JOURNAL OF
PEPTIDE SCIENCE

A partially structured region of a largely unstructured protein, *Plasmodium falciparum* merozoite surface protein 2 (MSP2), forms amyloid-like fibrils

XIAODONG YANG,[a] **CHRISTOPHER G. ADDA,**[b] **DAVID W. KEIZER,**[a,‡] **VINCE J. MURPHY,**[b] **MICHAEL M. RIZKALLA,**[c] **MATTHEW A. PERUGINI,**[d] **DAVID C. JACKSON,**[c] **ROBIN F. ANDERS**[b] and **RAYMOND S. NORTON**[a,*]

[a] The Walter & Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville 3050, Australia
[b] Department of Biochemistry, La Trobe University, Bundoora, 3086, Australia
[c] Department of Microbiology & Immunology, The University of Melbourne, Parkville 3010, Australia
[d] Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Parkville 3010, Australia

Abstract: Merozoite surface protein 2 (MSP2) from the human malaria parasite *Plasmodium falciparum* is expressed as a GPI-anchored protein on the merozoite surface. It has been implicated in the process of erythrocyte invasion and is a leading vaccine candidate. MSP2 is an intrinsically unstructured protein (IUP), and recombinant MSP2 forms amyloid-like fibrils upon storage. We have examined synthetic peptides corresponding to sequences in the conserved *N*-terminal region of MSP2 for the presence of local structure and the ability to form fibrils related to those formed by full-length MSP2. In a 25-residue peptide corresponding to the entire *N*-terminal region of mature MSP2, structures calculated from NMR data show the presence of nascent helical and turn-like structures. An 8-residue peptide from the central region of the *N*-terminal domain (residues 8–15) also formed a turn-like structure. Both peptides formed fibrils that were similar but not identical to the amyloid-like fibrils formed by full-length MSP2. Notably, the fibrils formed by the peptides bound both Congo Red and Thioflavin T, whereas the fibrils formed by full-length MSP2 bound only Congo Red. The propensity of peptides from the *N*-terminal conserved region of MSP2 to form amyloid-like fibrils makes it likely that this region contributes to fibril formation by the full-length protein. Thus, in contrast to the more common pathway of amyloid formation by structured proteins, which proceeds via partially unfolded intermediates that then undergo β-aggregation, MSP2 is an example of a largely unstructured protein with at least one small structured region that has an important role in fibril formation. Copyright © 2007 European Peptide Society and John Wiley & Sons, Ltd.

Supplementary electronic material for this paper is available in Wiley InterScience at http://www.interscience.wiley.com/jpages/1075-2617/suppmat/

INTRODUCTION

Malaria remains one of the most prevalent diseases in many developing countries, with an estimated 515 million clinical cases in 2002 [1] and more than one million deaths a year [2]. The development of a vaccine against *Plasmodium falciparum*, the cause of the most serious form of human malaria, remains one of the major priorities for addressing this global health burden. Merozoite surface proteins from the asexual blood stage of the parasite are considered important potential components of a vaccine against *P. falciparum* [1]. Merozoite surface protein 1 (MSP1) and apical membrane antigen 1 (AMA1) are two leading asexual blood stage vaccine candidates currently under development [3], but data from a field trial in Papua New Guinea indicated that merozoite surface protein 2 (MSP2) is also a promising candidate. In this trial a vaccine containing a combination of three antigens, including the 3D7 allelic form of MSP2, significantly reduced parasite densities in 6- to 9-year-old children living in an area where the transmission intensity of *P. falciparum* was high [4].

MSP2 is an integral-membrane protein of ~30 kDa anchored in the plasma membrane of the merozoite by a *C*-terminal GPI moiety [5,6]. Although highly polymorphic, the many allelic forms can be classified into two families, represented by the alleles from strains 3D7 and FC27 [7,8]. The cleavage of a predicted 19-residue

* Correspondence to: R. S. Norton, The Walter & Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, 3050, Australia; e-mail: ray.norton@wehi.edu.au
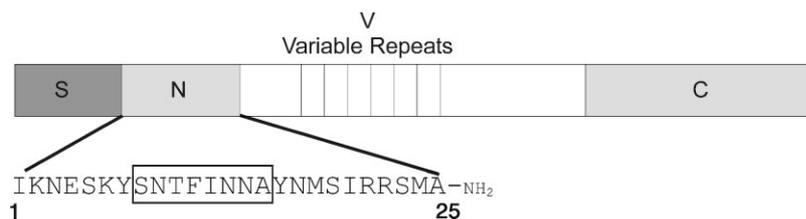‡ Present address: Bio21 Institute, University of Melbourne, Victoria, 3010, Australia.

WILEY
InterScience®
DISCOVER SOMETHING GREAT

Figure 1 Sequences of peptides used in this study and the relationships between them. S is the 19-residue signal peptide that is predicted to be cleaved to yield the mature protein. N is the 25-residue *N*-terminal region conserved between allelic families. V is the central repeat region that contains the highly polymorphic region of MSP2. C is the conserved 74-residue *C*-terminal domain. The $MSP2_{1-25}$ and $MSP2_{8-15}$ (boxed) peptides shown formed the basis of this study. All numbering of residues in this study is relative to mature full-length MSP2.

signal sequence generates the mature protein with a 25-residue *N*-terminal region that is perfectly conserved across both allelic families of MSP2 (Figure 1). The central variable region includes tandemly arrayed repeat sequences, the length, sequence and number of which differ between the different alleles [9]. The *C*-terminal region comprising 74 residues is also conserved across the allelic families. Despite the very different central regions in the dimorphic forms of MSP2, both are characterized by sequences of low complexity and biased amino acid composition. The polypeptides of both allelic types are unusually hydrophilic and deficient in nonpolar residues, which are typical characteristics of an IUP.

Conformational epitopes stabilized by intramolecular disulfide bonds are critical for the induction of protective antibody responses by AMA1 and the *C*-terminal fragments of MSP1 in clinical development [10–12]. Although MSP2 contains a single disulfide bond in the *C*-terminal conserved region, there is no evidence that it has an impact on the antigenicity or immunogenicity of MSP2. Indeed, the conserved *N*- and *C*-terminal regions appear to be poorly immunogenic as a relatively small proportion of the human antibody response recognizes epitopes that are common to both dimorphic forms of MSP2 [13–15]. The vaccine containing 3D7 MSP2 tested in Papua New Guinea limited the development of parasites with allelic forms of MSP2 belonging to the 3D7 family but not the FC27 family. Thus, the central variable region of MSP2 appeared to be the target of the induced immune response that limited parasite development.

