Reprinted from Nature, Vol. 325, No. 6099, pp. 84-86, 1 January 1987

Two variant surface glycoproteins of *Trypanosoma Brucei* of different sequence classes have similar 6 Å resolution X-ray structures

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

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA
* Merck, Sharp and Dohme Research Laboratories, PO Box 2000, Rahway, New Jersey 07065, USA

Antigenic variation in the African trypanosome is mediated through changes in the composition of the surface coat[1,2]. By controlling expression of the major surface protein, the variant surface glycoprotein (VSG), from a repertoire of perhaps 1,000 different genes[3] the organisms exhibit a series of antigenically distinct coats and evade the host's immune system. We have determined the 6 Å resolution structure of a *T. brucei* variant surface glycoprotein, ILTat 1.24, using X-ray crystallography. The crystallized protein consists of the N-terminal two-thirds of the intact VSG which has a relative molecular mass (M_r) of 60,000 (60K). The structure, which includes a 90 Å long α-helical bundle, is strikingly similar to that of the N-terminal fragment of VSG MITat 1.2 (ref. 4). Although most known VSG sequences show little similarity of primary sequence in the N-terminal domain, the similarity between the structure of a Class I (ILTat 1.24) and a Class II (MITat 1.2) VSG antigen suggests that VSGs may share a common tertiary structure.

VSG molecules are attached to the cell surface via a C-terminal glycolipid anchor[5-7]. The protein can be purified in a soluble form (sVSG), lacking the lipid, and is a dimer in solution[8,9]. Proteolysis experiments suggest the existence of two domains[10]. Significant sequence homologies between VSGs are confined to the C-terminal domain and are found largely in the last 50 amino acids[1,11,12]. Based on these homologies and two different patterns of cysteine residues in the C-terminal domain VSGs have been divided into two classes having C-terminal aspartic acid (Class I) or serine (Class II)[11,12]. The N-terminal, variable domain comprises two-thirds of the molecule. No amino-acid sequence homologies larger than two residues and no significant homology at the DNA level are evident in the first 350 amino acids of the N-terminal domains[12]. Nevertheless, three observations argue for an underlying conservation of structure: (1) the apparent conservation of four cysteines in several Class I VSGs[1], (2) the existence of sevenfold periodicities in hydrophobicity indicating stretches of interlocking alpha helices in four sequences[13], and (3) the fact that the first 40 amino acids of 18 N-terminal sequences have significant similarities, if conservative amino-acid substitutions are allowed[14]. The lack of agreement in secondary structure predictions for the amino-terminal domain of 9 sequences[15] has, however, been interpreted to suggest multiple folding patterns.

Sequence analysis of a partial complementary DNA clone of ILTat 1.24 showed that it belonged to the Class I family (M. Carrington, P. Liberator and M.T., unpublished). The complete sequence of the MITat 1.2 VSG has been deduced by sequence analysis of a cDNA clone, establishing it as a member of Class II (M. Carrington, I. Roditi and M.T., unpublished). ILTat 1.24 sVSG was prepared as described[16]. Purified sVSG was crystallized by vapour diffusion of a 40 mg ml^{-1} solution (50 mM MES, *p*H 6.5, 0.2% azide, 30% $(NH_4)_2SO_4$) against 60%-saturated ammonium sulphate in the same buffer. Protein recovered from the ILTat 1.24 crystals had a substantially lower M_r (46K) than the native purified protein (60K). The amino-terminal sequence of the crystalline fragment was found to be the same as the intact sVSG (T-FGVK).

