Nucleic Acids, Protein Synthesis, and Molecular Genetics:

High Diversity in Mucin Genes and Mucin Molecules in *Trypanosoma cruzi*

Javier M. Di Noia, Guido D. Pollevick, Marcia T. Xavier, José O. Previato, Lucia Mendoça-Previato, Daniel O. Sánchez and Alberto C. C. Frasch

J. Biol. Chem. 1996, 271:32078-32083.
doi: 10.1074/jbc.271.50.32078

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 50, Issue of December 13, pp. 32078–32083, 1996

Printed in U.S.A.

High Diversity in Mucin Genes and Mucin Molecules in *Trypanosoma cruzi**

(Received for publication, July 24, 1996)

Javier M. Di Noia‡, Guido D. Pollevick‡, Marcia T. Xavier§, José O. Previato§, Lucia Mendoça-Previato§, Daniel O. Sánchez‡, and Alberto C. C. Frasch‡¶

From the ‡Instituto de Investigaciones Bioquímicas, "Fundación Campomar," Antonio Machado 151, and Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, 1405 Buenos Aires, Argentina and the §Departamento de Microbiología Geral, Universidade Federal do Rio de Janeiro, 21944970 Cidade Universitária, Rio de Janeiro, RJ, Brasil

Mucins are highly *O*-glycosylated molecules which in mammalian cells accomplish essential functions, like cytoprotection and cell-cell interactions. In the protozoan parasite *Trypanosoma cruzi*, mucin-related glycoproteins have been shown to play a relevant role in the interaction with and invasion of host cells. We have previously reported a family of mucin-like genes in *T. cruzi* whose overall structure resembled that of mammalian mucin genes. We have now analyzed the relationship between these genes and mucin proteins. A monoclonal antibody specific for a mucin sugar epitope and a polyclonal serum directed to peptide epitopes in a MUC gene-encoded recombinant protein, detected identical bands in three out of seven strains of *T. cruzi*. Immunoprecipitation experiments confirmed these results. When expressed in eukaryotic cells, the MUC gene product is post-translationally modified, most likely, through extensive *O*-glycosylation. Gene sequencing showed that the central domains encoding the repeated sequences with the consensus T_8KP_2, varies in number from 1 to 10, and the number of Thr residues in each repeat could be 7, 8, or 10. A run of 16 to 18 Thr residues was present in some, but not all, MUC gene-derived sequences. Direct compositional analysis of mucin core proteins showed that Thr residues are much more frequent than Ser residues. The same fact occurs in MUC gene-derived protein sequences. Molecular mass determinations of the 35-kDa glycoproteins further extend the heterogeneity of the family to the natural mucin molecules. Difficulties in assigning each of the several MUC genes identified to a mucin product arise from the high diversity and partial sequence conservation of the members of this family.

Mucins are highly *O*-glycosylated molecules composed of up to 80% of carbohydrate attached to Thr and Ser residues of the core protein (1). Although sequence homology is low between different apomucin genes, the proteins they encode share a common structure with N and C termini domains flanking a variable number of tandem repeats (2). Repeat-containing domains are rich in Thr and Ser residues and have at least one Pro residue per repeat unit (3).

Highly *O*-glycosylated mucin-type molecules of *Trypanosoma cruzi*, the agent of the American trypanosomiasis (Chagas' disease), are suggested to be involved in the interaction of the parasite with and invasion of host mammalian cells (4, 5). Sialic acid (SA)[1] in the parasite mucins and in the host cell has also been implicated in these processes (reviewed in Refs. 6 and 7). *T. cruzi* is unable to synthesize SA, but it expresses trans-sialidase (TS), an enzyme that transfers SA from sialylated molecules in the host to surface parasite acceptors containing terminal β-galactosyl residues (reviewed in Ref. 8). All parasite SA acceptors described so far in the TS reaction are glycoconjugates containing a Thr/Ser- and Pro-rich protein core (9–11). The main SA acceptors in epimastigotes (insect vector replicative form of *T. cruzi*) and metacyclic forms (insect-derived infective stage) are mucin-type, glycophosphatidylinositol (GPI)-anchored, glycoproteins migrating in the 35–50-kDa range depending on the parasite strain (10, 12, 13). Mucins of both developmental stages have different biological properties even though they are recognized by the same mAbs (12). mAbs directed to parasite mucins, as well as the purified glycoconjugates, inhibit entry of metacyclic forms into host cells (14). Mucin carbohydrate structure is the same for both stages within the same strain, but they differ in the lipid nature of the GPI anchor (15). In the trypomastigote form (the cell-derived infective stage in the mammalian host), glycoproteins sharing the stage-specific epitope Ssp-3, which is dependent on parasite sialylation, are the main SA acceptors (8, 11). They were also shown to be of mucin-type and GPI-anchored to the parasite membrane (5). These molecules migrate as a broad band of 60–200 kDa on SDS-PAGE and are involved in cell adhesion and invasion too. While oligosaccharides from epimastigotes and metacyclics mucins are similar even among different strains, they clearly differ from those Ssp-3-containing mucins of trypomastigotes, even in the same strain (5, 15). The carbohydrate moieties of *T. cruzi* mucin-like glycoconjugates were found to be *O*-linked to threonine-rich peptides whose sequences are unknown. No studies beyond amino acid composition were done on the core proteins of these molecules.

We have recently identified in *T. cruzi* genes whose overall structure resembles those coding for mammalian mucins (16,

* This work was supported by grants from the UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (TDR), the Department for Research Cooperation (SAREC) from the Swedish International Development Cooperation Agency (Sida), the University of Buenos Aires, Argentina and Fundación Antorchas, Argentina and Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (PADCT/RHAE), Financiadora de Estudos e Projetos, and Conselho de Ensino e Pesquisa da Universidade Federal do Rio de Janeiro (CEPG-UFRJ). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U59482, U61532, U61533, and U62530.

