure propensity were being sampled. In the region around residues 14-21 and 140-150, C^α and CO resonances showed the largest deviation from random coil shifts, indicating a greater helical structure propensity than in the other regions of the polypeptide chain. In a consensus CSI analysis, the largest consecutive upfield deviations of H^α and HN chemical shifts occurred in roughly the same regions, supporting the indication of C^α and CO secondary chemical shift data. No consecutive consensus deviations larger than threshold values were observed in other regions of the polypeptide chain. Around residues 197-211 there were some relatively large deviations distributed irregularly, which might reflect local order caused by the single disulfide bond in this region (the oxidized state of the two cysteines was confirmed by mass spectrometry and chemical shift data). In summary, secondary chemical shift analysis indicated that FC27 MSP2 in 10 mM HOAc was not a random coil, but a largely unstructured protein containing limited regions of secondary structure, which were predominantly helical in nature.

Backbone dynamics

Despite severe overlap of resonances from the central unstructured region of the protein, R_1 , R_2 , and $\{^1\text{H}\}$ - ^{15}N NOE values were obtained for most assigned peaks. Data at 800 and 600 MHz are shown in Figure 6. These three relaxation parameters all displayed similar overall profiles and showed the expected dependence on magnetic field strength, with R_2 and $\{^1\text{H}\}$ - ^{15}N NOE values increasing, and R_1 decreasing with increasing magnetic field strength.³⁷ Residues at both ends of the polypeptide chain show obviously decreased R_1 , R_2 , and $\{^1\text{H}\}$ - ^{15}N NOE, reflecting their greater flexibility. In the conserved N- and C-terminal regions, consecutive and slightly higher values were observed for all three relaxation parameters, implying some motional restriction in these regions. In addition, there was a small degree of motional restriction in the region around residues 140-150. Except for the C-terminal region where the disulfide bond is located, the other two motionally restricted regions match very well with those regions of secondary structures indicated by chemical shift analyses. Moreover, in most portions of the polypeptide chain, R_2 presented similar profiles to R_1 and NOE, implying that there was little if any contribution from milli- to micro-second time scale motion to the relaxation. The largely disordered nature of FC27 MSP2 is particularly evident in the mostly negative values of $\{^1\text{H}\}$ - ^{15}N NOE at 600 MHz; the motionally restricted residues showed less negative or weakly positive values. Relatively larger R_2 values were observed for residues Ala38, Gly47/79, Ala131, Gly179, and Gly186 at 800 MHz, but not at 600 MHz; the larger field dependence for these residues compared with the bulk of the protein may reflect a contribution from conformational exchange, as discussed below.

The dynamic behavior of folded proteins is commonly interpreted within the framework of the model-free formalism, but, because of the wide range of local correlation times sampled by unstructured proteins such as MSP2 this approach is unlikely to provide an accurate analysis of the data.^{29,38} A reduced spectral density analysis avoids the assumption inherent in model-free analysis that the molecule tumbles with a single correlation time τ_m by directly mapping the spectral density function. $J(0.87\omega_H)$ reflects higher frequency motions and is largely determined by the heteronuclear NOEs. $J(\omega_N)$ is most sensitive to R_1 and, like R_1 , does not discriminate well between faster and slower motions. $J(0)$ is most sensitive to R_2 , reflecting slower (nanosecond) motions, with possible contributions from exchange processes on the milli- to micro-second time scale.^{29,38} Exchange contributions to R_2 , or slow milli- to micro-second motions, are reflected by elevated $J(0)$ values, with $J(0)$ expected to increase as the square of the magnetic field strength.^{37,39}

The spectral density functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$, obtained from data recorded at 600 and 800 MHz, are shown in Figure 7. As with R_1 , relatively larger values of $J(\omega_N)$ are observed in the regions indicated by secondary chemical shift data, signifying restriction of sub-nanosecond time scale motions. The variation of $J(0)$ along the sequence displays the same trend as $J(\omega_N)$, implying little or no contribution from low frequency motions. However, as for R_2 , $J(0)$ values for residues Ala38, Gly47/79, Ala131, Gly179, and Gly186 are relatively larger at 800 MHz, and show larger magnetic field dependence than other sites, indicating contributions from slower motions. Relative to $J(0)$ and $J(\omega_N)$, the

$J(0.87\omega_H)$ values are more uniform, reflecting the overall flexibility of the polypeptide chain (Figure 7).

Comparison with N-terminal peptides

In order to investigate whether the N-terminal region of MSP2 adopts the same conformation in an isolated peptide as in full-length protein, we compared the ^1H - ^{15}N HSQC spectrum for recombinant MSP2₁₋₂₅ peptide¹³ with that of FC27 MSP2. As shown in Figure 8, most peaks of corresponding residues in full-length MSP2 and the N-terminal peptide have similar chemical shifts, except for a few residues at the termini of the peptide. In addition, the H^α chemical shifts of N-terminal residues in full-length MSP2 show only minor differences from those of the recombinant peptide and a synthetic MSP2₁₋₂₅ peptide^{13,14} (Figure S8, Supplementary Material). This shows that the N-terminal region of full-length MSP2 samples a very similar range of conformations to those of the isolated peptides.

Effect of disulfide reduction

The C-terminus of MSP2 contains two Cys residues, which are covalently linked to form a disulfide bond in our samples, as confirmed by mass spectrometry and chemical shift data. To investigate the effect of local order associated with this disulfide bond on the overall conformation of the protein, ^1H - ^{15}N HSQC spectra were recorded before and after the addition of a 100-fold excess of the reducing reagent TCEP to ^{15}N -labelled FC27 MSP2. As shown in Figure S9, the peaks affected most significantly by disulfide bond reduction were from Ser198, Gln199, Glu201, Cys202, Thr203, Asp204, Gly205, Asn206, Glu208, Asn209, Cys210, Gly211, all of which are close to the disulfide bond. The peak from the C-terminal residue Ser221 was also shifted. Thus, the local structure in the vicinity of the disulfide bond in the conserved C-terminal region is affected by disulfide reduction but the rest of the protein is not.

Binding to DPC micelles

As MSP2 is a GPI-anchored merozoite surface protein and may play a role in parasite attachment to red blood cells, we investigated the effect of membrane lipids on the conformation of FC27 MSP2 by comparing ^1H - ^{15}N HSQC spectra in the presence and absence of DPC micelles. To exclude the influence of pH change on protein conformation (the pH of the solution changed from 3.5 to 4.9 after DPC was added), we compared ^1H - ^{15}N HSQC of the protein in the presence of DPC with that of the protein in HOAc solution at pH 4.9 rather than 3.5. Resonance assignments at pH 4.9 were obtained by titrating with HOAc to pH 3.5, where the assignments described above were made. ^1H - ^{15}N HSQC spectra of FC27 MSP2 at pH 4.9 in the presence and absence of DPC are superimposed in Figure 9. In the presence of DPC, most resonances remain unperturbed, with a few peaks shifted and several broadened, in some cases to the point of being unobservable. Most of these shifted and broadened resonances were found to be N-terminal residues, suggesting that it is this region of the protein that mediates interactions of the protein with the micelle. A few peaks from the C-terminal region also showed minor shifts. In the region of the spectrum containing side chain amide resonances, a few new peaks appear upon DPC addition,

implying an altered chemical environment of the side-chains of some residues. DPC binding was also examined under more physiological conditions. Similar to the observations at acidic pH, addition of DPC to FC27 MSP2 in phosphate buffer (pH 6.5) caused peak shifts and broadening of some peaks in the HSQC spectrum. However, relative to the case in 10 mM HOAc, fewer peaks disappeared and the intensities of the broadened peaks were slightly stronger in phosphate buffer. In addition, observable aggregates appeared in solution almost immediately after DPC was added in phosphate buffer, whereas this took weeks for the sample in 10 mM HOAc at low pH.

