

Trypanosoma congolense: Structure and Molecular Organization of the Surface Glycoproteins of Two Early Bloodstream Variants[†]

James E. Strickler,^{‡§} David A. Binder,^{‡||} James J. L'Italien,^{‡,⊥} Grant T. Shimamoto,^{‡,#} Suzanne W. Wait,[‡] Laura J. Dalheim,[‡] Jiri Novotny,^Δ Jeffrey A. Radding,[‡] William H. Konigsberg,[%] Martine Y. K. Armstrong,[×] Frank F. Richards,[‡] and Thomas M. Lalor^{*,‡}

Departments of Internal Medicine, Molecular Biophysics and Biochemistry, and Epidemiology and Public Health, The Yale-MacArthur Center for Molecular Parasitology, Yale University Parasitology and Tropical Medicine Program, New Haven, Connecticut 06510, and The Molecular and Cellular Research Laboratory, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114

Received May 6, 1986; Revised Manuscript Received October 2, 1986

ABSTRACT: The complete primary structures of two variant specific glycoproteins (VSGs) of the nannomonad *Trypanosoma (N.) congolense* are presented. These coat proteins subserve the function of antigenic variation. The secondary structure potentials of both VSGs have been calculated. The amino acid sequences and secondary structure potentials of these VSGs have been compared with the primary structures and secondary structure potentials of several *Trypanosoma brucei* complex VSGs. In homologous regions, the *T. brucei* complex VSGs show a pattern of sharply contrasting secondary structure potentials. It has been suggested previously that this pattern gives rise to different folding structures in different members of this polygene protein family. Thus, different short regions of the polypeptide sequence are exposed as antigenic "caps" on the solvent-exposed surface of intact trypanosomes. A sharply contrasting secondary structure potential pattern is also found in regions of the two *T. congolense* VSGs. However, there is little homology of primary structure between each of the two *T. congolense* VSGs and any member of the *T. brucei* complex VSG polygene family whose primary structure has been determined.

The variant specific glycoproteins (VSGs)¹ of the African trypanosomes comprise a highly antigenic surface coat covering the parasitic organism. Expression of alternative VSGs, having different primary structures, results in the phenomenon of antigenic variation. In relapsing waves of parasitemia, the majority of parasites in each wave is covered by a different, antigenically non-cross-reactive VSG. It is believed that antigenic variation, i.e., the ability to produce new populations of parasites not recognized by antibodies to previous trypanosome populations, is the key mechanism which permits trypanosomes to establish themselves in their mammalian host. In addition to primary structure alterations, there is preliminary evidence that antigenic variation is also accomplished by restricting the region of the protein which may be exposed to the solvent-accessible exterior of the organism and that the exposed region of the primary structure may differ in different serotypes (Lalor et al., 1984; Cohen et al., 1984).

It is known that the strategy of antigenic variation is not confined to trypanosomes and that prokaryotes such as spirochetes and salmonella are capable of changing their surface

coats (Stoenner et al., 1982; Silverman et al., 1979). We here present data on the amino acid sequences and secondary structure potentials of two *Trypanosoma congolense* variant specific glycoproteins. We discuss the structural basis of antigenic variation in light of these data.

MATERIALS AND METHODS

Trypanosomes and VSG Isolation. The Lister 1/148 stock of *T. congolense* and details of the cell cloning and fly passages have been previously described (Rosen et al., 1981). The *T. congolense* variants used have been given the acronym Yale (Nannomonas) antigen types 1.1 and 1.3 (YNat 1.1 and 1.3). An earlier set of papers (Rosen et al., 1981; Onodera et al., 1981; Bogucki et al., 1982) referred to YNat 1.1 and 1.2. The YNat 1.1 variant is the identical variant described in the cited papers. However, the YNat 1.2 variant was lost. The variant YNat 1.3 described in this paper was isolated independently from another BALB/c mouse which had been infected with YNat 1.1 and whose infection was allowed to relapse once only. Before the current study was begun, YNat 1.1 and YNat 1.3 were recloned by the triple-cloning procedure (Rosen et al., 1981). Following two expansions in irradiated rats, working stabilates of the clone were maintained in the vapor phase of a liquid nitrogen freezer. Thus, populations used for protein isolation were all derived from the third passage of the parasite

[†] This research was supported by USPHS Grants AI08614 (F.F.R.), AI15530 (F.F.R.), GM12607 (W.H.K.), AI06425 (T.M.L.), and AI06317 (G.T.S.) and by The MacArthur Foundation. J.E.S. was a fellow of the James Hudson Brown-Alexander B. Cox Foundation. J.A.R., F.F.R., and T.M.L. are investigators in the Consortium on the Biology of Parasitic Diseases of The MacArthur Foundation.

* Address correspondence to this author.

[‡] Department of Internal Medicine.

[§] Present address: Department of Molecular Genetics, Smith Kline and Beckman, Swedeland, PA 19479.

^Δ Present address: Emergency Medicine, University of Cincinnati, Cincinnati, OH 45221.

[⊥] Present address: Molecular Genetics, Minnetonka, MN 55343.

[#] Present address: Monsanto, St. Louis, MO.

^Δ The Molecular and Cellular Research Laboratory.

[%] Department of Molecular Biophysics and Biochemistry.

[×] Department of Epidemiology and Public Health.

¹ Abbreviations: VSG, variant specific glycoprotein; YNat, Yale Nannomonas antigen type; AnTat, Antwerp Trypanozoon antigen type; ILTat, International Laboratory Trypanozoon antigen type; MITat, Molteno Institute Trypanozoon antigen type; IoTat, Iowa Trypanozoon antigen type; Bicine, *N,N*-bis(2-hydroxyethyl)glycine; HPLC, high-performance liquid chromatography; PTH, phenylthiohydantoin; CNBr, cyanogen bromide; Tris-HCl, 2-amino-2-(hydroxymethyl)aminomethane hydrochloride; ATP, adenosine 5'-triphosphate; dNTP, deoxyadenosine, deoxyguanosine, thymidine, or deoxycytidine 5'-triphosphate; SDS, sodium dodecyl sulfate; APS, aminopolystyrene; PAGE, polyacrylamide gel electrophoresis; CD, circular dichroism; UTR, untranslated region; ds, double stranded; IEF, isoelectric focusing; bp, base pair(s).

grown in lethally X-irradiated outbred, male retired breeder Sprague-Dawley rats. Infected blood was obtained by either cardiac puncture or aortic cannulation and perfusion (Rosen et al., 1978) during late log phase of the parasite's growth. Trypanosomes were purified by glycerol lysis of rat red blood cells and DEAE-cellulose chromatography, as described previously (Rosen et al., 1978; Lanham & Godfrey, 1970), or by the Percoll method (Bwaye & Hirumi, 1980). The procedure was modified in that the DEAE-cellulose column was run in Bicine-saline-glucose buffer (50 mM Bicine, 7 mM glucose, 5 mM KCl, 4 mM MgCl₂, 50 mM NaCl, 55 mM sucrose, 20 mM CaCl₂, and 35 mM NaOH, pH 8.0) (Lanham & Godfrey, 1970). The purified trypanosomes were suspended at 5×10^9 cells/mL in 50 mM Tris-HCl, pH 7.4, 50 mM KCl, 5 mM MgCl₂, 1 mM tosyllysine chloromethyl ketone, and 0.8 mM phenylmethanesulfonyl fluoride and snap frozen in dry ice-ethanol. The frozen cells were stored at -70 °C. Lysis occurred on rapidly thawing the cell suspension at 37 °C. The lysate was further disrupted in a Dounce homogenizer. The VSG was isolated by lectin affinity chromatography and DEAE-cellulose chromatography as previously described (Bogucki et al., 1982; Strickler et al., 1978). Rabbit antisera raised against purified YNat 1.3 glycoprotein agglutinate living *T. congolense* organisms of the VSG 1.3 serotype and do not agglutinate living YNat 1.1 organisms. The agglutination observed is characteristic of VSG-mediated agglutination (data not shown).

Amino Acid Sequence Procedures. The general strategy of protein sequence analysis outlined previously (L'Italien & Strickler, 1982) was followed. Briefly, the reduced carboxamidomethylated protein was digested with CNBr, trypsin, *Staphylococcus aureus* protease, or chymotrypsin. Peptides were purified by reverse-phase HPLC on a Waters μ Bondapak C-18 column. The tryptic, *Staph. aureus* protease, and CNBr peptide maps for YNat 1.1 have been published as well as preliminary sequence data on some of the peptides (L'Italien & Strickler, 1982). Amino acid compositions were determined on a Beckman 121M amino acid analyzer. Proteins and peptides were immobilized on solid supports and subjected to automatic Edman degradation on a Sequemat Mini-15 solid-phase sequencer equipped with a P-6 autoconverter and a 65 min/cycle program (Laursen, 1971). PTH-amino acids were analyzed by HPLC on an Altex Ultrasphere ODS reverse-phase column (L'Italien & Strickler, 1982).

Construction of cDNA Libraries. The methods for preparation of RNA, poly(A) RNA, and plasmid cDNA libraries have been described in detail (Binder, 1984) and are outlined here. Total RNA was isolated by the urea-LiCl method of Auiffay and Rougeon (1980), with minor modifications. Poly(A) RNA was prepared by three cycles of enrichment on oligo(dT)-cellulose. The preparation of double-stranded copy DNA followed the methods outlined in Maniatis et al. (1982).

Fifty nanograms of purified dC-tailed ds cDNA was annealed with *Pst*I-cleaved, dG-tailed plasmid pUC 8 (Messing & Viera, 1982) and was used to transform *Escherichia coli* RR1 cells to ampicillin resistance. Separate plasmid libraries for YNat 1.1 and YNat 1.3 were constructed. Relative to transformation with undigested pUC8 plasmid, the efficiency of transformation of RR1 cells by the annealed cDNA-plasmid constructs was approximately 0.44% for both variants. White recombinant plasmid-containing colonies were distinguished from blue, intact vector-transformed colonies by use of the isopropyl β -D-thiogalactopyranoside and 5-bromo-4-chloro-3-indolyl β -D-galactopyranoside indicator system (Messing & Viera, 1982).

