Structure of the C-terminal Domain from *Trypanosoma brucei*
Variant Surface Glycoprotein MITat1.2*

Received for publication, September 20, 2004, and in revised form, November 22, 2004 Published, JBC Papers in Press, November 22, 2004, DOI 10.1074/jbc.M410787200

Anasuya Chattopadhyay‡§, Nicola G. Jones‡, Daniel Nietlispach‡, Peter R. Nielsen‡, H. Paul Voorheis‡, Helen R. Mott‡, and Mark Carrington‡**

From the ‡Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom and the §Department of Biochemistry, Trinity College, Dublin 2, Ireland

The variant surface glycoprotein (VSG) of African trypanosomes has a structural role in protecting other cell surface proteins from effector molecules of the mammalian immune system and also undergoes antigenic variation necessary for a persistent infection in a host. Here we have reported the solution structure of a VSG type 2 C-terminal domain from MITat1.2, completing the first structure of both domains of a VSG. The isolated C-terminal domain is a monomer in solution and forms a novel fold, which commences with a short α-helix followed by a single turn of 3₁₀-helix and connected by a short loop to a small anti-parallel β-sheet and then a longer α-helix at the C terminus. This compact domain is flanked by two unstructured regions. The structured part of the domain contains 42 residues, and the core comprises 2 disulfide bonds and 2 hydrophobic residues. These cysteines and hydrophobic residues are conserved in other VSGs, and we have modeled the structures of two further VSG C-terminal domains using the structure of MITat1.2. The models suggest that the overall structure of the core is conserved in the different VSGs but that the C-terminal α-helix is of variable length and depends on the presence of charged residues. The results provided evidence for a conserved tertiary structure for all the type 2 VSG C-terminal domains, indicated that VSG dimers form through interactions between N-terminal domains, and showed that the selection pressure for sequence variation within a conserved tertiary structure acts on the whole of the VSG molecule.

African trypanosomes have evolved a complex system of antigenic variation that facilitates the long term survival of a population in a mammalian host (see Ref. 1 for a recent review). The entire cell surface is covered with a densely packed monolayer of a single polypeptide, the variant surface glycoprotein (VSG),¹ which protects other cell surface components from effectors of the host immune system. The VSG is also the mediator of antigenic variation. At any time, only a single VSG gene is expressed, and antigenic variation results from a low frequency, stochastic event that causes a switch to the expression of a different VSG gene. During the course of an infection, there are successive clonal expansions of trypanosomes, each expressing an antigenically novel VSG.

In *Trypanosoma brucei* and the closely related species *T. evansi* and *T. equiperdum*, each VSG monomer has a molecular mass of 45–55 kDa and is made up of two or three domains (2, 3). The VSG N-terminal domain has 350–400 residues, which are followed by one or two C-terminal domains of 30–70 residues each. Less is known about the conformational changes in the VSG N-terminal domain (4). VSGs are dimeric (5, 6) and are attached to the plasma membrane via a covalent linkage from the C-terminal carboxyl group to a glycosylphosphatidylinositol anchor (7).

*VSGs* that are successively expressed by an infecting population are antigenically distinct as a result of extreme sequence variation (3, 8). The structures of the N-terminal domains from two different VSGs (MITat1.2 and ILTat1.24) have been solved and are remarkably similar despite little sequence identity (9, 10). It is believed that the conservation of structure is necessary for the protective function of the VSG, although this has not been proven. The N-terminal domain of VSG MITat1.2 (9) is elongated and is composed of seven α-helices, which comprise about 47% of the structure. Two major α-helices, each of ~70 Å, form a coiled coil that acts as a scaffold for the remaining smaller helices and loops. The structure of VSG ILTat1.24 (10) is homologous to that of MITat1.2. Although the structural identity is 60%, the sequences are only 16% identical over the same residues. This striking conservation of structure is likely to be present in other VSGs (4, 10–12) as well as some other cell surface proteins (13, 14).

The function of the VSG C-terminal domain remains obscure; it is possible that it elongates the VSG, resulting in a thicker surface coat, but this function is clearly not essential as neither *T. congolense* nor *T. vivax* VSGs contains an obvious equivalent of this domain. There is no obvious similarity between the VSG C-terminal domain sequence and any other protein, with the possible exception of an extracellular domain present in hexose transporters from trypanosomes and related

¹ The abbreviations used are: VSG, variant surface glycoprotein; TOCSY, total correlation spectroscopy; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; HSQC, heteronuclear single quantum correlation; HPLC, high performance liquid chromatography; MOPS, 4-morpholinepropanesulfonic acid.