¶ To whom correspondence should be addressed. Tel.: 54-1-863-4011; Fax: 54-1-865-2246.

[1] The abbreviations used are: SA, sialic acid; mAb, monoclonal antibody; GPI, glycophosphatidylinositol; TS, trans-sialidase; PAGE, polyacrylamide gel electrophoresis; MALDI-TOF-MS, Matrix-assisted laser desorption ionization/time of flight mass spectrometry.

17). *T. cruzi* mucin-like genes are characterized by a highly polymorphic central domain flanked by conserved 5′ and 3′ regions. Central domains are rich in codons for Thr and Pro residues organized in tandems of variable number of repeat units coding for the consensus sequence T_8KP_2. Other genes have central domains rich in codons for Thr, Ser, and Pro residues that are not organized in tandem repeats. Although these results point to the existence of a highly variable family of these genes in *T. cruzi*, there has been so far no biochemical association between the identified gene family and mucin-type glycoproteins. In this paper we present evidence that the products encoded by the putative mucin gene family are indeed the core proteins of *T. cruzi* mucin-type glycoproteins, and that diversity among these genes and glycoproteins is greater than previously thought.

MATERIALS AND METHODS

Parasites—*T. cruzi* epimastigotes were grown in brain heart infusion (BHT) medium (18). To purify mucin glycoproteins, BHT was supplemented with 10 mg/ml hemin and used without fetal calf serum, as described previously (13). Trypomastigotes were grown in Vero cell culture (19). Spontaneously differentiated metacyclic forms were purified using a DEAE-cellulose column (20). Strains used in this study were: RA (21), Y (22), CL-Brener (23), G (24), Tulahuén (24), cloned stock SylvioX-10/7 (kindly given by Dr. J Dvorak, National Institutes of Health), cloned stock CA-1/72 (25), Tulahuén 0 (obtained from Instituto Nacional de Chagas, "Dr. M. Fatala Chabén," Buenos Aires, Argentina), Tulahuén M (kindly given by Dr. J. Swindle, University of Tennessee).

Expression of Mucin Genes in Bacterial and Eukaryotic Cells—The repeat-containing and C terminus domains of MUC.M/76 gene (17) were fused in frame at the C terminus of SAPA repeats (26) and cloned in the appropriate reading frame for expression in two vectors, the pAT eukaryotic expression vector and pET-25b(+) (Novagen) *Escherichia coli* expression systems. The pAT vector is derived from pKG5 (a gift from Dr. Keith Gould) and contains a 250-base pair fragment encoding the peptide signal and 39 amino acids from the N-terminal of the mature antithrombin III (27). The vector pKG5 contains the pBR322 replication origin, the ampicillin resistance gene, the neomycin resistance gene expressed from the thymidine kinase promoter, and a polylinker downstream of the SV40 early promoter. Both chimeras were introduced in Vero cells by the calcium phosphate procedure (28) or in *E. coli* by calcium chloride method, respectively (29). In the case of transfected Vero cells, after 10 days of drug selection, neomycin-resistant colonies were pooled and seeded in 75-cm^2 flasks. For Western blot experiments, cells were grown on 10-cm plates.

Western Blotting—Parasites were resuspended in lysis buffer (150 mM NaCl, 1% Nonidet P-40, 50 mM Tris-HCl, pH 7.5, in the presence of 50 μM E64 and 1 mM phenylmethylsulfonyl fluoride), at a density of 10^6 parasites/μl. Lysates were extracted for 30 min on ice and then mixed with reducing cracking buffer before loading on gels. About 5×10^7 epimastigotes or 10^8 cell-derived or metacyclic trypomastigotes were loaded on each lane. Vero cells were scraped from culture plates (about 6×10^6 cells/plate) and directly dissolved in reducing cracking buffer with DNase (100 μg/ml). In all cases, conventional SDS-PAGE was performed (30) on 10% polyacrylamide gels, then transferred to nitrocellulose filters (Life Technologies, Inc.), reacted with the appropriate antisera, and developed with ^{125}I-Protein A (DuPont NEN).

Antibodies—mAb 10D8 is a monoclonal antibody directed to a carbohydrate epitope on the 35–50-kDa parasite mucins, kindly provided by Dr. Nobuko Yoshida (12, 14). Anti-MUC.M/76 is a polyclonal serum raised in rabbit against purified recombinant MUC.M/76 (17) expressed in pGEX-1 (Pharmacia Biotech Inc.). Anti-SAPA is a polyclonal serum raised in rabbit against the repeat-containing region of the SAPA protein (26) expressed in pGEX-1 (Pharmacia).

Immunoprecipitation—Epimastigotes of the CA-1/72 cloned stock of *T. cruzi* were resuspended in lysis buffer at a final density of 10^8 parasites/ml and incubated for 30 min on ice. The lysate was then clarified for 10 min at 10,000 × *g*, and aliquots of 1 ml were incubated for 2 h with 1/50 dilution of mAb 10D8. Immunoconjugates were precipitated using Protein A-agarose beads (Life Technologies, Inc.) following standard protocols (31) and analyzed by Western blot.

DNA Cloning and Sequencing—Polymerase chain reaction was performed using 100 ng of total genomic DNA of Y, CL-Brener, Tulahuén M, or Tulahuén 0 as template with oligonucleotides P1 and P2 as described previously (17). The resulting products were cloned in *Eco*RV-digested pBluescript KS(+) II (Stratagene) or pT7Blue (Novagen) and sequenced using synthetic primers by the dideoxy chain termination method (32) with a Sequenase 2.0 kit (U. S. Biochemicals).