Recovery of VSG cDNA Clones. White colonies were picked and used to prepare master plates for each library. Duplicate masterplates were blotted with nitrocellulose filters (Grunstein & Hogness, 1975); blotted filters were prepared for hybridization analysis and were prehybridized and incubated with cDNA probe in 50% recrystallized formamide, 5 \times Denhardt's solution, 5 \times SSPE, and 0.1% SDS. Filters from each YNat variant library were screened with probes prepared from mRNA of both YNat variants. Probe consisted of single-stranded cDNA synthesized from oligo(dT)-primed, poly(A) RNA from the respective variant, α -³²P-labeled dATP and dNTPs, and avian myeloblastosis virus reverse transcriptase (Rowecamp & Firtel, 1980). Putative VSG-encoding recombinant plasmids were selected by the differential hybridization method (Hoeijmakers et al., 1980). A total of 4000 colonies were screened for each library; 150 differentially reactive colonies were identified from the YNat 1.1 library, and 80 such colonies were identified in the YNat 1.3 library. After a second screening, 50 differentially reactive colonies from each library probe were selected for further analysis. Plasmid DNA from each colony was digested with *Pst*I and fractionated on 1% agarose gels. Recombinant plasmids with relatively long inserts were selected. Southern blotting analysis confirmed the differential reactivity with the two probes described above.

For the YNat 1.1 library (A series), plasmids A.08, A.09, and A.51 were mapped and subcloned into M13 mp8, mp9, mp18, or mp19 for DNA sequencing (Messing, 1983). Positive identification of VSG cDNA clones was confirmed by comparison of open reading frames inferred from DNA sequencing with known protein sequence obtained from analysis of VSG-derived peptides. The plasmid A.09 includes the entire coding region of YNat 1.1 VSG mRNA. For the YNat 1.3 library (B series), plasmids B.16, B.17, B.20, and B.23 were used for subcloning and sequence analysis. The B.17 plasmid includes the entire coding region of YNat 1.3 VSG mRNA.

None of these plasmids extended to the 5' end of the respective mRNA, as inferred from analysis of the corresponding mRNAs (data not shown).

DNA Sequence Analysis. The DNA sequencing methods of Sanger et al. (1977), Biggin et al. (1983), and Maxam and Gilbert (1980) were employed. For Maxam and Gilbert sequencing of inserts in M13 vectors, M13 universal primers (New England Biolabs) and synthetic primers (New England Biolabs) and synthetic primers were labeled at the 5' end with T₄ polynucleotide kinase and [γ -³²P]ATP, annealed with M13 single-stranded phage templates, and extended with the Klenow fragment of *E. coli* DNA polymerase I and dNTPs (Hong, 1981; Tschudi et al., 1985). After cleavage with appropriate endonucleases, extension products were run on an 8% polyacrylamide-7 M urea gel. Bands containing extension products were cut out and extracted with 0.3 M ammonium acetate. The labeled fragments were ethanol precipitated with carrier tRNA prior to DNA sequencing (Tschudi et al., 1985).

Calculation of Secondary Structure Potential. The analysis of secondary structure potential was by the methods of Chou and Fasman (1974a,b, 1978) and Garnier et al. (1978). In addition, the CHOU program of Novotny and Auiffay (1984) was used for the analysis of secondary structure potential, hydrophobicity, and charged residue distribution.

Homology Searches. Homologies between protein sequences were compared with the MATRIX program of Novotny (1982). This dot matrix analysis compares two sequences, *m* and *n* residues in length, in the form of a matrix composed of *m* rows and *n* columns. Each sequence element, e.g., amino

acid residue or base, of one sequence is compared to every element of the second sequence. The matrix element m with coordinates i and j , i.e., $m(i,j)$, is a dot if the corresponding i th and j th elements are identical. All computer-assisted computations were performed on a VAX 11/750 computer.

Circular dichroism measurements were performed on a Cary 16 spectrophotometer. A first approximation of the average value of α -helix content in VSG YNat 1.1 was calculated from the following relationship: $\% \alpha \text{ helix} = (\theta_{208\text{nm}} - 4000)/(33\,000 - 4000)$ where θ = mean residue ellipticity (Greenfield & Fasman, 1969).

RESULTS

The complete amino acid and DNA base sequences of the coding regions of YNat 1.1 and 1.3 are shown in Figure 1. The results are compiled from direct amino acid sequencing and DNA base sequence analysis of cloned cDNA fragments of YNat messengers. A partial restriction map indicating fragments and direction of DNA sequencing as well as the protein fragments used to assemble the data is shown in Figure 2.

Both YNat VSG messengers encode a putative N-terminal leader sequence. For the YNat 1.1 DNA sequence, the ATG codon for the initiator methionine of the leader sequence is 28 codons upstream of the GAT triplet that encodes the known N-terminal aspartic acid residue of the mature VSG. Similarly, the YNat 1.3 messenger encodes a putative 22-codon leader sequence upstream from the ACT codon for the N-terminal threonine residue of the mature YNat 1.3 VSG. The sizes of these putative leader sequences are within the range observed for the *T. brucei* complex VSG leader sequences (Donelson & Rice-Ficht, 1985). In YNat 1.1 cDNA clone A.09, there is an open reading frame starting at the initiator methionine of the leader, extending 419 codons and followed by a TAA stop codon. This reading frame is confirmed in 87% of the codons. The YNat 1.3 cDNA clone B.17 contains an open reading frame starting with the N-terminal leader, extending 413 codons, and followed by a TAA stop codon. The reading frame is confirmed for 24% of the residues by amino acid sequence analysis of peptides derived from N-terminal, middle, and C-terminal regions of the protein. Digestion of the respective VSGs with carboxypeptidase Y provided no evidence of a free carboxyl terminus for either VSG. For the YNat 1.1 variant, protein sequence studies confirm the reading frame to serine-372 (Figure 1), thus leaving 19 codons unconfirmed before the TAA stop. Ser-372 is tentatively assigned as the C-terminus of the mature protein since (i) by composition the tryptic peptide beginning with Ser-364 contained neither Lys nor Arg and (ii) every amino acid calculated from the composition is accounted for by the peptide Ser-364 to Ser-372. In three separate attempts, this tryptic peptide coupled in exceptionally low yield to APS (<10%), which would be consistent with a C-terminally blocked or partially blocked peptide. The YNat 1.3 reading frame is last confirmed at Arg-315 (Figure 1), thus leaving the last 75 codons unconfirmed. Multiple DNA sequencing methods have been applied to these regions. Also, in the case of YNat 1.3, DNA sequence analysis of three independent cDNA clones confirms the sequence in the C-terminal region.

The calculated molecular weight of the unglycosylated pre-protein, including the leader sequence, for YNat 1.1 is 44 825; the corresponding value for YNat 1.3 is 44 255. These values are somewhat higher than the experimental molecular weights recently reported for in vitro translated *T. congolense* VSG variants AmNat 1.1 and AmNat 3.1 of 39 000 and 43 000, respectively (Cook et al., 1985). SDS-PAGE analysis

of YNat 1.1 and 1.3 VSGs purified to homogeneity from freeze-thaw lysates, in the presence of protease inhibitors, gave apparent molecular weights of 53 000 and 58 000, respectively, for the mature, glycosylated forms (data not shown); a similar SDS-PAGE analysis of the mature AmNat 1.1 and AmNat 3.1 variants revealed apparent molecular weights of 51 700 and 49 900, respectively (Cook et al., 1985). Therefore, with respect to estimated polypeptide chain length and apparent molecular weight on SDS-PAGE, *T. congolense* VSGs as a group appear to be smaller glycoproteins than the VSGs of the *T. brucei* complex, which are composed of approximately 500 amino acids and have molecular weights of 55 000–65 000 in the mature forms (Rice-Ficht et al., 1981; Donelson & Rice-Ficht, 1985).

When the protein sequences of the YNat variants are aligned with respect to the known N-termini of the mature VSGs, and the distributions of selected residues are mapped (Figure 3), some possible amino acid residue and sequence conservation is suggested. Seven out of nine cysteines appear in similar positions in each variant. The cysteines at positions 13 and 14 in YNat 1.1 and YNat 1.3, respectively, are similar to the conserved cysteine which is seen near the N-terminus of *T. brucei* VSGs (Donelson & Rice-Ficht, 1985). There is a shared pattern of five cysteines in the middle of the sequences (YNat 1.1, positions 148, 165, 167, 180, and 215; YNat 1.3, positions 151, 173, 175, 187, and 217). This suggests that the cysteines may stabilize a domain structure. At least three of the six tryptophan residues in YNat 1.3 are in similar locations in both proteins. Each of the three tyrosine residues in the YNat 1.3 appears to have a counterpart in the YNat 1.1 sequence. The significance of conserved aromatic residues is unclear, although they may also be involved in stabilizing the protein structure.

Potential carbohydrate attachment sites of the (N-X-S/T) type are located in C- and N-terminal regions in both proteins, and in the middle region of the YNat 1.1 molecules (Figure 3). Amino acid sequence analysis revealed that three of the potential sites in YNat 1.1 are not utilized. Asn residues were positively identified at positions 116, 127, and 187, in yields comparable to those of the residues immediately preceding and following the respective Asn residues. No residue was identifiable at position 54, 330, or 351, consistent with glycosylation at these positions. Some of the attached carbohydrates may be on the surface of the molecule, since some *T. congolense* variants can be agglutinated with concanavalin A and other lectins (Jackson et al., 1978; Rautenberg et al., 1980). The carbohydrate side chain at Asn-351 of YNat 1.1 is sensitive to endoglucosaminidase H (L'Italien & Strickler, 1982) and is therefore probably of the high-mannose type, linked through a chitobiose unit to Asn. The sugar composition of the tryptic and staphylococcal protease derived peptides containing this residue is consistent with this interpretation (data not shown).

