Reprinted from Nature, Vol. 311, No. 5982, pp. 167-169, 13 September 1984

6 Å-Resolution X-ray structure of a variable surface glycoprotein from *Trypanosoma brucei*

D. M. Freymann, P. Metcalf, M. Turner* & D. C. Wiley

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 012138, USA
* MRC Biochemical Parasitology Unit, The Molteno Institute, Downing Street, Cambridge CB2 3EE, UK

The variable surface glycoprotein (VSG) is the predominant component of the surface coat of the African trypanosome[1]. The expression of antigenically distinct VSGs on minor populations during infection allows the parasite to escape the host immune response[2,3]. Purification of the protein is facilitated by the enzymatic release of a soluble form of VSG (sVSG) which occurs on cell lysis[4]. The soluble form is a dimer with an approximate molecular weight of 120,000–130,000[5]. Partial proteolysis of sVSG[6] reveals a protease-sensitive link between an amino-terminal domain which comprises about two-thirds of the molecule, and a C-terminal domain which contains the membrane attachment site[7]. We have obtained crystals suitable for high-resolution structural analysis from preparations of three sVSG: MITat 1.2, ILTat 1.25 and ILTat 1.22. The crystal structure of the dimer of the MITat 1.2 amino-terminal domain has been solved to 6 Å resolution. We report here that the dimer is an unusual 90 Å rod-like molecule composed of a helical bundle of at least four 80 Å-long α-helices.

sVSG was purified from *Trypanosoma brucei* as described previously[1]. MITat 1.2 was crystallized by vapour diffusion of a 40 mg ml^{-1} solution in 50 mM Tris-HCl/0.1% sodium azide, *p*H 7.5 against 9% polyethylene glycol (PEG) 8000/10% ethylene glycol in the same buffer. Tetragonal bipyramidal crystals, which diffract to 2.9 Å, appear within a week and grow to a maximum dimension of 1 mm. The space group is $P4_12_12$, with $a=96.3$, $c=111.3$ Å; the asymmetric unit contains one monomer of the amino-terminal domain. Details of the crystallization of ILTat 1.22 and ILTat 1.25 will be published elsewhere.

Protein recovered from MITat 1.2 crystals has an apparent molecular weight (MW) of 43,000 (43 K) on SDS-polyacrylamide gel electrophoresis (PAGE), co-migrating with the amino-terminal domain identified by Johnson and Cross[6] (Fig. 1). The amino-terminal sequence of the recovered protein is Ala-Ala-Glu-Lys, consistent with the previously published N-terminal sequence[6]. By using 'immuno-blots'[8], (data not shown) we have demonstrated that the protein recovered from the crystal does not contain the C-terminal 'cross-reacting determinant'[9] of the intact VSG. Thus, it seems that adventitious proteolysis from a contaminating protease releases an N-terminal crystallizable domain. On gel permeation HPLC (data not shown) protein recovered from MITat 1.2 crystals elutes with an approximate molecular weight of 80,000, clearly distinguishable from the intact native dimer at around 120,000, suggesting that the crystalline protein is a dimer of two 43K domains.

A mercury derivative was obtained after a 1-day soak in a 12% PEG solution (otherwise identical to the mother liquor) saturated with K_2HgI_4. Data from one native crystal and one Hg derivative crystal were collected by diffractometer. Data were corrected for Lorentz polarization effects and radiation damage, and an empirical absorption correction was applied[10]. Friedel equivalents were scaled together in blocks of reciprocal space[11]. Data collection statistics are summarized in Table 1.

Harker sections of isomorphous and anomalous difference Pattersons showed identical peaks, and were interpretable in terms of a single Hg site at 0.20, 0.19, 0.50. After 10 cycles of refinement of the heavy-atom parameters, including the anomolous scattering data, 1,211 reflections (89% of the 15–6 Å shell) were phased with a mean figure of merit of 0.72 (Table 1). A difference Fourier showed no additional Hg sites.