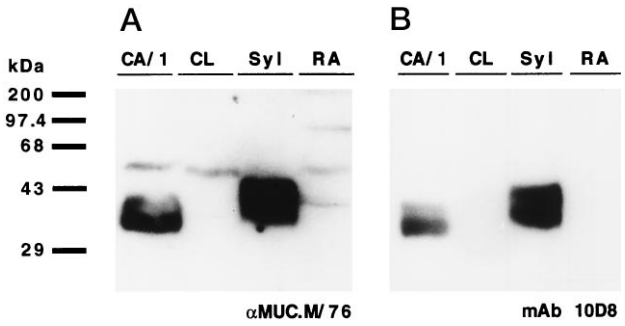

FIG. 1. **Western blot of total lysates from *T. cruzi* strains or cloned stocks.** About 5×10^7 epimastigotes were loaded in each lane. *A*, probed with anti-MUC.M/76 antisera. *B*, probed with mAb 10D8. *Lanes CA/1*, CA-1/72 cloned stock; *CL*, CL-Brener cloned stock; *Syl*, SylvioX-10/7 cloned stock; *RA*, RA strain. Molecular mass markers (Life Technologies, Inc.) are indicated on the *left*.

Purification of Mucin-like Glycoproteins from Y, G, CL, and Tulahuén Strains of T. cruzi—Frozen epimastigotes forms, grown without fetal calf serum, were thawed, extracted with cold water, and centrifuged, and the resulting pellet was extracted with 45% (v/v) aqueous phenol at 75 °C. The aqueous layer was dialyzed, freeze-dried, dissolved in water, and applied to a column (2 × 100 cm) of Bio-Gel P-100 (100–200 mesh). The excluded material was lyophilized, extracted three times with chloroform, methanol, water (10:10:3, v/v/v), and then boiled in 80% aqueous ethanol. The resulting insoluble product was dissolved in 10 mM sodium phosphate buffer (pH 7.2), containing 0.9% NaCl, and applied to a concanavalin A-Sepharose 4B column (1.5 × 30 cm) equilibrated with the same buffer. The bound material was eluted with 0.1 M methyl-α-D-mannopyranoside followed by chromatography on a Bio-Gel P-6 DG column (2.5 × 100 cm) (33). The mucin-like glycoproteins were analyzed by SDS-PAGE using a 15% acrylamide gel and stained by the periodate Schiff reagent (34).

Chemical and Physical Analysis of the Purified Mucin-like Glycoproteins—Mucin-like glycoproteins were analyzed for neutral sugars, *N*-acetylated amino sugars, phosphorus, and lipids, as described previously (9). Amino acid analysis of mucin-like glycoproteins was determined after acid hydrolysis in 5.6 N HCl, 110 °C, 22 h, in an autoanalyzer by the method of Fauconnet and Rochemont (35). Matrix-assisted laser desorption ionization/time of flight mass spectrometry (MALDI/TOF-MS) was performed to mucin-like glycoproteins of Y, CL, G, and Tulahuén strains. The positive and negative ion MALDI/TOF mass spectra were obtained for solution (2 μg/μl H_2O/CH_3CN, 7:3) of the mucin-like glycoprotein samples, diluted 1:10 in a solution (10 mg/ml H_2O/CH_3CN, 7:3) of 2,5-dihydroxybenzoic acid. One μl of this solution was deposed on the target and evaporated before analysis. Spectra were recorded with a Finnigan Mat Vision 2000 time of flight mass spectrometer equipped with a viewing unit, in the linear mode with an accelerating voltage of 30 kV. Laser was pulsed nitrogen laser (337 nm) with an irradiation time of about 3 ns.

RESULTS

The MUC-like Gene Family of T. cruzi Encodes Mucin-type Glycoproteins—In a previous work we have described a putative mucin gene family, which we called MUC-like genes in analogy with MUC genes from mucins in vertebrates (17). To identify the products of these genes we have now used an antiserum directed against a recombinant protein (anti-MUC.M/76) and a monoclonal antibody (mAb 10D8) specific for a carbohydrate epitope present in *T. cruzi* mucins (see "Materials and Methods"). Western blots of *T. cruzi* cell extracts from different parasite strains, as well as from the epimastigote, trypomastigote, and metacyclic forms, from a single parasite cloned stock, were probed with both antisera. Three out of seven *T. cruzi* strains or cloned stocks analyzed (CA-1/72, SylvioX-10/7, and Tulahuén 0) gave positive signals with both antibodies (Fig. 1 and not shown). The main signals obtained with anti-MUC.M/76 correspond with the bands detected by

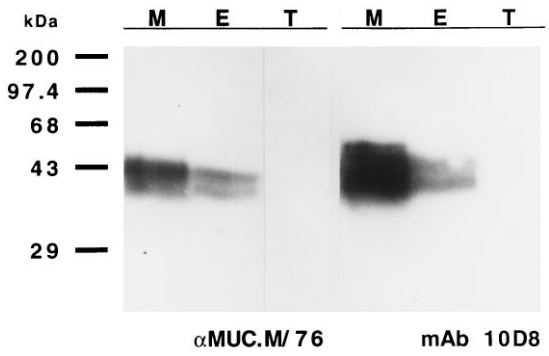

FIG. 2. **Western blot of different stages of *T. cruzi*.** Total lysates of SylvioX10/7 cloned stock probed with anti-MUC.M/76 (*left panel*) and mAb 10D8 (*right panel*). *M*, metacyclic stage (10^8 cells); *E*, epimastigote stage (5×10^7 cells); *T*, cell-derived trypomastigote stage (10^8 cells). Molecular mass markers (Life Technologies, Inc.) are indicated on the *left*.