VSG purified from YNat 1.1 gives a characteristic triplet banding pattern on isoelectric focusing gels (Rosen et al., 1981; Onodera et al., 1981). Each band was electroeluted and sequenced by using preparative flat-bed polyacrylamide gels. All three bands had identical N-terminal sequences through residue 15 (unpublished results). One possible explanation for this behavior is differential sialic acid content (Rautenberg et al., 1980). Qualitative carbohydrate analysis of selected 1.1 VSG peptides indicated the presence of sialic acid. We cannot, however, rule out the possibility that the variation in the IEF migration results from proteolytic cleavage at or near the C-terminus.

YNat 1.1

	ATG AAA OGT GTT TTA TCA AAT GTG TTA AAG GCT TGG ATA TTT ACA ATT GTT GCT TTT CAC AAC TTC TCC ACA TCG GTA ACA GCT	84
	Met Lys Arg Val Leu Ser Asn Val Leu Lys Ala Trp Ile Phe Thr Ile Val Ala Phe His Asn Phe Ser Thr Ser Val Thr Ala	
1	GAT GCA CCG GTC AAT GGC GCA GAG TAC AAT GGC CTT TGC GGC CTA TAT AAC ATA GGC AGA GGC GGT GAA GGA CTC AAA GAA GAA GAT TGG	174
	Asp Ala Pro Val Asn Ala Ala Glu Tyr Asn Ala Leu Cys Arg Leu Tyr Asn Ile Ala Arg Ala Gly Glu Gly Leu Lys Glu Glu Asp Trp	
31	CTG CCA TGC GCG GGT AAG GCA GGC TGT GAA AAG ACA GCT GGC TCA ATC GAA GAT GTC TTT ATG TCC AAG CTC AAN TTC TCT GAA CTT AGC GCA	264
	Leu Pro Cys Ala Gly Lys Ala Ala Cys Glu Lys Thr Ala Ala Ser Ile Glu Asp Val Phe Met Lys Leu Asn Phe Ser Glu Pro Ser Ala	
61	GTT GTG ACC ACT CTT GAC GGG ACT GGC GTC GAG CTG CAA AAT TCC GCG AGC ACT CGA ATC AAG OCG GCG AAG TTG GGC AAG GTA TTG GCG	354
	Val Val Thr Thr Leu Asp Gly Thr Arg Val Glu Leu Gln Asn Ser Ala Ser Thr Arg Ile Lys Arg Ala Lys Leu Ala Lys Val Leu Ala	
91	GCC GCT GAG CAG ATA AAA GCG CAG CAG CTT AAA TAC CAC GAA TCC TCG AAG AGC CTG TTG GAG AGT GCA AAG GCG AAC TTC ACG AAG GCG	444
	Ala Ala Glu Thr Ile Lys Ala Gln Gln Leu Lys Tyr His Glu Ser Ser Lys Ser Leu Leu Glu Ser Ala Lys Ala Asn Phe Thr Lys Ala	
121	ATT GTA GGG GGA TGG GGC AAC CCA ACG ACC CCC GAC GAG AGC GGA CTG CCG ACG ACA TTC AAA ACT AAT OCG GCG GAT GAT TGC AAA TTA	534
	Ile Val Gly Gly Trp Gly Asn Pro Thr Thr Pro Asp Glu Ser Gly Leu Pro Thr Thr Phe Lys Thr Asn Arg Ala Asp Asp Cys Lys Leu	
151	GCG GGA GGA AAC GCG GGA AAG TCA CTA GTA TTT GAT ATA GCA TGT CTA TGC ACG ACG AGT GAT TCC GCG AGC GCG TCC AAG TAT ACA TGT	624
	Ala Gly Gly Asn Gly Gly Lys Ser Leu Val Phe Asn Ser Ile Ala Cys Thr Cys Thr Ser Asp Ser Ala Ser Lys Thr Lys Thr Cys	
181	GGG CCC AAG TCA GCG GAC AAC GCG AGC GGC TGG TTA GAC AAC AAC GGG GAT AAC CAA GCG AAA OCG GCG GCG AAG GAG GCG TGG AAG AAC	714
	Gly Pro Lys Ser Gly Asp Asn Gly Ser Gly Trp Leu Asp Asn Asn Gly Asp Asn Gln Gly Lys Pro Ala Ala Lys Glu Ala Trp Lys Asn	
211	CTC CCG GCG GAC TGC CGA CGT CAA TCA GCT GGG GTT GCG GTA ACT CCG GAG CTA ATC AGC CAA TCA CTT GTT ATA TTC GAA GCG TTG ATT	804
	Leu Arg Ala Asp Cys Arg Gln Ser Ala Gln Ser Val Arg Val Thr Pro Glu Leu Ile Ser Lys Ser Leu Val Ile Phe Glu Gly Leu Ile	
241	GGA ACC CGA GCG GCA TCC GGG CAC GAC AAC GCA CGT TAC ATT TTT GGA ACC GCG ACC GCG CAA AGT TGT GGT CAT TCA ACC GCG ACA	894
	Gly Thr Arg Ala Ala Ser Gly His Asp Asn Ala Arg Tyr Ile Phe Gly Thr Val Ala Thr Ala Gln Ser Cys Gly His Ser Thr Ala Thr	
271	AAT AAG GGA TCG ATC GAC TAC AAG GCG AGC AAC GCG CAG CAA GCG GGG GAC ATA GAG TGG GAG AAG AAC CTG GCG ATG GCG GAG GGT GAC	984
	Asn Lys Gly Ser Ile Asp Tyr Lys Ala Ser Asn Ala Gln Gln Arg Gly Asp Ile Glu Trp Glu Lys Asn Leu Arg Met Ala Glu Gly Asp	
301	CTA CGA GGG CTT CTC ACT GCG AAA CAA CTC GTG GCG GCA CTT CAG GCT AGG GCT GAG CAT CTA GAA GAC GCT GCG TTC ACC ATT TTC AAC	1094
	Leu Arg Gly Leu Leu Thr Ala Lys Gln Leu Val Ala Ala Leu Cln Glu His Leu Glu Asp Ala Ala Phe Thr Thr Ile Phe Asn	
331	GAG TCT GTC TTA GAG ACT CAG ATA GCA TGG GAA TCA TCT CGT CCG OCT TCC ACT GAT GCA AAC ACT TCA CAA AAG GCG CCA CTC CAA AGG	1164
	Glu Ser Val Leu Glu Thr Gln Ile Ala Trp Glu Ser Ser Arg Pro Pro Ser Thr Asp Ala Asn Thr Ser Gln Lys Gly Pro Leu Gln Arg	
361	CCA GAA AAG TCG GGA GAA TCT TCC CAT TTA OCG TCA GGA AGT TCT CAT GGT ACT AAG GCA ATC CGA TCA ATA CTA CAC GTT GCG TTA CTT	1254
	Pro Glu Lys Ser Gly Glu Ser Ser His Leu Pro Ser Gly Ser Ser His Gly Thr Lys Ala Ile Arg Ser Ile Leu His Val Ala Leu Leu	
391	ATG TAA C A26C15 Met *	1302

