Biochemical and Biophysical Research Communications **287,** 98–103 (2001)
doi:10.1006/bbrc.2001.5545, available online at http://www.idealibrary.com on IDEAL®

Unusual Features of Poly[dT-dG]·[dC-dA] Stretches in CDS-Flanking Regions of *Trypanosoma cruzi* Genome

María Ana Duhagon, Bruno Dallagiovanna,[1] and Beatriz Garat[2]
Sección Bioquímica, Departamento de Biología Celular y Molecular, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay

Received August 6, 2001

In trypanosomatids, the mechanisms of gene expression regulation are not yet well understood. The genes are organized into long polycistronic transcription units separated by intergenic regions that may contain the signaling information for nucleic acid processing. Poly-dinucleotides are frequent in these regions and have been proposed to be involved in gene expression regulation. We analyzed their frequency in CDS-flanking sequences of sense strands in *Trypanosoma cruzi* and established that all but poly[dC-dC], poly[dC-dG], and poly[dG-dG] are significantly more frequent than expected by chance. Poly[dT-dG]·[dC-dA] is among the longest and most frequent poly-dinucleotides and shows a remarkable strand asymmetry. Furthermore, electrophoretic mobility shift assays using *T. cruzi* epimastigotes nuclear extracts demonstrated the existence of at least, one sequence specific single-strand binding activity for each strand. These results strongly suggest that poly[dT-dG]·[dC-dA] sequence is involved in regulatory mechanisms of relevance for the parasite biology.

Key Words: **poly-dinucleotide frequency; strand asymmetry; electrophoretic mobility shift assay; DNA-protein interactions; single-stranded DNA binding proteins; gene expression.**

The kinetoplastid parasite *Trypanosoma cruzi* is the etiological agent of Chagas disease. It constitutes a major health problem affecting several million people. Its endemic area expands along South and Central America and Mexico. Trypanosome protein-coding genes are arranged in large polycistronic transcriptional units from which individual mRNAs are generated by 5′ *trans*-splicing, involving the addition of a small conserved RNA called spliced leader (SL) and 3′ polyadenylation. During the past years, several groups have been working to elucidate the mechanisms that control gene expression in trypanosomatids (1).

Very few promoters for protein coding genes have been identified in trypanosomes. The variant surface glycoprotein (VSG) expression sites and the procyclic acidic repetitive protein (PARP) genes of *T. brucei* have been described as regulated at the level of transcription. These genes are transcribed by an α-amanitin resistant RNA polymerase (RNA pol) and may represent natural protein-coding genes transcribed by the RNA pol I, or an RNA pol I-like enzyme (2). The structure of the promoters for all other protein-coding genes transcribed by the conventional α-amanitin-sensitive RNA pol II has not yet been clearly defined. In fact, only a few of them have been reported (3, 4 and references therein), but their ability to increase transcription initiation is not clear. Concerning the transcription of the SL RNA, the experiments designed to analyze the putative promoter activity involved suggest that it is not a strong promoter (reviewed in 5). Very recently it has been demonstrated that the SL RNA is transcribed by an RNA pol II enzyme (6).

The intergenic regions of different polycistrons are capable of initiate the transcription of the downstream genes although no classical promoter elements have been identified in them. Besides, kinetoplastid transient transfection plasmids support promoter independent transcription. Likewise, the most common expression vectors used in *T. cruzi* place intergenic regions in front of the cloning site to initiate transcription (7). Recently, the complete sequence of the *Leishmania major* chromosome 1 has been reported. All its protein-coding genes are organized in two “head to head” units with absolute strand polarity (8). Interestingly, a similar arrangement has been observed in the longest *T. cruzi* contig published (9). From these evidences, two possible explanations for transcription initiation have been proposed: either the presence of a single RNA pol II promoter region upstream of each long polycistronic transcription unit or, alternatively, the existence of multiple sites somehow randomly distributed along the

[1] Present address: Instituto de Biologia Molecular do Parana, Rua Algacyr Munhoz Mader 3775, 81350010 Curitiba, PR Brazil.
[2] To whom correspondence should be addressed. Fax: (598 2) 525 86 17. E-mail: bgarat@fcien.edu.uy.