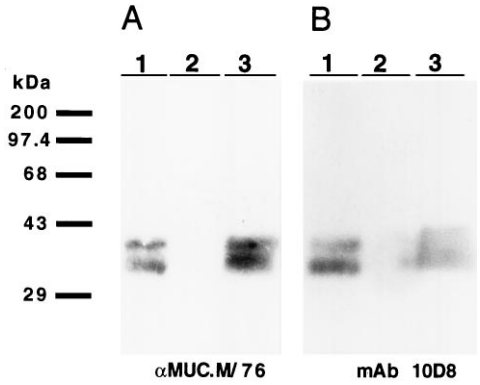

FIG. 3. **mAb 10D8 immunoprecipitation of total lysate from epimastigotes of CA-1/72 cloned stock of *T. cruzi*.** mAb 10D8 immunoprecipitated material corresponding to 10^8 cells was revealed with ^{125}I-Protein A, after SDS-PAGE separation and Western blotting. *A*, probed with anti-MUC.M/76 antisera. *B*, probed with mAb 10D8. *Lanes 1*, immunoprecipitated with mAb 10D8; *lanes 2*, immunoprecipitated with mouse normal serum; *lanes 3*, CA-1/72 epimastigotes total lysate. Molecular mass markers (Life Technologies, Inc.) are indicated on the *left*.

mAb 10D8. These broad bands are all in the range from 35 to 50 kDa, with small differences between strains. Both antibodies failed to detect 35–50 kDa bands in strains Y, CL-Brener, and RA (Fig. 1 and not shown). Strain Tulahuén M was the only case in which there was a positive signal with anti-MUC.M/76, but not with the mAb 10D8 (not shown). In these experiments, filters probed with anti-MUC.M/76 were exposed 10 times longer than those probed with mAb 10D8 to achieve the same signal intensities. Longer exposures of Western blots probed with anti-peptide antibodies are expected since most amino acid epitopes in mucins are shielded by their *O*-linked oligosaccharides.

Yoshida and co-workers (12) have previously shown that mAb 10D8 recognize the same 35–50-kDa mucin bands in epimastigote and metacyclic stages of *T. cruzi* and that these molecules are absent in bloodstream trypomastigote stage. So, we probed these three stages of the parasite with mAb 10D8 and anti-MUC.M/76. As can be seen in Fig. 2, both antibodies detected the same bands in extracts obtained from epimastigotes and metacyclic forms of the SylvioX/10 strain, but not in cell-derived trypomastigote extracts (Fig. 2).

A further indication that antibodies directed to the carbohydrate and protein epitopes recognize the same *T. cruzi* molecules was obtained from immunoprecipitation experiments. A total parasite extract of epimastigotes from the cloned stock CA-1/72 was immunoprecipitated with mAb 10D8. The immunoprecipitates were run in SDS-PAGE, transferred to nitrocellulose filters, and probed with anti-MUC.M/76. mAb 10D8-immunoprecipitated molecules were recognized with anti-MUC.M/76 serum. (Fig. 3*A*, *lane 1*). In a control experiment, mAb 10D8 was used to probe the Western blots of the immunoprecipitated material (Fig. 3*B*). In both cases, the immunoprecipitated bands coincided with those detected in total lysate (Fig. 3, *lanes 3*).

We conclude from these experiments that 10D8, a mAb specific for a carbohydrate epitope in a biochemically characterized group of *T. cruzi* mucins, and anti-MUC.M/76, a serum directed to a protein encoded by a MUC-like gene family member, recognize the same parasite glycoconjugates.

Expression of a Gene Encoding a Mucin from T. cruzi in Eukaryotic Cells Renders a Glycosylated Product—As further evidence that MUC-like genes encode potentially glycosylatable proteins, we expressed MUC.M/76 gene in eukaryotic cells. Two DNA constructs were made, one for control expression in *E. coli* and the other for expression in Vero cells. Both constructs contained an open reading frame with the Thr/Pro-rich amino acid repeat domain and the C terminus from the putative *T. cruzi* mucin gene MUC.M/76. These sequences are preceded by an in frame molecular tag made up of the highly antigenic amino acid repeats (SAPA) from the trans-sialidase gene (26). This tag was included to allow detection, in short times of exposure, of the expressed product in Vero cells.

Three different experimental results showed that the proteins expressed in Vero cells were modified as compared with the one expressed from the same DNA sequence in *E. coli*. First, antibodies directed to the tag detected the expected protein product both in bacteria and in Vero cells (Fig. 4*B*, *lanes 2* and *4*). However, the product in Vero cells has a considerably larger size and runs as a diffuse band, as expected for a highly glycosylated protein. Second, antiserum against the recombinant MUC.M/76 protein detected the protein expressed in bacterial extracts but not the one in transfected Vero cells (Fig. 4*A*, *lanes 2* and *4*). This result is expected if the protein epitopes are not accessible to the antibodies. Third, the protein expressed in Vero cells, but not the one in *E. coli*, remains in the aqueous phase after phenol extraction (Fig. 5), a method commonly used to isolate mucins. Altogether, these results strongly suggest that, in mammalian cells, this *T. cruzi* mucin core protein is highly glycosylated.

Diversity in Mucin Genes from Different Strains of T. cruzi—We have previously shown the existence of two different types of mucin genes in different strains of *T. cruzi* (17). One group whose central domain is composed of repetitive elements and a second group with nonrepetitive central domains. Interestingly, all those strains which were shown to have genes with repeats (from Southern blot or sequence data, Ref. 17 and not shown) were recognized in Western blots by mAb 10D8 and anti-MUC.M/76. On the contrary, those strains which do not seem to have repeat-containing mucin genes were negative in Western blots. To investigate whether there was a correlation between having Thr/Pro-rich repeats and a specific glycosylation pattern containing the 10D8 epitope, we further studied genes from some strains that did not react with both sera. Taking advantage of the conserved nature of the 5′- and 3′-noncoding regions of this gene family, we used primers annealing to those regions in polymerase chain reactions with genomic DNA as template. We identified and sequenced open reading frames from Y, Tulahuén M, CL-Brener, and Tulahuén 0 strains. Fig. 6 shows a comparative scheme of the proteins deduced from these genes with their relevant features. As can be seen, they do contain amino acid repeats rich in Thr and Pro. Consequently, we might exclude that detection of the same

