A novel strategy for the synthesis of the cysteine-rich protective antigen of the malaria merozoite surface protein (MSP-1)

Knowledge-based strategy for disulfide formation

JANE C. SPETZLER, CHANG RAO and JAMES P. TAM

Department of Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Received 9 June, accepted for publication 10 August 1993

The most promising antigen for a protective malaria vaccine is a cysteine-rich domain at the carboxyl terminus of the merozoite surface protein (MSP-1). Passive transfer of anti-MSP-1 antibody or immunization of MSP-1 against infection challenge confers protection in primate and rodent models. The antigen belongs to the three-disulfide epidermal growth factor (EGF) family based on the alignment of the six cysteines. In the K1 strain there are, however, only four cysteines corresponding to the four carboxyl cysteines of EGF. Furthermore, disulfide pairing would produce a non-EGF pattern. Because this cysteine-rich antigen is conformation-dependent, and reduction of the disulfide bonds abolishes antigenicity, we used a synthetic analog to investigate the probable disulfide pairing of this antigen. This paper describes the synthesis, folding and disulfide pairings of two 50-residue cysteine-rich peptides. One contains two disulfides (VK-50) derived from the native sequence of MSP-1 of the Thailand K1 strain (aa 1629-1679). The other contains an EGF-like, three-disulfide [Cys-9,14]VK-50 peptide. Both peptides were synthesized by a solid-phase method using Fmoc-chemistry. The crude peptide of VK-50 was folded, and the disulfide was oxidized by the DMSO method to obtain a structure with an expected disulfide pairing of 3-4, and 5-6. The specific pairing pattern of 1-3, 2-4 and 5-6 in [Cys-9,14]VK-50 was obtained using a 'knowledge-based' (KB) strategy for their formation. Purified VK-50 and [Cys-9,14]VK-50 had the correct molecular weight, as shown by CF-252 fission ionization mass-spectrometry. The disulfide pairings were confirmed by enzymatic digestion. These two peptides can be conjugated to the multiple peptide antigen core for immunization. The immunological results will allow us to conclude the correct disulfide pairing and the conformational importance of this antigen.

Key words: DMSO-mediated disulfide formation; epidermal growth factor-like domain; Fmoc synthesis; malaria vaccine; Reagent T; solid-phase peptide synthesis

Malaria infects 200 million humans each year with a fatality rate of 2–3%, and represents a major health problem in developing countries. To help solve this health problem, a viable vaccine candidate is needed for clinical trials. The major focus of malaria vaccine development has been the identification and character-
J. C. Spetzler et al.

Evidence strongly suggests the protective role is due to an antigen-specific MSP-1 via a humoral response and that MSP-1 would be promising vaccine candidate. MSP-1 is a 190–220 kDa glycoprotein with no determined function. The protective antigen has recently been mapped to a cysteine-rich region at the MSP-1 terminus (1–3, 6), specifically the amino half (aEGF) of the two tandem repeat EGF-like domains. This cysteine-rich region is a conformational epitope, since reduction of the disulfides abolishes binding to the protective polyclonal or mAbs. Moreover, protective mAbs do not bind to overlapping linear fragments of this region. The antigen is likely to be a discontinuous epitope contributed by amino acids arranged in a spatial relationship.

A major problem in developing MSP-1 or its smaller fragments containing the cysteine-rich region is the difficulty in obtaining recombinant proteins in bacterial or yeast expression systems that are equivalent in immunological potency to the affinity-purified native protein (7). The lack of a correct disulfide formation in the expression system may have contributed to this difficulty. In this paper we report on the synthesis of a 50-residue peptide, referred to as VK-50, encompassing the first EGF-like domain of MSP-1 of Plasmodium falciparum of the Thailand K1 strain (13) that corresponds to amino acid sequence 1629–1679 (Fig. 1). We have chosen to investigate K1 as a model to determine the disulfide pairings of the aEGF domain of MSP-1.