0006-291X/01 $35.00

chromosome which can adopt transient open conformations favoring the initiation of transcription (3, 10, 11).

DNA-protein interactions have been detected for the spliced leader (SL) promoter in trypanosomatids (5, 6) and in the VSG and PARP promoters in *T. brucei*. For the VSG and PARP genes, the binding activity to the core promoter region is mainly found when the DNA is in the single stranded form, both in the coding and in the noncoding strand (2, 12). This could imply that a melted form of DNA is necessary for transcription initiation. In addition for VSG promoter, double stranded DNA binding activity has been identified in bloodstream-form trypanosome nuclear extract (13). Factors from *T. cruzi* extract that interact with AP-1 sequences both in single and double stranded form have been also found (14).

Dinucleotides repeats are highly frequent in intergenic regions and have also been observed in *T. cruzi* (9). Several relevant biological roles have been ascribed to repetitive motives since they can adopt different structural conformations and may represent targets for protein recognition. Furthermore, they have been shown to be involved in regulation of gene expression.

In this work we studied the presence and distribution of dinucleotide repeats in CDS-flanking regions of *T. cruzi* genome. Firstly, we established that most types of poly-dinucleotide repeats are significantly more abundant than what is expected by chance. Secondly, we could determine that poly[dT-dG]·[dC-dA] exhibits asymmetrical strand distribution. Finally, we found out that this dinucleotide repeat constitutes a specific target for single-stranded binding proteins contained in nuclear extracts from *T. cruzi* epimastigotes.

MATERIALS AND METHODS

Computational analysis. Poly-dinucleotide occurrences in CDS-flanking regions were studied for each of the possible combinations of nucleotides. Two independent data sets were created for the analysis. One data set (C) was constructed using the three overlapping sequences that define a 93.4-Kb *T. cruzi* contig (9) (AF 052831, AF 052832, AF 052833). The other was obtained from the genome sequences registered in the GenBank using the *entrez* facilities. Reported sequences with a coding annotation (CDS), were selected (494 sequences). The CDS-flanking regions of these sequences were used to form the other data set (SI), and comprised 103,971 nt. The occurrence (O) and length (n) of perfect poly-dinucleotides $[dX\text{-}dY]_n$ with $n \geq 4$ were determined for both data sets (SI and C) and these data were used to define a parameter F ($F = n \cdot O$). For each poly-dinucleotide a unique value (D) consisting in the addition of every F ($D = \Sigma F(n)$) was obtained and normalized for the total number of bases in the corresponding data set (1.04 for SI and 0.93 for C). The occurrence (O) and length (n) of nonperfect poly-dinucleotides were also studied using the Tandem Repeat Finder program (*TRF*) (15) for $n \geq 10$ in both data sets (SI and C). In this case F also includes a factor (a), varying between 0 and 1, that measures the matching proportion of the corresponding perfect poly-dinucleotide ($F = n \cdot O \cdot a$). As above, for each poly-dinucleotide a unique value (D) consisting in the addition of every F was obtained and normalized.