Table 1 6 Å data collection and phase refinement for MITat 1.2 VSG

	Native		Hg	
Cell dimensions (Å)	96.3, 111.2		96.4, 111.6	
No. of reflections collected	2,781		2,812	
Friedel *R* factor		No. of pairs		No. of pairs
Centric	1.7%	381	1.4%	387
Acentric	2.2%	950	3.9%	979
No. of reflections phased	1,211			
Mean figure of merit	0.72			
Kraut *R*	7.39			
$\langle Fh\rangle/\langle E\rangle$	2.97			

Friedel *R* factor: $\text{sum}(|I^+ - I^-|)/\text{sum}(I^+ + I^-)$

Kraut *R*: $\dfrac{\text{sum}[|Fph(\text{obs}) - Fph(\text{calc})|]}{\text{sum}[Fph(\text{obs})]} \times 100\%$

Fph = derivative structure factor

$\langle Fh\rangle/\langle E\rangle$: $\dfrac{\text{r.m.s. heavy-atom structure factor}}{\text{r.m.s lack of closure error}}$

Before data collection, MITat 1.2 crystals were transferred to a solution containing 12% PEG, 10% ethylene glycol, 0.1% azide, 50 mM Tris HCl *p*H 7.5. Data were collected using a Wyckoff step scan[21] on a Syntex P3/F diffractometer equipped with a 50-cm helium-filled extension tube. About 3,000 reflections, including Friedel pairs, were collected in the 15–6 Å shell. Four reference reflections were collected every hundred reflections; the mean decay was 26% (native) and 21% (Hg) over the 3 days of data collection. The length of the *c* axis was observed to change by 1% during the data collection. Difference Patterson maps were calculated using coefficients $(Fph - Fp)^{**}2$ and $(\text{Kemp}^{**}2/4)(Fph^+ - Fph^-)^{**}2$, where Kemp[22] was 6.43. The heavy-atom position and occupancy were refined using the program PHARE3, which alternates cycles of parameter refinement and phase calculation[23]. The anomalous scattering information was included in the refinement[24], and was used to distinguish the correct space group, $P4_12_12$, from its enantiomer[25]. B was held at 25 Å$^{**}2$. The final mercury position, after 10 cycles, was $x/a=0.22$, $y/b=0.19$, $z/c=0.50$ with occupancy 0.58. The iodine atoms were not included in the refinement.

The native electron density map calculated using the 'best' phases[12] shows that the unit cell contains four dimers packed along the crystallographic 2-fold axes. At this resolution, rods of high electron density ~10 Å apart are clearly resolved, forming a bundle which extends about 80 Å. We interpret this as an α-helical 'core' of the dimeric molecule.

The dimer is elongated, with approximate dimensions of $40\times40\times90$ Å (Fig. 2*a*). When viewed down the dimer axis, the helical bundle is apparent (Fig. 2b–d). The connectivity of two α-helices in each monomer can be traced for about 80 Å. At one end (bottom of Fig. 2*a*) density connects the two helices, which form a 25 Å-long left-handed 6-fold bundle with another, unconnected, helical segment (Fig. 2*b*). The two helices from each subunit extend past this region to form a fairly straight four-α-helical bundle which extends about 30 Å (Fig. 2*c*). A third rod of high electron density begins adjacent to this structure. The strands then diverge to form two left-twisted structures, about 30 Å long, each of which clearly involves three helices, and possibly other adjacent density , at the 'top' of each subunit (Fig. 2*d*).

We estimate, by crude fitting of polyalanine α-helices, that about 175 residues are accounted for by the well defined helical regions of the low-resolution map. This is about 50% of the expected length of the polypeptide based on its migration on SDS-PAGE, and its consistent with circular dichroism results (M.J.T. and T. Bayley, unpublished data). We have not yet established the connectivity of additional density which occurs at the end of the dimer and around the helical bundle.