YNat 1.3

	ATG CTG GAT AAC TCG CGA GCG CCG TCA ATA GTA CAT TTG TTG ATT CTA TTA AAA GCG CAT GTG ATC	66
	Met Leu Asp Asn Ser Arg Ala Arg Ser Ile Val His Leu Leu Ile Leu Leu Lys Ala His Val Ile	
1	ACT CAA ATA ATT AAA AAC ACA CAA GAG TTT ACG AGT TTG TGT ACC TTC GTA AAA GTG ACA CTT AAG GCG ACC GAT GCG CTA ACG AGC GCA	156
	Thr Gln Ile Ile Lys Asn Thr Gln Glu Phe Thr Ser Leu Cys Thr Phe Val Thr Thr Lys Ala Thr Ser Gly Thr Thr Ala	
31	GCC TCA AAA TCG CAA ACA GAT TGG GCA CTT GGA GAA AAC CCA ACG TCC AGG ATA AAA AAA TTA ATC ACT GAA TTG GAA ACA TCC TCC GAC	246
	Ala Ser Lys Ser Gln Thr Asp Trp Ala Leu Gly Glu Asn Pro Thr Ser Arg Ile Lys Lys Leu Ile Thr Glu Leu Glu Thr Ser Ser Asp	
61	CGA ATT AGG CTA GGT GAG GAG CCA AAC CTG ACG ATA CAG CTT CCA GAA GCG GAC CCG AAG CAG CCG TTA AGA CCG AAG TTA GAA GTC TTC	336
	Arg Ile Arg Leu Glu Glu Glu Pro Asn Leu Thr Ile Gln Leu Pro Gln Gly Asp Pro Lys Gln Arg Leu Arg Lys Leu Thr Thr Phe	
91	CTG CCG CGT GCA AAG TAT ACG GAG GAG TTG GTG CGA CAA GCG CAA GGC GAT GTG GCG GGA AGG TGC AAT GAG GCA AAG GCT GAA TTG GAA	426
	Leu Ala Arg Ala Lys Tyr Thr Glu Glu Leu Val Arg Gln Ala Gln Gly Asp Val Gly Gly Arg Cys Asn Glu Ala Lys Ala Glu Leu Glu	
121	GAA GCT GTG ACG GCG AGG AAG GCG CCC GAT TTG GAG ACA CAA GCG GCG GCT GCG GCT GCG CTT CAC AAC AAA GCG CGA GGT ACC GCG	516
	Glu Ala Val Thr Gly Arg Lys Gly Pro Asp Leu Glu Thr Gln Ala Thr Ala Ala Leu His Asn Lys Lys Ala Thr Thr Thr Thr Ala	
151	TGT AAG GTA GCG GCG GCG ACG GAC ACC AAT TTT GCG GGT ACC TCA CTT GTT GCG GAC CTG ATG TGT TTG TGT GCG GCG GAG ACG AAT	606
	Cys Lys Val Ala Gly Ala Thr Thr Asp Thr Asn Phe Ala Gly Thr Ser Leu Val Ala Asp Leu Met Cys Leu Cys Ala Ala Glu Thr Asn	
181	TGG AGG GAA AAG CAT ATA TGC GGA TTC GAA TCG CAT GCT TCC GGA GTA TGG GCA AAT GCT GGA ACT AAT TCC AAC GCG GCG GAG ATT TGG	696
	Ser Arg Glu Lys His Ile Cys Gly Phe Glu Ser His Ala Ser Gly Val Trp Ala Asn Ala Gly Thr Asn Ser Asn Ala Gly Glu Ile Trp	
211	GGC AAA ATC CTA GCT TGT AAA AAC CGA GAG ATC CAA GTG GAA GTT ACA CCG CAG TTC TTG AGA ATT GCG ATC ACC AAG TTC GAA GCG	786
	Gly Lys Ile Leu Asp Ala Cys Lys Asn Arg Glu Ile Gln Val Glu Thr Gln Ala Thr Pro Gln Phe Leu Arg Ile Ala Thr Lys Phe Glu Gly	
241	CTA TTG GCG GCT CAG GCA CAC AAA CTG ACC TCA AAC GCG AAC GCG GCG TGG CTG CTC GCG TAC TCG ATG AAC GCG GCG AGC GTT ACT	876
	Leu Leu Gly Ala Gln Ala His Lys Leu Thr Ser Asn Gly Asn Ala Gly Ala Trp Leu Leu Gly Tyr Ser Met Asn Ala Gly Ser Val Thr	
271	TGC GAT GCG CAG TCA AGT ACA AAC GCG ATC TGC GTC GAT TAC AAA GCG AGT AGT GAT GCA AGA GGT CCA ATT GCT TGG TTG GGA CAT ATT	966
	Cys Asp Gly Gln Ser Ser Thr Asn Gly Ile Cys Val Asp Tyr Lys Lys Ser Ser Ser Asp Ala Arg Gly Pro Ile Ala Trp Leu Gly His Ile	
301	AAA AAT GCA ATC ACT GCG CTA GAG AAT GCG GAT AAA AAT TTA CAG AGG GTC AGA AAA CTG CAG CGA CAA GCA GAG GCA ATC CTC ATG AGT	1056
	Lys Asn Ala Ile Thr Ala Leu Glu Asn Arg Asp Lys Asn Leu Gln Arg Val Arg Lys Leu Gln Arg Gln Ala Glu Ala Ile Leu Met Ser	
331	GCA GAA GAT GCT CTA ATA GAG GCA AAT ATT TCT CTA GCG GGA AAG GAT ATG GTA CCG GCG AGC GAG GTC ACG GTA CCA AAC TCT TCC AAC	1146
	Ala Glu Asp Ala Leu Ile Glu Ala Asn Ile Ser Leu Gly Gly Lys Asp Met Val Pro Ala Ser Glu Val Thr Val Pro Asn Ser Ser Asn	
361	OCT ACC TCG AGA CAA AAT TCA GTG GTC CAA GAA CCA ACC ACC GTC AGC GCG GCG GCG ATC ACG CCG CTC ATC CTG CCA TGG ACG CTC CTC	1236
	Pro Thr Ser Arg Gln Asn Ser Val Val Gln Glu Pro Thr Thr Val Ser Ala Ala Ala Ile Thr Pro Leu Ile Leu Pro Trp Thr Leu Leu	
391	ATC TAA A13C8 Ile *	1263

FIGURE 1: Nucleotide base and amino acid sequences of *T. congolense* VSGs YNat 1.1 and YNat 1.3. YNat 1.1 base sequence from cDNA clone A.09; YNat 1.3 base sequence from cDNA clone B.17. Base are numbered on the right side, and amino acid residues are numbered on the left. Number 1 on the left marks the residue for the N-terminus of mature VSG. Numbering for bases initiates at the first base of the Met codon of the respective N-terminal leader peptides. Asterisks signify a stop codon. Amino acid residues underscored with a solid line were confirmed by peptide analysis; Asn residues underscored with a dashed line were blank cycles in peptide sequence analysis and are presumed or known (YNat 1.1, Asn-351) carbohydrate addition sites. The arrow at Ser-372 marks the postulated C-terminal residue of mature YNat 1.1 VSG.

Comparison of the aligned primary amino acid sequence of YNat 1.1 and 1.3 shows that there is some sequence homology but that this is not very striking (Figure 4). There is about

25% identity of amino acid residues. There are, however, homologous blocks of residues in proximity to the cysteine residues, suggesting that our cysteine alignment is justified.

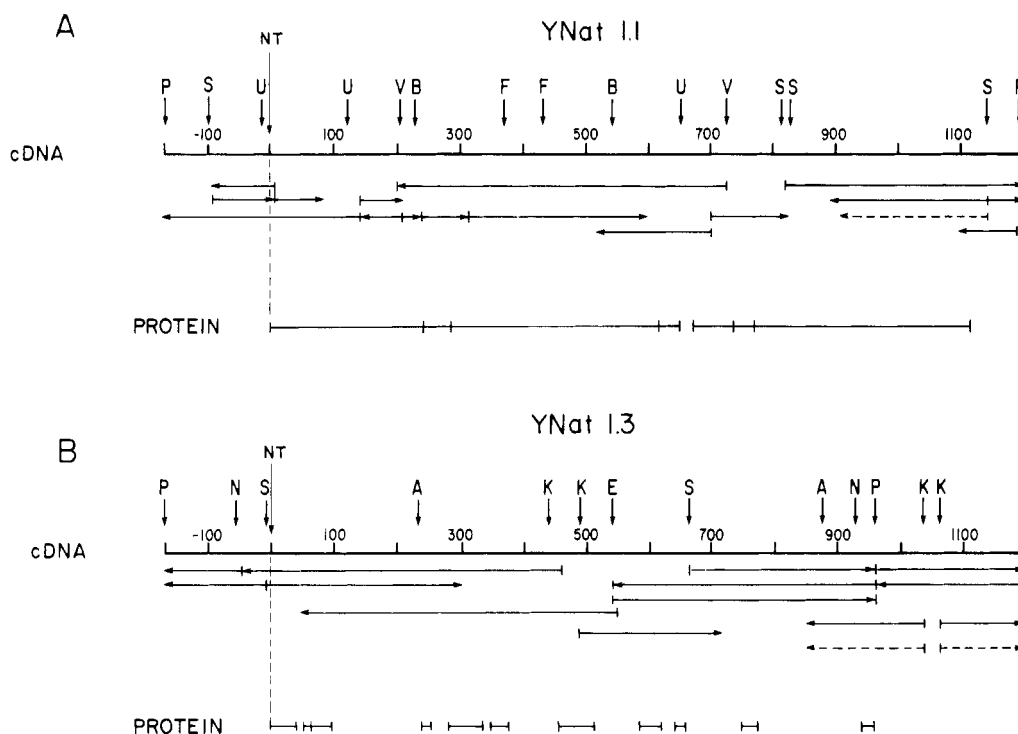


FIGURE 2: Diagram of the amino acid residue and nucleotide sequence strategy adopted for determining the primary structure of the *T. congolense* VSGs YNat 1.1 and 1.3. The cDNA lines show partial restriction maps of respective cDNA clones, and below them, solid lines with arrows indicate the extent and direction of dideoxy sequencing of M13 subclones; broken lines with arrows indicate the extent and direction of Maxam–Gilbert sequencing. The protein lines indicate the extent of overlap of all tryptic, chymotryptic, and acid protease peptides sequenced. Gaps in the peptide sequences are indicated by breaks in the line. Hatch marks on the protein line indicate where peptide information was not sufficient to give overlap information. Sequence information in all gaps and nonoverlapping regions was obtained by reference to the nucleotide base sequences. All nonoverlapping nucleotide sequencing restriction sites were in turn confirmed by peptide sequences. (A) YNat 1.1 cDNA line: restriction map of insert DNA of plasmid A.09. (B) YNat 1.3 cDNA line: restriction map of insert DNA of plasmid B.17. The C-terminal region of YNat 1.3 has been sequenced in four independent cDNA plasmid clones. Numbering of nucleotides: number 1 is the first base of the codon of the N-terminal amino acid of the respective mature VSG. Number -1 marks the third base of the last codon of the putative N-terminal leader peptide. Symbol code: NT = N-terminal amino acid residue of mature VSG; A = *Ava*II, B = *Bsp*I, E = *Eco*RI, F = *Fok*I, K = *Kpn*I, N = *Nru*I, P = *Pst*I, S = *Sau*3A, U = *Pvu*II, V = *Ava*I.

The homologous sequence blocks and single homologous residues are spread throughout the primary sequence but are somewhat more frequent in the C-terminal half of the VSG molecules, especially between residues 180 and 330 (YNat 1.1 numbering). Dot matrix homology analysis reveals a faint line of homology over the same region (180–330) (data not shown).

Comparison of amino acid sequences in *T. brucei* complex VSGs has shown the C-terminal 50 residues exhibit a class or isotypic pattern. Almost all reported sequences have strong homologies in this region, fitting them into one of two classes (Donelson & Rice-Ficht, 1985). The homology seen in the two *T. congolense* VSGs is not sufficient to discern any isotypic resemblance between the two proteins. We have used a dot matrix plotting program to locate similarities in amino acid sequences. The matrix program was set to determine matching amino acid residues in the two proteins compared and to print one dot for every match. The following *T. brucei* complex VSG sequences have been plotted against both YNat 1.3 and YNat 1.1 as well as against each other: AnTat 1.1, IoTat 1.2, ILTat 1.1, 1.2, 1.3, 1.4, and MITat 1.4. No dot matrices have been included, and a summary of our analyses follows.