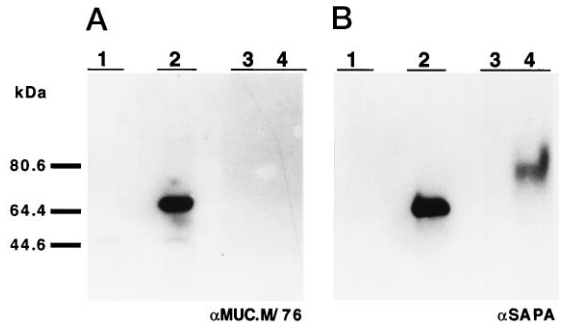

FIG. 4. **Expression of a gene encoding a putative mucin from *T. cruzi* in bacterial and eukaryotic cells.** Total lysates of *E. coli* and Vero cells were separated in SDS-PAGE, transferred to nitrocellulose and, after probing with different antisera, revealed with ^{125}I-Protein A. *A*, probed with anti MUC.M/76 serum. *B*, probed with anti-SAPA serum. *Lanes 1–4*, total lysates from: *lane 1*, *E. coli*; *lane 2*, MUC.M/76-SAPA transformed *E. coli*; *lane 3,* 6×10^6 Vero cells; *lane 4*, MUC.M/76-SAPA transfected 6×10^6 Vero cells.

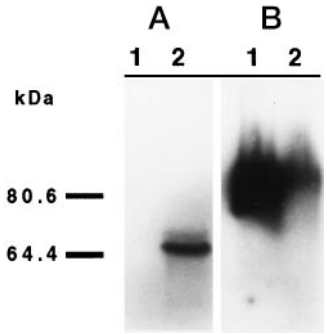

FIG. 5. **Phenolic extraction of MUC.M/76 expressed in bacterial and eukaryotic cells.** *A*, *E. coli* expressing recombinant MUC.M/76-SAPA (*lane 1*, aqueous phase of phenolic extraction corresponding to 10 times more cells than *lane 2*; *lane 2,* total lysate). *B*, 6×10^6 Vero cells expressing MUC.M/76-SAPA (*lane 1*, aqueous phase of phenolic extraction corresponding to 5 times more cells than *lane 2*; *lane 2*, total lysate). Both panels were probed with anti-SAPA.

parasite products by mAb 10D8 and anti-MUC.M/76 can be due to any correspondence between repeat-containing genes and the 10D8 epitope. Nevertheless, except for Tulahuén 0 gene (MUC.T0–1) which is very similar to MUC.M/76 gene, they have striking differences which make them unique. All isolated genes have a central region coding for a high number of potentially *O*-glycosylatable threonine residues organized in tandems of repeats, flanked by conserved 5′ and 3′ regions. Y strain gene (MUC.Y-1) has three repeat units in its central domain differing in the number of Thr residues (Fig. 6), at variance with most of the repeat-containing genes sequenced so far which have the consensus T_8KP_2. After the repeat units, MUC.Y-1 has a run of 16 Thr residues, that is not present in any previous sequences reported so far. Tulahuén M gene (MUC.TM-1) has a single canonical repeat, followed by a run of 18 Thr and 1 Ser. Finally, the gene from CL-Brener (MUC.CL-1) has one repeat with 7 Thr residues, one with 8 Thr, and a run of 17 Thr interrupted by 1 Ala. It is clear from these results that mucin gene diversity among *T. cruzi* strains is really much more considerable than described before (17).

Diversity in Mucin-type Glycoproteins from Different Strains of T. cruzi—Western blot probing of some strains, looking for the mAb 10D8 epitope, showed the occurrence of nonreacting strains (see above). In order to obtain further information about the nature and diversity of mucin products in these nonreacting strains, we analyzed purified mucin-type glycoproteins by MALDI/TOF-MS. We also determined the amino acid and sugar composition of these glycoproteins purified from Y and G strains, whose oligosaccharides and TS acceptor nature have been already described (9, 13), as well as CL-Brener and Tulahuén strains. MALDI/TOF-MS analysis of mucin-like glycoproteins indicated the presence of two major populations of highly glycosylated molecules for each strain. Their molecular mass distributions are centered around 13.4 and 14.9, 16.2 and 18.0, 15.9 and 17.5, and 15.8 and 17.4 kDa for Tulahuén, G, CL, and Y strains, respectively. In addition, minor populations having larger and smaller molecular mass were also detected. On SDS-PAGE analysis, these glycoproteins gave the broad bands that migrated in the range of 35–45 kDa (see Fig. 1), which were stained by periodate Schiff reagent, but not by Coomassie Brilliant Blue. The compositional analysis of these mucin-like glycoproteins indicated that their content in carbohydrate and protein was in the range of 61–67% and 24–28%, respectively. The apparent mass discrepancies of *T. cruzi* mucins revealed between MALDI/TOF-MS and SDS-PAGE are almost certainly due to the different hydrodynamic properties of highly glycosylated materials and protein molecular mass markers. Besides, the MALDI/TOF-MS values are consistent with the stoichiometry of the number of sugars and amino acid residues present in the mucin preparations, calculated in relation to 1.0 mol of *myo*-inositol per mol of mucin. This value was taken because *T. cruzi* mucins are GPI-anchored, and this cyclitol is present as a single unit in all GPI-anchored glycoproteins investigated to date (37). The monosaccharide composition observed for purified mucin-like molecules was Gal, Man, GlcNAc, Glc, SA, GalNAc, and Ins in an approximate molar ratio of 28.0, 15.0, 10.0, 8.0, 0.6, 1.0, and 1.0 for CL strain and 23.0, 10.0, 13.0, 14.0, 0.2, 0.1, and 1.0 for Tulahuén strain. These results are slightly different from those described for G and Y strains (9, 13).

