The Protozoan Trypanosoma cruzi Has a Family of Genes Resembling the Mucin Genes of Mammalian Cells*

(Received for publication, May 31, 1995, and in revised form, August 3, 1995)

Javier M. Di Noia†, Daniel O. Sánchez‡, and Alberto C. C. Frasch§

From the Instituto de Investigaciones Bioquímicas, “Fundación Campomar,” Antonio Machado 151, 1405 Buenos Aires, Argentina and Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina

*This work was supported by grants from the United Nations Development Program/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases; the International Atomic Energy Agency (Vienna, Austria); the Ministerio de Cultura y Educación de la República Argentina; the Universidad de Buenos Aires, Argentina; and Fundación Antorchas, Argentina. 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.

†Researchers from the Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

‡Researchers from the Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina.

§To whom correspondence should be addressed: Instituto de Investigaciones Bioquímicas Fundación Campomar, Antonio Machado 151, 1405 Buenos Aires, Argentina. Tel.: 54-1-865-4011 to 4019; Fax: 54-1-865-2246; E-mail: dsanchez@iris.lib.uba.ar.

Mucins are heavily O-glycosylated Thr/Ser/Pro-rich molecules. Given their relevant functions, mucins and their genes have been mainly studied in higher eukaryotes. In the protozoan parasite Trypanosoma cruzi, mucin-like glycoproteins were shown to play an important role in the interaction with the surface of the mammalian cell during the invasion process. We show now that this parasite has a family of putative mucin genes, whose organization resembles the one present in mammalian cells. Different parasite isolates have different sets of genes, as defined by their central domain. Central domains, rich in codons for Thr and/or Ser and Pro residues, are made up of either a variable number of repeat units in tandem or non-repetitive sequences. Conversely, 5'- and 3'-ends from different genes in different isolates have similar sequences, suggesting their common origin. Comparison of deduced amino acid sequences revealed that all members of the family have the same putative signal peptide on the N terminus and a putative sequence for glycosylphosphatidylinositol anchoring on the C terminus. The deduced molecular mass of the core proteins is small (from 17 to 21 kDa), in agreement with the 1-kilobase size of the mRNA detected. Putative mucin genes in *T. cruzi* are located on large chromosomal bands of about 1.6–2.2 megabase pairs.

Mucins are highly glycosylated proteins expressed by most secretory epithelial tissues in vertebrates. They consist of a core protein moiety where a number of carbohydrate chains are attached to serines and threonines by α-1–3 O-glycosidic bonds (1). The complex structure of these glycoproteins made the identification of genes encoding the protein moiety more difficult. However, several MUC-like genes have been isolated recently due to the fact that they have a defined basic structure and sequence, which allows their inclusion in a gene family. MUC-like genes in vertebrates are essentially composed of a central domain and 5'- and 3'-flanking sequences (2, 3). The central domains, comprising up to 70% of the coding sequences, are composed by tandemly repeated units enriched in codons for Ser and Thr, which are the target sites for O-glycosylation in the protein product, as well as Pro residues (4). Sequences flanking the central domain, on the 5'- and 3'-ends of mucins genes, lack repeated sequences. The percentage of amino acid identities among different mucin core proteins are low. No substantial identities were found among the repeats in different molecules. They are unique in size and sequence for each member of the mucin family, even though they contain many Ser and Thr residues, suggesting that their only function is to serve as a scaffold for O-linked glycans (3). Furthermore, different individuals have a variable number of repeated units in homologous core proteins, making the loci coding for mucins highly polymorphic among individuals (5, 6). Partial sequence identities were found in defined regions of the N and C termini. For example, significant identities were observed between the deduced amino acid sequence from MUC2 and putative MUC5 human mucins (7), the porcine and bovine submaxillary mucins (8, 9), and the cysteine-rich C-terminal regions of rat intestinal mucin-like and human MUC2 peptides (10–12).

Genes encoding molecules that have mucin-like features in lower eukaryotes have been detected in Leishmania major (13) and Trypanosoma cruzi (14). Particularly in *T. cruzi*, the ethiological agent of Chagas disease, much work has been done on the biochemical and functional characterization of mucin-like surface glycoconjugates (15). These heavily O-glycosylated molecules are Thr-, Ser-, and Pro-rich and are attached to membrane by glycosylphosphatidylinositol anchor (16). Mucins in *T. cruzi* are the major acceptors of sialic acid in a reaction catalyzed by trans-sialidase (15, 16). Recent evidence suggests that these molecules are involved in the cell invasion process, probably mediating adhesion of the parasite to the mammalian cell surface (17, 18). We have previously identified a putative mucin gene in *T. cruzi* (14) having a small size and encoding five repeat units with the consensus sequence TgKP2. In this work, we show that *T. cruzi* has, in fact, a putative mucin gene family resembling the one present in vertebrate cells. Their members have a Thr/Ser/Pro-rich central domain, which might or might be not organized in repetitive units, and highly conserved non-repetitive flanking domains.