PFG diffusion experiments were performed for the samples in 10 mM HOAc in the absence and presence of DPC. Before DPC was added, FC27 MSP2 had a diffusion coefficient of $7.22 \times 10^{-11} \text{ m}^2/\text{s}$, 1/17.8 that of dioxane. After DPC was added, the protein had a smaller diffusion coefficient, $6.04 \times 10^{-11} \text{ m}^2/\text{s}$, 1/19.6 that of dioxane. In the solution without FC27 MSP2, DPC had a diffusion coefficient of $1.22 \times 10^{-10} \text{ m}^2/\text{s}$, and this was reduced slightly to $1.14 \times 10^{-10} \text{ m}^2/\text{s}$ following the addition of protein, although it was still larger than for MSP2 in the same solution. The concentration of DPC in these solutions was 100 mM. If we assume that the number of DPC molecules per micelle was around 60,40 then the mass of the micelles was roughly 23.4 kDa and the effective concentration around 1.7 mM, 19-fold that of MSP2 (0.09 mM). Thus, the measured average diffusion coefficient for DPC in the presence of MSP2 was dominated by the value for free DPC micelles. The diffusion coefficient of MSP2 was already less than that of free micelles, reflecting the slightly greater mass and, more importantly, its extended flexible structure. The reduction in the diffusion coefficient of MSP2 in the presence of micelles reflects the formation of a complex, but one whose diffusion rate is dominated by the extended flexible structure of MSP2 rather than the compact, largely spherical structure of the micelle.

Discussion

Previous studies have shown that MSP2 is an intrinsically unstructured protein which is prone to form amyloid-like fibrils in aqueous solution.¹² The present study was undertaken in order to characterise the hydrodynamic properties, solution conformation and backbone dynamics of MSP2. These properties are discussed in relation to the use of MSP2 as a vaccine and the state of MSP2 on the parasite surface.

Residual secondary structure

The deviations of backbone chemical shifts from random coil values are good indicators of secondary structure.^{29,38,41–44} The residual secondary structure suggested by chemical shift analyses for FC27 MSP2 is consistent with secondary structure prediction results for the N-terminal region, which showed some helical propensity. Previous studies of synthetic and recombinant peptides corresponding to the conserved N-terminal region of MSP2 found evidence of turn-like and nascent helical structures in the central and distal regions.^{13,14} Comparison of the chemical shifts of N-terminal residues in full-length MSP2 with those of these N-terminal fragments showed very little difference, implying that the N-terminal region in full-length protein samples a similar range of conformations to the isolated peptides. It appears, therefore, that the N-terminus of MSP2 is hardly affected by the rest of

the protein and not involved in any significant or stable interactions with other regions. It is important to note, however, that long-range interactions in the IUP α -synuclein_{45–47} were detected only using approaches such as paramagnetic relaxation enhancement (PRE) measurements on spin-labelled protein and not via chemical shift perturbations.^{46,47} Nonetheless, it seems unlikely that a conformational change of the type described for the IUP α -synuclein_{45–47} is necessary for the initiation of amyloid formation by MSP2. A significant contributor to this different behaviour is likely to be the marked dearth of hydrophobic residues in MSP2.

In addition to the N-terminal region, there is another helical region suggested by secondary chemical shift data, which is not predicted by any sequence-based secondary structure prediction algorithm. This region, residues 140–150, is located in the central variable region, which differs between the 3D7 and FC27 isoforms of MSP2 (Figure 1).¹⁰ As these residues are mostly hydrophilic, local hydrophobic clustering is unlikely to be responsible for the nascent helical structure. Previous studies have shown that 3D7 MSP2 forms fibrils much more slowly than FC27 MSP2,¹² which must reflect differences in this central region.

There is no evidence that MSP2 forms disulfide-bonded oligomers and the two cysteine residues in the C-terminal region of MSP2 are assumed to form an intramolecular disulfide bond in the protein on the merozoite surface, but what role it plays in MSP2 function is not clear. Our results show that the local order caused by the disulfide bond has no obvious correlation with the rest of the protein; except for residues immediately adjacent to the disulfide bond, the chemical shifts of most residues in the protein hardly changed after disulfide bond reduction (Figure S9).

Dynamics

The regions of MSP2 displaying motional restriction correlate well with residual secondary structures indicated by chemical shift deviations as well as the region around the disulfide bond. Moreover, the conserved N-terminal region of full-length MSP2 samples a similar range of conformations to those of free peptides corresponding to this region. These observations imply that the observed motional restrictions in FC27 MSP2 in 10 mM HOAc are a consequence of residual local structure rather than long-range hydrophobic or other interactions.

The backbone dynamics of many unfolded states of proteins have been investigated by NMR.^{25–27,37,48} In comparison with the synucleins,^{25,26} one of which, α -synuclein, is a well studied amyloidogenic IUP, FC27 MSP2 shows similar dynamics characteristics, with mean values of 1.37 and 2.54 s⁻¹ for R_1 and R_2 , respectively, at 800 MHz (corresponding values at 600 MHz are 1.36 and 1.97 s⁻¹). Most heteronuclear NOEs of FC27 MSP2 fall well below 0.5 at 800 MHz, clearly indicating the absence of rigid structure and restricted dynamics in the ps–ns timescale; this is also supported by the relatively uniform $J(0.87\omega_H)$ (Figure 7). The high degree of flexibility is also indicated by the fact that the R_2 values are slightly greater than R_1 . Relative to the synucleins,^{25,26,47} FC27 MSP2 displays slightly smaller R_2 and heteronuclear NOEs. Moreover, in the synucleins, R_2 relaxation rates show larger variations than R_1 , whereas very similar profiles were observed for R_1 and R_2 in the case of FC27 MSP2. The above differences may be explained by the presence of greater

order and long-range interactions in the synucleins. Compared with these proteins, FC27 MSP2 is more flexible and less compact.

Implications for function and vaccine development

The conformation of merozoite surface proteins being assessed as vaccine components is considered critical to the induction of antibodies with the fine specificity required to inhibit parasite growth. Hence, considerable attention has been paid to ensuring that native intramolecular bonds are in place in AMA1 and MSP1₁₉, two other leading *P. falciparum* asexual blood-stage vaccine candidates.⁴⁹ The intrinsically unstructured nature of MSP2, with motional restriction in only very limited regions of the recombinant protein, could explain why it has been difficult to induce polyclonal antibodies to MSP2 that inhibit merozoite invasion *in vitro* (Anders et al., unpublished). The antibody response to MSP2 may be characterized by specificities to many of the alternative conformations sampled by the molecule in solution. This would be expected to reduce the effectiveness of these antibodies for blocking merozoite invasion, particularly if the ensemble of conformations adopted by MSP2 on the merozoite surface is more restricted because of protein-protein or protein-lipid interactions involved in the formation or function of the merozoite surface coat.