When the members of the known *T. brucei* complex VSG family are compared, the expected pattern emerges in the C-terminal regions. There is strong homology within isotopes and modest homology between isotopes. In the remainder of the molecule, the degree of homology varies; some VSGs have strong similarity (AnTat 1.1 and ILTat 1.3), and others show little or no discernible similarity (AnTat 1.1 and ILTat 1.1). If, however, the whole set of *T. brucei* complex VSG is considered, there is always some regional or general similarity

among its members. In contrast, when dot matrix diagrams are plotted between either of the two *T. congolense* VSGs and each of the *T. brucei* complex VSGs, no discernible homology is found. In other words, even though the degree of similarity varies widely among *T. brucei* complex VSGs, when any given *T. brucei* VSG is compared to all others, modest or strong homology can be seen. The two *T. congolense* VSGs, in contrast, show no discernible similarity with any *T. brucei* complex VSG.

We have calculated the secondary structure potential of the *T. congolense* VSGs (Figure 5). These calculations express the potential for forming α -helical structures, β turns, and β -stranded or extended forms of the polypeptide backbone. We emphasize that we do not predict the *actual presence* of secondary structures but only the potentials for forming them. When two different proteins contain a region having identical secondary structure potential, the two proteins may or may not express a common secondary structure in that region. If, however, a region of protein has high potential for one type of structure in one region whereas in another protein a comparable region has a different potential, then it is unlikely that these two regions will show a common secondary structure.

Both YNat 1.1 and YNat 1.3 have a high percentage of α -helix potential; our programs give strongly predicted α -helical regions for approximately 60% of the primary structure. A strongly predicted region is one where the predicted secondary structure potential is at least twice as likely as the next most likely secondary structure (Lalor et al., 1984). As expected, the potential is a conservative estimate since only strongly predicted regions are included, and it is likely that

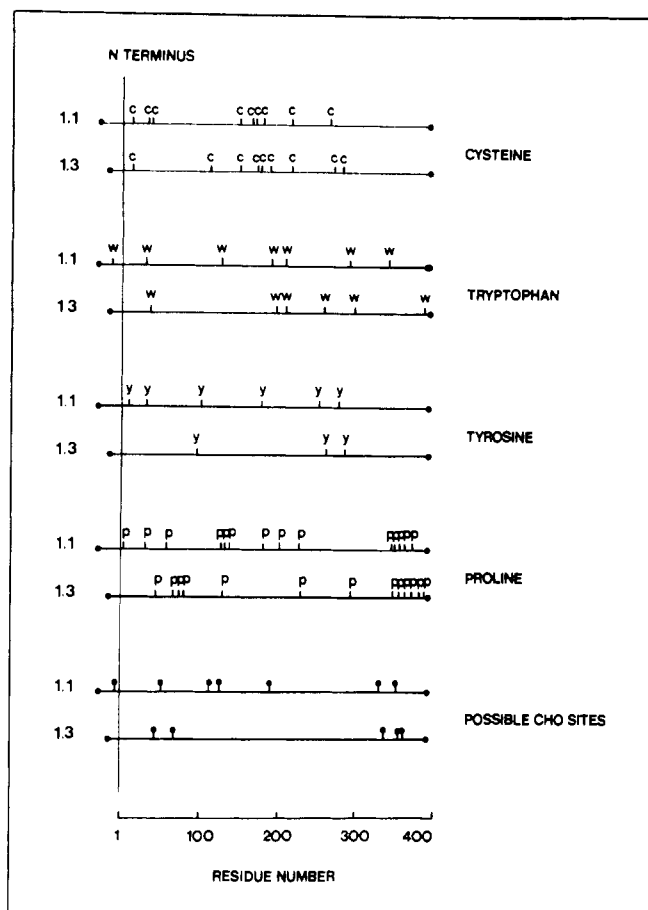


FIGURE 3: Diagram outlining the distribution of cysteine, tryptophan, tyrosine, and proline amino acid residues in the *T. congolense* VSGs YNat 1.1 and YNat 1.3. Potential carbohydrate attachment sites (N-X-S/T) are indicated by the symbol (•). Amino acid analysis indicated that Asn residues at positions 54, 330, and 351 in YNat 1.1 are modified and therefore probably carbohydrate addition sites. The VSG sequences are aligned by the N-terminal amino acid residues of the mature proteins. The amino-terminal residue of each VSG is marked by "N TERMINUS" in the figure, and its position is numbered 1 on the scale at the bottom of the figure. The letters C, W, Y, and P are single-letter amino acid code designations for cysteine, tryptophan, tyrosine, and proline, respectively.

some less strongly predicted regions will also be α helical. In line with this expectation, the average value for the α helix as calculated from circular dichroism measurements of YNat 1.1 was 65%. It is interesting to note that X-ray crystallographic studies at 6-Å resolution have revealed that the N-terminal 60% of the *T. brucei* complex VSG M.Tat 1.2 contains two long α helices and other segments with helix structures (Freyman et al., 1984).

When the locations of regions of α -helical, β -turn, or β -strand potential are compared in the two YNat proteins (Figure 5), a pattern emerges that is reminiscent of that observed when a similar comparison is made for the *T. brucei* complex VSGs (Lalor et al., 1984). In both families, there is a similar kind of diversity in the arrangement of breaks in the segments of α -helical potential, and in the locations of strong β -turn and β -strand potentials. Differences in the extent of α -helical potential between these *T. congolense* VSGs are particularly marked in the region from residue 120 to 250 (numbering as in Figure 1). Over this region, the YNat 1.1 protein contains more strongly helix-breaking proline residues (which cause the polypeptide chain to turn through a 110° angle) and more other helix-breaking residues than does the YNat 1.3 protein. Most important, there are regions where the two proteins show strongly divergent potential. In the

region centering around residue 178 in YNat 1.1, this protein has strong β -turn potential, and YNat 1.3 has strong α -helical potential. Other regions of contrasting potential center on residues 124 and 218 (YNat 1.1 numbering).

The C-terminal regions of these proteins exhibit similarities in primary structure and secondary structure potential. A cluster of strongly helix-breaking proline residues is contained in the C-terminal region of each VSG (Figure 3). Secondary structure potential in the corresponding regions (residues 340–380) is characteristic of extended, open structures (Figure 5, and unpublished results). Interestingly, some known or possible N-linked carbohydrate addition sites are located in this region (Figure 3). In each protein, the region from residue 305 to residue 340 has strong α -helical potential (Figure 5). Additionally, these two VSGs contain similar hydrophobic sequences at the extreme C-termini, reminiscent of the putative membrane-associated hydrophobic tail of the *T. brucei* complex VSGs. Thus, there is evidence for conservation of structure in the C-terminal regions of the YNat VSGs. Such conservation may indicate that these regions form structures required for important conserved functions, e.g., intermolecular VSG–VSG interaction sites, intramolecular helix packing, or domains associated with membrane attachment.

DISCUSSION

There are stringent conditions imposed on the families of structurally related proteins which mediate antigenic variation. First, a variant protein must have new antigenic determinants, and second, it must lose all old surface determinants present on previous variants. Cross-reactivity is not simply a matter of having exactly the same determinants on two proteins; it can occur with structurally dissimilar determinants and with determinants having only partial resemblance (Richards et al., 1975). In the latter instance, although the K_a of the interaction might be reduced, it could still interfere with antigenic variation. Furthermore, since the cell surface coat presents a repeating array of VSG epitopes to divalent and multivalent host antibodies, avidity effects would increase the effective K_a of antibody–surface coat interactions.

Comparison of the published *T. brucei* complex VSG primary structures has revealed the existence of a C-terminal region of approximately 25 amino acid residues which shows a class or isotypic structure. So far, two classes have been identified. There is considerable homology among members of an isotype, but even between isotypes, there is some homology. Bordering on the N-terminal end of the isotypic region, there is approximately 100 amino acid residues which typically shows less homology (Donelson & Rich-Ficht, 1985). In addition, comparison of the secondary structure potentials in this region shows considerable similarity in all *T. brucei* complex VSGs analyzed so far (Lalor et al., 1984). It has been suggested that these regions both are in contact with the membrane of the trypanosomes and are involved in the interaction which is known to exist between VSG molecules (Strickler & Patton, 1982). Such common structural requirements may account for the conserved secondary structure. The middle region of the molecule containing approximately 200 amino acid residues has many regions of sharply contrasting potential, where, for instance, there is α -helical potential in one VSG molecule and β -turn potential in another (Lalor et al., 1984). This has prompted the idea that this region mediates antigenic variation by allowing folding differences in the polypeptide backbone of VSG so that different caplike regions of the primary structure are exposed to the solvent–surface in different VSG serotypes. Thus, new antigenic determinants would appear, and old ones would dis-