The randomly expected frequencies for each perfect poly-dinucleotide were determined dividing the product of each nucleotide occurrences (X) by the size (N) of the data set, $[(X/N)\cdot(Y/N)]^n$, for $n = 4$. These frequencies multiplied by the length of the corresponding data base (N) gave the number of poly-dinucleotides expected by chance (E). For C N = 93,435 nt; A, 24,114; C, 21,734; G, 23,541; T, 24,046. For SI N = 103,971 nt; A, 27,613; C, 21,035; G, 24,360; T, 30,963. The significance of the deviation of observed (O) from expected (E) values was measured using the χ^2 test calculated as $[(O\text{-}E)^2/E]$ and was then compared to tabulated values for one degree of freedom. Here the observed values (O) were considered as the number of poly-dinucleotides with $n \geq 4$, irrespectively of the length. To analyze the existence of a significant difference in the distribution of the complementary perfect poly-dinucleotides a difference of two proportion hypothesis test (Z) was performed ($Z = (O_1 - O_2) - (E_1 - E_2)/[O_1(1 - O_1)/N + O_2(1 - O_2)/N]^{1/2}$) and the results were contrasted to tabulated values. For this purpose database C was evaluated considering hypothetical coding strand switch at base 34,000 (9).

Nuclear extracts. Nuclear extracts were prepared as reported for *T. brucei* (2), with some modifications. *T. cruzi* CL Brener strain epimastigotes were grown at 28°C in liver infusion tryptose (LIT) liquid medium supplemented with 10% heat inactivated fetal calf serum (Sigma) to late-log phase and 10^9 parasites were centrifuged 10 min at 1000*g*. Cells were washed three times in PBS and the final pellet was resuspended in buffer Tris–HCl 10 mM pH 7.6, 2 mM $MgCl_2$, 5 mM KCl, 2 mM $CaCl_2$, 0.5 mM DTT, 1 mM EDTA, 1 mM spermidine, and 6% PEG. After 10 min on ice, NP-40 was added to a final concentration of 1.5%. All buffers used contain 1 mM PMSF, 1 μM pepstatin, 0.6 μM leupeptin and all steps were carried out at 4°C. The cell suspension was homogenized in a Tri-R Stir-R homogenizer (Model K41) at 3000 rpm. One volume of 0.64 M sucrose, 40 mM Tris–HCl pH 7.6, 60 mM KCl, 1 mM EDTA, 1 mM EGTA, 0.5 mM DTT, and 1 mM spermidine was added. The disrupted cells were centrifuged for 10 min at 1500*g* at 4°C. The supernatant was discarded and the pellet resuspended in 1 ml of buffer 20 mM Hepes pH 7.9, 100 mM KCl, 0.2 mM EDTA, 20% glycerol, 0.5 mM DTT, 1 mM PMSF. The suspension was centrifuged for 1 h at 25,000*g* at 4°C. The supernatant was removed and the pellet was resuspended in 500 μl of the former buffer and homogenized with 20 strokes of a microfuge tube homogenizer (K Deltaware). Soluble nuclear fractions were stored at −80°C. Protein concentration was determined by Lowry's method using bovine serum albumin as standard.

Binding reaction and electrophoretic mobility shift assays. DNA–protein gel retardation analysis was performed according to standard methods. Reactions were carried out in 20 μl reaction volume containing binding buffer (10 mM Tris–HCl, 10 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, 0.1 M EDTA), 5 mM spermidine and 2 μg (poly[dI-dC]·poly[dI-dC]) as a nonspecific competitor. The oligodeoxynucleotides ($[dT\text{-}dG]_{15}$ or $[dC\text{-}dA]_{15}$, Operon Technologies) end-labeled with [γ-^{32}P]dATP and T4 polynucleotide kinase were used. Reactions were incubated for 20 min at room temperature and loaded onto a native 6% polyacrylamide gel. For competition assays a 15 nt poly[dT] oligodeoxynucleotide, two different oligodeoxynucleotides named CR4 (TAATCCGCATGCGTTGCACAGGTAGT) and MP1 (CGCCGGGCGTGTGCTGCCAAACCGCTGACC) and a 300-bp *Nco*I fragment which contains a 28-bp poly[dT-dG] repeat obtained from the intergenic region between the *tcpgp*2 and *ptr1* gene of *T. cruzi* (16) were used. Specific competitors were added at room temperature 10 min before the addition of the labeled probe.