MSP2 is a major component of the merozoite surface coat and there is some evidence that it is oligomeric in this location,¹² but neither its state on the merozoite surface nor its function is known. Both isoforms of MSP2 form fibrils *in vitro*, and the N-terminal region is a prominent component of the protease-resistant core of these fibrils.¹² As a result, our initial studies have focused mainly on the correlation between residual structure in the N-terminal region and fibrillogenesis. Studies on free peptides corresponding to the conserved N-terminal region showed that they readily form fibrils *in vitro*, and structure calculations indicated that they contained nascent helical and turn-like structures.^{13,14} The present study confirmed that the N-terminal region in full-length FC27 MSP2 samples a similar range of conformations to those of the free peptides. As the N-terminal region has been found to be involved in the core structure of amyloid fibril formed by MSP2,¹² the residual secondary structures in this region may play a key role in the onset of fibril formation. Characteristics of β -aggregation prone sequences include high hydrophobicity, low charge, and a propensity to form β -sheet rather than α -helical structure.¹⁹ Thus, mutations that change the above three properties of the N-terminal fragment will disturb the fibril formation of MSP2. Indeed, an N-terminal 28-mer peptide mutant in which two hydrophobic residues, Phe11 and Ile12 were substituted with Ala showed much slower fibril formation (unpublished results). It is possible that transient formation of amphipathic helical structure in the N-terminal region could stabilize intermolecular interactions through hydrophobic contacts, promoting oligomerization.⁵⁰

Besides their possible role in fibril formation, preformed structural elements in IUPs could serve as initial contact points, the binding of which facilitates interactions with the binding partner and, in some cases, enables formation of extensive protein interfaces.^{23,51} MSP2 may play a role in merozoite attachment to and invasion of the red blood cell. As a first step towards understanding the state of MSP2 on the merozoite membrane and possible interactions with the red blood cell membrane, we investigated its interaction with the

membrane mimetic DPC (since both red blood cell and parasite membranes are rich in phosphatidylcholine containing lipids).⁵² We found that the N-terminal region was most affected by DPC micelles, with most N-terminal residues shifting position and broadening or becoming unobservable. This resonance broadening is attributed to preferential interaction of the N-terminal region with the micelles, leading to a significant enhancement in the transverse relaxation rate of those residues in direct contact with the lipid. Other regions of the protein not in contact with the lipid would be free to undergo internal motion and thus continue to yield sharp resonances. Upon mixing with DPC, a few new peaks appeared in the side chain region of the spectrum, reflecting changes in residues close to the lipid but still flexible enough to give sharp lines. The type of ordered structure formed by the N-terminal domain when MSP2 binds to DPC micelle remains to be investigated. However, as in the case of α -synuclein, in which residual helical structure in the free protein accurately reflects well-formed helical structures of the micelle-bound form,⁵⁰ the nascent helical structures in the N-terminal region of MSP2 may be stabilized upon membrane binding. Indeed, preliminary CD studies of full-length MSP2 and a recombinant peptide corresponding to the conserved N-terminal domain of MSP2 in DPC micelles showed a clear increase in helical content for the peptide (Figure S11), although little change for the full-length protein. These results are consistent with our NMR data, which show that the major interaction with DPC micelles is through the N-terminal domain, and that the bulk of the protein is not involved in membrane binding. The stabilization of helical structures may promote intermolecular hydrophobic interactions and thus the oligomerization of the protein. Oligomeric forms of MSP2 may, as for α -synuclein,⁵³ cause membrane disruption and thereby facilitate invasion of the merozoite into the red blood cell.

The N-terminal region of MSP2 could also interact with the merozoite surface membrane. As MSP2 is anchored to the membrane through the C-terminal GPI moiety, this would make both the N- and C-terminal regions proximal to the membrane, with the central variable region of the molecule distal to the membrane and more accessible to the external environment. Such an arrangement is consistent with the observation from immunofluorescence analyses that epitopes in the variable region of MSP2 are more accessible on the merozoite surface than epitopes in the conserved N- and C-terminal regions (Adda et al. unpublished results). Further studies of the conformation and self-association of MSP2 on membrane surfaces are clearly required, and will have to include model membranes where several MSP2 molecules are present on the same membrane. In the DPC micelles studied here it is most likely that each micelle harbours just one MSP2 molecule. Nonetheless, our results provide interesting insights into the conformational preferences of this protein and its possible behaviour on the merozoite cell membrane.

Materials and Methods

Protein expression and purification

Many of the studies were carried out with 3D7 and FC27 isoforms of MSP2 that were produced at Novozymes GroPep Limited (Thebarton, SA, Australia) with funding from the PATH Malaria Vaccine Initiative. The antigens were expressed in *E. coli* with C-terminal hexa-His tags to facilitate purification by metal chelate chromatography. After this initial

purification step the antigens were further purified by anion-exchange and reversed-phase chromatography. The purified proteins were buffer-exchanged into 10 mM acetic acid (~2 mg/mL) and then stored at -70°C before being used in the studies reported below. The full details of the process used for the production of the clinical batches of MSP2 will be reported elsewhere (Hobba et al., in preparation).

Untagged full-length FC27 MSP2 was expressed in *E. coli* strain BL21 after cloning the PCR amplified sequence into the *EcoR*I site of the expression vector pQE9 (Qiagen Pty Ltd). The stock strain was inoculated into 1000-fold volume of Luria-Bertani medium containing 100 $\mu\text{g/mL}$ ampicillin and incubated overnight to saturation. The overnight culture was diluted 1:10 in super broth medium and grown to $\text{OD}_{600}\approx 1.2$, then IPTG was added to a final concentration of 1 mM to induce expression. After 4 h, the induced cells were harvested by centrifugation at $5500 \times g$ and 4°C for 20 min and then frozen at -80°C . The target protein was purified using a strategy specific for recombinantly expressed unstructured proteins.⁵⁴ The cell pellets containing FC27 MSP2 were thawed and incubated in a boiling water bath for 10 min. Immediately following heating, the cell lysate was cooled by incubating in ice/water/salt mixture for more than 15 min. Supernatant was then collected by centrifugation at $13000 \times g$ and 4°C for 20 min, and purified with a Source 15 (Amersham Biosciences, 16 mm $\times$ 80 mm) followed by a C18 reverse phase column (Vydac 218TP510, 10 mm $\times$ 200 mm). Finally the purified product was lyophilized. ^{15}N -labelled and ^{15}N - ^{13}C double labelled FC27 MSP2 were prepared in the same way except that super broth was replaced by M9 medium with ^{15}N ammonium chloride (1 g/L) or/and ^{13}C glucose (2 g/L). The proteins were $> 95\%$ pure as judged by SDS-PAGE and reverse phase HPLC. The identity of the construct was confirmed by DNA sequencing as well as electrospray mass spectrometry. Recombinant N-terminal 28-mer peptide was prepared as described previously.¹³

Circular dichroism spectroscopy

Circular dichroism (CD) spectra of 3D7-6H and FC27-6H MSP2 were recorded on a Jasco J-815 CD spectrophotometer over the wavelength range 195–250 nm at a temperature of 20°C . Samples (0.15 mg/mL) were dissolved in 10 mM acetic acid, pH 3.4 $\pm$ 150 mM NaCl, 10 mM sodium acetate, pH 6.2 $\pm$ 150 mM NaCl, and 10 mM potassium phosphate, pH 7.4 $\pm$ 150 mM NaCl. Data were collected using a step size of 0.5 nm, a slit bandwidth of 1.0 nm and a signal averaging time of 2.0 s in a 1 mm pathlength quartz cuvette. CD spectra were fitted using a database comprised of 48 proteins (SDP48) employing the CONTINLL algorithm included in the CDPro software package⁵⁵ available at <http://amar.colostate.edu/~sreeram/CDPro/>.