Sequences obtained from different strains or species have shown that six cysteines are in alignment with those corresponding to EGF (14, 15). Since EGF has a disulfide pairing of 1–3, 2–4, 5–6, the expectation is that the aEGF domain will contain the same disulfide pattern. However, in aEGF of K1 strains, the amino terminal cysteines are missing and the four carboxyl end cysteines are found to correspond to positions 3, 4, 5 and 6 of EGF. Thus, the most comparable EGF-like disulfide isomer of this K1 strain would have a disulfide pairing pattern of 3–4, 5–6, which is different from that of EGF. The implication would be that other aEGF domains in MSP-1 would have a nonEGF disulfide pairing pattern. Alternatively, the K1 sequence derived from the cDNA sequence is incorrect. To explore these possibilities, [Cys-9,14]VK-50 was prepared with an EGF-like disulfide pairing pattern. In this paper we report the synthesis and characterization of VK-50 and [Cys-9,14]VK-50. In preparing [Cys-9,14]VK-50 we used a new disulfide formation scheme that specifically produced the EGF-like disulfide pairing pattern. We refer this method as the ‘knowledge-based’ (KB) strategy for disulfide formation. Because the protective antigen of MSP-1 is conformation-dependent, these two peptides may be able to distinguish the structure by various polyclonal and monoclonal antibodies, which in turn could allow us to arrive at a conclusion of the disulfide pairing of the EGF-like domain in MSP-1.

**EXPERIMENTAL**

**Reagents**

Fmoc-amino acids were purchased from Bachem (Torrance, California). DMF and DCM (both biotechnology grade) were obtained from Fisher Scientific; DCC and EDT (Fluka Chem.); piperidine, thioanisole, thiophenol and anisole (Aldrich Chem.); and TFA (Halocarbon). All other chemicals were the purest grade available.

**Solid-phase peptide synthesis of VK50 and [Cys-9,14]VK50**

Both peptides were synthesized by solid-phase method using 4-alkoxybenzyl alcohol resin (0.9 mmol/g substitution). All amino acids were protected with Nα-9-fluorenylmethoxycarbonyl (Fmoc). Side-chain protecting groups were Arg(Pmc), Asn(Trt), Asp(OtBu), Cys(Trt), Cys(Acm), Gln(Trt), Glu(OtBu), Lys(Boc), Ser(tBu), Thr(tBu) and Tyr(tBu). Attachment of the first amino acid to the resin was performed in DCC (2.5–5 equiv.) and dimethylaminopyridine (10 mol%) in DCM DMF (1:1 v/v), and the unreacted hydroxymethyl end was capped by acetic anhydride. The entire synthesis was performed on a Beckman 990 synthesizer and consisted of 49 cycles, each comprising (i) 20 min

---

**FIGURE 1**

Amino acid sequence of the N-terminal EGF domain of MSP-1 of the Thailand K1 strain (top). The homologous sequence of the MAD20 strain (Papua New Guinea) with three disulfide bonds is shown for comparison (middle). The probable disulfide pairings of the K1 sequence and the would-be K1-like structure for MAD20 vs. the EGF-like structure (bottom) are shown for comparison.
Synthesis of protective malaria antigen

Deprotection with 20% piperidine in DMF; (ii) washing of the resin with DCM and DMF and (iii) coupling with DCC (3 equiv.) alone for 1 h in DCM/DMF. Coupling of Fmoc-Asn(Trt), Fmoc-Gln(Trt) and Fmoc-Arg(Pmc) was performed in HBTU (3 equiv.) and DIEA (3 equiv.) in DMF/DCM. All couplings were monitored by the quantitative ninhydrin test (22).

Cleavage
The protected peptide-resin (400 mg) was stirred in 80 mL TFA/thioanisole/thiophenol/ethanedithiol/water (82.5:5:5:2:5:5 v/v) for 2 h at room temperature. The solution was filtered and the resin was washed with 95% TFA (20 mL). TFA was removed under vacuum. After precipitation with ethyl ether the crude peptide mixture was extracted with 20 mL of 6 M urea and 0.05 M glycine in 0.1 M Tris buffer, pH 8.5. For comparison, Reagent R (TFA/thioanisole:EDT:anisole, 90:5:3:2 v/v) was also used in the same manner (23).