© 2005 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in U.S.A.
species (13). The C-terminal domains from T. brucei-like VSGs are related to one another but can be split into four types based on sequence homology (3). In the structure of VSG MITat1.2, the N-terminal domain ends with an α-helix (helix S) which is followed by a single type 2 C-terminal domain in the native VSG. Type 2 C-terminal domains are characterized by 4 conserved cysteine residues. A similar arrangement of cysteines is present in type 1 C-terminal domains, where they are disulfide-bonded in a 1 to 3 and 2 to 4 pattern (15). When different type 2 C-terminal domains are compared, there is little sequence conservation other than the cysteines, a conserved hydrophobic (often aromatic) residue after the third cysteine and an N-linked glycosylation site very close to the C terminus (Fig. 1).

Progress in understanding the mechanism by which the VSG monolayer protects the trypanosome cell surface is largely dependent upon improving our knowledge of the complete structure of the VSG. This goal has been hindered by difficulties in obtaining high quality crystals of a complete VSG (16, 17). Here, we have reported the solution structure of the type 2 C-terminal domain from VSG MITat1.2 and thus the completion of the first structure of both domains of a VSG. The C-terminal domain has a novel structure and forms a compact fold of 42 residues, with a core comprised of 2 disulfide bonds and 2 conserved hydrophobic residues. Other divergent type 2 C-terminal domains have been modeled using the structure, indicating that the conservation of tertiary structure against a background of sequence variation is likely to be present in the VSG C-terminal domain as well as in the VSG N-terminal domain. The C-terminal domain structure has been used to produce a model of the VSG on the trypanosome cell surface that is consistent with a role for the C-terminal domain in the diffusion barrier function of the VSG.

**Fig. 1. Sequence alignment of VSG type 2 C-terminal domains.** The numbers along the top refer to the residue numbers of M1.2C. The secondary structure of M1.2C is shown along the top: α-helices are represented by yellow cylinders; the 3_10-helix by a gray cylinder, and β-strands by yellow arrows. The conserved cysteines are boxed and highlighted in yellow; conserved hydrophobic residues are highlighted in green; conserved charged residues are blue (basic) or red (acidic). The C-terminal Asn that is glycosylated is boxed. For each VSG the label indicates (i) the species (Tbr, T. brucei; Tev, T. evansi; Teq, T. equiperdum) and the N- and C-terminal domain types (3); (ii) the EMBL accession number, and (iii) the VSG name, if available. The figure was produced using Alscript (37).
**EXPERIMENTAL PROCEDURES**

**Protein Expression and Purification**—A construct designed to express residues 359–433 of mature VSG MITat1.2 (EMBL accession number X56762) was cloned into the XhoI and BamHI sites of pET15b. The polypeptide was expressed either in *Escherichia coli* AD494(DE3) or in *E. coli* BL21(DE3) in a minimal medium based on MOPS buffer, containing 5% [15N]-labeled Celton and [13C]-labeled ammonium chloride or 5% [13C,15N]-labeled ammonium chloride and [13C]glucose (Spectra Stable Isotopes). The C-terminal domain (M1.2C) was purified by affinity chromatography using a Ni²⁺ column followed by digestion with thrombin to remove the N-terminal His tag. After thrombin cleavage, 5 residues (GSHML) from the vector remained at the N terminus. M1.2C was further purified by gel filtration using a Superdex 30 column (Amersham Biosciences), followed by reverse-phase HPLC using a 25 × 1-cm Hichrom-5 C8 column and a 0–20% acetonitrile gradient in 0.1% trifluoroacetic acid. The presence of free sulfhydryl groups in M1.2C was measured using the Ellman assay (18) by reaction with 5,5'-dithio-bis(2-nitrobenzoate). The relative molecular mass of M1.2C was determined by electrospray ion trap mass spectrometry.