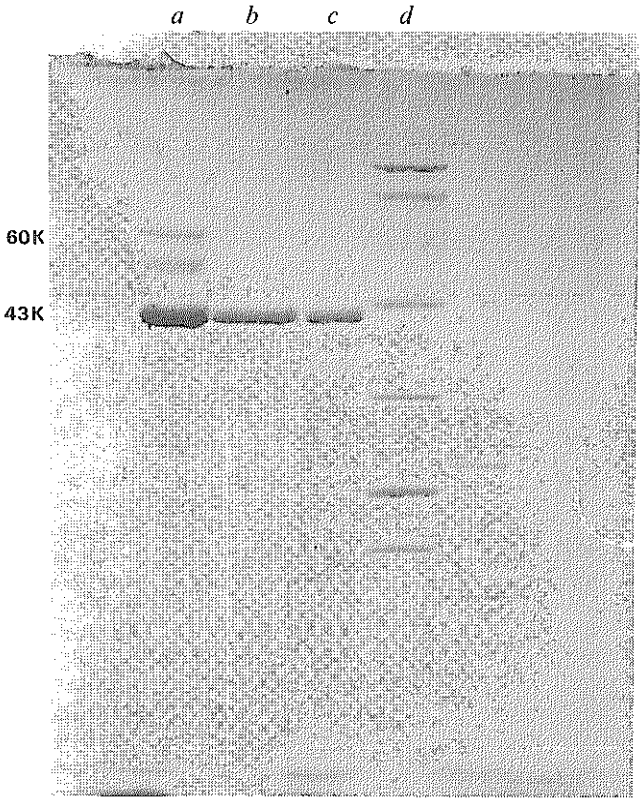

Fig. 1 Protein recovered from MITat 1.2 crystals migrates with an apparent MW of 43,000 (43K) on a 12% SDS polyacrylamide gel. *a*, 5 μg MITat 1.2 from a solution used for crystallization. The intact polypeptide migrates with an apparent MW of 60,000 (60K); the fragment at 43K corresponds to protein recovered from the crystal. *b*, Two large (0.5, 1 mm) crystals were washed in 2 ml 12% PEG, 10 mM TrisHCl, 0.1% azide, *p*H 7.5, and dissolved in a capillary with 30 μl 10 mM TrisHCl *p*H 7.5. The dissolved protein was boiled for 3 min after addition of 90 μl sample buffer; an aliquot of 20 μl was loaded on the gel. *c*, 5 μg MITat 1.2 was digested with trypsin, using the conditions of Johnson and Cross[6]; the digestion was stopped by adding excess phenylmethanesulphony fluoride and boiling after addition of sample buffer. The amino-terminal protease-resistant fragment co-migrates with the crystallized protein. *d*, Bio-Rad molecular weight standards: 92.5K, 66.2K, 45.0K, 31.0K, 21.5K and 14.5K.