DMSO-folding and disulfide bond formation
The folding and DMSO oxidation was performed according to Tam et al. (18). The peptide solution was dialyzed (Spectrapor 6, MW cutoff 2000) against 2 L each of deacrated and N2-purged solutions consisting of 6, 4 and 2 M urea, all in 0.1 M Tris-HCl, glycine (0.05 M), pH 8.5, for 6–12 h. The peptide solution after dialysis was diluted to about 200 mL with 10–20% DMSO in 1 M urea solution in 0.1 M Tris-HCl, glycine (0.05 M), pH 8.5 and stirred slowly at room temperature until the folding was completed. All peptides were purified by preparative high-performance liquid chromatography (HPLC; Vydac C18 reversed-phase column, 2.5 x 30 cm; Waters Associates, Bedford, MA) eluted with 0.045% TFA/CH3CN at a flow rate of 10 mL/min. The major fraction was collected, lyophilized and identified.

Removal of Acm protecting groups
The peptide (6 mg, 1 mmol) in H2O/MeOH (6:1, pH 3, 9.6 mL) was saturated with N2. Then I2 in MeOH (10 mM) was added dropwise to the peptide solution under nitrogen until the color turned yellow (25). The solution was stirred for 25 min at room temperature under nitrogen. The peptide solution was cooled to 0°C in an ice-water bath and treated dropwise with 10 mM sodium thiosulfate until the color was removed. In addition, 50% excess of 1 M sodium thiosulfate was added. Methanol was removed by vacuum. The peptide was purified by semi-preparative HPLC (C18 reversed-phase column) under isocratic conditions at a flow rate of 3 mL/min.

Characterization
Amino-acid analysis: amino acid hydrolysis of purified peptides was performed in 5.7 N HCl at 110 °C for 24 h: [Cys-9,14]VK50 (2 disulfide pairs): Asp/Asn (5) 4.83, Ser (2) 2.30, Glu/Gln (8) 8.33, Gly/Thr (6) 5.73, Val/Met (4) 4.06, Leu (3) 2.79, Tyr (1) 0.55, Phe (2) 1.79, His (2) 1.86, Lys (6) 5.48, Arg (3) 2.79, Ile (1). 1. The peptides were analyzed on a HPLC C18-reversed-phase column (Vydac, 5 mm, 0.4 x 25 cm) eluted with 0.045% TFA/CH3CN at a flow rate of 1.5 mL/min (Figs. 2 and 3). The synthetic peptides were analyzed by Cf-252 fission ionization mass spectrometry (26, 27): [Cys-9,14] (2 disulfide pairs with a 4-hydrazinobenzoyl modification): (cal/found)

**FIGURE 2**
C18 reversed-phase HPLC profiles of two cleavage reactions using Reagent T (A and B) and Reagent R (C and D). Crude products after dialysis and before oxidation of disulfides (A and C) and after oxidation with 10–20% DMSO for 24 h (B and D). The shaded peak in each chromatogram is the desired product.
J.C. Spetzler et al.

6079.2/6079, VK-50 (with a 4-hydrazinobenzoyl modification) (calc./found) 6000.78/6000. The purified peptide (11.5 nmol) was digested with thermolysin (14.6 mg) in 0.1 M pyridine/acetic buffer, pH 6.5 (187 mL) for 5 h at 40 °C. The sample was filtered and chromatographed by C18 reversed-phase HPLC. All the peaks were collected and subjected to amino acid analysis. [Cys-9,14] (2-disulfide pairs): Arg(1) 0.77, Glu(2) 2.38, Gly(1) 1.3, Cys(2) 1.76, His(1) 0.5, Leu(1) 1, Lys(1) 0.56, Phe(1) 1.2, Ser(1) 1.03. VK50: Arg(1) 1.07, Glu(3) 2.76, Gly(1) 1.3, Cys(3) 2.94, Leu(1) 1, Lys(2) 1.61, Ser(1) 1.3, Val(1) 1.5 (Table 1). Cys was determined as cysteic acid after performic acid oxidation.