**Analytical Ultracentrifugation**—Equilibrium sedimentation analysis was performed in a Beckman XL-1 ultracentrifuge. Three two-sector cells were filled with 100 μl of 29, 30, or 37 μl M1.2C in 50 mM sodium phosphate, pH 6.0, and 150 mM NaCl; 110 μl of the same buffer was used on the reference side of each cell. The cells were placed in an An50-Ti rotor. For each sample, the equilibrium concentration gradients were measured for angular velocities of 25000, 33000, and 40000 rpm. Concentrations were measured as the absorbance at 280 nm (and using the Raleigh interference optics). The measured data were analyzed by global fitting using the program WinNonlin (19). The nine data sets were fitted simultaneously, with the molecular mass as a global parameter and with two local parameters for each (the concentration at the reference position and the baseline offset). The solvent density used was 1.080 g/ml at 5 °C, and the protein partial specific volume at 5 °C was 0.6998 ml/g. Both values were calculated using Sednterp (T. Laue, University of New Hampshire). An extinction coefficient for the protein of 9098 M⁻¹ cm⁻¹ was obtained using amino acid analysis.

**NMR Spectroscopy**—NMR experiments were performed using [15N]- and [13C,15N]-labeled samples of 0.7–1.0 mM M1.2C in 50 mM sodium phosphate, pH 6.5, 150 mM sodium chloride, 10% D₂O. Spectra were recorded at 25 °C on Bruker DRX 500 and DRX 600 spectrometers. The experiments recorded were 15N HSQC, 13C HSQC, 13C HSQC, 15N-separated NOESY (150 ms mixing time), [13C,15N]-separated TOCSY (67.9 ms DIPSI mixing), HNCA, HNCOCA, HNCA, HCCCH, TOCSY, HBBH, HBBH, HBBH, HBBH, TOCSY, 13C HSQC, and 13C-separate NOESY (150 ms mixing time). TALOS (20) was used to estimate backbone torsion angles based on the CA-CO-CB-N, and HA chemical shifts. CLEANEX (21) experiments were recorded to analyze solvent exchange rates and deduce hydrogen bonding information. Backbone dynamics were assessed using a heteronuclear NOE experiment (22). Spectra were processed using the program AIZARA (W. Boucher, University of Cambridge) and analyzed with ANSIG (23).

**Structure Calculations**—Structure calculations were performed using ARIA (24) interfaced to CNS 1.0 (25). The structures were calculated with 90 ps of high temperature dynamics and 30 ps of cooling from 1,000 to 500 K in torsion angle space. This was followed by 24 ps of high temperature dynamics and 30 ps of cooling from 1,000 to 50 K using dynamics in Cartesian space. Figs. 3, 5, and 6 were produced using Molscript (26) and Raster3D (27).

**Homology Modeling**—The structures of the C-terminal domains of MITat1.1 (EMBL accession number X56761) and ILTat1.21 (EMBL accession number X56766) were modeled using residues Ala-375 to Glu-417 of the closest structure to the mean of MITat1.2. The sequence alignment shown in Fig. 1 was used as the input for MODELLER (28). The model with the lowest MODELLER objective function was used for subsequent analysis after the stereochemistry was checked using PROCHECK (29).

**Docking**—The whole VSG structure was docked using the program 3D-Dock (30). The N-terminal domain dimer from MITat1.2 (Protein Data Bank code 1YVS) was docked with the core structure of the C-terminal domain (residues 375–416) on the basis of surface complementarity, and the resulting structures were then ranked on the basis of an empirical pair potential matrix (30).

**RESULTS**

**Purification of the MITat1.2 C-terminal Domain**—A construct was designed to express a polypeptide containing the MITat1.2 C-terminal domain plus a short overlap with the N-terminal domain. The polypeptide (M1.2C) started at Glu-359, which is within helix S of the N-terminal domain (9), and ended at the mature C terminus, Ser-433. M1.2C was expressed in *E. coli* AD494(DE3) and *E. coli* BL21(TrxB/DE3); both strains carry a mutation in the *trxB* gene that allows the formation of protein disulfide bonds in the *E. coli* cytoplasm. M1.2C prepared from these strains contained no free sulfhydryl groups as determined by reaction with 5,5'-dithio-bis(2-nitrobenzoic acid). After thrombin cleavage, a relative molecular mass of 8730.61 ± 0.35 was determined by electrospray ion trap mass spectrometry, in agreement with the expected mass of 8730.4 for fully disulfide-bonded M1.2C.