There is currently no information on the three-dimensional structural features of MSP2 that are important for inducing a protective immune response. Adda *et al.* [16] have shown recently that full-length MSP2 has the characteristics of an IUP and that it forms amyloid-like fibrils. Upon digestion of this polymeric form of MSP2 with proteinase K, an MSP2 fragment that was resistant to degradation was obtained. This fragment was subsequently identified as the conserved *N*-terminal region of MSP2. Here we report that the conserved *N*-terminal region of MSP2 contains nascent helical and turn-like structures, which are preserved in an 8-residue synthetic fragment. Peptides

corresponding to the entire *N*-terminal domain and the 8-residue fragment both formed fibril structures with the characteristics of amyloid and some similarities to those formed by the full-length protein. It appears likely that the conserved *N*-terminal region of MSP2 contributes to the formation of MSP2 polymers with a cross-β structure.

MATERIALS AND METHODS

Peptide Synthesis and Sample Preparation

Peptides used in this study (Figure 1) were synthesized using Fmoc chemistry. The correct masses of the peptides were verified by mass spectrometry, and the overall purities (>90%) were confirmed by reverse-phase HPLC. Samples were prepared for NMR by dissolving 2–3 mg of the appropriate peptide in 95% H_2O/5% ^2H_2O containing 10 mM acetic acid, pH 3.4 ($MSP2_{1-25}$) or 4.7 ($MSP2_{8-15}$). pH was measured at room temperature and no allowance was made for isotope effects. Peptide samples for analyses of fibril formation were dissolved in PBS at pH 7.4.

NMR Spectroscopy

Two-dimensional homonuclear COSY, TOCSY (spin-lock time 37.5 ms) and NOESY (mixing times 100 and 250 ms) spectra were acquired on a DRX-600 spectrometer (Bruker Biospin). Water signals were suppressed using the WATERGATE pulse sequence [17]. Backbone amide exchange was measured by dissolving the peptides in ^2H_2O at pH 3.4 and recording a series of 1D and TOCSY spectra at 5 °C. Once the exchange was complete, E-COSY and 250-ms NOESY spectra were acquired. All amide proton signals except those of Phe11, Ile12 and Ile20 disappeared within 30 min, indicating that the protons were in fast exchange. The amide proton signal of Ile12 in $MSP2_{1-25}$ was still visible in a 2D TOCSY acquired 24 h after ^2H_2O addition. A pulsed-field gradient (PFG) longitudinal eddy-current delay pulse sequence was used for diffusion measurements [18–20]. All spectra were collected at 5 °C unless otherwise stated and were referenced to an impurity peak present at 0.15 ppm. Spectra were processed with TopSpin, version 1.3 (Bruker Biospin) and analyzed with XEASY, version 1.3.13 [21]. Chemical shifts for the *N*-terminal region of MSP2 have been deposited in the BioMagResBank(BMRB) [22] with accession number 6698.

Structure Calculations

Structures for MSP2$_{1-25}$ and MSP2$_{8-15}$ were calculated using distance restraints from NOESY spectra acquired at 600 MHz. Intensities of NOE cross-peaks measured in XEASY were calibrated using the CALIBA macro from the program CYANA [23]. Initial structures were calculated using torsion angle dynamics and simulated annealing protocols in CYANA, and structures were optimized for a low target function. The same constraint set was also used to calculate a new family of 100 structures using XPLOR-NIH [24]. The 50 lowest-energy structures were selected for energy minimization in a box of water. A final family of 20 lowest-energy structures was chosen for analysis using PROCHECK-NMR [25] and MOLMOL [26]. The 30 lowest-energy structures before water refinement had no experimental distance violations >0.2 Å or dihedral angle violations >5°. For the final 20 structures after water refinement, no distance violations >0.2 Å or dihedral angle violations >5° were present. In some cases the structures derived from CYANA were subjected directly to unrestrained energy minimization in a box of water using GROMACS [27]. Structural figures were prepared using MOLMOL [26] and PyMol (http://www.pymol.org).

CD Spectroscopy

CD spectra of MSP2$_{8-15}$ were measured at a peptide concentration of 0.15 mg ml^{-1} dissolved in 10 mM Na acetate, pH 3.4; 10 mM Na acetate, pH 4.7; and 10 mM Na phosphate, pH 7.4. Spectra were recorded on a Jasco J815 CD spectrometer over the wavelength range 190–250 nm at a temperature of 20 °C. Data was collected using a step size of 0.5 nm, signal averaging time of 2.0 s and a slit bandwidth of 1.0 nm in a 1-mm path length quartz cuvette. CD spectra were fitted using a database comprising 22 soluble proteins (SP22×) employing the CONTIN algorithm [28], which is included in the CDPro software package available at http://lamar.colostate.edu/~sreeram/CDPro/.

Electron Microscopy

Peptide and protein samples (10 µl) were applied to 400-mesh copper grids coated with a thin layer of carbon. Excess material was removed by blotting, and samples were negatively stained twice with 10 µl of a 2% (w/v) uranyl acetate solution (Electron Microscopy Services). The grids were air-dried and viewed using a Jeol JEM-2010 transmission electron microscope operated at 80 kV. Peptide samples for EM were prepared within a week of the peptides being reconstituted in PBS for the dye-binding assays described below. Recombinant FC27 MSP2 lacking any *N*- or *C*-terminal polyhistidine tags was expressed in *Escherichia coli* and purified using a procedure that exploited the intrinsically unstructured characteristics of MSP2 [29]; this material was incubated in PBS at room temperature for 3 days prior to analysis by EM and dye-binding assays.

Congo Red and Thioflavin T Binding Assays

For Congo Red assays, samples containing 200 µg ml^{-1} protein or peptide and 2.5 µM Congo Red in PBS were incubated for 5 min before being scanned at wavelengths between 400 and 600 nm using a Cary 1E UV-visible spectrophotometer. The difference spectrum was created by subtracting the absorbance of 2.5 µM Congo Red in PBS for each MSP2 sample at each wavelength. Thioflavin T assays were performed on samples containing 100 µg ml^{-1} protein or peptide and 10 µM Thioflavin T in PBS. Samples were scanned at wavelengths between 440 and 600 nm and emission spectra were collected with an excitation wavelength of 417 nm using a Perkin Elmer luminescence spectrometer model LS 50 B. Peptide samples used for these assays were dissolved in PBS on the day the assays were conducted.