** Experimental Procedures

Parasites—*T. cruzi* epimastigotes (insect vector replicative stage) were grown in liquid medium (19). Trypomastigotes (vertebrate bloodstream infective stage), were grown in Vero cell culture (20). Strains used in this study were RA (21), CanII, NIH2, Y, CL-Brener (Ref. 22 and references therein), SylvioX-10/7, and the CA1/72 cloned stock (23).

Oligonucleotide Sequences—Sequences are as follows: P1, 5'-CCAT- GTCCTCAGTCAGTAG-3'; P2, 5'-ACATCGGACCACCGTGAAG-3'.
Putative mucin gene (14). Theduced sequences of the two other clones were almost identical to those of MUC.CA-3 (see below). All three MUC.CA-deduced

sequences showed a central domain made up of tandemly repeated units and non-repetitive flanking domains on the N and C termini. The repetitive units were very similar in all clones

RESULTS AND DISCUSSION

A Family of Putative Mucin Genes Differing in the Number of Repetitive Units—Southern blot analysis of T. cruzi DNA probed with a gene having a mucin-like structure revealed several bands (Fig. 1, panel 2), suggesting the presence of more than one gene in the T. cruzi genome. To study this gene family, a cloned stock of T. cruzi (CA1/72) was used. 12 positive clones were identified by hybridization with a probe containing the nucleotide tandem repeats ("repeats" probe) present in a putative mucin gene (14). Three groups of clones were detected according to the size of the inserts. One clone of each group was selected for sequencing and named MUC.CA-1, MUC.CA-2, and MUC.CA-3. The deduced amino acid sequence from one of them (MUC.CA-3) is shown in Fig. 2 and compared with the sequence deduced from the gene previously isolated from the Miranda/76 clone (14). The deduced sequences of the two other clones (MUC.CA-1 and MUC.CA-2) were almost identical to those of MUC.CA-3 (see below). All three MUC.CA-deduced

...
repeat) for MUC.CA-1, -2, and -3, respectively. The non-repetitive N- and C-terminal domains were almost identical in all three MUC.CA clones and very similar to the flanking sequences in the original Miranda/76 clone (Fig. 2). Southern blot experiments using restriction endonuclease enzymes that trim sequences in the original Miranda/76 clone (Fig. 2). Size markers used were Saccharomyces cerevisiae chromosomes (Sigma).

Members of the Putative Mucin Gene Families Might Largely Differ in Sequence—To know if genes containing repeats were widely distributed among T. cruzi parasites, several strains and clones were analyzed. Homologous sequences to the repeats present in MUC.CA genes were detected in a second parasite clone (SylvioX-10/7) but not in others (RA stock and CL-Brener clone) (Fig. 1, panel 1). In fact, only 4 out of 13 strains and clones of T. cruzi tested showed sequences homologous to the repeated domain (data not shown). However, the complete MUC.CA gene probe, which contains the 5' and 3'-flanking sequences in addition to the repeats, revealed several bands in the strains previously mentioned (Fig. 1, panel 2) and in another 9 strains and clones of T. cruzi tested (data not shown). Patterns compatible with a gene family were observed in all strains analyzed.

These results suggest that different isolates and clones of T. cruzi might have related sequences sharing 5' and 3'-ends but differing in their central regions. To test this possibility, one strain of T. cruzi (RA), which did not hybridize with the repeated region of the MUC.CA genes, was analyzed. 18 recombinant clones were obtained by PCR as described under “Experimental Procedures.” Two of them, named MUC.RA-1 and MUC.RA-2, were selected for sequencing since they cross-hybridized weakly in Southern blot experiments. Comparison of their deduced amino acid sequences revealed two highly homologous regions having 78 and 81% of identity in the N and C termini, respectively, and two degenerate repeats similar to those present in MUC.CA sequences (Fig. 2). Between these two conserved domains, both MUC.RA-1 and -2 genes lack repetitive units and differ almost completely. These results indicate that a single parasite might have a family composed of highly divergent members. Furthermore, since MUC.RA and MUC.CA sequences greatly differ and are specific of each parasite stock, it might be proposed that different parasites have a different putative mucin gene family.