RESULTS AND DISCUSSION

High Abundance and Asymmetrical Strand Distribution of Poly[dT-dG]·[dC-dA] in CDS-Flanking Regions of T. cruzi

In order to study the peculiar dinucleotide abundance in the genome of *T. cruzi*, we analyzed the na-

ture and frequency of poly-dinucleotide in CDS-flanking regions. Two data sets (C) and (SI) concerning sense strand were constructed. Since there are only very few sequences with defined untranslated regions (UTRs), the term "flanking" in this work includes both UTRs and intergenic regions. The data base SI contains flanking sequences of variable length adjacent to the coding regions meanwhile the data set C contains complete intergenic plus UTR sequences. Therefore SI is enriched in regions surrounding CDS compared to C. A computational analysis for poly-dinucleotide content was carried out. A parameter (D) comprising both occurrence (O) and length (n) for the different repetitive lengths was determined for each perfect poly-dinucleotide and normalized for the total number of bases in the corresponding data set. The results obtained for each poly-dinucleotide are shown in Fig. 1A. The longest and most frequent poly-dinucleotides are poly[dA-dA], poly[dT-dT], poly[dA-dT], and poly[dT-dG]. Though this appreciation is valid for both data sets, a remarkable difference in relative abundance is observed. There is a higher amount of poly[dA-dA], poly[dT-dT] and poly[dA-dT] in C than in SI. Deletion of an AT rich region with no ORFs in the contig did not modify this finding (data not shown). Strikingly, the poly[dT-dG] is the poly-dinucleotide with the highest D value in the regions surrounding the CDS (SI). Subsequently, we carried out the analysis of the frequency of nonperfect poly-dinucleotides repeats using the Tandem Repeat Finder program (15) and calculated the parameter D for $n \geq 10$. Similar results were obtained (Fig. 1B).

The high abundance of poly[dA-dA], poly[dA-dT] and poly[dT-dT] both as perfect or nonperfect repeats could be related to a conformational role in DNA dynamics due to their propensity to adopt a bending structure. The presence of A and T rich regions in the 3′ flanking regions of CDS has been related to mRNA stabilization in different organisms and very recently in *T. cruzi* (17). Besides, polydA·polydT sequences have been shown to preset chromatin structure accessible to transcription factors by means of nucleosome destabilization (18).

In order to determine the statistical significance of the abundance of the poly-dinucleotide repeats an analysis of the expected frequency of poly-dinucleotides with $n = 4$ was carried out attending the base composition of the corresponding data set. The observed and expected frequencies were compared using a χ^2 test. Coincidentally with qualitative appreciation, all perfect poly-dinucleotides except poly[dC-dC], poly[dC-dG] and poly[dG-dG], display significant deviations from the expected frequency ($P < 0.0001$) in both data sets (Table 1).

Taking into account the presence of a long tract of the dinucleotide (dT-dG) in the intergenic regions between the developmentally regulated genes of *T. cruzi* H-like *locus* (16, 19, 20) and the relevance of CDS-flanking regions in the regulation of the stage specific expression (1), we analyzed the location of perfect