1.3	ACT	CAA	ATA	ATT	AAA	AAC	ACA	CAA	GAG	TTT	ACG	AGT	TTG	TGT	ACC	TTC	GTA	AAA	GTG	ACA	-	-	-	-	-	-	CTT	AAG	GCG	ACC	GAT	25
1.1	T	Q	I	I	K	N	T	Q	E	F	T	S	L	C	T	F	V	K	V	T	-	-	-	-	-	-	L	K	A	T	D	29
	-	GAT	GCA	CCG	GTC	AAT	GCC	GCA	GAG	TAC	AAT	GCC	CTT	TGC	CGC	CTA	TAT	AAC	ATA	GCC	AGA	GCC	GGT	GAA	GGA	CTC	AAA	GAA	GAA	GAT		
1.3	GGG	CTA	ACG	AGC	GCA	GCC	TCA	AAA	TGC	CAA	ACA	GAT	TGG	GCA	CTT	GGA	GAA	AAC	OCA	ACG	TCC	AGG	ATA	AAA	AAA	TTA	ATC	ACT	GAA	TTG	55	
1.1	G	L	T	S	A	A	S	K	S	Q	T	D	W	A	L	G	E	N	P	T	S	R	I	A	K	K	L	I	T	E	L	58
	TGG	CTG	CCA	TGC	GCG	GGT	-	AAG	GCA	GCC	TGT	GAA	AAG	ACA	GCT	TCA	ATC	ATC	ATC	GAT	GTC	TTT	ATG	AAG	CTC	AAC	TTC	TCT	GAA	OCT		
1.3	GAA	ACA	TCC	TCC	GAC	CGA	ATT	AGG	CTA	GGT	-	-	GAG	GAG	OCA	AAC	CTG	ACG	ATA	CAG	CTT	CCA	CAA	GGG	GAC	CCG	AAG	CAG	GCG	TTA	83	
1.1	E	S	T	S	D	R	I	D	G	T	R	V	E	E	P	N	N	L	P	Q	L	P	K	R	G	D	V	G	Q	R	L	88
	AGC	GCA	GTT	GTG	ACC	ACT	CTT	GAC	GGG	ACT	CGC	GTC	GAG	CTG	CAA	TCC	GCG	AGC	ACT	OGA	ATC	AAG	CCG	GGG	CCG	AAG	TTG	GCC	AAG	GTA		
1.3	AGA	CGG	AAG	TTA	GAA	GTC	TTC	CTG	GCG	GGT	-	GCA	AAG	TAT	ACG	GAG	TTG	GTG	OGA	CAA	GCC	CAA	GGG	GAC	CCG	GTG	GCG	OGA	AGG	TGC	112	
1.1	L	R	R	K	L	E	V	F	I	K	A	Q	Q	L	K	Y	H	E	S	K	S	L	E	S	-	-	-	G	A	K	A	115
	TTG	GCG	GCC	GCT	GAG	ACG	ATA	AAA	GCG	CAG	CAG	CTT	AAA	TAC	CAC	GAA	TCC	TOG	AAG	AGC	CTG	TTG	GAG	AGT	-	-	-	GCA	AAG	GCG		
1.3	AAT	GAG	GCA	AAG	GCT	GAA	TTG	GAA	GAA	GCT	GTG	ACG	GCG	AGG	AAG	GCG	CCC	GAT	TTG	GAG	ACA	CAA	GCG	ACA	GCG	GCT	GCG	GCT	GCC	CTT	142	
1.1	N	E	K	K	A	I	V	G	G	W	G	N	P	T	-	T	P	D	L	E	S	G	L	P	T	-	-	-	-	-	139	
	AAC	TTC	ACG	AAG	GCG	ATT	GTA	GGG	GGA	TGG	GCC	AAC	CCA	ACG	-	ACC	CCC	GAC	GAG	AGC	GGA	CTG	CCG	ACG	ACA	-	-	-	-	-		
1.3	CAC	AAC	AAA	GCC	OGA	GGT	ACC	GCG	TGT	AAG	GTA	GCC	GGG	GCG	ACG	ACG	GAC	ACC	AAT	TTT	GCG	GCT	ACC	TCA	CTT	GTT	GCC	GAC	CTG	ATG	172	
1.1	H	N	K	A	R	G	T	D	C	K	L	A	G	G	N	G	G	K	N	F	-	-	-	-	S	L	V	F	D	I	M	164
	FTC	AAA	ACT	AAT	OGG	GCG	GAT	GAT	TGC	AAA	TTA	GCG	OGA	OGA	AAC	GCG	OGA	AAG	-	-	-	-	-	-	TCA	CTA	GTA	TTT	GAT	ATA	GCA	
1.3	TGT	TTG	TGT	GCG	GCG	GAG	ACG	AAT	TCG	-	AGG	GAA	-	AAG	CAT	ATA	TGC	OGA	TTG	GAA	TCG	CAT	GCT	-	-	TCC	GGA	GTA	TGG	-	197	
1.1	C	L	C	T	N	S	D	S	S	A	R	E	S	K	H	I	C	G	G	F	E	S	H	A	-	S	G	-	-	-	192	
	TGT	CTA	TGC	ACG	ACG	-	AGT	GAT	TCC	GCC	AGC	GGC	TCC	AAG	TAT	ACA	TGT	GGG	CCC	AAG	TCA	GGC	GAC	AAC	GGC	AGC	GGC	-	TGG	TTA		
1.3	GCA	AAT	GCT	GGA	ACT	AAT	TCC	AAC	GCC	GGG	GAG	ATT	-	-	-	TGG	GCG	AAA	ATC	CTA	GAC	GCT	TGT	AAA	AAC	OGA	GAG	ATC	CAA	GTG	224	
1.1	A	N	N	G	T	N	S	N	A	G	E	I	-	-	-	W	K	K	I	L	D	A	C	C	K	N	R	E	I	Q	V	222
	GAC	AAC	AAC	GGG	GAT	AAC	CAA	GCG	AAA	CCG	GCC	GCC	AAG	GAG	GCG	TGG	AAG	AAC	CTC	CGG	GCG	GAC	TGC	OGA	OGT	CAA	TCA	GCT	GGG	GTT		
1.3	GAA	GTT	ACA	CCC	CAG	TTC	TTG	AGA	ATT	GCC	ATC	ACC	AAG	TTC	GAA	GGG	CTA	TTG	GGG	GCT	CAG	GCA	CAC	AAA	CTG	ACC	TCA	AAC	GCG	AAC	254	
1.1	E	V	T	P	Q	F	L	R	I	A	I	T	K	F	E	G	L	L	I	G	A	Q	A	H	K	L	T	S	N	G	N	250
	CGC	GTA	ACT	CCG	GAG	CTA	ATC	AGC	CAA	TCA	CTT	GTT	ATA	TTC	GAA	GGC	TTG	ATT	OGA	ACC	CGA	GCC	GCA	TOC	GGG	CAC	GAC	-	-	-	AAC	
1.3	GGG	GGG	GGG	TGG	CTG	CTC	GGG	TAC	TCG	ATG	AAC	GCC	GGC	AGC	GTT	ACT	-	-	TGC	GAT	GGC	CAG	TCA	AGT	ACA	AAC	GGC	ATC	TGC	GTC	282	
1.1	A	G	A	W	L	L	G	Y	S	M	N	A	G	S	V	T	-	-	C	D	G	Q	S	S	T	N	G	I	C	V	275	
	GCA	GGT	-	-	-	-	-	TAC	ATT	TTT	GGA	ACC	GTG	GCC	ACC	GCC	CAA	AGT	TGT	GGT	CAT	TCA	ACG	GCC	ACA	AAT	AAG	GGA	TOG	ATC		
1.3	GAT	TAC	AAA	GCG	AGT	AGT	GAT	GCA	-	AGA	GGT	CCA	ATT	GCT	TGG	TTG	OGA	CAT	ATT	AAA	AAT	GCA	ATC	ACT	GCG	CTA	GAG	AAT	GCG	GAT	311	
1.1	D	Y	K	G	S	S	D	A	-	R	G	P	I	E	W	L	K	H	I	K	N	A	I	T	T	A	E	N	R	L	D	305
	GAC	TAC	AAG	GCC	AGC	AAC	GCC	CAG	CAA	CCC	GGG	GAC	ATA	GAG	TGG	GAG	AAG	AAC	CTG	CGC	ATG	GCC	GAG	GGT	GAC	CTA	OGA	GGG	CTT	CTC		
1.3	AAA	AAT	TTA	CAG	AGG	GTC	AGA	AAA	CTG	CAG	OGA	CAA	GCA	GAG	GCA	ATC	CTC	ATG	AGT	QCA	GAA	GAT	GCT	CTA	ATA	GAG	GCA	-	AAT	ATT	340	
1.1	K	N	L	Q	R	V	R	K	L	Q	R	Q	A	E	A	I	L	M	S	A	E	D	A	A	L	I	E	A	-	N	E	331
	ACT	GCC	AAA	CAA	CTC	GTG	GGA	GCA	CTT	CAG	GCT	AGG	GCT	GCT	AAG	GAT	-	-	-	-	GAA	GAC	GCT	GCC	TTC	ATT	TTT	AAC	GAG			
1.3	TCT	CTA	GGC	GGA	AAG	GAT	ATG	GTA	CCC	GCC	AGC	GAG	GTC	ACG	GTA	CCA	AAC	TCT	TCC	AAC	OCT	ACC	TOG	AGA	CAA	AAT	TCA	GTG	GTC	CAA	370	
1.1	S	L	G	K	D	M	V	P	A	S	T	GAG	V	T	V	P	N	S	T	S	N	P	T	S	Q	K	G	P	L	Q	R	361
	TCT	S	V	L	E	T	Q	I	A	W	E	A	S	OGT	OCG	OGT	TOC	ACT	GAT	GCA	AAC	ACT	TCA	AAG	GGG	CCA	CTC	CAA	AGG	CCA		
1.3	GAA	CCA	-	-	-	-	-	-	-	-	-	-	ACC	ACC	GTC	AGC	S	A	A	GCG	GCC	ATC	ACG	CCC	CTC	ATC	CTG	CCA	ACG	CTC	390	
1.1	E	P	-	-	-	-	-	-	-	-	-	-	T	T	V	S	A	A	K	A	A	I	R	S	-	I	L	H	V	A	L	390
	GAA	AAG	TCG	GGA	GAA	TCT	TCC	CAT	CCA	OGG	TCA	GGA	AGT	TCT	CAT	GGT	ACT	AAG	GCA	ATC	OGA	TCA	-	ATA	CTA	CAC	GTT	GGG	TTA	CTT		
1.3	ATC	TAA	AAA	AAA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	391
1.1	I	*	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	391
	ATG	TAA	CAA	AAA	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	

FIGURE 4: Alignment of amino acid and nucleotide sequences of mature YNat 1.1 and YNat 1.3 VSGs. The protein sequences were aligned with emphasis to increase homology about apparently conserved cysteine, tryptophan, and tyrosine residues. Gaps have been introduced to maximize homology. Boxed areas highlight amino acid identities in this alignment. Numbers on the right give the amino acid residue number of the last residue on that line for the respective protein sequence, where number 1 marks the N-terminal residue of the respective mature protein. Top nucleotide and amino acid sequence are of YNat 1.3, marked by 1.3; the lower sequences are of YNat 1.1, marked by 1.1.

appear (Lalor et al., 1984). The following evidence is in favor of this concept. Freymann et al. (1984) have shown in their low-resolution X-ray crystallographic map of one *T. brucei* complex VSG fragment that indeed caplike structures occur at the ends of two α -helical rods which are a part of the N-terminal 60% of MITat 1.2. Cohen and her collaborators (Cohen et al., 1984) reported the presence of heptad repeats in *T. brucei* complex VSG primary structures, suggestive of a coiled-coil structure in the α -helical stalk of the molecule.