Sedimentation velocity

Sedimentation velocity experiments were performed in a Beckman Coulter model XL-I analytical ultracentrifuge at 20°C . Samples (380 μL) and reference solutions (400 μL) were loaded into a conventional double sector quartz cell and mounted in either a Beckman Coulter An-60 Ti 4-hole or An-50 Ti 8-hole rotor. MSP2 samples (3D7-6H and FC27-6H) at an initial concentration of 0.5 mg/mL dissolved in 10 mM acetic acid, 150 mM NaCl, pH 3.4, 10 mM sodium acetate, 150 mM NaCl, pH 6.2, and 10 mM potassium phosphate, 150

mM NaCl, pH 7.4 were centrifuged at a rotor speed of 40,000 rpm. Data were collected in continuous mode at a wavelength of 235 nm using a time interval of 300 s and a step-size of 0.003 cm without averaging. Solvent densities, solvent viscosities, the partial specific volume of 3D7-6H (0.687 mL/g) and FC37-6H (0.689 mL/g) and estimates of the axial ratio (a/b) assuming a prolate ellipsoid were computed using the program SEDNTERP.⁵⁶ Sedimentation velocity data at multiple time points were fitted to a single discrete species or a mixture of up to four non-interacting species⁵⁷ using the program SEDFIT, which is available at www.analyticalultracentrifugation.com.

Sedimentation equilibrium

Sedimentation equilibrium experiments were conducted at 20 °C in a Beckman model XL-I analytical ultracentrifuge using an An-60 Ti rotor. Sample (100 µL, 0.15 mg/mL) and reference (120 µL) were loaded into conventional double-sector quartz cells and centrifuged at 25,000 and 35,000 rpm. Data were collected at a wavelength of 235 nm using a step size of 0.001 cm with ten averages until sedimentation equilibrium was attained (~ 24 h). As for the sedimentation velocity experiments, samples were dissolved in 10 mM acetic acid, 150 mM NaCl, pH 3.4, 10 mM sodium acetate, 150 mM NaCl, pH 6.2, or 10 mM potassium phosphate, 150 mM NaCl, pH 7.4. To estimate the signal due to non-sedimenting contaminants, high-speed depletion was performed at 40,000 rpm for up to 5 h following sedimentation equilibrium, and the resulting baseline offset (E) was calculated by averaging the absorbance over a radial range of 0.1 cm in the plateau region adjacent to the sample meniscus. For all samples, the values for E represented < 5 % of the initial absorbance. Sedimentation equilibrium data at multiple speeds were globally fitted to single discrete species model by employing the program SEDPHAT,^{57,58} which is also available at www.analyticalultracentrifugation.com.

NMR spectroscopy

Samples used for 1D NMR and translational diffusion measurements by pulsed field gradient NMR were prepared by dissolving 1 mg recombinant MSP2 in 500 µL of buffer containing 5% ²H₂O and 0.05 µL dioxane, which was added as an internal reference. The sample used for backbone ¹⁵N relaxation measurements as well as the recording of ¹⁵N-NOESY-HSQC and ¹⁵N-TOCSY-HSQC spectra was prepared by dissolving lyophilized uniformly ¹⁵N-labelled FC27 MSP2 in 10mM HOAc containing 5% ²H₂O (pH 3.5) at a concentration of 8 mg/mL (0.35 mM). The sample used for 3D heteronuclear triple-resonance experiments recorded to make resonance assignments (HNCACB, CBCA(CO)NH, HNCO, HNCA, and HN(CO)CA) was prepared in the same way except that the protein was labelled with both ¹³C and ¹⁵N and the concentration was 10 mg/mL (0.44 mM). The samples for TCEP reduction, DPC binding, and TMAO-induced folding were prepared with ¹⁵N-labelled protein as above, at a protein concentration of 2 mg/mL (0.09 mM). TCEP (Sigma-Aldrich), DPC (Cambridge Isotope Laboratories, Andover, MA), and TMAO (Sigma-Aldrich) were added as solid powder. In the case of DPC, addition of MSP2 to a micelle solution produced the same spectral changes as addition of DPC to a MSP2 solution. The sample of recombinant N-terminal peptide, MSP₁₋₂₅, was prepared at 0.8 mg/mL (0.25 mM). All the samples were centrifuged at 13,000 rpm (16,000 g) for 10 min to remove potential precipitates before transfer to the NMR tube. The protein concentration

was determined by absorption at 280 nm with an extinction coefficient of $3105 \text{ M}^{-1} \text{ cm}^{-1}$ calculated from the amino acid composition (using the program ProtParam at the EXPASY server, available at www.expasy.ch/tools).

NMR spectra were acquired at 25 °C on Bruker Avance800 and Avance500 spectrometers, both equipped with cryoprobes, and a Bruker DRX-600 spectrometer. ^1H chemical shifts were referenced with dioxane set as 3.751 ppm, and ^{15}N and ^{13}C chemical shifts were referenced indirectly from ^1H using the ratios $\gamma_{\text{N}}/\gamma_{\text{H}} = 0.101329118$ and $\gamma_{\text{C}}/\gamma_{\text{H}} = 0.251449453.59$

Two-dimensional ^1H - ^{15}N HSQC spectra were acquired with 2048 (^1H) and 256 (^{15}N) points and spectral widths of 9 (^1H) and 24 ppm (^{15}N), respectively. In all 3D spectra, the spectral widths along the ^1H and ^{15}N dimensions were the same as in 2D ^1H - ^{15}N HSQC spectra, and the number of points in F3 for most experiments was 2048. A ^{15}N TOCSY-HSQC spectrum was acquired with 48 and 144 points in the ^{15}N (F2) and ^1H (F1) dimensions, respectively, and a ^{15}N NOESY-HSQC spectrum was acquired with 44 (^{15}N) and 144 (^1H) time points (the ^1H spectral width for the third dimension of both spectra was 9 ppm). A 60 ms mixing time was used in the ^{15}N TOCSY-HSQC experiment, and mixing times of 120 and 250 ms were used in ^{15}N NOESY-HSQC spectra recorded at 800 and 500 MHz, respectively. CBCA(CO)NH and HNCACB spectra were recorded at 600 and 800 MHz, respectively, with spectral width of 70 ppm (^{13}C) and 48 (^{13}C) and 128 (^{15}N) time points. An HNCO spectrum at 600 MHz was recorded with spectral width of 6 ppm (^{13}C) and 64 (^{13}C) and 128 (^{15}N) time points. HNCA spectra were recorded at 800 MHz and with 48 (^{13}C) and 110 (^{15}N) time points and at 600 MHz with 48 (^{13}C) and 80 (^{15}N) time points. An HN(CO)CA spectrum was recorded at 600 MHz with 48 (^{13}C) and 80 (^{15}N) time points. Both HNCA and HN(CO)CA spectra had a ^{13}C spectral width of 30 ppm.