Two-dimensional [1H] TOCSY and NOESY experiments were recorded on unlabeled material that had been expressed in *E. coli* AD494(DE3). Some of the resonances in these spectra were accompanied by a second, weaker peak of about one-tenth the intensity. We concluded that these shadow cross-peaks were due to an alternative disulfide-bonded form of M1.2C because: (i) the electrospray ion trap mass spectrometry data indicated a homogeneous polypeptide of the expected mass, (ii) no free sulfhydryl groups were detected, and (iii) M1.2C appeared homogeneous by reducing SDS-PAGE but migrated as two close but distinct bands on non-reducing SDS-PAGE at a similar relative intensity to the two sets of resonances observed in the NMR spectra (data not shown). In subsequent experiments, the ratio of the two forms varied with the *E. coli* host strain, growth medium, and temperature of expression. The isotopically labeled samples were produced in *E. coli* BL21(TrxB/DE3), and it was necessary to purify the major form using reverse-phase HPLC. After HPLC purification, the intensity ratio of the resonances in a [15N] HSQC spectrum was 11:1.

**FIG. 2.** The equilibrium concentration gradients for M1.2C. The curves represent the values calculated based on the values obtained fitting to a single ideal species. The upper panel shows the residuals when subtracting the data from the fitted curve. The even distribution of these around zero indicates that the model describes the data well.
**M1.2C Is a Monomer in Solution**—VSGs are dimeric, and the N-terminal domain contains a large dimerization interface that buries ~8,000 Å² (9). Equilibrium sedimentation analysis was used to determine whether the C-terminal domain is also a dimer (Fig. 2). The best fit of the model to the data gave a molecular mass of 8795 g/mol. When compared with the mass calculated from the amino acid composition of 8730.4 g/mol, it can be seen that M1.2C is a monomer in solution under the conditions used. The highest concentration used was 37 μM, so the $K_d$ for any dimer formation would be >0.5 mM.

**Three-dimensional Structure of M1.2C—Backbone resonances were assigned from HCNO, HN(CO)CA, HNCACB, and CBCA(CO)NH experiments using standard triple resonance techniques. Side chain assignments were completed using the $^{13}$N TOCSY, HCCH-TOCSY, HBHA(CBCACO)NH, and (H)C-C(CO)NH-TOCSY experiments. Assignment of NOEs was based on the $^{15}$N-separated NOESY (502 unambiguous, 100 ambiguous NOEs) and 13C-separated NOESY (509 unambiguous, 65 ambiguous NOEs). These NOEs were translated by ARIA into 719 unambiguous and 121 ambiguous, non-degenerate distance restraints. 24 pairs of dihedral restraints were derived from the backbone chemical shifts using TALOS-based predictions. 20 slowly exchanging amides were identified in the CLEANEX experiment, which were translated into hydrogen bond restraints within the secondary structure by manual inspection of structures generated by early rounds of structure calculations. The hydrogen bonds were treated as unambiguous NOEs, with distance restraints of 2.3 Å (O to H) and 3.3 Å (O to N). At the end of 8 iterations, there were 772 unambiguous and 97 ambiguous NOE-derived distance restraints. In the final iteration, 100 structures were calculated and the 50 lowest energy structures were used for further analysis.

The NMR data defined a structured core with unstructured residues at both the N and C termini. The structured part of M1.2C contained 42 residues from Ala-375 to Asp-416 (Fig. 3 and Table I). Each residue in this core region has a positive value for the heteronuclear NOE (Fig. 4). Outside the structured core, no long range distance restraints were observed for residues from Glu-359 to Ala-374 or from Glu-417 to Ser-433. The heteronuclear NOE values for these residues were less than 0.1 (Fig. 4), indicating that they are disordered in solution.

The initial structure calculations were performed with no disulfide restraints to attempt to determine the disulfide-bonding pattern by determining the distances between the S atoms in the cysteine residues (data not shown). This approach was inconclusive as all 4 cysteine residues were clustered together in the core (Figs. 3 and 5). The disulfide bonding pattern was subsequently elucidated by the identification of a strong NOE between the $\beta_C$ and $\beta_S$ interatomic distance of 3.497 Å in the structure closest to the mean, a value close to the canonical distance for two cysteines linked by a disulfide bond (31). No NOEs were observed between Cys-389 and Cys-404; their absence may be due to conformational exchange that leads to line broadening, as the backbone amide resonance of Cys-404 could not be observed in any NMR spectra. Interestingly, however, Cys-389 NH is observed and its heteronuclear NOE is typical of a structured residue. As there are no free thiols in the monomeric M1.2C and the relative molecular mass corresponds to the expected

---

**FIG. 3. The structure of M1.2C.** a, the backbone trace between Glu-376 and Asp-416 of the ensemble of 50 lowest energy structures of M1.2C. b, ribbon representation of the structure closest to the mean over the same residue range. The disulfide bonds are shown in yellow, and the residue numbers of the Cys residues involved are indicated.