Moreover, the breaks in the distribution of heptad repeats occurred in different primary sequence regions in different VSGs, prompting these authors to suggest, as we did, the radical folding differences might make an important contribution in the structural basis of antigenic variation. Additionally, these investigators reported that electron microscopy of two isolated VSG molecules of different serotypes showed differences in the size of caplike structures (Cohen et al., 1984). Lastly, preliminary evidence suggests that radioactive chemical

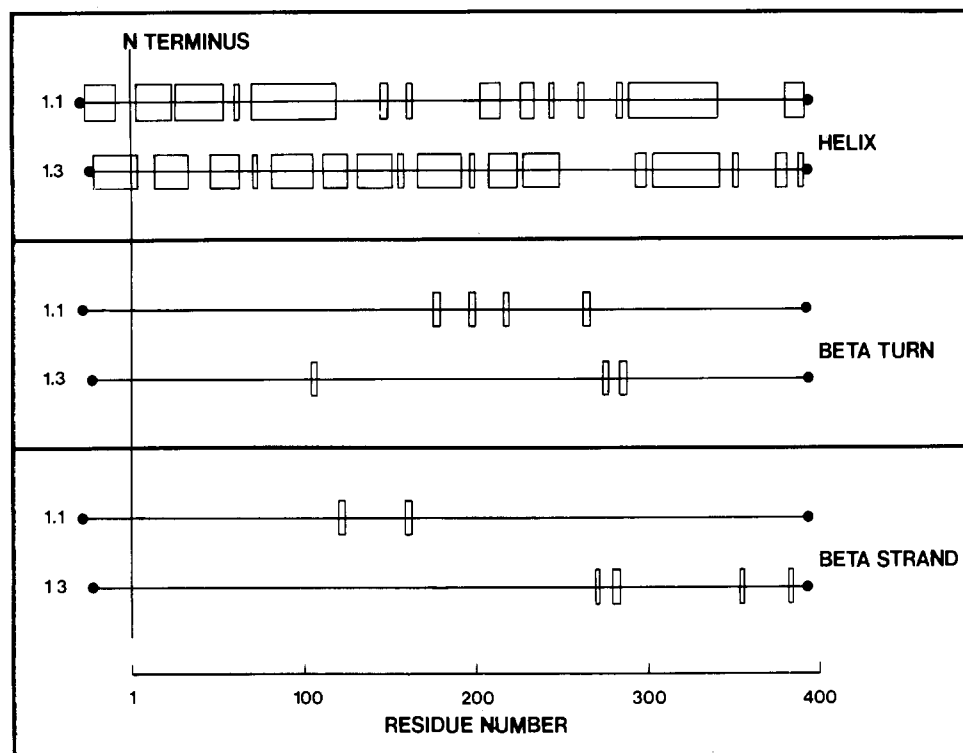


FIGURE 5: Strongly predicted secondary structure potentials of the *T. congolense* VSGs YNat 1.1 and 1.3, as calculated with the DELPHI program (Lalor et al., 1984). In a strongly predicted region, designated by open boxes, the indicated secondary structure potential is at least twice as the next most likely secondary structure. β -Turn regions were mapped only when at least four consecutive β turns calls occurred. Both VSGs are strong in α -helical potential but show differences in the distribution of strongly predicted α -helical segments. Strongly predicted β -turn and β -strand potentials vary widely in their distribution in the two proteins.

agents which modify surface lysine or tyrosine residues on intact, live trypanosomes label different regions of the primary structure in different serotypes. This labeling has been shown to occur at or close to regions where α -helical potential regions become regions of open structure potentials (manuscript submitted for publication). These studies support our idea but do not constitute definitive proof. High-resolution X-ray crystallographic studies of VSGs having different folding potential will provide the rigorous test of the multiple folding pattern hypothesis.

The question now arises whether other organisms which exhibit antigenic variation in a single surface protein, such as the nannomonad *T. congolense* or the spirochete *Borrelia hermsii*, show similar phenotypic mechanisms. This paper analyzes this question with regard to *T. congolense*. There are some morphological, physical, and physiological differences between *T. brucei* complex and *T. congolense* trypanosomes, and the two organisms were placed into a different subgenera on the basis of these differences. This suggests that there may be some evolutionary distance between trypanosomes of the *T. brucei* complex and *T. congolense*.

The results of our study show that certain of the organizational features found in the *T. brucei* complex VSGs are present in *T. congolense* VSGs. The *T. congolense* VSGs not only have a high α -helical potential but from the CD study we know that in YNat 1.1 this potential is expressed. The middle regions of the YNat molecules show the same pattern of broken α -helical potential with the breaks in different regions of the primary structure. The YNat glycoproteins also contain sequences with coiled-coil related, heptadic periodicities. Strongly predicted α -helical segments of the YNat 1.1 protein sequence which exhibit heptad repeat potential are (numbering as in Figure 1) as follows: 4–28, 30–54, 69–120, 228–244, and 292–340. Similar potential in the strongly predicted α -helical segments of YNat 1.3 sequence is found in 12–33, 47–63, 112–126, 132–158, 167–192, 208–227, and

294–342. It is interesting to note that both proteins exhibit coiled-coil potential in the N-terminal regions (within residues 1–60) and a strong potential in the C-terminal one-third of the molecules (290–342). This suggests the possibility that the N- and C-terminal regions may associate in a coiled-coil helical bundle to form a stemlike structure in analogy to the *T. brucei* complex VSGs (Cohen et al., 1984; Freyman et al., 1984). Also, the distributions of remaining coiled-coil potential in the two variants are different, particularly in the middle region. Further, the middle regions contain sharply contrasting primary structure potential, suggesting the same kind of pattern that is found in the *T. brucei* complex VSGs (Lalor et al., 1984).

The comparison of the primary structure homology shows considerable differences. The dot-matrix plots demonstrated (data not shown) that in nearly every instance there is homology between the members of the *T. brucei* complex VSG family. Sometimes homology is extensive and sometimes it is modest, but with one noted exception relatedness is always apparent. A modest degree of scattered homology (approximately 25% overall) also exists between the two *T. congolense* VSGs, and this falls in the same range which is observed between some members of the *T. brucei* complex VSG family. However, when each of the two *T. congolense* VSGs is compared in dot-matrix plots to every sequenced member of the *T. brucei* complex VSG family, there is no homology found at all even when the conditions of the comparison are sensitive to distant homologies.

Examination of the last 20 amino acids of the YNat 1.1 and YNat 1.3 variants reveals that the sequence of each variant contains tri- and tetrapeptides which also appear as consensus sequence elements in the last 17–23 amino acid residues of *T. brucei* complex VSGs. For instance, the YNat 1.3 sequence includes the tetrapeptide VSAA; the tripeptide SAA is a consensus tripeptide in isotype I of the *T. brucei* complex VSGs, and VSAA is found 8/9 times in that group (Donelson

& Rice-Ficht, 1985). The sequence TPLXL defines a consensus element among isotype II of the *T. brucei* complex VSGs (Donelson & Rice-Ficht, 1985) and is found in the same YNat 1.3 sequence as TPLIL. The YNat 1.1 C-terminal sequence VALLM is closely related to the VALLF sequence, which is found with high frequency at the C-terminus of the members of isotype I of *T. brucei* complex VSGs (Donelson & Rice-Ficht, 1985). Because of shortness of the sequences and the limited data base, the significance of short homologous peptides to the question of the evolutionary relationship between *T. congolense* and *T. brucei* complex VSGs is not clear.

The lack of homology between *T. brucei* complex and *T. congolense* VSGs is striking. These two related hemoflagellates may use variants of the same ur-protein as the antigenic variation locus. It can be argued that rapid gene evolution has produced so many mutations that these VSGs no longer resemble each other. A second possibility is that each of the two parasitic hemoflagellates originally has adopted a different gene family to code for its respective VSG polygene family. The similar secondary structure in the face of a different primary structure may then be the result of the common evolutionary pressure leading to antigenic variation. According to this view, independent adaptations, using different surface proteins as sources of parasite-covering variant-specific glycoproteins, have occurred. However, the limited sequence data base does not allow us to distinguish between these two possibilities.

As noted above, the last 20 or so residues of YNat 1.1 and 1.3, as predicted from the respective cDNA clones, are similar to the hydrophobic tail sequences of *T. brucei* complex VSGs. The tail sequence is cleaved from most, if not all, mature *T. brucei* complex VSGs, and a glycopospholipid structure is added to the new C-terminal amino acid through an amide linkage (Ferguson et al., 1985). Given the similarities at the extreme C-termini of the two sets of VSGs, it likely that similar posttranslational modifications may occur with the *T. congolense* VSGs. Such glycopospholipid structures are thought to mediate membrane attachment of cell surface molecules and are found in various eukaryotes (Tse et al., 1985).

We see no common isotype structure in the C-terminal region of the two *T. congolense* VSGs, although there is more homology in the C-terminal region than elsewhere in the two molecules. However, the analysis of two structures is not sufficient to draw any conclusions about class structure. It is interesting to note that Majiwa et al. (1985) found that the VSG cDNA probe from *T. congolense* ILNat 2.1 did not hybridize with genomic DNA of strains from geographically separated *T. congolense* serodemes, suggesting a possible absence of a conserved isotypic C-terminal region.

That the YNat cDNA clones appear to contain little or no 3' untranslated region (UTR) raises a number of questions. Assuming that the cDNA clones accurately reflect the structure of the respective mRNA molecules, the lack of a substantial 3' UTR represents a marked difference from other eukaryotic mRNAs in general and, in particular, from the *T. brucei* complex VSG mRNAs. In that case, the 3' UTRs, are approximately 75 bp in length, though variable, and contain considerable sequence conservations, including two consensus sequences (Borst & Cross, 1982). It was suggested (Borst & Cross, 1982) that the conserved sequences could serve an important function, such as in recombinational events in VSG gene switching. Thus, the lack of a 3' UTR in the *T. congolense* VSG mRNAs may indicate a difference in the nature and location of nucleic acid structural elements involved in *T. congolense* VSG duplication/transposition events. Our studies on the genetic rearrangements in the *T. congolense*

VSG system will be the subject of another publication (manuscript in preparation).

