Polymorphisms of the genes encoding cruzipain, the major
cysteine proteinase of \textit{Trypanosoma cruzi}, in the region
encoding the C-terminal domain

Javier Martínez a,b*, Jan Henriksson b, Maria Ridåker b, Ulf Pettersson b, Juan José Cazzulo a,c

\textit{a} Instituto de Investigaciones Bioquímicas “Luis F. Leloir”, Fundación Campomar, CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, A. Machado 151, 1405 Buenos Aires, Argentina
\textit{b} Department of Medical Genetics, Biomedical Center, University of Uppsala, Uppsala, Sweden
\textit{c} Instituto de Investigaciones Biotecnológicas, Universidad Nacional de General San Martín, Av. General Paz y Albarellos, C.C. 30, 1650 San Martin, Prov. Buenos Aires, Argentina

Received 20 September 1997; revised 20 November 1997; accepted 22 November 1997

Abstract

Forty-eight cDNA clones obtained from different developmental stages of \textit{Trypanosoma cruzi} and all encoding the C-terminal domain of the major cysteine proteinase (cruzipain) have been sequenced. A number of polymorphisms were detected, seven of them resulting in amino acid replacements. The predicted pI values of the corresponding gene products varied between 7.05 and 8.12. These changes in amino acid sequence, together with previously reported variations in carbohydrate composition at the only N-glycosylation site in the C-terminal domain, may account for most of the heterogeneities found in the mature enzyme. © 1998 Federation of European Microbiological Societies. Published by Elsevier Science B.V.

Keywords: \textit{Trypanosoma cruzi}; Cruzipain; Cysteine proteinase; Isoform; cDNA clone

1. Introduction

Cruzipain is the major cysteine proteinase (CP) present in \textit{Trypanosoma cruzi}, the parasitic protozoan causing the American trypanosomiasis, Chagas Disease [1]. The enzyme is expressed in all stages of the parasite’s life cycle, although it is quantitatively more important in the epimastigote stage [1]. Cruzipain contains, like all the Type I CPs from trypanosomatids [2] and at variance with the other CPs reported so far, in addition to a catalytic domain highly homologous to some of the mammalian cathepsins also a characteristic C-terminal domain, the function of which is unknown [1]. Cruzipain is purified from epimastigotes as a complex mixture of isoforms, detectable, although not cleanly separable so far, by ionic exchange chromatography, isoelectrofocusing, reversed-phase HPLC and SDS-PAGE in substrate-containing gels [3]. The reason for this microheterogeneity is not known, although it may be hypothesized to depend on several concurring phenomena, namely the presence of a number of post-
translational modifications [4], including carbohydrate heterogeneity [5], as well as mutations leading to amino acid replacements [6–8]. With the exception of the most divergent CP reported so far in *T. cruzi* (cruzipain 2 [8]), which has only 86% identity with the sequences predicted from other cruzipain genes, most if not all of the non-conservative amino acid replacements seemed to be present in the C-terminal domain. In order to provide further information on this subject, we have obtained and sequenced 48 cDNA clones, isolated from epimastigotes of the Tul-2 strain and from the four major developmental stages of the RA strain. The results give support to the idea that, although other causes can not be excluded, the simultaneous expression of several slightly different genes may be the origin of the heterogeneities reported.

2. Materials and methods

2.1. Parasites and culture

Epimastigotes of the Tul-2 and RA strains were grown in axenic medium and harvested as previously described [9]. Metacyclic trypomastigotes were obtained by spontaneous differentiation at 28°C of epimastigotes of the RA strain, followed by purification by DEAE-cellulose chromatography. Amastigotes and trypomastigotes were obtained by infection of Vero cell monolayers with trypomastigotes of the RA strain [10]. Trypomastigotes were obtained free of cellular debris by leaving them to swim off the centrifuged pellet for 1 h at 37°C [10].

2.2. RNA purification and preparation of cDNAs

Total RNA was purified from parasites as described [11]. Reverse transcription was performed on 1 to 5 μg of RNA, using the First-Strand cDNA Synthesis Kit (Pharmacia Biotech, Cat. Nr. 27-9261-01) according to the manufacturer’s instructions.

2.3. PCR and cloning procedures

The PCR reactions were carried out on 2–3 μl of the reverse transcription product, denaturated for 5 min at 93°C, in a Perkin Elmer 480 thermocycler using 30 cycles under the following conditions: 94°C, 1 min; 55°C, 1 min; 72°C, 2 min. The final extension was performed at 72°C for 7 min.

The primers used were designed based on genomic DNA sequences of the cruzipain gene [6]: CTA: 5′ TGCCCTTGTCAGGAGGAGGCAGC 3′; CTB: 5′ CCATAAGCTTCAGCATCAGAAACAAAGAGTGCCC 3′. CTA corresponds to the end of the catalytic domain, whereas CTB corresponds to the non-coding 3′ region, downstream of the cruzipain stop codon. The primers used, therefore, should allow the amplification of the complete coding region for the C-terminal domain, except in cases where the stop codon might have been replaced by an amino acid-encoding one, or if the 3′ UTRs are highly divergent.

The amplified fragment was ligated to vector T and transformed into competent *Escherichia coli* HB101.

Colony PCR on lysed bacteria was performed using the same primers, and similar amplification conditions, except for the annealing temperature (60°C), which was increased to diminish the probability of unspecific amplification.

2.4. Sequencing of DNA

DNA for sequencing was amplified by PCR from lysed recombinant bacteria, using biotinylated primers corresponding to the vector T and purified using streptavidin bound to magnetic beads. The sequence reactions using fluorescent primers were performed, and the sequences were obtained in an ABI373 (Applied Biosystems) automatic sequencer run according to the manufacturer’s instructions [12].