| Table I | Experimental restraints and structural statistics |
|---------|--------------------------------------------------|
| Number of experimental restraints | | |
| Unambiguous | 772 |
| Ambiguous | 97 |
| Dihedral restraints | 48 |
| Hydrogen bonds | 40 |
| Root mean square deviations of backbone atoms (375–416) (Å) | 0.65 ± 0.17 |
| Root mean square deviations of heavy atoms (375–416) (Å) | 1.19 ± 0.17 |
| NOE distances (Å) | 0.033 ± 0.0033 |
| Dihedral angles (°) | 0.35 ± 0.29 |
| From idealized geometry | |
| Bonds (Å) | 0.0040 ± 0.00015 |
| Angles (°) | 0.536 ± 0.020 |
| Improps (°) | 1.56 ± 0.122 |
| Final energy | $-495.99 ± 8.85$ |
| Ramachandran analysis (375–416) | |
| Residues in most favoured regions | 79.3% |
| Residues in additionally allowed regions | 15.5% |
| Residues in generously allowed regions | 2.3% |
| Residues in disallowed regions | 2.8% |

$^a$ (SA) represents the average root mean square deviations for the ensemble.

$^b$ (SA), represents values for the structure that is closest to the mean.

$^c$ The Lennard-Jones potential was not used at any stage in the refinement.
value for the fully disulfide-bonded form, Cys-389 and Cys-404 must also be forming a disulfide bond. Thus, M1.2C has a disulfide bonding pattern of 1 to 3 and 2 to 4, the same as that present in VSG type 1 C-terminal domains (15).

The secondary structure of M1.2C (Fig. 2) begins with a short α-helix between residues Glu-378 and Lys-383. This is followed by a single turn of $3_{10}$-helix between residues Glu-386 and Glu-388, which is connected by a short loop to the first strand of the β-sheet from Cys-393 to His-396. The two short strands of the β-sheet result in a typical network of interstrand NOEs and hydrogen bonds. The second strand of the β-sheet runs from Lys-403 to Leu-406 and leads to a longer α-helix stretching from Lys-408 to Asp-416. As well as the usual intrahelix NOE pattern, this terminal α-helix has a limited number of long range NOEs to the rest of the core region, in agreement with the lower heteronuclear NOE values for this region (Fig. 4).

The loop, Asn-397 to Lys-402, connecting the two β-strands is not well defined (at the top of the structure in Fig. 3) with the heteronuclear NOE values ranging between 0.3 and 0.6, lower than for most of the structured core (Fig. 4). The line widths of the resonances corresponding to residues within this loop were also larger than the average line width in the core, implying that there is motion on the millisecond time scale within this region.

The heteronuclear NOE values for the C-terminal α-helix (Lys-408 to Asp-416) are positive but decrease toward the terminus, suggesting that there is an increase in the amplitude of helix motion. The majority of residues in the α-helix are charged (Fig. 1), and it appears to be stabilized by a number of salt bridges (Fig. 5) both within the helix and between the helix and the rest of the domain. In the middle of the helix, Lys-412 makes a salt bridge with Asp-416 and Lys-413 is sandwiched between Glu-409 and Glu-417. The spacing between these residues conforms to the i, i + 4 pattern that has been observed to stabilize isolated α-helices (32). Two sets of ionic interactions between the C-terminal α-helix and other residues in the structured core pin the N terminus of the helix to the rest of the core: Lys-394 in the first β-strand interacts with Asp-407 and Glu-410, whereas Glu-378 in the first α-helix interacts with Lys-408 (Fig. 5).

The 4 cysteines in M1.2C lie buried in the core of the structure (Fig. 5). The remainder of the core is composed of 2 hydrophobic residues that flank the disulfide bonds, Trp-395 and Leu-406. These residues, as well as the cysteines, are conserved in many VSG type 2 C-terminal domains (Fig. 1); in other type 2 C-terminal domains hydrophobic residues are usually found at an equivalent location in the sequence. Most of the remainder of the domain is made up of small polar and charged residues, with the exception of Pro-392, which also makes hydrophobic contacts with Ala-411 and Val-414 in the C-terminal helix. None of these 3 residues is conserved in type 2 C-terminal domains.