RESULTS

Although MSP2 is an IUP, proteinase K digestion studies on the amyloid-like fibrils formed by MSP2 identified a protected fragment that may contain local structure [16]. This fragment was identified by *N*-terminal sequencing as the conserved *N*-terminal region of MSP2 (Figure 1). In this study we have examined the properties of two synthetic peptides corresponding to sequences in the conserved *N*-terminal region of MSP2. One peptide, MSP2$_{1-25}$, corresponded to the entire *N*-terminal region that is conserved in all MSP2 alleles, commencing at the predicted *N*-terminus after cleavage of the signal peptide. The second was the central octapeptide SNTFINNA (MSP2$_{8-15}$), which was chosen because, as discussed below, it corresponds to a region that was found to induce protective immunity in mice and was identified as likely to form fibrils.
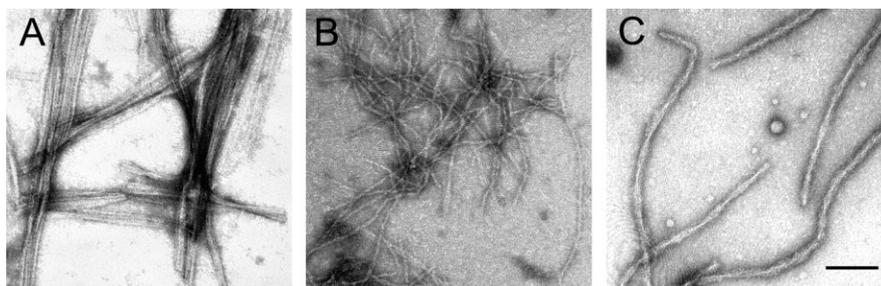

Figure 2 Electron micrographs of the fibrils formed by (A) MSP2$_{8-15}$ peptide, (B) MSP2$_{1-25}$ peptide and (C) recombinant full-length FC27 MSP2. Peptides and protein were incubated in PBS at 1 mg ml^{-1}. Samples were negatively stained and viewed using transmission electron microscopy. The scale bar represents 100 nm.

Fibril Formation

MSP2$_{1-25}$ and MSP2$_{8-15}$ peptides formed amyloid-like fibrils that differed in some characteristics from those formed by full-length MSP2 (Figure 2). When the lyophilized peptides were resuspended at 1 mg ml^{-1} in water, amyloid-like fibrils were found to be present. Dissociation and removal of those fibrils was achieved by heating the peptides to 80 °C and filtering them through a 0.02-μm syringe filter. The fibrils re-formed when the monomeric peptide was incubated in PBS, pH 7.4 (Figure 2), with the rate of fibril formation increasing with peptide concentration (data not shown). The results of representative fibril formation assays, as monitored by Thioflavin T fluorescence at peptide concentrations of 0.5 and 0.8 mM are shown in Figure S1 of the Supplementary Material. Fibrils formed by the MSP2$_{8-15}$ peptide comprised several protofilaments, each with a diameter of 3–4 nm (Figure 2(A)), which associated laterally to form rigid bunches of fibers that vary in length, up to 5 μm. The protofilaments formed by the MSP2$_{1-25}$ peptide appear to have diameters of 4–6 nm but, unlike the MSP2$_{8-15}$ fibrils, appear to have more flexibility, and only a few fibrils associate laterally to form twisted bunches that vary in diameter (Figure 2(B)). The fibrils formed by the peptides lack the regular twists apparent in the full-length MSP2 fibrils. Fibrils formed by recombinant full-length MSP2 have a diameter of about 13 nm and twists with a periodicity of about 75 nm, but the protofilaments are not clearly seen in the images of these fibrils (Figure 2(C)).

Fibrils formed by both peptides bound Congo Red and exhibited a red shift in the absorbance spectrum similar to that seen when fibrils formed by full-length MSP2 bound this dye (Figure 3(A)). An atypical feature of fibrils formed by full-length 3D7 MSP2 was their failure to bind Thioflavin T. In contrast, both MSP2$_{1-25}$ and MSP2$_{8-15}$ fibrils bound Thioflavin T, giving a fluorescence spectrum similar to that obtained when Thioflavin T bound to fibrils formed by the Aβ_{1-42} peptide (Figure 3(B)). These results suggest that the fibrils formed are amyloid in nature. In full-length MSP2, the binding site for Thioflavin T [30] may be blocked by the central or *C*-terminal domain of the protein.

NMR Spectroscopy

NMR experiments were performed on samples in 10 mM acetic acid at pH 3.4 at either 5 or 25 °C, although other pH values were also examined. Acidic conditions were used for most of the NMR experiments because fibril formation was slower at low pH than pH 7.4, as found for full-length MSP2 (unpublished data). Translational diffusion coefficients [20] of 1.2×10^{-10} and 2.2×10^{-10} m^2 s^{-1} were obtained at 5 °C for MSP2$_{1-25}$ and MSP2$_{8-15}$, respectively. Values for MSP2$_{1-25}$ compare well with values of 1.24×10^{-10} m^2 s^{-1} for F1, a
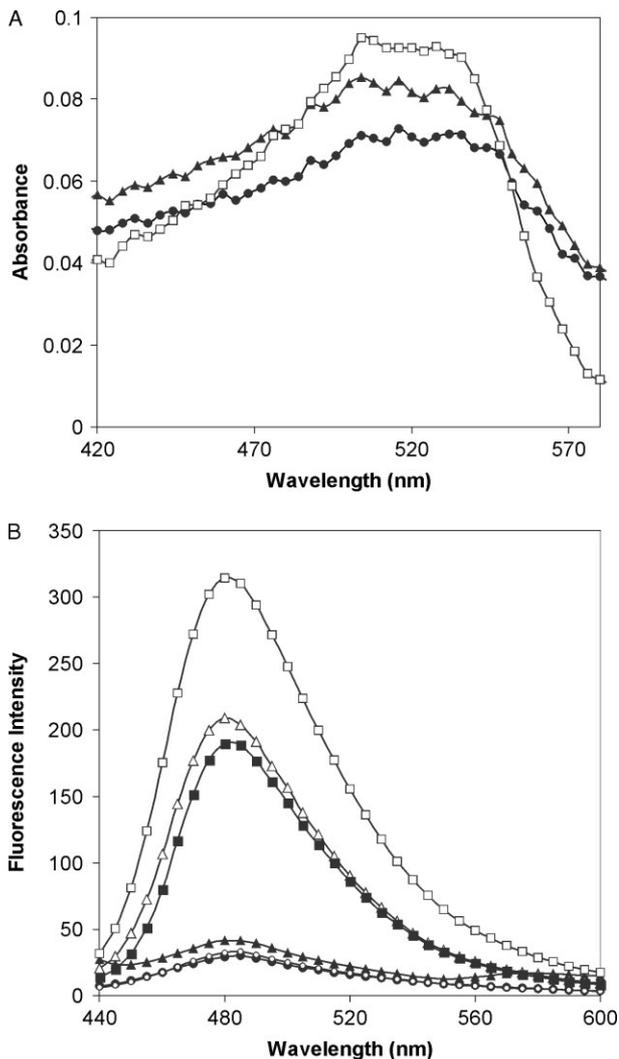
Figure 3 Characterization of fibrils formed by peptides from MSP2. (A) Difference spectrum of Congo Red binding to MSP2$_{8-15}$ (●), MSP2$_{1-25}$ (▲) and full-length 3D7 MSP2 fibrils (□). (B) Interactions of Thioflavin T. Thioflavin T alone (●), full-length 3D7 MSP2 monomer (○) and polymer (▲) plus Thioflavin T, MSP2$_{8-15}$ (△), MSP2$_{1-25}$ (■) and Aβ_{1-42} (□) peptides plus Thioflavin T.