Chromosomal Localization and RNA Blot Analysis of Putative Mucin Genes—Filters containing T. cruzi chromosomal bands fractionated by size using pulse field gel electrophoresis were hybridized with either the 3'-end or the “repeats” probe of MUC.M/76 sequences. The conserved 3'-region revealed one or two bands in all of the strains analyzed (Fig. 3, panel 1), while the “repeats” probe only revealed bands in the CA1/72 cloned stock (Fig. 3, panel 1). These results are in agreement with the idea that central domains in the putative mucin gene family greatly differ among parasites while flanking sequences are conserved. In all cases, bands were between the 1.6- and 2.2-megabase pair markers, showing that this gene family is not dispersed throughout the genome but restricted to few chromosomal bands. An interesting result from these hybridizations was observed in the CA1/72 stock. While the 3'-probe hybridized with a unique band (Fig. 3, panel 2), the “repeats” probe lit up an additional band (Fig. 3, panel 1). This observation suggests that some genes with tandem repeats do not have conserved flanking sequences, at difference with the pattern described in the MUC.CA and MUC.RA sequences studied. This might further increase the number of variants in the T. cruzi putative mucin gene family.

Northern blots were carried out to determine the size of the transcripts. A complete MUC.CA-2 probe detected a broad band around 1 kilobase in the four parasite strains and clones tested (Fig. 4 and data not shown). The epimastigote and trypomastigote parasite stages, respectively.

![Fig. 3. Genomic organization of putative mucin genes.](image)

**Fig. 3. Genomic organization of putative mucin genes.** The same five strains or cloned stocks were used for pulsed field gel electrophoresis experiments in panels 1 and 2. Filters were hybridized with probe “repeats” from MUC.M/76 (panel 1) or 3'-MUC.M/76 (panel 2). Size markers used were Saccharomyces cerevisiae chromosomes (Sigma).

![Fig. 4. Northern blot analysis.](image)

**Fig. 4. Northern blot analysis.** Total RNA from RA and CL-Brener strains was hybridized with complete MUC.CA-2 probe. Size markers are from an RNA ladder (Life Technologies, Inc.). E and T indicate epimastigote and trypomastigote parasite stages, respectively.

![Fig. 5. Schematic scale representation of predicted regions for the different mucin-like deduced polypeptides.](image)

**Fig. 5. Schematic scale representation of predicted regions for the different mucin-like deduced polypeptides.** The deduced protein domains from genes MUC.M/76, MUC.CA-3, MUC.CA-2, MUC.RA-1, and MUC.RA-2 genes are represented. Percentage of similarities with MUC.M/76 deduced amino acid sequences are indicated for regions between dashed lines. Calculations were done using the program DNASTAR (DNASTAR Inc.) with sequences aligned by the Jotun Hein method (34). To reduce the number of figures, location of oligonucleotides used for PCR and probes are indicated above/their corresponding polypeptide site.
General Structure of the Putative Mucin-Deduced Amino Acid Sequences—A schematic representation of the structure of MUC.CA and MUC.RA amino acid deduced sequences and their percentage of identity is shown in Fig. 5. Interestingly, domains similar to those present in mammalian mucins can be identified, including a putative signal peptide, a non-repetitive mature N terminus, a central domain rich in Thr-Ser-Pro residues, and the C-terminal non-repetitive domain. The deduced sequences on the N termini are very similar in all MUC.CA and MUC.RA sequences studied, comprising the first 20–25 amino acids in the different members. This region has the three common structural features present in signal sequences (28): an n-region with net positive charge, located in the first 10 or 5 amino acids in MUC.CA and MUC.RA, respectively; then follows a central hydrophobic h-region, located up to 10 amino acids, 6 of which are Leu and 2 of which are Val residues; and finally a polar c-region 5 amino acids long, 3 of which are Cys residues (Fig. 2). Since all described mucins in T. cruzi have a melanoma antigen gene with sequence homology to neu and a melanoma antigen gene with sequence homology to Plasmodium falciparum, the causative agent of malaria. These genes are homologous over their N and C termini and even over the flanking non-coding regions, but their central regions are formed of so different repeats that they do not even cross-hybridize on Southern blots performed at low stringency (33).

Acknowledgments—We thank Drs. B. Franke de Cazzulo and J. J. Cazzulo for filters containing size-fractionated chromosomes; we also thank Dr. J. J. Cazzulo for critical reading of the manuscript.