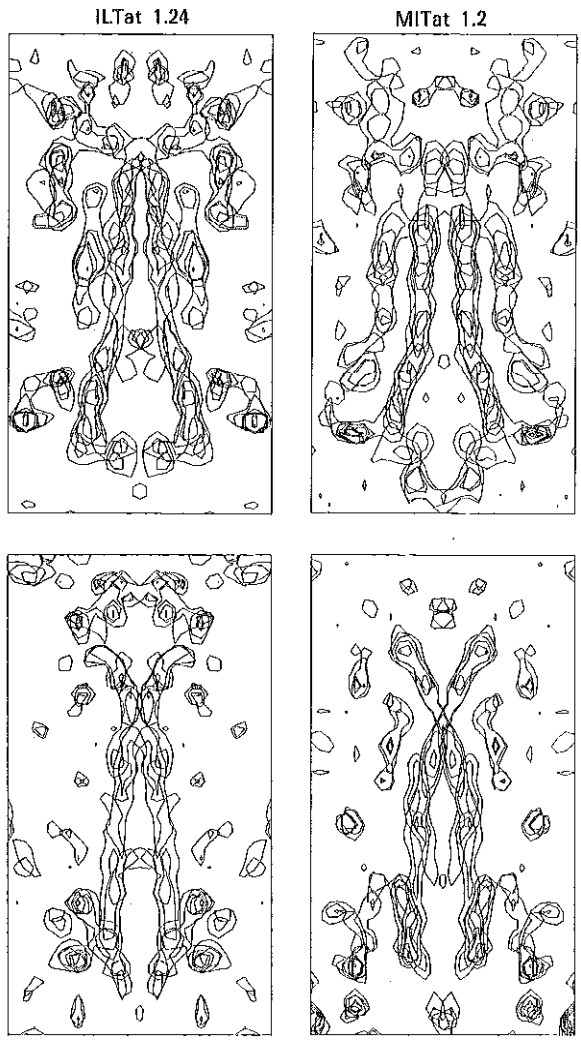

Fig. 1 Central lateral sections showing the α-helical core of the molecules: Left ILTat 1.24, right MITat 1.2. Each diagram (110 Å × 60 Å) consists of 5 superimposed 2 Å sections of the 6 Å resolution electron density map contoured at 1 and 2 standard deviations above the mean density. Noise in low density regions of the electron density map was reduced using an interative solvent flattening algorithm[21]. The relative orientation of the twofold averaged ILTat 1.24 map with respect to the MITat 1.2 map was determined by maximizing the cross-correlation of their electron densities (see Table 1).

The initial 6 Å electron density map (crystallographic data are summarized in Table 1) showed strong features suggesting a rod-shaped molecule packed with its long axis at ~40° from the c axis in the b, c plane. The electron density sectioned perpendicular to this molecular axis showed features recognizable as the sixfold and fourfold helical bundle regions seen in the low resolution structure of MITat 1.2 (ref. 4). Comparison of the map to the current model of the MITat 1.2 VSG structure (D.F. *et al.*, unpublished) allowed a more accurate determination of the dimer's non-crystallographic twofold axis. Orientation of this axis was refined by maximizing the twofold electron density correlation in a cylindrical region of the map and the density was then averaged about the refined axis.

The MITat 1.2 and ILTat 1.24 N-terminal dimers are both elongated molecules with a length of ~100 Å and a maximum width of 60 Å. The twofold axes of the dimers are parallel to the long axes of the molecules. The dominant features of both maps (Figs 1, 2) are long rods of high electron density, which are characteristic of α-helices at this resolution. The arrange-

Table 1 Summary of 6 Å data from ILTat 1.24 VSG crystals

	Native	Hg
Cell dimensions (Å)	55.1, 98.8, 172.9	55.1, 98.6, 172.4
No. of reflections collected	11452	11577
Symmetry R factor	0.037	0.055
No. of unique reflections in R factor	2394	2472
No. of reflections phased	2470	
Mean figure of merit	0.68	
Kraut R	11.8	
$\langle Fh\rangle/\langle E\rangle$	1.98	
$CC_{twofold}$	0.28	
CC_{cross}	0.27	

$$\text{Symmetry R factor} = \frac{\sum_{hkl}\sum_{obs}(I_{obs} - \langle I\rangle_{hkl})}{\sum_{hkl}\sum_{obs} I_{obs}};$$

$$\text{Kraut R} = \frac{\sum[|Fph\,(obs) - Fph\,(calc)|]}{\sum[Fph\,(obs)]} \times 100,$$

where Fph is the derivative structure factor;