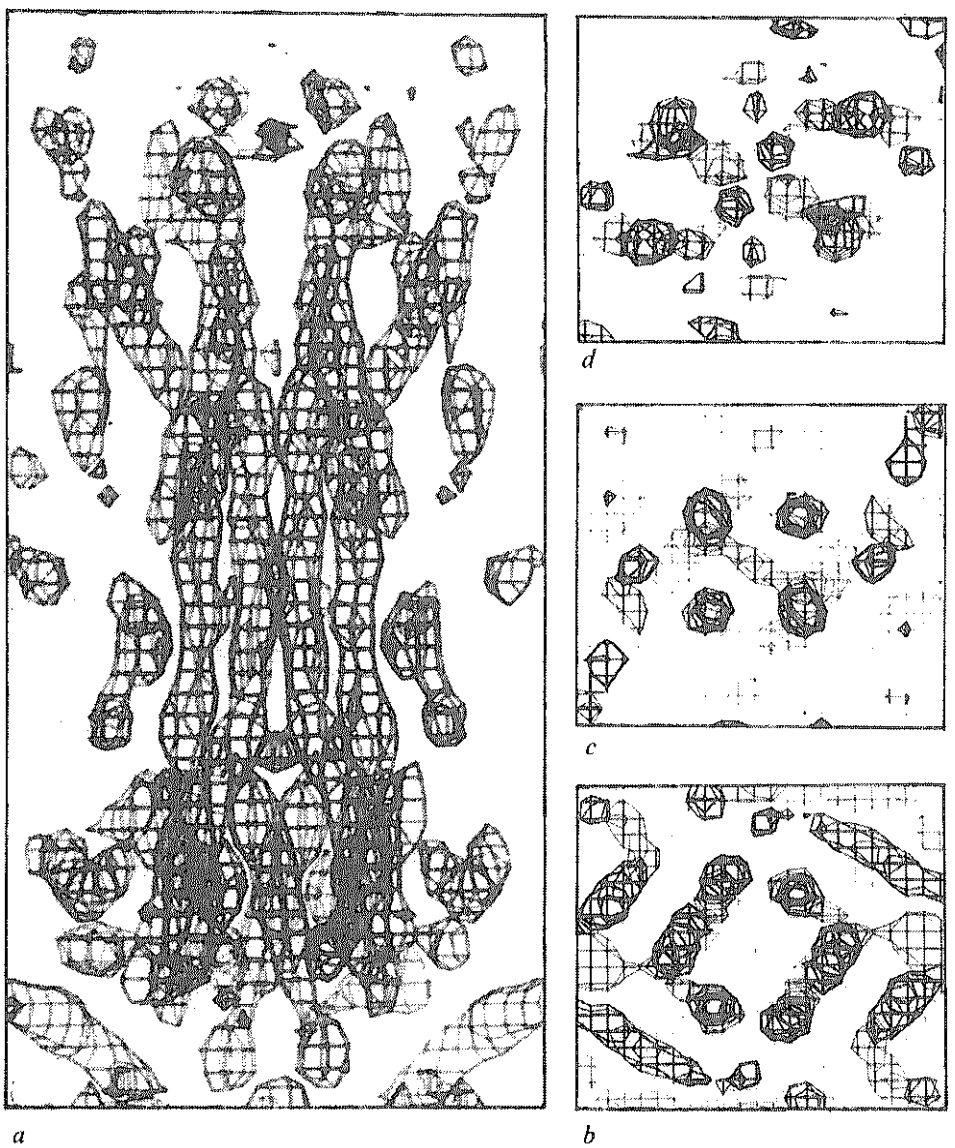

Fig. 2 *a*, View of the electron density map perpendicular to the dimer axis, which is vertical in the figure. The map was contoured two standard deviations above mean density and displayed on an Evans and Sutherland PS300 using the FRODO molecular graphics program by J. W. Pflugrath, M. A. Saper, B. Bush and A. Jones. At the bottom is the 6-fold bundle extending about 25 Å along the dimer axis. At the top, the density diverges, forming a 'two-headed' structure. (Symmetry-related density is seen in the lower corners.) *b–d*, Regions of the map shown in *a*, viewed along the dimer 2-fold axis. *b*, Approximately 20 Å from the 'bottom'. The 6-fold bundle of electron-dense rods is apparent. The distance between rods of density is ~10 Å. Unconnected density occurs adjacent to the bundle. *c*, Approximately 50 Å from the 'bottom'. The 4-fold bundle is the continuation of density which begins in *b*. *d*, Approximately 80 Å from the 'bottom'. Density from each monomer diverges and appears to form two distinct bundles.