Table I shows the amino acid composition of several mucins isolated from different strains of *T. cruzi*. The similarities among the composition of the major amino acids of all these preparations suggest that these glycoconjugates belong to the same group of molecules. Amino acid analysis of mucins isolated from G, CL, Y, and Tulahuén epimastigote forms of *T. cruzi*, indicates that approximately 50% of residues are Thr and only 1% are Ser, showing that Thr is almost the exclusive target for *O*-glycosylation. This finding is consistent with the deduced amino acid sequence of MUC.M/76 and other MUC gene family members, where Thr residues are much more frequent than Ser residues.

DISCUSSION

Mucins are believed to play essential functions in metazoa, like cytoprotection of epithelia against chemical products and microorganisms. In recent years it became evident that mucins also play important roles in cell-cell interactions in multicellular organisms (reviewed in Ref. 38). An example is the adhesive phenomena occurring between rolling leukocytes and the endothelial cells in the vascular surface during inflammation. In this process, adhesion molecules with lectin activity, termed selectins, interact with carbohydrate ligands present on mucin-like molecules (39). Unicellular parasites interact with host cells when invading multicellular organisms, and it is becoming evident that mucins from the host and/or the parasite are involved in these interactions. For example, the galactose-specific adherence protein from *Entamoeba histolytica* allows the parasite to adhere to human colonic mucins and epithelial cells (40).

In the case of the intracellular parasite *T. cruzi*, it has been known for years that it produces a number of molecules rich in carbohydrates. It has been suggested recently that these glycoconjugates, that might be involved in the interaction with and/or invasion of mammalian host cells, are of mucin-type (8).

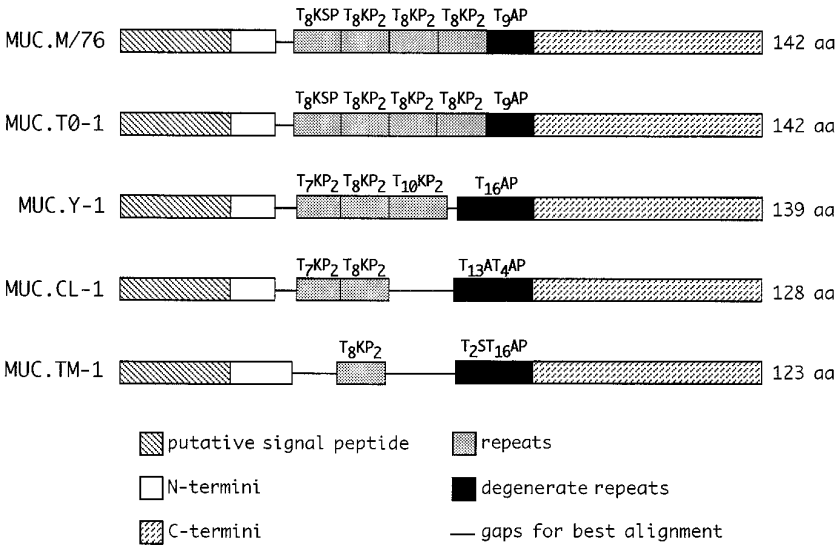

FIG. 6. **Schematic scale representation of proteins derived from newly isolated MUC genes of *T. cruzi*.** Sequence of canonical and degenerate repeats are written above their corresponding *boxes* as well as for long Thr runs. Nucleotide sequences were submitted to GenBank™/EBI Data Bank under the accession numbers indicated in the footnote.

TABLE I

Amino acid composition of mucin-like proteins isolated from different strains and stages of T. cruzi

Amino acid content is indicated in number of residues per 100 residues. Above each lane is indicated the name given to the described molecule and, in parentheses, the reference where amino acid composition was determined.

Amino acid	*T. cruzi* strains								
	Y and G[a] (F2/3 (Ref. 5))	G[b] (35/50 antigens (Ref. 10))	Y[c] (MUC.Y-1)	CA1[c] (MUC.CA-2 (Ref. 17))	RA[c] (MUC.RA-2 (Ref. 17))	Y[d] (Mucin-type protein)	Tul[d] (Mucin-type protein)	CL[d] (Mucin-type protein)	G[d] (Mucin-type protein)
Thr	20.8	31.0	54.5	56.1	16.0	45.9	50.5	53.8	50.3
Asp/Asn	9.2	11.2	4.5	3.5	8.8	11.2	12.9	10.0	12.8
Ala	8.6	9.9	5.7	4.4	8.9	10.1	10.6	8.7	10.3
Pro	9.6	7.6	11.4	14.0	9.5	8.4	6.3	7.8	9.3
Ser	9.6	7.1	4.6	4.4	10.7	1.1	0.9	0.9	0.8
Glu/Gln	13.8	6.8	5.7	3.5	11.9	6.9	5.7	5.4	5.2
Lys	5.1	6.7	3.4	5.3	4.1	4.4	2.1	0.6	0.6
Gly	8.7	6.4	0.0	0.0	8.9	8.5	7.8	6.8	7.1
Arg	2.6	3.5	4.6	2.6	3.6	0.4	0.0	0.3	0.0
Val	3.1	2.6	1.1	2.6	7.7	1.6	2.3	3.2	2.3
Met	0.0	1.9	0.0	0.0	0.6	0.0	0.0	0.0	0.0
Ile	1.9	1.6	3.4	2.6	4.1	0.7	0.3	0.8	0.4
Tyr	0.7	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Leu	3.8	1.5	0.0	0.0	2.4	0.5	0.5	0.5	0.5
His	0.9	0.4	0.0	0.0	1.8	0.3	0.0	0.5	0.4
Phe	1.6	0.4	0.0	0.0	0.6	0.0	0.0	0.0	0.0

[a] Cell culture trypomastigotes.
[b] Metacyclic trypomastigotes.
[c] Composition deduced from the gene sequence (putative signal and GPI anchor sequences were withdrawn for calculations).
[d] Culture epimastigotes.

T. cruzi mucins present in epimastigote and metacyclic forms have been studied by different groups in several strains and named differently. It is now evident that molecules described as: A, B, and C (41), GP24/GP31/GP37 (42), 35–50-kDa antigens (14), or 38–43/40–45-kDa sialoglycoproteins (9, 13) are probably all members of the same family.