Diffusion measurements were performed using a pulsed field gradient longitudinal eddy-current delay pulse sequence,^{60,61} as implemented by Yao et al;⁶² a series of 12 spectra was acquired with the strength of the diffusion gradient varying from 3.44 to 36.22 G cm^{-1} . Backbone ^{15}N T_1 and T_2 relaxation times and steady-state $\{^1\text{H}\}$ - ^{15}N NOEs were measured at both 800 and 600 MHz at 25 °C. For the T_1 measurements at 800 MHz, 12 time points were collected with relaxation durations of 10, 50, 100, 150, 280, 420, 560, 700, 850, 1100, 1500 and 2400 ms and duplicate spectra were recorded at 50 and 280 ms to allow estimation of errors. The same relaxation durations were used for measurements at 600 MHz except for 850 and 1500 ms. For T_2 , relaxation durations of 17, 34, 51, 68, 85, 102, 136, 170, 204 and 255 ms (duplicate spectra at 34 and 136 ms) were measured at 800 MHz and relaxation durations of 15.4, 30.8, 46.2, 61.7, 77, 82.5, 123.3, 154, 184.8, 231, 338.8 and 616 ms (duplicate spectra at 30.8 and 123.3 ms) were measured at 600 MHz. Recycle delays were 3 and 2.8 s for both T_1 and T_2 measurements at 800 and 600 MHz, respectively. The steady-state $\{^1\text{H}\}$ - ^{15}N NOE was measured from duplicate pairs of ^1H - ^{15}N spectra recorded with and without amide proton saturation.

NMR data analysis

Spectra were processed using TOPSPIN (Version 1.3, Bruker Biospin) and analyzed using XEASY (Version 1.3.13),⁶³ Sparky (Version 3.111)⁶⁴, and NMRViewJ (Version 6.23).⁶⁵

Secondary chemical-shift data were calculated using the chemical-shift index (CSI) in NMRViewJ, and random coil reference values derived from model peptides in acidic 8 M urea.⁶⁶

Diffusion coefficients, D , were obtained by fitting the intensities of a number of peaks in the amide and aliphatic region (for acidic pH samples) or just the aliphatic region (for physiological pH samples) to the following equation using the T_1/T_2 relaxation module in TOPSPIN: $I = I_0 \exp\{-\gamma^2 g^2 \delta^2 (-\delta/3) D\}$, where γ is the gyromagnetic ratio and g , δ and δ are the amplitude, duration and separation of the single pair of gradient pulses, respectively. The values of the hydrodynamic radius of the protein were calculated from: $R_h^{\text{prot}} = R_h^{\text{ref}}(D_{\text{ref}}/D_{\text{prot}})$, where R_h^{prot} is the protein hydrodynamic radius and R_h^{ref} is the radius of the reference molecule (2.12 Å for dioxane), D_{prot} and D_{ref} represent the measured diffusion coefficients of the protein and reference molecule.³³

^{15}N R_1 and R_2 relaxation rates were estimated by least squares minimization fitting of the measured peak heights to a two-parameter equation:

$$I(t) = I_0 \exp(-tR_{1,2}) \quad (1)$$

where $I(t)$ is the intensity of the peak after the relaxation delay t and I_0 is the peak intensity at zero time. The calculations were performed with SigmaPlot software (version 9.0). The $\{^1\text{H}\}$ - ^{15}N NOE values were calculated as follows:

$$\text{NOE} = I_{\text{noe}} / I_{\text{st}} \quad (2)$$

where I_{noe} is the peak intensity with proton presaturation and I_{st} without. All calculations were performed with Excel (Microsoft).

Reduced spectral density functions were calculated with the following equations as described by Farrow et al. ³⁹ using an in-house Excel program (Microsoft):

$$\sigma_{\text{NH}} = R_1(\text{NOE} - 1)\gamma_{\text{N}}/\gamma_{\text{H}} \quad (3)$$

$$J(0) = (6R_2 - 3R_1 - 2.72\sigma_{\text{NH}})/(3d^2 + 4c^2) \quad (4)$$

$$J(\omega_{\text{N}}) = (4R_1 - 5\sigma_{\text{NH}})/(3d^2 + 4c^2) \quad (5)$$

$$J(0.87\omega_{\text{H}}) = 4\sigma_{\text{NH}}/(5d^2) \quad (6)$$

where $d = \mu_0 h \gamma_{\text{N}} \gamma_{\text{H}} \langle r_{\text{NH}}^{-3} \rangle / (8\pi^2)$, $c = 3^{-1/2} \omega_{\text{N}}$, μ_0 is the permeability of free space, h is Planck's constant; γ_{N} and γ_{H} are the gyromagnetic ratios of ^{15}N and ^1H , respectively; $r_{\text{NH}} = 1.02$ Å is the N-H bond length; ω_{N} and ω_{H} are the Larmor frequencies of ^{15}N and ^1H , respectively; $\sigma = -160$ ppm. Relaxation measurements were undertaken at two fields to permit a finer mapping of the spectral density and more specifically to test the assumptions inherent in the reduced spectral density analysis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank David Keizer for assistance with preliminary experiments and Xiaodong Yang for helpful discussions. This work was supported in part by the National Institutes of Health (R01AI059229) and the Australian Research Council (DP0664723). R.S.N. acknowledges fellowship support from the NHMRC.

Abbreviations

3D7-6H	3D7 isoform of MSP2 with a C-terminal hexa-histidine tag
FC27-6H	FC27 isoform of MSP2 with a C-terminal hexa-histidine tag
CD	circular dichroism
CSI	chemical shift index
DPC	dodecylphosphocholine- ² H ₃₈
GPI	glycosyl-phosphatidyl-inositol
HPLC	high pressure liquid chromatography
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl β-D-thiogalactopyranoside
IUP	intrinsically unstructured protein
MSP2	merozoite surface protein 2
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PBS	phosphate buffered saline
PFG	pulsed field gradient
TCEP	tris(2-carboxyethyl)phosphine hydrochloride
TMAO	trimethylamine N-oxide

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature*. 2005; 434:214–217. [PubMed: 15759000]
2. Greenwood B. Between hope and a hard place. *Nature*. 2004; 430:926–927. [PubMed: 15318230]
3. Balu B, Adams JH. Functional genomics of *Plasmodium falciparum* through transposon-mediated mutagenesis. *Cell Microbiol*. 2006; 8:1529–1536. [PubMed: 16984409]
4. Genton B, Betuela I, Felger I, Al-Yaman F, Anders RF, Saul A, Rare L, Baisor M, Lorry K, Brown GV, Pye D, Irving DO, Smith TA, Beck HP, Alpers MP. A recombinant blood-stage malaria vaccine reduces *Plasmodium falciparum* density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. *J Infect Dis*. 2002; 185:820–827. [PubMed: 11920300]