The importance of the cysteines for the structural integrity of M1.2C was tested by making mutants in which the cysteines were changed to alamines. Three mutants were made: (i) C381A and C393A, (ii) C389A and C404A, and (iii) all four cysteines mutated to alanine. The proteins were expressed, purified, and preliminary 1H NMR spectra collected. The spectra of all three mutants showed considerably less dispersion of resonances (data not shown), indicating that none of them retains the structure of wild-type M1.2C.

Comparison with Other Structures—The structure of the M1.2C does not bear direct homology to any known structure in the Protein Data Bank, according to a search performed using the Dali server (33). Comparison with other small, disulfide-rich structures suggests that the C-terminal domain is similar to Knottins and Trefoil fold proteins. The Knottins are a diverse group of proteins with a common structural motif comprising a β-hairpin with a cysteine on each side that forms disulfide bonds with cysteines outside the hairpin (34). In Knottins, however, the cysteine residues in the hairpin are usually directly opposite each other, whereas in M1.2C the equivalent cysteines (Cys-393 and Cys-404) are staggered (Figs. 3 and 5). A similar staggered arrangement only occurs in the Trefoil fold (for example Protein Data Bank 1e9t and 1ps2). Although the Trefoil fold contains three disulfide bonds, it appears to be the nearest, but remote, structural neighbor of M1.2C.

Modeling of the Structures of Other VSG Type 2 C-terminal Domains—The structure of M1.2C was used to model the VSG type 2 C-terminal domains from VSGs MITat1.1 and ILTat1.21.
The VSG C-terminal domain is comprised of a compact core of 42 residues, Ala-375 to Asp-416, flanked by two unstructured terminal domains that are increasingly divergent from the sequence of MITat1.2 (Fig. 1). The C-terminal domain of MITat1.1 has a higher degree of identity to M1.2C than ILTat1.21, which is among the most divergent from M1.2C. The divergence is most apparent in the region of ILTat1.21 corresponding to the second &alpha;-helix in M1.2C; only 2 of the 9 residues are charged in ILTat1.21 compared with the 6 in M1.2C. It is not clear whether this region of ILTat1.21 will form an &alpha;-helix as is found in M1.2C. Both of the domains modeled have the 2 hydrophobic residues that stabilize each side of the core, performing a function equivalent to Trp-395 and Leu-406 in MITat1.2 (Figs. 1 and 5), Phe-407 and Leu-416 in MITat1.1, Tyr-397 and Pro-406 in ILTat1.21. The models were of good stereochemical quality, as determined by PROCHECK (data not shown).

In the model of MITat1.1, the C-terminal helix is likely to be stabilized in a similar way to that of M1.2C (Fig. 6). In several places along the helix, the charges are reversed with respect to those in M1.2C, and in these cases the charge of the interacting residue is also reversed. For example, in MITat1.2, Lys-412 and Asp-416 are close together on the surface of the helix, whereas in MITat1.1 these residues are replaced by Glu-422 and Lys-426, respectively. The model predicts that the C-terminal helix in MITat1.1 will be stabilized by interactions between Lys-423 and Asp-419/Glu-427, between Lys-417 and Asp-416, and between Lys-426, respectively. The model predicts that the C-terminal helix in MITat1.1 will be stabilized by interactions between Lys-423 and Asp-419/Glu-427, between Lys-417 and Asp-419, and between Glu-384 and Lys-418 (Fig. 6b). There are also two hydrophobic side chains within this helix in MITat1.1, Val-420 and Leu-424, which are on the same side of the helix. These residues are stabilized by a hydrophobic interaction with Leu-400 (Fig. 6b) within the loop before the first &beta;-strand, which is longer in MITat1.1 than MITat1.2, so that Leu-400 corresponds to a gap in the alignment (Fig. 1). These considerations suggest that MITat1.1 can form the same structured core as MITat1.2. The C-terminal helix can be stabilized by ion pairs in a similar way, with extra stabilization coming from hydrophobic interactions.

In contrast, in ILTat1.21 the C-terminal helix has almost no charged residues, although those that are present can make stabilizing interactions (Fig. 6c). Thus, Lys-408 and Glu-411 could interact, as could Glu-396 and Lys-407. The remainder of this helix is made up of small, uncharged residues. The charged residues that are present are at the N terminus of the helix, so it is likely that ILTat1.21 will form a shorter helix here, of a single turn. The interaction between Lys-408 and Glu-411 would favor 3_10-helix formation as this would bring these side chains into close proximity.