The chemical structures of *O*-linked oligosaccharides of mucins isolated from G and Y strain epimastigote forms have recently been elucidated (9, 13). They are dissimilar. The substitution of GlcNAc by both Gal*f* and Gal*p* units observed in the G strain does not occur in the Y strain, nor in the CL strain.[2] In the Y and CL strains, the major *O*-GlcNAc-linked oligosaccharides have a core structure containing GlcNAc units substituted on O-4 positions by Gal*p* residues. These findings agree with Y and CL strain mucin-like glycoproteins being nonreactive with the mAb 10D8, specific for the carbohydrate moiety of G strain mucins (see Fig. 1). These mucins also show some differences in their amino acid composition (see Table I) reflecting the heterogeneity of mucin core proteins.

The relationship we have obtained between this mucin family and MUC gene products is based on different lines of evidence. The monoclonal antibody 10D8 that specifically recognizes the carbohydrate moiety of the 35–50-kDa group of mucins and a polyclonal serum to a recombinant protein encoded in MUC.M/76 gene detected the same *T. cruzi* products in Western blots of total lysate. Since the 35–50-kDa mucins might be composed of different glycoproteins, it is always possible to argue that both antibodies recognized different products within the same molecular size range. However, the almost perfect correlation between the presence or absence of products with both antibodies in the different developmental stages and strains of the parasite and the immunoprecipitation experiments, strongly suggests that this is not the case. As some degree of heterogeneity was demonstrated for epimastigote mucins by amino acid composition and molecular mass determinations, it is likely that even though all of them share the 10D8 epitope, only some of them were codified by the MUC genes identified before. Given the method we used to clone MUC genes, we are biasing isolation toward those genes with

[2] A. R. Todeschini, J. O. Previato, C. Jones, R. Wait, and L. Mendoça-Previato, unpublished results.

both 5′ and 3′ regions highly conserved, so we cannot exclude the existence of other less conserved genes encoding other members of these glycoprotein groups. When introduced into mammalian cells, the MUC.M/76 gene product acquired some properties expected from a mucin molecule. The MUC.M/76 gene deduced protein has only one *N*-glycosylation site. This single glycosylation event could not explain the aqueous phase partition of the modified product after phenolic extraction. So it can be assumed that, when expressed in Vero cells, *T. cruzi* mucins are *O*-glycosylated. Thus, we may conclude that the MUC gene family encodes some proteins belonging to the group of 35–50-kDa mucin-type glycoproteins of metacyclic and epimastigote forms of *T. cruzi*.

Interestingly, the SAPA amino acid repeats, whose consensus sequence is DSSAHGTPSTPV, have 64 potential *O*-glycosylation sites, as predicted by the algorithm of Hansen *et al.* (36). However, unlike MUC.M/76 region, the SAPA region of the fusion is nonglycosylated since it is recognized by its specific antibodies (see Fig. 4) and, when expressed alone in Vero cells, runs as a sharp band of the expected apparent molecular weight.[2]

We have previously identified two groups of genes according to the sequence of their central domains. One group contains tandem repeats and the second does not (17). The results presented in this work show that the heterogeneity among strains in terms of mucin genes is much larger than previously thought. In fact, genes obtained from three strains are different and also differ from the two groups previously identified. Thus, it is likely that different parasite stocks might have their own set of mucin genes. It should be mentioned that *T. cruzi* populations are essentially clonal in origin and that the parasites duplicate by binary fission with very little, if any, sexual interchange of genetic material during its life cycle (43). The genes reported in this paper have novel structural characteristics, like the presence of repeats with different lengths in the same gene and the long runs of threonine residues. This fact could also explain, at least partially, the differences found in amino acid composition of mucins in different parasite strains (see Table I).

The estimated molecular masses of the peptide domain of *T. cruzi* mucins isolated from epimastigote forms of Y, G, CL, and Tulahuén strains are in the range of 3.7 to 4.1 kDa while those from epimastigotes of the Peru strain and metacyclic forms of G strain are 4.0 and 7.0, respectively (44, 12). The molecular masses of protein domains deduced from MUC-like gene sequences, after trimming putative signal and GPI sequences, range from 7.4 to 17.1 kDa. Future work will be necessary to establish whether these differences are due to further protein processing.

Identification of the core proteins of the mucins present in *T. cruzi* will contribute to analysis of the variety of molecules and their possible biological relevance in the different developmental stages and strains of the parasite. Since the number of genes obtained points to a large diversity in mucins, it would now be possible to know whether this is related or not with differences in glycosylation patterns and with different functions. The developmental stages and strains of the parasite might use different mucins for a number of functions, like adhesion to and/or invasion of mammalian cells, cellular tropism phenomena, protection against lysis by complement, attachment/detachment from epithelial cells of insect host gut to permit differentiation. Further work along these lines of research will establish the correlation between individual genes and isolated glycoproteins, the possible function of each of the parasite mucins, and the identification of the cell receptors for *T. cruzi* glycoconjugates.

Acknowledgments—We thank Yves Leroy and Colette Brassart at the Laboratoire de Chimie Biologique, UMR 111 du CNRS, USTL, Lille, France (Director André Verbert), for MALDI/TOF-MS analysis, Dr. J. J. Cazzulo for critical reading, and Liliana Sferco for helpful technical assistance. We thank Dr. N. Yoshida for providing us with the mAb 10D8.