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

$$CC = \frac{\langle\rho_1\rho_2\rangle - N\langle\rho_1\rangle\langle\rho_2\rangle}{[(\langle\rho_1^2\rangle - N\langle\rho_1\rangle^2)(\langle\rho_2^2\rangle - N\langle\rho_2\rangle^2)]^{1/2}}$$

where for $CC_{two\text{-}fold}$, ρ_1, ρ_2 are electron densities at positions related by rotation of 180° around the dimer axis; for CC_{cross}, ρ_1 = ILTat 1.24 electron density, ρ_2 = MITat 1.2 electron density. The space group is $P2_12_12_1$; unit cell, a = 55.1, b = 98.9, c = 172.9 Å. Bipyramidal crystals grow to $0.4\times0.3\times0.3$ mm^3 and diffract to 2.5 Å. The asymmetric unit of the crystal contains one copy of the N-terminal dimer. An isomorphous mercury derivative was obtained by soaking the crystals for 6 h in 65% ammonium sulphate, 5 mM MES *p*H 6.5, 0.2% azide and 0.1 mM K_2HgI_4. The 4.2 Å data were collected on a multiwire area detector (Nicolet/Xentronics, Madison, WI) mounted 12 cm from the crystal, using a rotating anode X-ray source (Elliot GX-13) with 100 mm focus and Franks double mirror optics[17,18]. Each data set consists of 1,500 data frames (0.083 degree oscillations of the crystal exposed for 90 s) which were collected from crystals at 4 °C over a 50 h period and processed on-line[18]. The diffraction data were divided into groups consisting of integrated spot intensities from 50 consecutive frames and these groups were scaled to each other[19]. Harker sections from the isomorphous and anomalous difference Patterson maps were interpreted as being consistent with a single Hg site at fractional unit cell coordinates 0.076, 0.190, 0.082. The heavy atom parameters were refined[20] and single isomorphous replacement and anomalous scattering phases calculated for reflections of resolution less than 6 Å. Correlation coefficients were calculated inside a cylinder with radius of 30 Å and length of 110 Å oriented along the dimer axis.

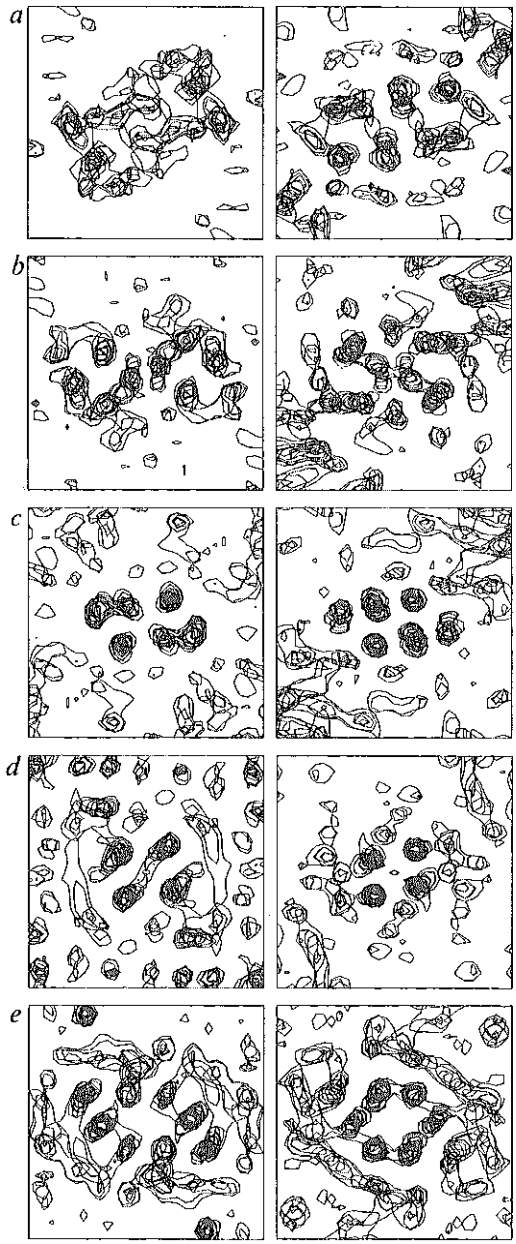

Fig. 2 Axial sections: Left ILTat 1.24, right MITat 1.2 *a*, 80–96 Å; *b*, 64–80 Å; *c*, 48–64 Å, *d*, 32–48 Å; *e*, 16–32 Å from the bottom of the region shown in Fig. 1. Each diagram (60 Å × 60 Å) consists of 8 superimposed 2 Å sections each contoured at 1 and 2 standard deviations above the mean density.

ments of these helices, which form the central core of the molecules, are similar: two helices extend nearly 90 Å, forming a fourfold bundle around the dimer axis in the middle third of the molecule (Fig. 2*c*, *d*). A third helix is packed against the long helices to form a sixfold bundle in the bottom third of the molecule (Fig. 2*e*). A fourth helix is packed against the top of the fourfold bundle (Fig. 2*b*, *c*). These features account for at least 40% of the polypeptide chain.