15-residue peptide that binds to apical membrane antigen 1 [31], and 1.1×10^{-10} m^2 s^{-1} for J1, a 19-residue peptide that mimics an epitope on the same protein [32], implying that both MSP2$_{8-15}$ and MSP2$_{1-25}$ are monomeric under the solution conditions used in this study. This was confirmed by sedimentation equilibrium analysis (Figure S2 in Supplementary Material). Diffusion coefficient measurements were repeated regularly throughout the study to ensure that samples remained monomeric.

Spectra of MSP2$_{8-15}$ were recorded over the pH range 3–8 at a concentration of 0.1 mM in water at 5 °C. At pH 3.0, 3.6 and 4.7, the chemical shifts were essentially identical. At pH 6.1 the Asn9 backbone amide had disappeared, consistent with it being the second

residue (as exchange of the backbone amide of the second residue with solvent is catalyzed by the *N*-terminal ammonium moiety in peptides and proteins). At pH 7.3 the chemical shifts of most backbone and side-chain resonances were unchanged, although the backbone amides of Thr10 and Phe11 had moved slightly upfield (by 0.04–0.05 ppm) and become broader. At pH 8.0 the only backbone amide still visible was that of Ile12; presumably it is protected from solvent exchange by the turn-like structure (see below). This is consistent with the results of amide exchange experiments, which showed that Ile12 was the most slowly exchanging amide. These results indicate that the conformation of MSP2$_{8-15}$ is invariant over the pH range 3–7.3, as expected, given that there are no titratable groups in the peptide with p*K*a values in this range (the *C*-terminus was amidated). NMR data for structure calculations were therefore acquired at pH 4.7, where all backbone amide resonances could be observed. It should be noted that aggregation and fibril formation do occur at this pH, albeit more slowly than at pH 7.4 (the signal intensity from MSP2$_{8-15}$ dropped to roughly half its initial value over 20 days at 4 °C owing to the formation of fibrillar aggregates). CD spectra of MSP2$_{8-15}$ over the pH range 3.4–7.4 also showed no pH dependence (see below).

A detailed study of the pH dependence of the spectrum of MSP2$_{1-25}$ was not undertaken in view of the results with MSP2$_{8-15}$. MSP2$_{1-25}$ has only one additional side chain with a p*K*a in the range 3–8, and this residue, Glu4, is in the least structured region of the peptide (see below) near the *N*-terminus.

Chemical shift assignments were completed for all backbone protons and most of side-chain protons for both peptides at 5 or 25 °C using standard 2D techniques (Tables S1–S2 in Supplementary Material). The fingerprint region of a representative NOESY spectrum of MSP2$_{8-15}$ at 5 °C is also shown in the Supplementary Material (Figure S3). Figure 4 shows the deviations from random coil chemical shifts for C$^{\alpha}$H and backbone amide resonances of both peptides at 5 °C. The C$^{\alpha}$H shifts were within about 0.2 ppm of the random coil values except at the *N*-terminus. In MSP2$_{1-25}$ the backbone amide chemical shifts were downfield of random coil values at the *N*-terminus but upfield in the central and *C*-terminal regions, whereas in MSP2$_{8-15}$ they were consistently upfield except for the second residue (Asn9). Indeed, the magnitude and direction of the amide deviations are similar for residues 10–15 of the two peptides.

CD Spectroscopy

In order to support the NMR data, CD spectra of MSP2$_{8-15}$ were recorded over the range 190–250 nm (Figure 5). The CD spectra at pH 3.4, 4.7, and 7.4 were very similar, as indicated by the significant overlap
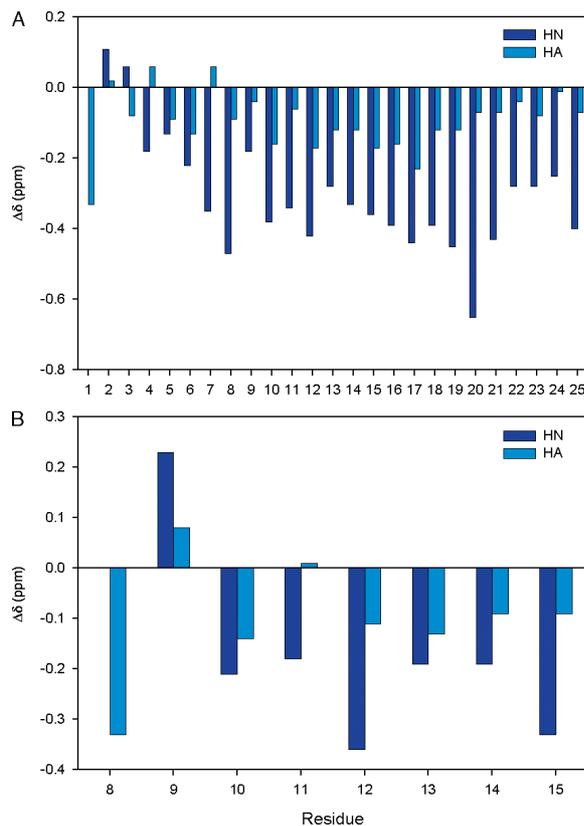

Figure 4 Deviation of chemical shifts from random coil values at 5 °C. Chemical shifts of backbone amides and C$^{\alpha}$H resonances from (A) MSP2$_{1-25}$ and (B) MSP2$_{8-15}$ peptides were compared to the random coil values of Merutka *et al.* [33]; random coil values were subtracted from the observed peptide chemical shifts.

in the spectra (Figure 5(A)). The minimum at approximately 200 nm was indicative of significant random or unordered structure, while the broad minimum at 210–230 nm suggested that the peptide contains some helical and/or β-structure. This assertion was confirmed when the data at pH 3.4 were fitted by nonlinear regression using the CONTIN algorithm and SP22× reference set [28], yielding a best fit to 18% helix, 20% β-structure and 62% turn and/or unordered structure (Figure 5(B)) with an r.m.s.d. 0.240. Nonlinear analyses were also conducted with the CONTIN algorithm and different protein reference sets, yielding different proportions of secondary structure; for example, nonlinear analysis using the SDP48 reference set resulted in a best fit to 3.9% helix, 16.1% strand and 80% turn and/or unordered structure with an r.m.s.d. of 0.255 (fit not shown). Similar nonlinear least-squares fits were obtained for MSP2$_{8-15}$ at pH 4.7 and 7.4 (fits not shown).