REFERENCES

1. Devine, P. L., and McKenzie, I. F. C. (1992) *BioEssays* **14,** 619–625
2. Gum, J. R., Jr. (1992) *Am. J. Respir. Cell. Mol. Biol.* **7,** 557–564
3. Gum, J. R., Jr. (1995) *Biochem. Soc. Trans.* **23,** 795–799
4. Ruiz, R. C., Rigoni, V. L., Gonzales, J., and Yoshida, N. (1993) *Parasite Immunol. (Oxf.)* **15,** 121–125
5. Almeida, I. C., Ferguson, M. A. J., Schenkman, S., and Travassos, L. R. (1994) *Biochem. J.* **304,** 793–802
6. Schenkman, S., and Eichinger, D. (1993) *Parasitol. Today* **9,** 218–222
7. Cazzulo, J. J., and Frasch, A. C. C. (1992) *FASEB J.* **6,** 3259–3264
8. Schenkman, S., Eichinger, D., Pereira, M. E. A., and Nussenzweig, V. (1994) *Annu. Rev. Microbiol.* **48,** 499–523
9. Previato, J. O., Jones, C., Xavier, M. T., Wait, R., Travassos, L. R., Parodi, A. J., and Mendoça-Previato, L. (1995) *J. Biol. Chem.* **270,** 7241–7250
10. Schenkman, S., Ferguson, M. A. J., Heise, N., Cardoso de Almeida, M. L., Mortara, R. A., and Yoshida, N. (1993) *Mol. Biochem. Parasitol.* **59,** 293–304
11. Schenkman, S., Jiang, M.-S., Hart, G. W., and Nussenzweig, V. (1991) *Cell* **65,** 1117–1125
12. Mortara, R. A., da Silva, S., Araguth, M. F., Blanco, S. A., and Yoshida, N. (1992) *Infect. Immun.* **60,** 4673–4678
13. Previato, J. O., Jones, C., Gonçalves, L. P. B., Wait, R., Travassos, L. R., and Mendoça-Previato, L. (1994) *Biochem. J.* **301,** 151–159
14. Yoshida, N., Mortara, R., Araguth, M. F., Gonzales, J. C., and Russo, M. (1989) *Infect. Immun.* **57,** 1663–1667
15. Serrano, A. A., Schenkman, S., Yoshida, N., Mehlert, A., Richardson, J. M., and Ferguson, M. A. J. (1995) *J. Biol. Chem.* **270,** 27244–27253
16. Reyes, M. B., Pollevick, G. D., and Frasch, A. C. C. (1994) *Gene (Amst.)* **140,** 139–140
17. Di Noia, J. M., Sánchez, D. O., and Frasch, A. C. C. (1995) *J. Biol. Chem.* **270,** 24146–24149
18. Cazzulo, J. J., Franke de Cazzulo, B. M., Engel, J. C., and Cannata J. J. B. (1985) *Mol. Biochem. Parasitol.* **16,** 329–343
19. Zingales, B., Katzin, A. M., Arruda, M. V., and Colli, W. (1985) *Mol. Biochem. Parasitol.* **16,** 21–34
20. De Sousa, M. A. (1983) *Mem. Inst. Oswaldo Cruz Rio J.* **83,** 123–133
21. González Cappa, S. M., Katzin, A. M., Añasco, N., and Lajmanovich, S. (1981) *Medicina (Buenos Aires)* **41,** 549–555
22. Silva, L. H. P., and Nussenzweig, V. (1953) *Folia Clin. Biol.* **20,** 191–203
23. Brener, Z., and Chiari, E. (1963) *Rev. Inst. Med. Trop. Sao Paulo* **5,** 220–224
24. Carreira, J. O., Jones, C., Wait, R., Previato, J. O., and Mendoça-Previato, L. (1996) *Glycoconj. J.* **13,** in press
25. González Cappa, S. M., Chiale, P., Del Prado, G. E., Katzin, A. M., De Martini, G. W., Isola, E. D., Abramo Orrego, L., and Segura, E. L. (1980) *Medicina (Buenos Aires)* **40,** 63–68
26. Pollevick, G. O., Affranchino, J. L., Frasch, A. C. C., and Sánchez, D. O. (1991) *Mol. Biochem. Parasitol.* **47,** 247–250
27. Prochownik, E. V., Markham, A. F., and Orkin, S. H. (1983) *J. Biol. Chem.* **258,** 8389–8394
28. Graham, F. L., and Van Der Eb, A. J. (1973) *Virology* **52,** 456–467
29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
30. Laemmli, U. K. (1970) *Nature* **227,** 680–685
31. Harlow, E., and Lane, D. (1988) *Antibodies: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
32. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
33. Previato, J. O., Andrade, A. F. B., Pessolani, M. C. M., and Mendoça-Previato, L., (1985) *Mol. Biochem. Parasitol.* **16,** 85–96
34. Fairbanks, G., Stek, T. L., and Wallach, D. F. H. (1971) *Biochemistry* **10,** 2606–2617
35. Fauconnet, M., and Rochemont, J. (1978) *Anal. Biochem.* **91,** 403–409
36. Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J. O., Hansen, J. E. S., and Brunak, S. (1995) *Biochem. J.* **308,** 801–813
37. McConville, M. J., and Ferguson, M. A. J. (1993) *Biochem. J.* **294,** 305–324
38. Van Klinken, B. J., Dekker, J., Buller, H. A., and Einerhand, A. W. (1995) *Am. J. Physiol.* **269,** 913–627
39. McEver, R. P., Moore, K. L., and Cummings, R. D. (1995) *J. Biol. Chem.* **270,** 11025–11028
40. Ravdin, J. I., Shain, D. C., and Kelsall, B. L. (1993) *Vaccine* **11,** 241–246
41. Alves, M. J. M., and Colli, W. (1975) *FEBS Lett.* **52,** 188–190
42. Ferguson, M. A. J., Snary, D., and Allen, A. K. (1985) *Biochim. Biophys. Acta* **842,** 39–44
43. Tibayrenc, M., and Ayala, J. M. (1987) *Parasitol. Today* **3,** 189–190
44. Singh, B. N., Lucas, J. J., Beach, D. H., and Costello, C. E. (1994) *J. Biol. Chem.* **269,** 21972–21982