The top of the central helical core is capped by a region containing what appears to be non-α-helical density in both molecules. This cap extends ~20 Å and, viewed from above, has a rectangular cross-section of ~30 Å × 50 Å (Fig. 2*a*). At this resolution we cannot interpret the cap region density or establish the connectivity of the α-helices.

The extent of structural homology between the two VSGs at 6 Å suggests that there may be an underlying structural motif within the VSGs. The primary selection pressure for the survival of VSG sequences may be their ability to conform to the constraints of this structure[22,23]. This similarity also leaves open the question of what significance, if any, the classification of VSGs, based on C-terminal homologies, has to the structure of the N-termini. Fragments of four other VSGs have been crystallized[24] and it is likely that at least somé of these will eventually yield further structures. Should the sequences of ILTat 1.24 and MITat 1.2 prove to be as dissimilar as expected from our current knowledge of VSG sequences, the similarity of their structures suggests the possibility of creating a useful structural database from the large number, possibly 1,000, of disparate VSG sequences that may correspond to a single protein fold.

We thank Judith Creighton for technical assistance in the preparation of VSGs and Anne Gurnett and Michael Silver for their interest and help. D.C.W. acknowledges support from NIH (AI-21324-02) and NSF (CHE-8509574). This work was initiated by the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (project No. 810278 and No. 800328).

Received 23 July; accepted 17 October 1986.

1. Cross, G. A. M. *Phil. Trans. R. Soc. Lond.* B307, 3–12 (1984).
2. Turner, M. J. *Phil. Trans. R. Soc. Lond.* B307, 27–40 (1984).
3. Van der Ploeg, L. H. T. *et al. Nucleic Acids Res.* **10,** 5905–5923 (1982).
4. Freymann, D., Metcalf, P., Turner, M. J. & Wiley, D. *Nature* **311,** 167–169 (1984).
5. Cardoso de Almeida, M. L. & Turner, M. J. *Nature* **302,** 349–352 (1983).
6. Ferguson, M. A. J. & Cross, J. A. M. *J. biol. Chem.* **259,** 3011–3015 (1984).
7. Ferguson, M. A. J., Haldar, K. & Cross, J. A. M. *J. biol. Chem.* **260,** 4963–4968 (1985).
8. Auffret, C. A. & Turner, M. J. *Biochem. J.* **193,** 647–650 (1981).
9. Gurnet, A. M., Raper, J. & Turner, M. J. *Molec. biochem. Parasitol.* **18,** 141–153 (1986).
10. Johnson, J. G. & Cross, G. A. M. *Biochem. J.* **178,** 689–697 (1979).
11. Holder, A. A. & Cross, G. A. M. *Molec. biochem. Parasitol.* **2,** 135–150 (1981).
12. Rice-Ficht, A. C., Chen, K. K. & Donelson, J. E. *Nature* **294,** 53–57 (1981).
13. Cohen, C. *et al. Nature* **311,** 169–171 (1984).
14. Olafson, R. W. *et al. Molec. biochem. Parasitol.* **12,** 287–298 (1984).
15. Lalor, T. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81,** 998–1002 (1984).
16. Cross, G. A. M., *J. Cell Biochem.* **24,** 79–90 (1983).
17. Durbin, R. M. *et al. Science* **232,** 1127–1132 (1986).
18. Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Cryst.* (in the press).
19. Fox, G. C. & Holmes, K. C., *Acta cryst.* **20,** 886–891 (1966).
20. Dickerson, R. E., Weinzierl, J. E. & Palmer, R. A. *Acta cryst.* B24, 997–1003 (1968).
21. Wang, B. C. *Meth. Enzym.* **115,** 90–112 (1985).
22. Turner, M. J., *Biochem. Soc. Symp.* **49,** 169–181 (1984).
23. Turner, M. J., *Annls Inst. Pasteur, Paris (Immunol.)* **136C,** 41–49 (1985).
24. Metcalf, P. *et al.* (manuscript in preparation).

Printed in Great Britain by Turnergraphic Limited, Basingstoke, Hampshire