Solution Structures

Defining the conformations of linear peptides lacking disulfide or other covalent constraints requires some
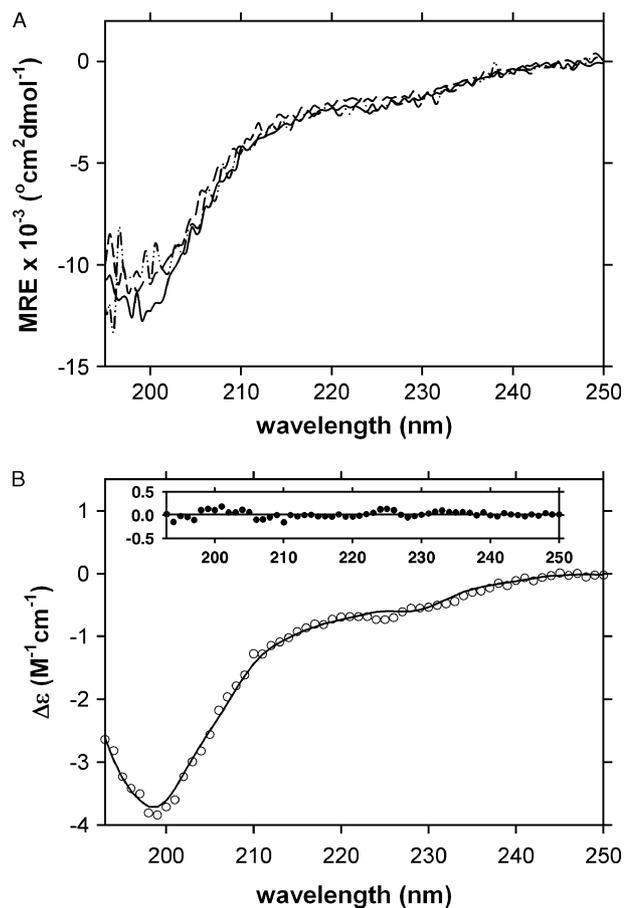
A

B

Figure 5 Circular dichroism spectroscopy of MSP2$_{8-15}$. (A) Mean residue ellipticity (MRE) is plotted as a function of wavelength for the peptide at a concentration of 0.15 mg ml^{-1} in 10 mM Na acetate, pH 3.4 (solid line), 10 mM Na acetate, pH 4.7 (dashed line), and 10 mM Na phosphate, pH 7.4 (dash-dot-dot line). Data were collected at a temperature of 20 °C using a Jasco J-815 CD spectrometer in step mode with a 0.5 nm stepsize, a response time of 2 s and using a 1-mm quartz cuvette. (B) Data generated for the peptide at pH 3.4 converted to $\Delta\varepsilon$ as a function of wavelength (open symbols) overlaid with the nonlinear least-squares best fit using the CONTIN algorithm and SP22× reference set [28]. The best fit was obtained to 18% helix, 20% β-structure, and 62% turn and/or unordered structure (r.m.s.d. 0.240). The inset shows residuals for the nonlinear best fit plotted as a function of wavelength (nm).

caution, as the presence of just a few medium- or long-range restraints can bias the calculated structures when the vast majority of restraints is from short-range interactions (in contrast to the situation for folded proteins, in which there are numerous medium- and long-range restraints). We have calculated structures for both MSP2$_{1-25}$ and MSP2$_{8-15}$ because their chemical shifts suggested the presence of nonrandom structure and their NOESY spectra showed several medium-range NOEs. In an attempt to assess how well the NMR data defined the structures, we have also

calculated them in different ways using the same NMR restraint set. Figure 6 shows stereo views of the families of structures for MSP2$_{8-15}$ calculated using CYANA (Figure 6(A)), CYANA followed by unrestrained energy minimization in GROMACS (Figure 6(B)) and CYANA followed by restrained simulated annealing and energy minimization in X-PLOR (Figure 6(C)). Figure 6(D) shows the closest-to-average structures from each set overlaid with each other. Clearly, the derived structures are influenced by the calculation procedure, indicating that they are not adequately defined by the NMR data. Nonetheless, they concur in showing a chain reversal in the vicinity of Phe11 and Ile12, even though the exact nature of the turn was not defined.

Similar sets of structure calculations were carried out on MSP2$_{1-25}$. In this case, interpretation of the NMR data was further compromised by peak overlap. The three families of structures, superimposed over residues 9–14 and 17–24, two regions of better-defined local structure according to angular order parameter analyses, are shown in the Supplementary Material (Figure S4). Superpositions of the closest-to-average structures from each set of structure calculations are shown in Figure 7. It is clear that there is a chain reversal in the vicinity of Phe11 and Ile12 but the nature of the turn and its exact position are not well defined. In addition, there are turn-like structures closer to the *C*-terminus, which the backbone chemical shifts (Figure 4) suggest has the properties of a nascent helix. Table S3 in the Supplementary Material contains a summary of structural parameters and statistics for structures of MSP2$_{1-25}$ and MSP2$_{8-15}$ calculated in X-PLOR.