Model of a Complete VSG Polypeptide—The program 3D-Dock (30) was used to produce a model of the complete VSG polypeptide (Fig. 7). This model was selected by filtering the output for (i) the surface complementarity and an empirical likelihood that certain residues interact, (ii) the position of the C-terminal domain at the membrane-proximal end of the N-terminal domain, (iii) a structure that was consistent with the envelope determined for the C-terminal domain of VSG ILTat1.24 (17), and (iv) a structure that had the C terminus pointing down toward the location of the plasma membrane and the glycosylphosphatidylinositol anchor. The model shown was the highest ranked that met these criteria.

**FIG. 6. Models of the C-terminal domains of MITat1.1 and ILTat1.21.** a, the structure of MITat1.2. b, model of MITat1.1. c, model of ILTat1.21. The disulfide bonds are colored yellow. The residues that may stabilize the C-terminal helix are red (acidic), blue (basic), and pink (hydrophobic). The helix in ILTat1.21 is unlikely to continue after Glu-411, so although residues C-terminal to this were modeled as helix, they are colored white.

**DISCUSSION**

The VSG monolayer that covers the cell surface of trypanosomes in a mammal is the molecular interface between host and parasite and has evolved to facilitate the long term persistence of an infection. VSGs are subject to two evolutionary pressures, sequence variation for antigenic variation and conservation of tertiary structure, which is assumed necessary for the protective function of the VSG. An understanding of the molecular mechanism by which the VSG is able to protect other cell surface components from the host immune system is dependent on knowledge of the VSG structure.

Here, we have presented the first structure of a VSG C-terminal domain. The structure is novel and has at its core a pair of disulfide bonds. In the absence of either pair of cysteines, the domain does not fold correctly, and thus both disulfide bonds are required to form the structure. The requirement for both disulfide bonds is consistent with the absolute conservation of 4 cysteines in all other VSG type 2 C-terminal domains. The isolated domain is a monomer in solution, so the native VSG dimerizes through interactions between the N-terminal domains of the monomers. However, it is not possible to rule out interactions between the C-terminal domains at the exceptionally high concentrations of VSG present on the trypanosome surface.

The VSG C-terminal domain is comprised of a compact core of 42 residues, Ala-375 to Asp-416, flanked by two unstructured
The second finding in this study was the demonstration that the structure of M1.2C can be used to produce models of other VSG type 2 C-terminal domains. The different type 2 C-terminal domains have the 4 conserved cysteines and 2 conserved hydrophobic residues that make up the core of the domain; otherwise they have little conservation of sequence except around the N-linked glycosylation site close to the C terminus (Fig. 1). The conservation of tertiary structure mirrors that previously found in the N-terminal domain and is presumably necessary for the function of the VSG C-terminal domain.

It has been proposed that the function of the VSG C-terminal domain is to increase the ability of the VSG monolayer to protect underlying proteins from host immunoglobulins, possibly by simply increasing the thickness of the monolayer (36). However, the C-terminal domain is a compact structure and the N and C termini of the domain are relatively close to each other. Thus, the domain is not particularly extended as might be expected of a domain with such a structural role in increasing the thickness of the VSG monolayer. The model of the VSG (Fig. 7) suggests that the VSG is not particularly extended by the inclusion of the C-terminal domain but does suggest that the C-terminal domain contributes to a densely packed polypeptide layer proximal to the plasma membrane.

The VSG coat has evolved to ensure the long term survival of an infecting population by functioning both in antigenic variation and in the formation of a protective macromolecular diffusion barrier. The results presented herein suggest that the C-terminal domain contributes to the diffusion barrier by creating a region densely packed with polypeptide adjacent to the plasma membrane.

Acknowledgments—We thank Alexey Murzin and Dave Barry for helpful discussions. Mass spectrometry and initial disulfide analysis was carried out by the Protein and Nucleic Acid Chemistry group, Department of Biochemistry, University of Cambridge. This was a contribution from the Cambridge Center for Molecular Recognition, supported by the Biotechnology and Biological Sciences Research Council and the Wellcome Trust.