RESULTS

Protecting-group scheme and analysis
Syntheses of VK-50 and [Cys-9,14]VK-50 (Fig. 1) were carried out by solid-phase peptide synthesis (16, 17) using Fmoc-chemistry (19, 20, 28) on Wang resin (21). A maximum protecting group scheme was used for all side chains including those amides in Asn and Gln that were protected as trityl derivatives. In VK-50, the side chains of all four cysteines were protected as trityl. In [Cys-9,14]VK-50 two sets of thiol protecting groups were used in the protecting-group scheme (see next section). In the first set, four Cys at position 9, 20, 32 and 43, forming disulfide pairs of 1–3 and 5–6, were protected as trityl, whereas in the second set, Cys-14 and Cys-30, which were to form the disulfide pair 2–4, were protected as acetamido derivatives (2, 4). All couplings were mediated by DCC alone with satisfactory results, except for Arg(Pmc), Asn(Trt) and Gin(Trt), in which HBTU was used.

Cleavage was mediated by a mixture of TFA, thioanisole, thiophenol, 1,2-ethanediethiol (EDT) and water in a volume ratio of 82.5:5:5:2.5:5. This mixture was rich in thiol and was designed by our laboratory (31, 32) for the purpose of scavenging the long-lived carbonium ions generated from TFA cleavage. The trityl, tert-butyl and aryl sulfonyl cations resulting from cleavage of the side-chain protecting groups are excellent alkylating agents for thiols and sulfides, and would be very damaging to cysteine-rich peptides such as VK-50.

Furthermore, anisole is a potential alkylating agent source in which the protonated anisole transfers the methyl group to a sulfide in low acidity and in the presence of a weak base nucleophile such as thioanisole. The presence of both would be counterproductive, and anisole was not used in our reagent mixture. Acidity was lowered by the addition of water to further minimize the alklylation reaction, which is enhanced by a more acidic condition. We refer to our cleavage reagent as Cleavage Reagent T (T for thiol!). For comparison, we compared Reagent T with an established reagent Reagent R which contains TFA, anisole, thioanisole and EDT (Fig. 2). The result showed that Reagent T gave a superior result with the desired product shown as a major peak. In contrast, cleavage by Reagent R gave a substantially lower yield and did not provide a major peak.

DMSO-mediated disulfide formation of VK-50
After TFA cleavage, the crude peptide of VK-50 was dialyzed sequentially in descending concentrations of urea in 0.1 M Tris buffer at pH 8.5 to allow the peptide to fold into a native conformation. At 1 M urea, the solution was diluted and 10–20% DMSO was added to oxidize the thiols to disulfides. Concentrations of DMSO lower than 10% gave slow oxidation rates, probably due to the inhibitive effect of thiol scavengers carried over from the cleavage reaction. DMSO has been found to be an excellent oxidant for protein disulfide formation (18) because it is a water-miscible solvent and a weak oxidizing agent which does not oxidize Met to Met(O). In general, DMSO was added directly to the crude mixture after dialysis to initiate the disulfide oxidation process. The use of DMSO also eliminates the need of highly diluted solution. Disulfide formation was completed within 24 h as monitored by reversed-phase HPLC. Only one of the three possible disulfide isomers was detected. The disulfide pairings were (3–4) and (5–6) (EGF disulfide nomenclature, Cys numbering starting from the N-terminus, see Fig. 4) and were the intended product. It also showed that the product with these disulfide pairings was the most thermodynamically stable of the three possible disulfide isomers. The peptide was then purified by preparative C18 reversed-phase HPLC.

Knowledge-based (KB) strategy in disulfide formation of [Cys-9,14]VK-50
There were two disulfide linkages in VK-50. An orthogonal protecting group scheme such as trityl vs. acetamido groups could have been used. However, we were interested in obtaining the thermodynamic product and were confident that the isomeric disulfide product could be influenced using the solvent chaperone properties of DMSO (30). In [Cys-9,14]VK-50 there were six Cys, and an orthogonal protecting group scheme to form each of the three disulfides selectively

---

**TABLE 1**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Fragment</th>
<th>Disulfide pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>VK-50</td>
<td>SGCFRHLK</td>
<td>Cys30-Cys30</td>
</tr>
<tr>
<td></td>
<td>ACm</td>
<td>Cys32-Cys43</td>
</tr>
<tr>
<td>[Cys-9,14]VK-50</td>
<td>LCKCEERE</td>
<td>Cys32-Cys43</td>
</tr>
<tr>
<td></td>
<td>GSKCV</td>
<td></td>
</tr>
</tbody>
</table>