3. Results and discussion

cDNAs were obtained from total RNA from epimastigotes of the Tul-2 strain, and the fragments encoding the C-terminal extension were amplified by PCR and sequenced as described in Section 2. Thirty-four clones arising from two independent PCR reactions (17 each) were sequenced (GenBank Accession Numbers AF023623 to AF023656). In or-
order to discard possible PCR artifacts, only those fragments showing identical sequences in both PCR reactions, or those having a homolog among the sequences already reported [6-8] (29 clones in total) were included in the sequence comparison (Fig. 1). These clones could be divided into eight groups; it is noteworthy that within the two most abundant groups (Ct1 and Ct2, consisting of eight clones each), two sequences, differing in one silent mutation, were obtained in each case, with identical frequency (not shown). Therefore, at least ten different cruzipain genes, from a total number of about 130 [6], were concomitantly transcribed. Considering all the clones, seven replacements of amino acid residues were detected; the number of replacements per C-terminal extension ranged from one (Ct2) to six (Ct11) (Fig. 1). Since five of these changes involve charged residues, these replacements would result in eight isoforms differing in predicted pI (Fig. 1). In three groups of clones (Ct3, Ct6 and Ct11), as well as in cruzipain 2 [8], the loss of Cys-318 (residue numbering of the mature cruzipain) probably results in the loss of one of the disulfide bridges [4]. Cys-292 was reported to be replaced by Ser in czeP [8]. It is tempting to speculate that these two Cys residues might be either free, or involved in the formation of the same disulfide bridge. The other six Cys residues in Ct1 (244, 249, 253, 263, 276 and 306) are conserved in all the sequences reported so far by ourselves or other authors. We have previously reported, from chemical evidence, that most if not all the Cys residues present in the C-terminal domain

---

**Fig. 1.** Comparison of amino acid sequences at the C-terminal extension of cruzipain. Residue numbering corresponds to the mature form of the enzyme. Identical residues are indicated by points, and only amino acid replacements (taking as the basic sequence that of group Ct1) are indicated. The composition of the groups (clone numbering as in GenBank Accession Numbers AF023623 to AF023656) is as follows: Ct1, a5, a9, a10, a11, a37, b12, b14 and b16; Ct2, a6, a8, a15, a21, a49, a50, b42 and b43; Ct3, a13, a56, b1 and b35; Ct4, b17; Ct5, a36; Ct6, b22 and b32; Ct7, b5; Ct8, b6; Ct9, b3; Ct10, a11, b27 and b34; Ct11, a36; Ct12, a38 and Ct13, b4. Groups Ct4, Ct5, Ct7, Ct12 and Ct13, each consisting of only one clone predicting an amino acid replacement neither found in other clones in this study nor reported before, are not included in the comparison, since they could have arisen from PCR mistakes. Published sequences are taken from the respective references [6,8,13]; in those from [8], only those including the complete C-terminal extension are shown.
form disulfide bridges [4]. The only potential N-glycosylation site in the C-terminal domain, Asn-255, as well as the seven Thr residues placed at the beginning of the domain, in the proposed ‘hinge’ region [1] (Thr residues 220 and 224 to 229) are also conserved in all reported cruzipain sequences (Fig. 1 and [8]); this suggests that most post-translational modifications characteristic of this domain may be conserved. The changes detected in predicted pl, as well as the probable loss of one disulfide bridge in some isoforms, might explain the previously noted changes in processing of the only oligosaccharide chain present in the C-terminal extension, at Asn-255 [5].

In order to see whether similar heterogeneities were present in different developmental stages of the parasite, we amplified by PCR and sequenced 14 clones from different stages of the RA strain, three from epimastigotes, three from amastigotes, two from metacyclic trypomastigotes and six from cell culture derived trypomastigotes. With only one exception, which might be due to a PCR artefact since no identical clones were found, the deduced amino acid sequences were identical to that of the Ct10 group of clones from Tul-2 epimastigotes (not shown). Since the RA strain is known to contain a lower number of cruzipain genes, about 50 [6], than Tul-2, and the isoform pattern detected in gelatin-SDS-PAGE gels is simpler than that shown by Tul-2 (V. Duschak and J.J. Cazzulo, unpublished results), it is possible that a far more restricted number of genes is simultaneously expressed by this strain, even in different developmental stages. Lima et al. [8], using other parasite stocks, reported two complete and five partial cDNA sequences encoding the C-terminal extension, obtained from RNAs from different parasite stages, and showed some differences. This probably reflects differences among different T. cruzi strains.

The present results, together with those previously reported [6–8], indicate that not only many polymorphic cruzipain genes are present in the genome of at least some parasite strains, but that a number of them might be expressed simultaneously. It is noteworthy that, with the exception of cruzipain 2 [8], although amino acid residue replacements have also been found in the catalytic domain, all of them seem to be conservative, not leading to changes in the predicted pl values [1]. Although our comparisons do not include the catalytic domain, we propose that most of the heterogeneities found in cruzipain may be essentially due to variation in amino acid sequence, structure of the oligosaccharide moiety, and probably also of post-translational modification, at the C-terminal extension. On the other hand, the non-conservative substitutions reported in the catalytic moiety for cruzipain 2 [8] are likely to result in changes in substrate specificity, as recently reported for the Type I CPs of Leishmania mexicana [14].

Acknowledgments

We are indebted to Professor A.C.C. Frasch for helpful discussions. J.J.C. is a member of the Research Career of the National Research Council of Argentina (CONICET), and J.M. was a Research Fellow from the University of Buenos Aires. This work was aided by grants from the University of Buenos Aires and the Swedish Agency for Research Cooperation with Developing Countries (SAREC/Sida).

References