REFERENCES
1. McCulloch, R. (2004) Trends Parasitol. 20, 117–121
2. Johnson, J. G., and Cross, G. A. M. (1979) Biochem. J. 178, 689–697
3. Carrington, M., Miller, N., Blum, M., Roditi, I., Wiley, D., and Turner, M. (1991) J. Mol. Biol. 221, 823–835
4. Bussler, H., Linder, M., Linder, D., and Reinwald, E. (1998) J. Biol. Chem. 273, 32582–32586
5. Auffret, C. A., and Turner, M. J. (1981) Biochem. J. 193, 647–650
6. Strickler, J. E., and Patton, C. L. (1982) Exp. Parasitol. 53, 117–132
7. Ferguson, M. A., Homans, S. W., Dwek, R. A., and Rademacher, T. W. (1988) Science 239, 753–759
8. Cross, G. A. M. (1975) Parasitology 71, 393–417
9. Freyman, D., Down, J., Carrington, M., Roditi, I., Turner, M., and Wiley, D. (1990) J. Mol. Biol. 216, 141–160
10. Blum, M. L., Down, J. A., Garnett, A. M., Carrington, M., Turner, M. J., and Wiley, D. C. (1993) Nature 362, 603–609
11. Mehler, A., Bond, C. S., and Ferguson, M. A. J. (2002) Glycobiology 12, 607–612
12. Campill, N., and Carrington, M. (2003) Mol. Biochem. Parasitol. 127, 79–84
13. Carrington, M., and Boothroyd, J. (1996) Mol. Biochem. Parasitol. 81, 119–126
14. Salmon, D., Hanoq-Queretier, J., Paturiaux-Hanoq, F., Pays, A., Tehabi, P., Nolan, D. P., Michel, A., and Pays, E. (1997) EMBO J. 16, 7272–7278
15. Allen, G., and Garnett, L. P. (1988) Biochem. J. 209, 481–487
16. Down, J., Garman, S., Garnett, A., Turner, M., and Wiley, D. (1991) J. Mol. Biol. 218, 679–683
17. Garman, S. C. (1994) The X-ray Crystallographic Analysis of an Intact VSG from the Trypanosome. Ph.D. thesis, Harvard University, Cambridge, Massachusetts
18. Sediak, J., and Lindsay, R. M. (1968) Anal. Biochem. 25, 192–205
19. Johnson, M. L., Correira, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) Biophys. J. 36, 575–588
20. Corbalan, G., Delaglio, F., and Bax, A. (1999) J. Biomol. NMR 13, 289–302
21. Hwang, T. L., van Zijl, P. C. M., and Mori, S. (1996) J. Biomol. NMR 11, 221–226
22. Farrow, N. A., Zhang, O. W., Fornan-Kay, J. D., and Kay, L. E. (1995) Biochemistry 34, 866–876
23. Kraulis, P. J., Domsma, P. J., Campbell-Burk, S. L., Aken, T. V., and Laue, E. D. (1994) Biochemistry 33, 3515–3531
24. Linge, J. P., O'Donoghue, S. J., and Nilges, M. (2001) Methods Enzymol. 339, 71–90

Fig. 7. Model of the complete VSG MITat1.2 dimer. The two N-terminal domains are shown in green and pink and the two C-terminal domains in blue. The dashed white line in one of the monomers represents the 15-residue protease-sensitive linker between the two domains. There are an additional 16 residues containing an N-linked glycosylation site between the C terminus of the model and the ethanolamine of the glycosylphosphatidylinositol anchor.
VSG C-terminal Domain Structure

25. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Koszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
26. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 627–633
27. Merritt, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
28. Marti-Renom, M., Stuart, A., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) Annu. Rev. Biophys. Biomol. Struct. 29, 291–325
29. Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. (1993) J. Appl. Crystallogr. 26, 283–291
30. Moont, G., Gabb, H. A., and Sternberg, M. J. E. (1999) Proteins Struct. Funct. Genet. 35, 364–373
31. Klaus, W., Broger, C., Gerber, P., and Senn, H. (1993) J. Mol. Biol. 232, 897–906
32. Lyu, P. C., Gans, P. J., and Kallenbach, N. R. (1992) J. Mol. Biol. 223, 343–350
33. Holm, L., and Sander, C. (1998) Nucleic Acids Res. 26, 316–319
34. Gelly, J. C., Gracy, J., Kaas, Q., Le-Nguyen, D., Heitz, A., and Chiche, L. (2004) Nucleic Acids Res. 32, D156–D159
35. Zamze, S. E., Ashford, D. A., Wooten, E. W., Rademacher, T. W., and Dwek, R. A. (1991) J. Biol. Chem. 266, 20244–20261
36. Ziegelbauer, K., and Overath, P. (1993) Infect. Immun. 61, 4540–4545
37. Barton, G. J. (1993) Protein Eng. 6, 37–40