354
Synthesis of protective malaria antigen

FIGURE 3
C_{18} reversed-phase HPLC profiles of \([\text{Cys-9,14}]\text{VK-50}\) after cleavage with Reagent T (A), after oxidation with 20% DMSO for 24.5 h (B) and purified product after treatment with I\(_2\) (C).

would be a complex undertaking. Concurrent oxidation of all six sulphhydryls would produce 15 disulfide iso- 
ers, some of which were likely to be stable, which would make purification and identification a complica-
ted task. In EGF, the disulfide pairing pattern is 1–3, 2–4, 5–6. From our experience in the synthesis of many 
analogs of EGF and EGF-like domains, we found that the major misfolded isomer contains the disulfide pairing of 1–2, 3–4, 5–6. This pairing pattern was found in the synthetic VK-50. In the synthesis of an EGF-like domain of factor IX, we obtained the main product which contained the disulfide pairing pattern similar to 
VK-50, i.e. 1–2, 3–4, 5–6 (29). Thus, in the 15 possible disulfide isomers, the two most thermodynamically 
stable isomers, the two most thermodynamically stable disulfide isomers obtained from oxidation of the reduced form are the EGF- and VK-50-like isomers. With this knowledge, a disulfide-forming strategy was formulated to obtain either the correct or misfolded isomer. We therefore referred to this as ‘KB’ strategy for disulfide for-
mation. It is shown in Fig. 5.

In this strategy a two-stage disulfide formation using two sets of protecting groups was used. Trityl was used 
for four cysteines in one group and acetamido for two cysteines in another group. The trityl groups were first 
removed, and the number of possible disulfide isomers were reduced from 15 to 3 during this stage of disulfide 
formation. These three isomers were further reduced to one by strategically placing the acetamido groups to 
block the formation of the unwanted isomers during the first disulfide formation mediated by DMSO and by 
making use the knowledge of the most thermodynamically favorable isomer. In this way, the possibility of 
forming the misfolded isomer is eliminated.

EGF is rich in \(\beta\)-strands, and the disulfide formation between Cys 1 and 2 or 3 is an interstrand disulfide 
formation that appears to be equally favorable. By blocking Cys 2, the only thermodynamic isomer would 
be the formation of Cys 1–3 and 5–6. In the present case, the acetamido group was placed at Cys 2 and Cys 4, 
thus preventing the disulfide formation between Cys 1–2 and allowing the disulfide formation between Cys 
1–3. Using this strategy, selective disulfide formation was achieved. Similar to VK-50, \([\text{Cys-9,14}]\text{VK-50}\) was 
treated with 20% DMSO for the formation of two disulfides, 1–3 and 5–6. The reaction as monitored by 
HPLC (Fig. 3) was completed in 24 h. The remaining Acm group was removed, and the disulfide at position 
2–4 was formed by treatment with I\(_2\) to give the three-disulfide product.

Characterization

Both peptides were purified to homogeneity by \(\text{C}_{18}\) reversed-phase preparative HPLC. The peptides gave
good agreement with amino acid analysis and the expected composition. The molecular masses of VK-50 and \([\text{Cys-9,14}]\text{VK-50}\) were obtained as determined by \(^{232}\text{Cl}\) fission ionization mass spectrometry (26, 27). The correct mass values show that these peptides are monomeric and all the thiols are folded as disulfides. To verify the disulfide pairing, the peptides were treated with thermolysin to obtain fragments that were purified by C18 reversed-phase HPLC and identified. Fragments from \([\text{Cys-9,14}]\text{VK-50}\) contained the disulfide pair Cys5-Cys6, which was the crucial disulfide linkage in the three-disulfide structure. A fragment corresponding to Cys2-Cys4 was obtained for VK-50, and thus confirmed the structure contained a disulfide pairing pattern of 3,4 and 5,6 (Table 1).