5. Genton B, Al-Yaman F, Betuela I, Anders RF, Saul A, Baea K, Mellombo M, Taraika J, Brown GV, Pye D, Irving DO, Felger I, Beck HP, Smith TA, Alpers MP. Safety and immunogenicity of a three-component blood-stage malaria vaccine (MSP1, MSP2, RESA) against *Plasmodium falciparum* in Papua New Guinean children. *Vaccine*. 2003; 22:30–41. [PubMed: 14604568]
6. Fenton B, Clark JT, Khan CM, Robinson JV, Walliker D, Ridley R, Scaife JG, McBride JS. Structural and antigenic polymorphism of the 35- to 48-kilodalton merozoite surface antigen (MSA-2) of the malaria parasite *Plasmodium falciparum*. *Mol Cell Biol*. 1991; 11:963–971. [PubMed: 1990294]
7. Smythe JA, Coppel RL, Brown GV, Ramasamy R, Kemp DJ, Anders RF. Identification of two integral membrane proteins of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A*. 1988; 85:5195–5199. [PubMed: 3293051]
8. Felger I, Steiger S, Hatz C, Smith T, Beck HP. Antigenic cross-reactivity between different alleles of the *Plasmodium falciparum* merozoite surface protein 2. *Parasite Immunol*. 2003; 25:531–543. [PubMed: 15053774]
9. 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; 75:131–143. [PubMed: 8992312]
10. Smythe JA, Coppel RL, Day KP, Martin RK, Oduola AM, Kemp DJ, Anders RF. Structural diversity in the *Plasmodium falciparum* merozoite surface antigen 2. *Proc Natl Acad Sci U S A*. 1991; 88:1751–1755. [PubMed: 2000383]
11. Smythe JA, Peterson MG, Coppel RL, Saul AJ, Kemp DJ, Anders RF. Structural diversity in the 45-kilodalton merozoite surface antigen of *Plasmodium falciparum*. *Mol Biochem Parasitol*. 1990; 39:227–234. [PubMed: 2181307]
12. Adda CG, Murphy VJ, Sunde M, Waddington LJ, Vingas K, Anderson V, Masciantonio R, Howlett GJ, Hodder AN, Anders RF. *Plasmodium falciparum* merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. 2008 Submitted.
13. Low A, Chandrasekaran IR, Adda CG, Yao S, Sabo JK, Zhang X, Soetopo A, Anders RF, Norton RS. Merozoite surface protein 2 of *Plasmodium falciparum*: expression, structure, dynamics, and fibril formation of the conserved N-terminal domain. *Biopolymers*. 2007; 87:12–22. [PubMed: 17516503]
14. Yang X, Adda CG, Keizer DW, Murphy VJ, Rizkalla MM, Perugini MA, Jackson DC, Anders RF, Norton RS. A partially structured region of a largely unstructured protein, *Plasmodium falciparum* merozoite surface protein 2 (MSP2), forms amyloid-like fibrils. *J Pept Sci*. 2007; 13:839–848. [PubMed: 17883245]
15. Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem*. 2002; 269:2–12. [PubMed: 11784292]
16. Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002; 11:739–756. [PubMed: 11910019]
17. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*. 2005; 18:343–384. [PubMed: 16094605]
18. Feng ZP, Zhang X, Han P, Arora N, Anders RF, Norton RS. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. *Mol Biochem Parasitol*. 2006; 150:256–267. [PubMed: 17010454]
19. Chiti F, Dobson CM. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*. 2006; 75:333–366. [PubMed: 16756495]
20. Gebbink MF, Claessen D, Bouma B, Dijkhuizen L, Wosten HA. Amyloids--a functional coat for microorganisms. *Nat Rev Microbiol*. 2005; 3:333–341. [PubMed: 15806095]
21. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci*. 2002; 27:527–533. [PubMed: 12368089]
22. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. 2005; 6:197–208. [PubMed: 15738986]
23. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol*. 2007; 17:3–14. [PubMed: 17250999]

24. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. *Chem Rev.* 2004; 104:3607–3622. [PubMed: 15303830]
25. Bertoncini CW, Rasia RM, Lamberto GR, Binolfi A, Zweckstetter M, Griesinger C, Fernandez CO. Structural characterization of the intrinsically unfolded protein β -synuclein, a natural negative regulator of α -synuclein aggregation. *J Mol Biol.* 2007; 372:708–722. [PubMed: 17681539]
26. Sung YH, Eliezer D. Residual structure, backbone dynamics, and interactions within the synuclein family. *J Mol Biol.* 2007; 372:689–707. [PubMed: 17681534]
27. Shojania S, O'Neil JD. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. *J Biol Chem.* 2006; 281:8347–8356. [PubMed: 16423825]
28. Thapar R, Mueller GA, Marzluff WF. The N-terminal domain of the Drosophila histone mRNA binding protein, SLBP, is intrinsically disordered with nascent helical structure. *Biochemistry.* 2004; 43:9390–9400. [PubMed: 15260482]
29. Yao J, Chung J, Eliezer D, Wright PE, Dyson HJ. NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry.* 2001; 40:3561–3571. [PubMed: 11297422]
30. Bai Y, Chung J, Dyson HJ, Wright PE. Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions. *Protein Sci.* 2001; 10:1056–1066. [PubMed: 11316886]
31. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* 1996; 266:540–553. [PubMed: 8743705]
32. Kneller DG, Cohen FE, Langridge R. Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol.* 1990; 214:171–182. [PubMed: 2370661]
33. Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry.* 1999; 38:16424–16431. [PubMed: 10600103]
34. Yao J, Dyson HJ, Wright PE. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* 1997; 419:285–289. [PubMed: 9428652]
35. Zhang O, Forman-Kay JD, Shortle D, Kay LE. Triple-resonance NOESY-based experiments with improved spectral resolution: applications to structural characterization of unfolded, partially folded and folded proteins. *J Biomol NMR.* 1997; 9:181–200. [PubMed: 9090132]
36. Seavey BR, Farr EA, Westler WM, Markley JL. A relational database for sequence-specific protein NMR data. *J Biomol NMR.* 1991; 1:217–236. [PubMed: 1841696]
37. Buevich AV, Shinde UP, Inouye M, Baum J. Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J Biomol NMR.* 2001; 20:233–249. [PubMed: 11519747]
38. Dyson HJ, Wright PE. Elucidation of the protein folding landscape by NMR. *Methods Enzymol.* 2005; 394:299–321. [PubMed: 15808225]
39. Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. Spectral density function mapping using ^{15}N relaxation data exclusively. *J Biomol NMR.* 1995; 6:153–162. [PubMed: 8589604]
40. le Maire M, Champeil P, Møller JV. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta.* 2000; 1508:81–111.
41. Spera S, Bax A. Empirical correlation between protein backbone conformation and C^α and C^β nuclear magnetic resonance chemical shifts. *J Am Chem Soc.* 1991; 113:5490–5492.
42. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J Mol Biol.* 1991; 222:311–333. [PubMed: 1960729]
43. Wishart DS, Sykes BD. Chemical shifts as a tool for structure determination. *Methods Enzymol.* 1994; 239:363–392. [PubMed: 7830591]
44. Schwarzsinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc.* 2001; 123:2970–2978. [PubMed: 11457007]
45. Bernado P, Bertoncini CW, Griesinger C, Zweckstetter M, Blackledge M. Defining long-range order and local disorder in native α -synuclein using residual dipolar couplings. *J Am Chem Soc.* 2005; 127:17968–17969. [PubMed: 16366524]