β-Aggregation Propensity

As these peptides contained elements of ordered structure and formed amyloid-like fibrils, it was of interest to evaluate whether these properties were predicted by currently available algorithms. Predictions of secondary structure for MSP2$_{1-25}$ were performed using several programs (Figure S5, Supplementary Material). Four of the five programs predicted a helix in the *C*-terminal of the peptide, while two predicted short extended strand segments. Predictions could not be performed for the MSP2$_{8-15}$ peptide with these programs because of its short length. TANGO, a program that predicts the propensity of a sequence to participate in cross β-aggregation [34] was applied to both peptides (Figure S6, Supplementary Material). Neither was predicted to aggregate in this manner at 25 °C, but some potential for β-aggregation was observed at 5 °C. Within MSP2$_{1-25}$ a 10-residue sequence was predicted to have a propensity for β-aggregation, which corresponded to the MSP2$_{8-15}$ peptide with a Tyr residue at either terminus. This
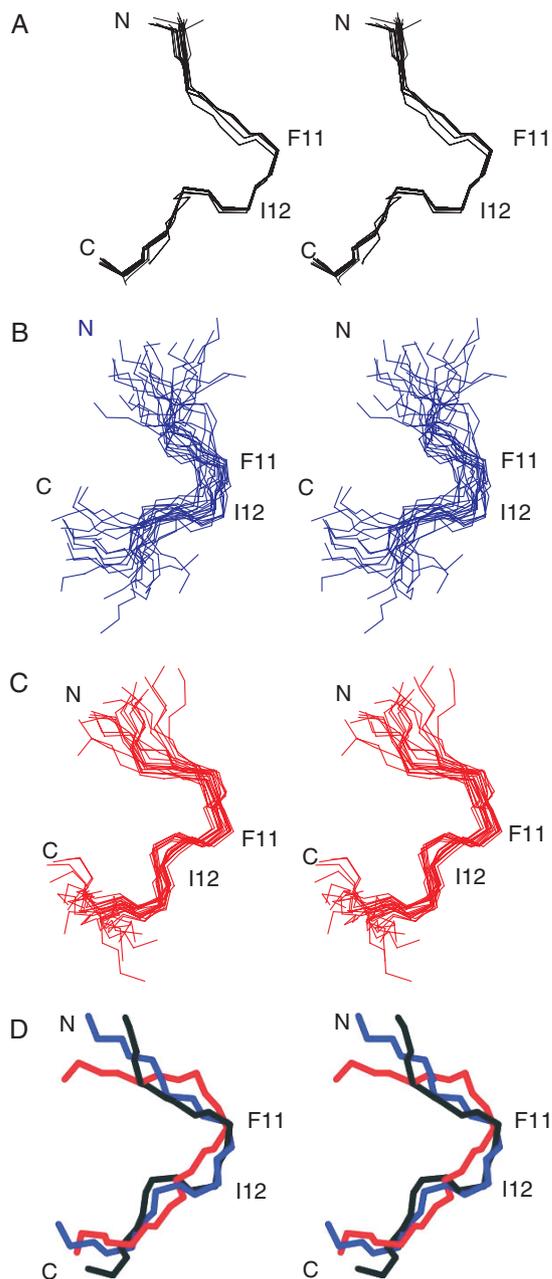
Figure 6 Structure of MSP2$_{8-15}$ at 5 °C. (A) Stereo view of superposition of the 20 lowest-energy structures for MSP2$_{8-15}$ calculated in CYANA; structures were superimposed over the backbone heavy atoms (N, C$^{\alpha}$ and C) of residues Asn9–Asn14. (B) As in A but following unrestrained energy minimization in GROMACS. (C) As in A but following restrained simulated annealing and energy minimization in X-PLOR (D) Closest-to-average structures from each set of structure calculations (black, CYANA; blue, GROMACS; red, X-PLOR).

Figure 7 Structure of MSP2$_{1-25}$ at 5 °C. Stereo views of superposition of the closest-to-average structures from each set of structure calculations (black, CYANA; blue, GROMACS; red, X-PLOR). Structures were superimposed over the backbone heavy atoms (N, C$^{\alpha}$ and C) of residues Asn9–Asn14 (A) and Asn17–Met24 (B).

10-mer sequence (MSP2$_{7-16}$) alone had the highest β-aggregation propensity. A peptide with this sequence was produced synthetically but could not be analyzed by solution NMR because it formed a gel within 2 h at 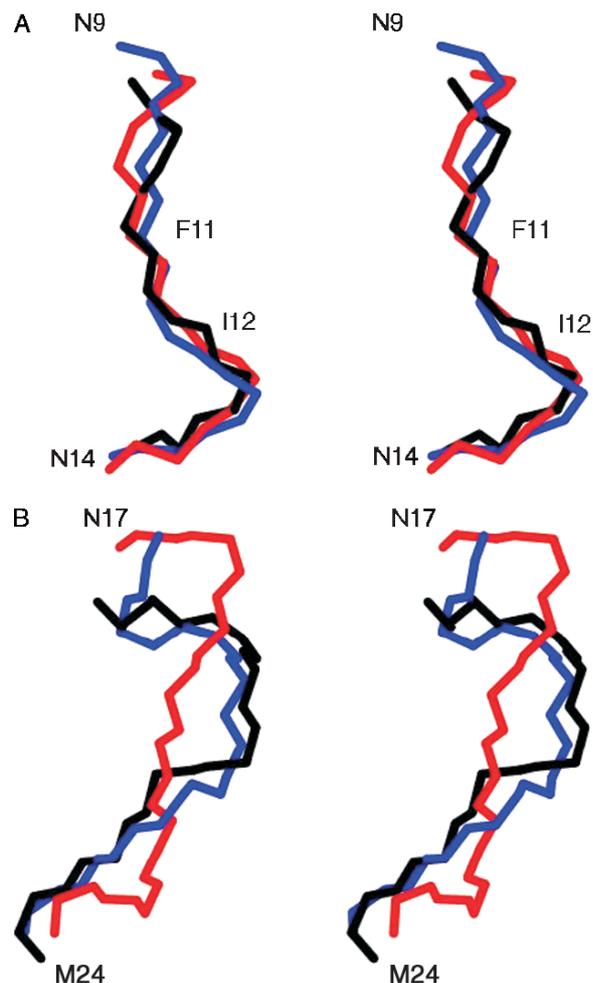pH 3.4 and temperatures between 5 and 35 °C. Figure S6 also shows that the propensity for β-aggregation is more pronounced at higher pH. This increased propensity was shown experimentally for the formation of amyloid-like fibrils for full-length MSP2 (data not shown). It should be noted that TANGO does not predict β-hairpin or amyloid fibril formation [34].

DISCUSSION

Although full-length MSP2 is predicted to be disordered [16], regions of local structure may exist, particularly in regions in which the amino acid sequence is conserved among all known isolates of *P. falciparum*. This is supported by the observation of Adda *et al.* [16] that MSP2 fibrils digested with proteinase K resulted in a stable fragment representing the *N*-terminal region. In this study, we have investigated fibril formation by the conserved *N*-terminal region of MSP2 and a shorter

8-residue sequence within this region and characterized their conformations in aqueous solution. Both peptides contained turn-like structures. The solution structure determined for MSP2$_{8-15}$ was better defined than that of MSP2$_{1-25}$, probably as a consequence of reduced spectral overlap, and showed a turn-like structure centered on residues 10–12. The longer peptide also showed turn-like or nascent helical structure towards the *C*-terminus, around residues 17–24. Clearly, these peptides do not adopt a single unique folded structure; rather, they exist as a mixture of partially structured states in rapid dynamic equilibrium with each other, but dominated by turn-like structures in the central and *C*-terminal regions. As these structural elements are determined by local interactions, we predict that they will also be present in full-length MSP2 and may help promote the intermolecular interactions that are necessary for the formation of the cross-β structure of amyloid-like fibrils.