**DISCUSSION**

One of the intended goals in peptide synthesis is to confirm structures. Even with advances in the efficiency of recombinant DNA technologies and the great accuracy of protein sequencing, it is still necessary to verify the disulfide structure in peptides with multiple disulfides. In the EGF-like domain of MSP-1, the disulfide pairing is in doubt because of the sequence of the K1 strain, which contains four Cys instead of six and which forms a structure with disulfide pairing different than that of the EGFs. We synthesized this K1 EGF-like domain as well as an analog with six disulfide bonds similar in disulfide pairing to EGFs to determine the correctness of the disulfide structure. These disulfide structures would be distinguished by a panel of conformation-dependent monoclonal antibodies specific for the K1 antigen. Because the disulfide arrangements of VK-50 and \([\text{Cys-9,14}]\text{VK-50}\) differ, it is expected that their conformation or shape would be different and these mAbs would be able to distinguish between them.

The synthesis of multiple disulfide peptides is still a challenge. The selective formation of three disulfides in \([\text{Cys-9,14}]\text{VK-50}\) could have been accomplished by three different sets of thiol protecting groups with careful planning and the investment of a great length of time. However, we sought a more efficient approach to providing a good yield that eliminated several sequential deprotection and purification steps attendant to selective disulfide formation when three orthogonal protecting groups are used. Our approach, the KB strategy for disulfide formation, uses two stages and two orthogonal disulfide protecting groups, and is guided by knowledge of the misfolded disulfide isomer. In this way the disulfides are oxidized in two stages to reduce the 15 possible disulfide isomers to one. First, four cysteines of the crude peptide, after being deblocked in the TFA cleavage step, are oxidized by DMSO to give three possible isomers. The DMSO method is more efficient than the conventional approach, and the reaction is completed within 24 h. The air or mixed glu-
tathione method in the conventional approach would have required a longer time, based on our experience working with many similar EGF-like analogs. However, with the knowledge of the thermodynamically stable isomer, the formation of the two misfolded isomers is blocked by the Acm protecting groups. In the next step, treatment with I2 removes the Acm group and concurrently forms the third disulfide and only one isomer. Enzymatic digestion confirms that the sole product formed in this efficient scheme contains the EGF-like disulfide structure. Our success in obtaining proteins with three or more disulfides (Fig. 6). A criterion in applying this strategy is that the peptides and proteins are β-strand rich, because interstrand disulfide formation is more favorable than intrastrand, particularly when the spacing between the two Cys is greater than three amino acid residues, \( C-(X)_n-C, n < 3 \). The application of the strategy in the Fmoc chemistry will be as follows (Fig. 6):  
1. find three strands that form interlocking disulfide bonds e.g. strands 1, 2 and 3 in Fig. 6;  
2. find the common strand that forms disulfide bond with both strand 1 and 3, e.g. S2 with two cysteines;  
3. block one of the two Cys on S2 with one of the more stable orthogonal protecting group (e.g. Acm);  
4. find and Acm-protect the second Cys on S3 that forms the disulfide with Cys(Acm) on S2;  
5. block all others with Trt (or other suitable protecting groups);  
6. alternatively, when interlocking disulfide structure is absent, block the middle disulfide with Acm so that the formation of the first disulfide is as far as possible from the third disulfide.  

Finally, the synthesis of these multiple disulfide peptides provides a useful contribution to the Fmoc-chemistry because these peptides are usually prepared by Boc-chemistry (33). Many improvements, particularly the protecting group and coupling strategies, have been introduced by other laboratories (19, 20). In our synthesis, we introduced a thiol-rich reagent mixture, Reagent T, for the final cleavage reaction. Reagent T was found to be suitable for cleaving these cysteine-rich complex peptides because of the power and abundance of its thiol scavengers. These thiol scavengers are needed to quench long-lived carbocation and thus minimize side reactions to the sulffhydrils. We anticipate that Reagent T would be equally useful in other cleavage reactions without cysteines.

ACKNOWLEDGEMENT

This work was supported in part by grants from USPHS CA 36544 and AI 28701.
J.C. Spetzler et al.


Address:

James P. Tam
Department of Microbiology and Immunology
A5321 MCN
Vanderbilt University
Nashville, TN 3723
USA
Tel: 615-343-1465
Fax: 615-343-1467