46. Bertonecini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, Jovin TM, Zweckstetter M. Release of long-range tertiary interactions potentiates aggregation of natively unstructured α -synuclein. *Proc Natl Acad Sci U S A*. 2005; 102:1430–1435. [PubMed: 15671169]
47. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. Mapping long-range interactions in α -synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J Am Chem Soc*. 2005; 127:476–477. [PubMed: 15643843]
48. Farrow NA, Zhang O, Forman-Kay JD, Kay LE. Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry*. 1997; 36:2390–2402. [PubMed: 9054544]
49. Vekemans J, Ballou WR. *Plasmodium falciparum* malaria vaccines in development. *Expert Rev Vaccines*. 2008; 7:223–240. [PubMed: 18324891]
50. Eliezer D, Kutluay E, Bussell R Jr, Browne G. Conformational properties of α -synuclein in its free and lipid-associated states. *J Mol Biol*. 2001; 307:1061–1073. [PubMed: 11286556]
51. Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett*. 2005; 579:3346–3354. [PubMed: 15943980]
52. Sherman IW. Biochemistry of *Plasmodium* (malarial parasites). *Microbiol Rev*. 1979; 43:453–495. [PubMed: 94424]
53. Zhu M, Li J, Fink AL. The association of α -synuclein with membranes affects bilayer structure, stability, and fibril formation. *J Biol Chem*. 2003; 278:40186–40197. [PubMed: 12885775]
54. Kalthoff C. A novel strategy for the purification of recombinantly expressed unstructured protein domains. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2003; 786:247–254.
55. Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal Biochem*. 2000; 287:252–260. [PubMed: 11112271]
56. Laue, TM.; Shah, BD.; Ridgeway, TM.; Pelletier, SL. Computer-aided interpretation of analytical sedimentation data for proteins; Analytical Ultracentrifugation in Biochemistry and Polymer Science; The Royal Society of Chemistry, Cambridge. 1992.
57. Schuck P. Sedimentation analysis of non-interacting and self-associating solutes using numerical solutions to the Lamm equation. *Biophys J*. 1998; 75:1503–1512. [PubMed: 9726952]
58. 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]
59. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR*. 1995; 6:135–140. [PubMed: 8589602]
60. Dingley AJ, Mackay JP, Chapman BE, Morris MB, Kuchel PW, Hambly BD, King GF. Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2. *J Biomol NMR*. 1995; 6:321–328. [PubMed: 8520223]
61. Gibbs SJ, Johnson CS. A PFG NMR experiment for accurate diffusion and flow studies in the presence of Eddy currents. *J Magn Reson*. 1991; 93:305–402.
62. Yao S, Howlett GJ, Norton RS. Peptide self-association in aqueous trifluoroethanol monitored by pulsed field gradient NMR diffusion measurements. *J Biomol NMR*. 2000; 16:109–119. [PubMed: 10723990]
63. Bartels CH, Xia TH, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomolecular NMR*. 1995; 5:1–10.
64. Goddard, TD.; Kneller, DG. SPARKY 3. University of California; San Francisco, CA: 2004.
65. Johnson BA. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol*. 2004; 278:313–352. [PubMed: 15318002]
66. Schwarzsinger S, Kroon GJ, Foss TR, Wright PE, Dyson HJ. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR*. 2000; 18:43–48. [PubMed: 11061227]

67. Merutka G, Dyson HJ, Wright PE. 'Random coil' ^1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J Biomol NMR*. 1995; 5:14–24. [PubMed: 7881270]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Domain structure of FC27 and 3D7 MSP2 isoforms based on amino acid sequence analysis (GenBank accession numbers J03828, M28891, M28892, M59765-68).7,10,11

Figure 2.

Circular dichroism spectroscopy of MSP2 in aqueous solution. (a,b) Mean residue ellipticity ($^{\circ}\text{cm}^2\text{dmol}^{-1}$) plotted as a function of wavelength (nm) for 3D7-6H (panel A) and FC27-6H (panel B) solubilised at an initial concentration of 0.15 mg/mL in buffer at pH 3.4 and $t = 0$ (dashed line), pH 3.4 and $t = 168$ h (solid line), pH 6.2 and $t = 0$ (solid line + circle), pH 6.2 and $t = 168$ h (solid line + triangle), pH 7.4 (dash-dot-dot line), and pH 7.4 and $t = 168$ h (long-short-short dashes). (c) Change in ellipticity, $\epsilon(\text{M}^{-1}\text{cm}^{-1})$ plotted as a function of wavelength (nm) for FC27-6H at pH 3.4 and $t = 0$ (circles) and $t = 7$ days post incubation at

room temperature (diamonds). The nonlinear least squares best fits employing the CDPro software package⁵⁵ with the SDP48 reference set and CONTINLL algorithm are shown as solid lines. The $t = 0$ data yielded a nonlinear best-fit to 7.4% α -helix, 1.6% β -strand, 8.9% turn and 82.1% unordered structure. By contrast, the $t = 7$ days data resulted in a nonlinear best-fit to 27.9% α -helix, 16.8% β -strand, 20.6% turn and 34.7% unordered structure

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3.

Sedimentation velocity analyses of 3D7-6H and FC27-6H MSP2. (a,b) The absorbance at 235 nm is plotted as a function of radius (cm) at 12 min intervals for 3D7-6H at pH 3.4 (panel a) and FC27-6H at pH 3.4 (panel b). The raw absorbance data (symbols) is overlaid with the nonlinear least-squares best-fits (solid lines) to a single non-interacting species (panel a) and a mixture of 4 species (panel b) fitted using the program SEDFIT. All nonlinear fits yielded rmsd values <0.0061 and Runs test Z values <7.40 . *Top panels* – Residuals for the nonlinear least squares best fits described below plotted as a function of radial position (cm) from the axis of rotation. (c) The absorbance at 235 nm is plotted as a function of radius for freshly prepared FC27-6H (i.e. at $t = 0$) at pH 3.4 and FC27-6H incubated at room temperature and pH 3.4 for $t = 24$ h, $t = 48$ h, $t = 72$ h and $t = 168$ h as indicated. The data shown represents the initial absorbance profile as a function of radius

from the axis of rotation in the XL-I analytical ultracentrifuge at 40,000 rpm. (d) The average absorbance signal due to monomeric MSP2 plotted as a function of time following incubation at room temperature of 3D7-6H at pH 7.4 (open squares + dashed line), 3D7-6H at pH 6.2 (open triangles + dashed line), 3D7-6H at pH 3.4 (open circles + dashed line), FC27-6H at pH 7.4 (solid squares + solid line), FC27-6H at pH 6.2 (solid triangles + solid line), and FC27-6H at pH 3.4 (solid circles + solid line). Each data point represents the mean absorbance value at 235 nm averaged over radii values of 6.1 to 7.1 cm (as shown for example in panel c for FC27-6H at pH 3.4) enclosed by the standard deviation (error bars).