Secondary structure prediction programs are usually designed to find motifs in well-ordered globular proteins and are not well suited to predicting the β-strand propensity for amyloid-forming molecules [35]. TANGO, a statistical mechanics algorithm, predicts the propensity for regions of a protein to be involved in β-aggregates rather than standard parallel or antiparallel β-sheets [34]. The factors that determine whether a protein has the propensity to form β-sheets or β-aggregates partially overlap, but are not necessarily the same [34]; those proteins with high β-sheet potentials will not necessarily proceed to form β-aggregates. TANGO analysis predicted that neither full-length MSP2 nor either of the MSP2$_{1-25}$ and MSP2$_{8-15}$ peptides would form β-aggregates at 25 °C. However, when predictions were performed on MSP2$_{1-25}$ at 5 °C, a propensity for β-aggregation was observed over a 10-residue segment that included the MSP2$_{8-15}$ peptide (Figure S6, Supplementary Material). Predictions for MSP2$_{8-15}$ also showed a propensity for β-aggregation at 5 °C. A third sequence with an extra Tyr residue at either terminus was predicted to have an even higher aggregation potential, suggesting a possible role of aromatics in stabilizing these fibrils (see below). The predicted increase in β-aggregation potential at physiological pH or at higher ionic strengths is consistent with observations made during studies of full-length MSP2 (unpublished data).

It was shown in the mid-1990s that even short peptide sequences could form amyloid fibrils. The peptide KLVFF was responsible for the propensity of a truncated version of Aβ to form amyloid fibrils [36], and many diverse peptide fragments have since been shown to form amyloid fibrils. There was an overabundance of aromatic residues, particularly Phe, in these sequences [37], leading to the suggestion that π–π interactions may be important in stabilizing amyloid formation through favorable energy contributions from the π-orbital stacking and the ability to provide directionality to the growing fibril [37]. Several recent papers have also noted important roles for aromatic residues in amyloid formation [38,39] although one proposed a more limited role [40]. Both of the peptides examined in this study contained a Phe that could assist in stabilizing interactions in this manner. The 10-residue peptide (MSP2$_{7-16}$) contains two additional Tyr residues, potentially explaining its higher propensity for aggregation compared with MSP2$_{8-15}$. Pawar *et al.* [41] recently published a list of aggregation propensities for individual amino acids, in which Trp, Phe, and Tyr occupied the first, second and fourth positions respectively. Hosia *et al.* [42] showed that the structural context in which a potential fibril-forming sequence is present can prevent fibril formation by favoring self-limiting oligomerization over polymerization. Whether turn-like structures promote or inhibit the formation of amyloid structures is context dependent. Examples such as Aβ_{21-30} form a β-turn structure yet remain monomeric in solution [43], while others such as OspA contain a β-turn that controls β-hairpin alignment and subsequent amyloid fibril formation [44]. Different types of β-turns have differing propensities for β-sheet nucleation with type I' and II' having the highest propensity [45].

The amyloid-like nature of the fibrils formed by the MSP2$_{1-25}$ and MSP2$_{8-15}$ peptides in this study was confirmed by the observation that they bound Congo Red and Thioflavin T. Krebs *et al.* [30] recently proposed that Thioflavin T binds to naturally occurring channels in the β-sheet. If this is the case, the lack of Thioflavin T binding observed for full-length MSP2 may be related to the bulk of the protein hindering the access of dye molecules to this binding site.

Many proteins that form amyloid fibrils are either amphipathic or highly hydrophobic [46–48]. More diverse examples of peptides that can form amyloid-like fibrils have emerged recently; for example, a 23-residue polar peptide based on a β-hairpin from the *Borrelia* protein OspA has been shown to form a single-layer β-sheet while retaining its β-hairpin conformation [44]. The MSP2-derived peptides in this study are an example of peptides that are neither highly hydrophobic nor amphipathic yet are able to form turn-like structures and can form amyloid-like fibrils.

IUPs are well known to form fibrils. Some of these, for example α-synuclein and tau [41], are associated with neurodegenerative diseases. MSP2, however, is atypical in that it contains a small, partially structured, core that appears to have a role in amyloid-like fibril formation. In this respect its behavior is also distinct from that of natively folded proteins, in which amyloid formation proceeds via partially unfolded intermediates which subsequently undergo β-aggregation. It remains to be seen whether other IUPs that can form amyloid fibrils do so via regions of local structure, as appears to be the case for MSP2.

How are the fibrils observed *in vitro* related to MSP2 on the merozoite surface? The most direct evidence that MSP2 forms oligomers on the merozoite surface has come from the glutaraldehyde cross-linking of erythrocytic late-stage malaria parasites [16]. Indirect evidence has been obtained from immunofluorescence experiments with the mAb 6D8, which preferentially recognizes monomeric MSP2. MAb 6D8 failed to bind late-stage malaria parasites, in contrast to mAb 11E1, which recognizes both monomeric and polymeric MSP2 [16]. The conformational constraints imposed by the fibril structure may explain the loss of the 6D8 epitope in polymeric MSP2. Alternatively, the 6D8 epitope may be inaccessible in polymeric MSP2 and on the merozoite surface because it is buried in the cross-β structure that is generated by intermolecular interactions involving the conserved *N*-terminal sequence of MSP2.

Most naturally occurring antibodies in human populations are targeted to the central variable region of MSP2 [9,13–15] and the importance of this region as a target of protective responses is highlighted by the results of a vaccine trial in Papua New Guinea [4]. Nevertheless, conserved *N*-terminal sequences have been reported to induce antibodies that block invasion of *P. falciparum* merozoites [49] and the 8-residue sequence MSP2$_{8-15}$ has induced protection in immunized mice against the rodent parasites *P. chabaudi* [50] and *P. yoelii* [49]. As these rodent parasites lack MSP2 orthologs, this protection may reflect cross-reactions between conformational epitopes generated by these sequences in MSP2 and structural homologs of MSP2 in the other *Plasmodium* species. This would be consistent with the formation of aggregates with a cross-β structure on the merozoite surface.

Supplementary Material

Supplementary electronic material for this paper is available in Wiley InterScience at: http://www.interscience.wiley.com/jpages/1075-2617/suppmat/

Acknowledgements

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.

2. Greenwood B. Between hope and a hard place. *Nature* 2004; **430**: 926–927.

3. Richie TL, Saul A. Progress and challenges for malaria vaccines. *Nature* 2002; **415**: 694–701.

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.

5. 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.

6. 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.

7. 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.

8. Fenton B, Clark JT, Wilson CF, McBride JS, Walliker D. Polymorphism of a 35–48 kDa *Plasmodium falciparum* merozoite surface antigen. *Mol. Biochem. Parasitol.* 1989; **34**: 79–86.