The long rod-like α-helical core of the VSG is similar, in principle, to the trimeric α-helical coiled-coil in the stem region of the influenza virus haemagglutinin (HA)[13]. The occurrence of helical motifs in other surface glycoproteins has been discussed[14]. The connectivity observed here between the two long helices in the MITat 1.2 amino-terminal domain implies a helical region extending about 120 residues with only one break. If such a structure were common to VSGs of diverse sequences, it might be revealed by analysis of secondary structure predictions. Alternatively, this feature might readily be recognized in the low-resolution structure of a second VSG. Although interesting sequence homology among VSGs has been observed in the C-terminal region[15–18], the extent of structural homology is unknown[19].

Recent work[20] has shown that the N-terminal domain of another VSG, MITat 1.6, binds most monoclonal antibodies against the intact VSG molecule; that only a subset of the monoclonal antibodies which bind to isolated intact VSG will interact with the VSG when it is in the trypanosome surface coat; and that the subset of monoclonal antibodies which bind to the trypanosome coat also bind to the N-terminal domain. These results suggest that the N-terminal domain contains the regions of the VSG which are responsible for the antigenic variation recognized by antisera against live trypanosomes. Thus the three-dimensional structure of the N-terminal domain should contribute to establishing which parts of the sequence are implicated in the generation of antigenically distinct trypanosomes and to determining the structure of the variable epitope(s).

We thank Judith Creighton for technical assistance in the preparation of VSGs. Both laboratories have received support from the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (Project No. 810278 to D.C.W. and Project No. 800328 to M.J.T.). M.J.T. also received support from ILRAD, Nairobi, Kenya).

Received 27 April; accepted 5 July 1984.

1. Cross, G. A. M. *Parasitology* **71**, 393–417 (1975).
2. Vickerman, K. *Nature* **273**, 613–617 (1978).
3. Turner, M. J. *Adv. Parasit.* **21**, 69–153 (1982).
4. Cardoso de Almeida, M. L. & Turner, M. J. *Nature* **302**, 349–352 (1983).
5. Auffret, C. A. & Turner, M. J. *Biochem J.* **193**, 647–650 (1981).
6. Johnson, J. G. & Cross, G. A. M. *Biochem. J.* **178**, 689–697 (1979).
7. Ferguson, M. A. J. & Cross, G. A. M. *J. biol. Chem.* **259**, 3011–3015 (1984).
8. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
9. Barbet, A. F., Musoke, A. J., Shapiro, S. Z., Mpimbaza, G. & McGuire, T. C. *Parasitology* **83**, 623–637 (1981).
10. North, A. C. T., Phillips, D. C. & Mathews, F. S. *Acta crystallogr.* **A24**, 351–359 (1968).
11. Matthews, B. W. & Czerwinski, E. W. Acta crystallogr. **A31**, 480–487 (1975).
12. Blow, D. M. & Crick, F. H. C. *Acta crystallogr.* **12**, 794–802 (1959).
13. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
14. Cohen, C. & Phillips, G. N. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5303–5304 (1981).
15. Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **294**, 53–57 (1981).
16. Bridgen, P. J., Cross, G. A. M. & Bridges, J. *Nature* **263**, 613–614 (1976).
17. Mathyssens, G., Michiels, F., Hamers, R., Pays, E. & Steinert, M. *Nature* **293**, 230–233 (1981).
18. Holder, A. A. & Cross, G. A. M. *Molec. biochem. Parasit.* **2**, 135–150 (1981).
19. Lalor, T. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 998–1002 (1984).
20. Miller, E. N., Allan, L. M. & Turner, M. J. *Molec. biochem. Parasit.* (in the press).
21. Wyckoff, H. W. *et al. J. molec. Biol.* **27**, 563–578 (1967).
22. Matthews, B. W. *Acta crystallogr.* **20**, 230–239 (1966).
23. Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. *Acta crystallogr.* **B24**, 997–1003 (1968).
24. Matthews, B. W. *Acta crystallogr.* **20**, 82–86 (1966).
25. Blow, D. M. & Rossmann, M. G. *Acta crystallogr.* **14**, 1195–1202 (1961).

Printed in Great Britain by Macmillan Production Ltd., Basingstoke, Hampshire