Figure 4. ^1H - ^{15}N HSQC (800 MHz) for FC27 MSP2 in 10 mM HOAc at 25 °C. Peaks are labeled with residue numbers in the native protein; overlapped resonances are labeled with numbers separated by slashes.

Figure 5.

Secondary chemical shifts HN (a), H^{α} (b), C^{α} (c), and CO (d) along the sequence of FC27 MSP2 in 10 mM HOAc at 25 °C. Dotted lines represent threshold values (the residues with chemical shift deviations beyond threshold values have a propensity to adopt ordered secondary structure). The positions of the disulfide bond and helical regions inferred from the CSI consensus results are shown on the top panel. Random coil chemical shift references are from Schwarzsinger et al.;⁶⁶ we used these random coil values as our spectra were obtained under acidic conditions. We also undertook a CSI analysis using another set of random coil chemical shifts,⁶⁷ which were derived under non-denaturing conditions and only for proton atoms. The result was consistent with the analysis documented in this figure, except that additional residues in the N-terminal region now appeared to be involved in helical structure (not shown).

Figure 6. ¹⁵N relaxation parameters of FC27 MSP2 in 10 mM HOAc at 25 °C. (a) R_1 , (b) R_2 , and (c) steady-state {¹H}-¹⁵N NOE. Relaxation parameters measured at 600 and 800 MHz are shown in filled and open circles, respectively. A schematic representation of helical regions inferred from the CSI consensus plot and the disulfide bond location is shown at the top.

Figure 7. Reduced spectral density functions of FC27 MSP2. (a) $J(0)$, (b) $J(\omega_N)$, and (c) $J(0.87\omega_H)$ calculated from relaxation parameters measured at 600 MHz (filled circles) and 800 MHz (open circles) using Eqs 3–6. A schematic representation of helical regions inferred from the CSI consensus plot and the disulfide bond location is shown at the top.

Figure 8.

Superposition of backbone amide region of ^1H - ^{15}N HSQC spectra (500 MHz) of full-length FC27 MSP2 (red, 2048 $\times$ 256 points) with that of the N-terminal 28-mer peptide MSP2₁₋₂₅ 13 (green, 1024 $\times$ 128 points), both in 10 mM HOAc at 25 °C. The peaks labelled with negative numbers were residues from the vector prior to the start of the native sequence. In addition, Lys6 was not observed in the spectrum of full-length protein.

Figure 9.

Superposition of ^1H - ^{15}N HSQC spectra (500 MHz) of FC27 MSP2 at pH 4.9 in the absence (green) and presence (red) of 100 mM DPC at 25 °C. (a) Peaks that showed the largest shifts and/or broadening are labelled. The upper panel (b) shows the spectral region containing resonances from side chains of Asn, Gln and other residues, the arrows highlight side chain peaks that appear in the presence of DPC.

Table 1
Hydrodynamic properties of monomeric 3D7-6H and FC27-6H MSP2 determined by analytical ultracentrifugation

pH	$s_{20,w}^0$ (Svedberg, S)	M^I (kDa)	$D^I \times 10^{-7}$ (cm ² /s)	a/b^2	M_{eq}^3 (kDa)
3D7-6H					
3.4	1.60	24.7	4.92	17.2	20.4
6.2	1.68	24.5	5.10	15.1	22.1
7.4	1.64	25.7	4.85	16.0	23.8
FC27-6H					
3.4	1.62	25.1	4.96	16.2	ND ⁴
6.2	1.65	24.5	4.95	15.6	22.7
7.4	1.64	25.3	4.93	16.0	23.2

¹ Determined from nonlinear least squares best-fit to a single non-interacting species.
² Axial ratio (a/b) assuming a prolate ellipsoidal structure calculated using the v method.⁵⁶
³ Equivalent molar mass (M_{eq}) calculated from the global nonlinear least squares best-fit assuming a single species to sedimentation equilibrium data obtained at 25,000 and 35,000 rpm.
⁴ ND = not determined due to significant aggregation during the experiment.

Table 2

Nonlinear least squares best fit parameters to sedimentation velocity data of freshly prepared FC27-6H MSP2 in solution buffered at pH 3.4

Model	s^1 (S)	M^2 (kDa)	Proportion ³ (%)	Statistics
Non-interacting mixture of 4 species	1.57	24.2	71	rmsd = 0.00604 & Runs test Z = 7.39
	101	1,340	5.0	
	246	2,220	16	
	338	6,750	8.0	

¹_s = sedimentation coefficient

²*M* = apparent molecular mass

³ Concentration of each species as a percentage of total.

Table 3

Hydrodynamic properties of FC27-6H MSP2 (229 residues, 23.8 kDa) under different conditions, measured using PFG-NMR

pH (buffer)	3.6 (10 mM HOAc)	7.0 (PBS)	7.25 (10 mM phosphate)
$D(\text{protein})^I$ m ² /sec	(6.97±0.15)E-11	(7.41±0.18)E-11	(7.97±0.23)E-11
$D(\text{dioxane})$ m ² /sec	1.29E-9	1.25E-9	1.27E-9
$R(\text{protein})/R(\text{dioxane})= D(\text{dioxane})/D(\text{protein})^I$	18.4	16.9	16.0
$R(\text{dioxane})$ nm		0.212	
$R(\text{protein})$ nm	3.91±0.10	3.57±0.08	3.39±0.11
R_F nm		2.29	
R_U nm		4.89	
R_{MG} nm		2.56	
$R_{IU(PMG)}$ nm		3.35	
$R_{IU(coil)}$ nm		4.05	

^I D and R are abbreviations for diffusion coefficient and hydrodynamic radius respectively. $R_{IU(PMG)}$, $R_{IU(coil)}$, R_F , R_{MG} , and R_U represent the theoretically predicted hydrodynamic radii of intrinsically unstructured pre-molten globule, intrinsically unstructured coil, folded, molten globule, and fully unfolded states, respectively. The experimental values were determined for 2 mg/mL vaccine samples using PFG NMR method. The theoretically predicted values were calculated as described by Uversky,¹⁵ except that the predicted R_U is estimated with the equation $R_h=2.21N^{0.57\text{\AA}}$,³³ where R_h represents hydrodynamic radius and N the residue number.

Table 4

Hydrodynamic properties of 3D7-6H MSP2 (237 residues, 24.2 kDa) under different conditions, measured using PFG-NMR

pH (buffer)	3.6 (10 mM HOAc)	7.0 (PBS)	3.5 (10 mM HOAc + 135 mM NaCl)
$D(\text{protein})^I$ m ² /sec	(6.58±0.13)E-11	(7.80±0.16)E-11	(7.61±0.12)E-11
$D(\text{dioxane})$ m ² /sec	1.27E-9	1.26E-9	1.25E-9
$R(\text{protein})/R(\text{dioxane})= D(\text{dioxane})/D(\text{protein})$	19.3	16.2	16.5
$R(\text{protein})$ nm	4.10±0.10	3.43±0.07	3.49±0.06
R_F nm		2.30	
R_U nm		4.99	
R_{MG} nm		2.58	
$R_{IU(PMG)}$ nm		3.37	
$R_{IU(coil)}$ nm		4.07	

^I Refer to Table 3 for abbreviations and details of calculated values shown here. $R(\text{dioxane}) = 0.212$ nm