The location and nature of the poly(A) addition signal sequences for the *T. brucei* complex mRNAs have not yet been determined with certainty, though the sequence AAAATTPyT has been proposed (Hasan et al., 1984). The apparent lack of a 3' UTR in the YNat VSG messages may imply that the poly(A) addition signal sequences are located in the coding region of the VSGs; however, in other more complicated schemes, the signal sequences would map to portions of the messenger precursors which are processed off at the addition of the poly(A) tail.

ACKNOWLEDGMENTS

We thank Dr. Joseph Coleman and Dr. Robert Yu for their valued advice and assistance. Dr. John Newel kindly made computing facilities available to us. We thank Dr. Morten Kjeldgaard for his advice and the DELPHI computer program. We thank Louise Camera for providing her excellent technical skills and Charlene Sullivan for bringing order to chaos.

Registry No. DNA (*Trypanosoma congolense* variant surface glycoprotein YNat 1.1 messenger RNA complementary), 106160-76-3; glycoprotein YNat 1.1 (*Trypanosoma congolense* variant surface precursor protein moiety reduced), 106160-78-5; glycoprotein YNat 1.1 (*Trypanosoma congolense* variant surface protein moiety reduced), 106160-79-6; DNA (*Trypanosoma congolense* variant surface glycoprotein YNat 1.3 messenger RNA complementary), 106160-77-4; glycoprotein YNat 1.3 (*Trypanosoma congolense* variant surface precursor protein moiety reduced), 106160-80-9; glycoprotein YNat 1.3 (*Trypanosoma congolense* variant surface protein moiety reduced), 106160-81-0.

REFERENCES

- Auffray, C., & Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
- Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80, 3963-3965.
- Binder, D. A. (1984) M.D. Thesis, Yale University School of Medicine.
- Bogucki, M. S., Onodera, M., Rosen, N. L., Lifter, J., Kohnsberg, W. H., & Richards, F. F. (1982) *Exp. Parasitol.* 53, 1-10.
- Borst, P., & Cross, G. A. M. (1982) *Cell (Cambridge, Mass.)* 29, 291-303.
- Bwaye, J. J., & Hirumi, H. (1980) *Proc. 2nd Annu. Med. Conf. Kenya Med. Res. Inst. Kenya Trypanosomiasis Res. Inst. (Nairobi)*.
- Chou, P. Y., & Fasman, G. D. (1974a) *Biochemistry* 13, 211-222.
- Chou, P. Y., & Fasman, G. D. (1974b) *Biochemistry* 13, 222-245.
- Chou, P. Y., & Fasman, G. D. (1978) *Annu. Rev. Biochem.* 47, 251-276.
- Cohen, C., Reinhardt, B., Parry, D. A., Roelants, G. E., Hirsch, W., & Kanwe, B. (1984) *Nature (London)* 311, 169-171.
- Cook, G. A., Honigberg, B. M., & Zimmermann, R. A. (1985) *Mol. Biochem. Parasitol.* 15, 281-293.
- Donelson, J. E., & Rice-Ficht, A. C. (1985) *Microbiol. Rev.* 49, 107-125.
- Ferguson, M. A. J., Haldar, K., & Cross, G. A. M. (1985) *J. Biol. Chem.* 260, 4963-4968.
- Freyman, D. M., Metcalf, P., Turner, M., & Wiley, D. C. (1984) *Nature (London)* 311, 167-169.
- Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol.* 120, 97-120.

- Greenfield, N., & Fasman, G. D. (1969) *Biochemistry* 8, 4108-4111.
- Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 3961-3965.
- Hasan, G., Turner, M. J., & Cordingley, J. S. (1984) *Cell (Cambridge, Mass.)* 37, 333-341.
- Hoeijmakers, J. H. J., Borst, P., Van der Burg, J., Weissmann, C., & Cross, G. A. M. (1980) *Gene* 8, 391-417.
- Hong, G. F. (1981) *Biosci. Rep.* 1, 243-252.
- Jackson, P. R., Honigberg, B. M., & Holt, S. C. (1978) *J. Protozool.* 25, 471-481.
- Lalor, T. M., Kjeldgaard, M., Shimamoto, G. T., Strickler, J. E., Konigsberg, W. H., & Richards, F. F. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 998-1002.
- Lanham, S. M., & Godfrey, D. G. (1970) *Exp. Parasitol.* 28, 521-534.
- Laursen, R. A. (1971) *Eur. J. Biochem.* 20, 89-102.
- L'Italien, J. J., & Strickler, J. E. (1982) *Anal. Biochem.* 127, 198-212.
- Majiwa, P. A. O., Matthyssens, G., Williams, R. O., & Hamers, R. (1985) *Mol. Biochem. Parasitol.* 16, 97-108.
- Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning (A Laboratory Manual)*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
- Messing, J. (1983) *Methods Enzymol.* 101, 20-78.
- Messing, J., & Vieira, J. (1982) *Gene* 19, 269-276.
- Novotny, J. (1982) *Nucleic Acids Res.* 10, 127-131.
- Novotny, J., & Auffray, C. (1984) *Nucleic Acids Res.* 12, 243-255.
- Onodera, M., Rosen, N. L., Lifter, J., Hotez, P., Konigsberg, W. H., & Richards, F. F. (1981) *Exp. Parasitol.* 52, 427-433.
- Rautenberg, P., Reinwald, E., & Risse, H. J. (1980) *Parasitology* 80, 113-122.
- Rice-Ficht, A. C., Chen, K. K., & Donelson, J. E. (1981) *Nature (London)* 294, 53-57.
- Richards, F. F., Rosenstein, R. W., Konigsberg, W. H., & Varga, J. M. (1975) *Science (Washington, D.C.)* 187, 130-134.
- Rosen, N. L., Onodera, M., Patton, C. L., Lipman, M., & Richards, F. F. (1978) *Exp. Parasitol.* 47, 378-383.
- Rosen, N. L., Onodera, M., Bogucki, M. S., Cross, G. A. M., Patton, C. L., Konigsberg, W. H., & Richards, F. F. (1981) *Exp. Parasitol.* 52, 210-215.
- Rowcamp, W., & Firtel, R. A. (1980) *Dev. Biol.* 79, 409-418.
- Sanger, F., Wicklen, S., & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
- Silverman, M., Zieg, J., Hillman, M., & Simon, M. (1979) *Proc. Natl. Acad. Sci. U.S.A.* 76, 391-396.
- Stoenner, H. G., Dodd, T., & Larsen, C. (1982) *J. Exp. Med.* 156, 1297-1311.
- Strickler, J. E., & Patton, C. L. (1982) *Mol. Biochem. Parasitol.* 5, 117-131.
- Strickler, J. E., Mancini, P. E., & Patton, C. L. (1978) *Exp. Parasitol.* 46, 262-276.
- Tschudi, C., Young, A. S., Ruben, L., Patton, C. L., & Richards, F. F. (1985) *Proc. Natl. Acad. Sci. U.S.A.* 82, 3998-4002.
- Tse, A. G. D., Barclay, A. N., Watts, A., & Williams, A. F. (1985) *Science (Washington, D.C.)* 230, 1003-1008.

The M_r 93 000 Polypeptide of the Postsynaptic Glycine Receptor Complex Is a Peripheral Membrane Protein[†]

Bertram Schmitt, Petra Knaus, Cord-Michael Becker, and Heinrich Betz*

ZMBH, Zentrum für Molekulare Biologie, Universität Heidelberg, 6900 Heidelberg, Federal Republic of Germany

Received August 1, 1986; Revised Manuscript Received October 9, 1986

ABSTRACT: The glycine receptor of mammalian spinal cord is an oligomeric membrane protein that, after affinity purification on aminostrychnine-agarose or immobilized antibody, contains three polypeptides of M_r 48 000, 58 000, and 93 000. Here, the association and the properties of the polypeptides of the rat glycine receptor were investigated. Upon phase partitioning in the nonionic detergent Triton X-114, the three receptor polypeptides behaved as a hydrophilic protein complex exhibiting phospholipid binding. Sucrose gradient centrifugation or gel filtration in the presence of dithiothreitol and Triton X-100 separated the M_r 93 000 polypeptide from the M_r 48 000 and 58 000 polypeptides, which harbor the antagonist binding site of the glycine receptor. Alkaline or dimethylmaleic acid anhydride treatment of crude synaptic membrane fractions resulted in extraction of the M_r 93 000 polypeptide. Lectin binding was observed for the M_r 48 000 and 58 000 glycine receptor subunits but not the M_r 93 000 polypeptide. These results indicate that the M_r 93 000 polypeptide is a peripheral membrane protein that is located at the cytoplasmic face of the postsynaptic glycine receptor complex.

Glycine is a major inhibitory neurotransmitter in spinal cord and other regions of the central nervous system (Werman et al., 1967; Curtis et al., 1968; Aprison & Daly, 1978). Binding of this amino acid to the postsynaptic glycine receptor (GlyR)¹

increases the chloride conductance of the neuronal membrane and thus produces a hyperpolarization, i.e., inhibition of the postsynaptic neuron. The convulsive alkaloid strychnine an-

[†] This work was supported by the Deutsche Forschungsgemeinschaft, the Bundesministerium für Forschung und Technologie, and the Fonds der Chemischen Industrie. C.-M.B. is a fellow of the Boehringer Ingelheim Foundation.

¹ Abbreviations: DMMA, dimethylmaleic acid anhydride; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GlyR, glycine receptor; HRP, horseradish peroxidase; IgG, immunoglobulin G; mAb, monoclonal antibody; SDS, sodium dodecyl sulfate; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.