9. Franks S, Baton L, Tetteh K, Tongren E, Dewin D, Akanmori BD, Koram KA, Ranford-Cartwright L, Riley EM. Genetic diversity and antigenic polymorphism in *Plasmodium falciparum*: extensive serological cross-reactivity between allelic variants of merozoite surface protein 2. *Infect. Immun.* 2003; **71**: 3485–3495.

10. Crewther PE, Matthew ML, Flegg RH, Anders RF. Protective immune responses to apical membrane antigen 1 of *Plasmodium chabaudi* involve recognition of strain-specific epitopes. *Infect. Immun.* 1996; **64**: 3310–3317.

11. Anders RF, Crewther PE, Edwards S, Margetts M, Matthew ML, Pollock B, Pye D. Immunisation with recombinant AMA-1 protects mice against infection with *Plasmodium chabaudi*. *Vaccine* 1998; **16**: 240–247.

12. Ling IT, Ogun SA, Holder AA. Immunization against malaria with a recombinant protein. *Parasite Immunol.* 1994; **16**: 63–67.

13. Lawrence N, Stowers A, Mann V, Taylor D, Saul A. Recombinant chimeric proteins generated from conserved regions of *Plasmodium falciparum* merozoite surface protein 2 generate antiparasite humoral responses in mice. *Parasite Immunol.* 2000; **22**: 211–221.

14. Thomas AW, Carr DA, Carter JM, Lyon JA. Sequence comparison of allelic forms of the *Plasmodium falciparum* merozoite surface antigen MSA2. *Mol. Biochem. Parasitol.* 1990; **43**: 211–220.

15. Taylor RR, Smith DB, Robinson VJ, McBride JS, Riley EM. Human antibody response to *Plasmodium falciparum* merozoite surface protein 2 is serogroup specific and predominantly of the immunoglobulin G3 subclass. *Infect. Immun.* 1995; **63**: 4382–4388.

16. Adda CG, Murphy VJ, Sunde M, Waddington LJ, Culvenor JG, 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. *Mol. Biochem. Parasitol.* 2007 (submitted).

17. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR* 1992; **2**: 661–665.

18. 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**: 395–402.

19. 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.

20. 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.

21. Bartels C, Xia TH, Billeter M, Güntert P, Wüthrich K. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* 1995; **6**: 1–10.

22. Seavey BR, Farr EA, Westler WM, Markley JL. A relational database for sequence-specific protein NMR data. *J. Biomol. NMR* 1991; **1**: 217–236.

23. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* 2002; **319**: 209–227.

24. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 2003; **160**: 65–73.

25. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 1996; **8**: 477–486.

26. Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 1996; **14**: 51–55, 29–32.

27. Lindahl E, Hess B, van der Spoel D. GROMACS: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* 2001; **7**: 306–317.

28. 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.

29. 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.

30. Krebs MR, Bromley EH, Donald AM. The binding of thioflavin-T to amyloid fibrils: localisation and implications. *J. Struct. Biol.* 2005; **149**: 30–37.

31. Keizer DW, Miles LA, Li F, Nair M, Anders RF, Coley AM, Foley M, Norton RS. Structures of phage-display peptides that bind to the malarial surface protein, apical membrane antigen 1, and block erythrocyte invasion. *Biochemistry* 2003; **42**: 9915–9923.

32. Sabo JK, Keizer DW, Feng ZP, Casey JL, Parisi K, Coley AM, Foley M, Norton RS. Mimotopes of apical membrane antigen 1: Structures of phage-derived peptides recognized by the inhibitory monoclonal antibody 4G2dc1 and design of a more active analogue. *Infect. Immun.* 2007; **75**: 61–73.

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

34. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* 2004; **22**: 1302–1306.

35. Yoon S, Welsh WJ. Detecting hidden sequence propensity for amyloid fibril formation. *Protein Sci.* 2004; **13**: 2149–2160.

36. Tjernberg LO, Naslund J, Lindqvist F, Johansson J, Karlstrom AR, Thyberg J, Terenius L, Nordstedt C. Arrest of β-amyloid fibril formation by a pentapeptide ligand. *J. Biol. Chem.* 1996; **271**: 8545–8548.

37. Gazit E. A possible role for π-stacking in the self-assembly of amyloid fibrils. *FASEB J.* 2002; **16**: 77–83.

38. Tartaglia GG, Cavalli A, Pellarin R, Caflisch A. The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* 2004; **13**: 1939–1941.

39. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. U.S.A.* 2005; **102**: 315–320.

40. Tracz SM, Abedini A, Driscoll M, Raleigh DP. Role of aromatic interactions in amyloid formation by peptides derived from human amylin. *Biochemistry* 2004; **43**: 15901–15908.

41. Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* 2005; **350**: 379–392.

42. Hosia W, Bark N, Liepinsh E, Tjernberg A, Persson B, Hallen D, Thyberg J, Johansson J, Tjernberg L. Folding into a β-hairpin can prevent amyloid fibril formation. *Biochemistry* 2004; **43**: 4655–4661.

43. Lazo ND, Grant MA, Condron MC, Rigby AC, Teplow DB. On the nucleation of amyloid β-protein monomer folding. *Protein Sci.* 2005; **14**: 1581–1596.

44. Ohnishi S, Koide A, Koide S. The roles of turn formation and cross-strand interactions in fibrillization of peptides derived from the OspA single-layer β-sheet. *Protein Sci.* 2001; **10**: 2083–2092.

45. Gibbs AC, Bjorndahl TC, Hodges RS, Wishart DS. Probing the structural determinants of type II′ β-turn formation in peptides and proteins. *J. Am. Chem. Soc.* 2002; **124**: 1203–1213.

46. Zhang S, Holmes T, Lockshin C, Rich A. Spontaneous assembly of a self-complementary oligopeptide to form a stable macroscopic membrane. *Proc. Natl. Acad. Sci. U.S.A.* 1993; **90**: 3334–3338.

47. Kelly JW. Alternative conformations of amyloidogenic proteins govern their behavior. *Curr. Opin. Struct. Biol.* 1996; **6**: 11–17.

48. Mihara H, Takahashi Y. Engineering peptides and proteins that undergo α-to-β transitions. *Curr. Opin. Struct. Biol.* 1997; **7**: 501–508.

49. Lougovskoi AA, Okoyeh NJ, Chauhan VS. Mice immunised with synthetic peptide from N-terminal conserved region of merozoite surface antigen-2 of human malaria parasite *Plasmodium falciparum* can control infection induced by *Plasmodium yoelii yoelii* 265BY strain. *Vaccine* 1999; **18**: 920–930.

50. Saul A, Lord R, Jones GL, Spencer L. Protective immunization with invariant peptides of the *Plasmodium falciparum* antigen MSA2. *J. Immunol.* 1992; **148**: 208–211.