REFERENCES
1. Devine, P. L., and McKenzie, I. F. C. (1992) BioEssays 14, 619–625
2. Gilm, J. R. (1992) Am J Respir Crit Care Med 145, 557–564
3. Hilkens, J., Ligtenberg, M. J. L., Vos, H. L., and Litvinov, S. V. (1992) Trends Biochem. Sci. 17, 359–363
4. Khatri, I. A., Forstner, G. G., and Forstner, J. F. (1993) Biochem. J. 294, 391–399
5. Gendler, S., Taylor-Papadimitrou, J., Duhig, T., Rothebaur, J., and Burchell, J. (1988) J. Biol. Chem. 263, 12820–12823
6. Swallow, D. M., Gendler, S., Griffiths, B., Czorny, G., Taylor-Papadimitrou, J., and Bramwell, M. E. (1987) Nature 328, 82–84
7. Meerzaman, D., Choplin, P., Dascal, E., Polymopoulos, M. H., Martin, B. M., and Rose, M. C. (1994) J. Biol. Chem. 269, 12932–12939
8. Erickson, A. E., Timp, C., Abraham, J. L., Zhao, Y., and Hill, R. L. (1991) J. Biol. Chem. 266, 9678–9686
9. Bhargava, A. K., Wootch, J. T., Davidson, E. A., and Bhaananda, V. P. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 6796–6802
10. Xu, G., Huan, L., Khatri, I. S., Sajjan, U. S., McCool, D., Wang, D., Jones, C., Forstner, G., and Forstner, J. F. (1992) Biochem. Biophys. Res. Commun. 183, 678–682
11. Xu, G., Huan, L. J., Khatri, I. A., Wang, D., Bennett, A. F., Fahim, R. E. F., Forstner, G. G., and Forstner, J. F. (1992) J. Biol. Chem. 267, 5401–5407
12. Gueye, R. I., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E., and Kim, Y. S. (1992) J. Biol. Chem. 267, 21375–21383
13. Murray, P. J., and Spithill, T. W. (1991) J. Biol. Chem. 266, 24477–24484
14. Reyes, M. B., Polivich, G., and Frasch, A. C. C. (1994) Gene (Amst.) 140, 139–140
15. Schenkman, S., Eichinger, D., Pereira, M. E. A., and Nussenzweig, V. (1994) Annu. Rev. Microbiol. 48, 499–523
16. 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
17. Ruiz, R. C., Rigoni, V. L., Gonzales, J., and Yoshida, N. (1993) Parasite Immunol. 15, 121–125
18. Yoshida, N., Mortara, R., Araguth, M. F., Gonzales, J., and Russo, M. (1989) Infect. Immun. 57, 1663–1667
19. Cazzulo, J. J., Franke de Cazzulo, B. M., Engel, J. C., and Cannata, J. J. (1985) Mol. Biochem. Parasitol. 16, 329–343
20. Zingales, B., Katzin, A. M., Arruda, M. V., and Colli, W. (1985) J. Biol. Chem. 260, 21375–21383
21. Gonza´lez Cappa, S. M., Katzin, A. M., An ˜asco, N., and Lajmanovich, S. (1981) Annu. Rev. Biochem. 50, 285–320
22. Brener, Z., and Chiari, E. (1963) Rev. Inst. Med. Trop. Sao Paulo 184, 99–105
23. Dufosse, J., Porchet, N., Audie, J. P., Guyonnet Duperat, V., Laine, A., Morot-Gaudry, V., Sedel, M. E., and Brunak, S. (1995) Biochem. J. 308, 801–813
24. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25. Forney, R. M., Miyakoshi, J., Day, R. S., III, and Paterson, M. (1978) Focus 10, 5–9
26. Henriksson, J., Åslund, L., Macina, R. A., Frankede Cazzulo, B., Cazzulo, J. J., Frash, A. C. C., and Pettersson, U. (1989) Mol. Biochem. Parasitol. 41, 213–224
27. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467
28. von Heijne, G., and Nilsson, T. O. (1995) Trends Biochem. Sci. 20, 1–10
29. Forstner, G. G., and Forstner, J. F. (1992) J. Biol. Chem. 267, 12932–12939
30. Hansen, J. E., Lund, O., Engerbruch, J., Bohr, H., Nielsen, J. O., Hansen, J. E., and Brunak, S. (1995) Biochem. J. 308, 801–813
31. Hei, J. (1990) Methods Enzymol. 183, 626–645
32. Forstner, G. G., and Forstner, J. F. (1992) J. Biol. Chem. 267, 5401–5407
33. Frash, A. C. C., and Pettersson, U. (1990) Mol. Biochem. Parasitol. 36, 329–343
34. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467
35. von Heijne, G., and Nilsson, T. O. (1995) Trends Biochem. Sci. 20, 1–10
36. Lehmann, J. M., Thiemann, G., and Joelson, J. P. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 9891–9895
37. Hansen, J. E., Lund, O., Engerbruch, J., Bohr, H., Nielsen, J. O., Hansen, J. E., and Brunak, S. (1995) Biochem. J. 308, 801–813
38. Forstner, G. G., and Williams, A. F. (1988) Annu. Rev. Biochem. 57, 285–320
39. Dufosse, J., Pochet, N., Audie, J. P., Guyonnet Duperat, V., Laine, A., Van-Seeuningen, I., Marrakachi, S., Degrand, P., and Aubert, J. P. (1993) Biochem. J. 259, 293–317
40. Cowman, A. F., Saint, R. B., Copple, R. L., Brown, G. V., Anders, R. F., and Kemp, D. J. (1985) Cell 40, 757–763
41. Hei, J. (1990) Methods Enzymol. 183, 626–645

Trypanosoma cruzi Mucin-like Genes and Gene Families