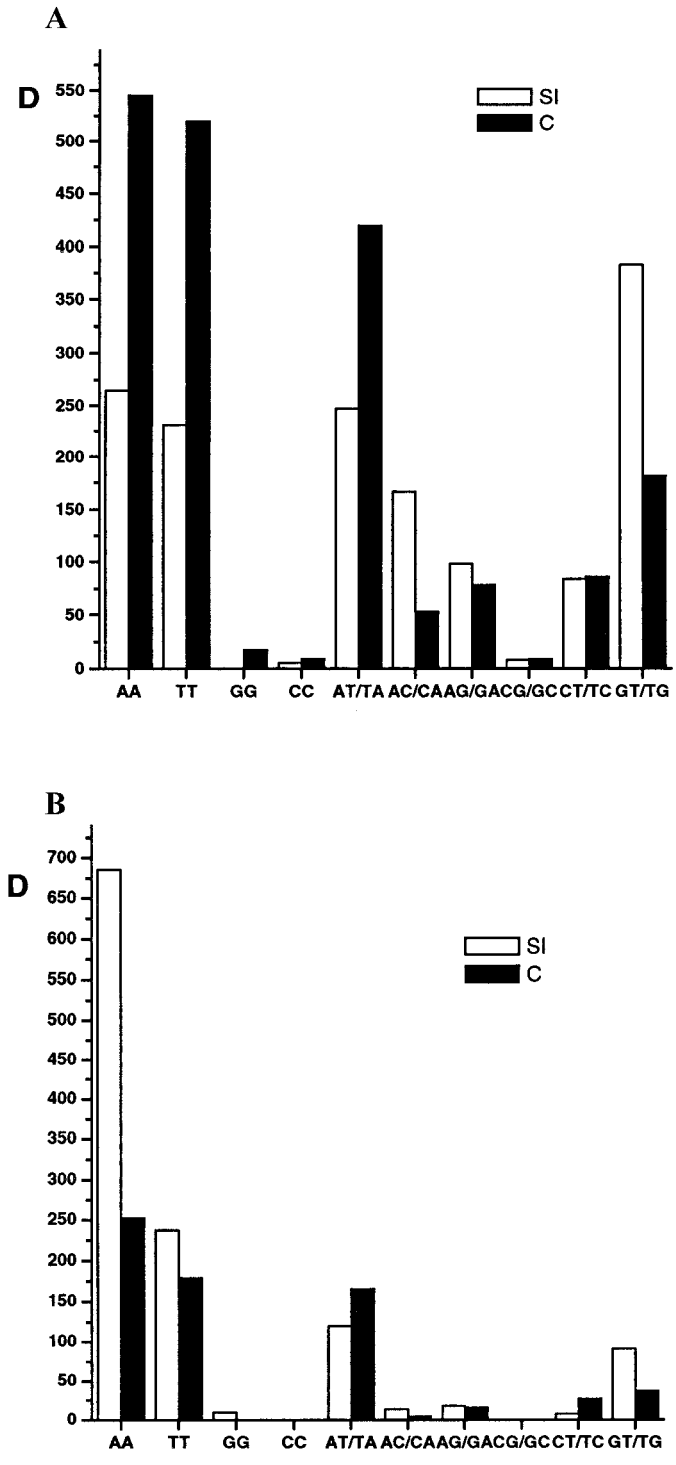

FIG. 1. Analysis of poly-dinucleotide repeats occurrence in CDS-flanking regions of *T. cruzi*. (A) Abundance parameter (D) obtained for perfect poly-dinucleotides $(XY)_n$ with $n \geq 4$ in SI (no color) and in C (black). (B) Abundance parameter (D) obtained for non-perfect poly-dinucleotides with $n \geq 10$ in SI (no color) and C (black). Poly-dinucleotide sequences are indicated in the abscissa.

TABLE 1

Observed and Expected Values of Poly-dinucleotides [(XY)n] for n ≥ 4 in CDS-Flanking Regions of *T. cruzi*

		AA*	AC/CA*	AG/GA*	AT/TA*	CC	CG/GC	CT/TC*	GG	GT/TG*	TT*
	O	81	9	13	58	2	2	13	3	27	84
C	E	2	2	2	2	2	2	2	2	2	2
	O	46	37	17	34	1	2	15	0	77	42
SI	E	3	3	3	4	3	3	3	3	3	7

Note. O represents the observed number of poly-dinucleotides with n ≥ 4, E the expected values and * indicates significant deviations from the expected values ($P < 0.0001$) by χ^2 test.

poly[dT-dG] relative to the CDS for the SI data set. The genes bearing these repeats are listed (Table 2). We were unable to find out a clear pattern of location. However, the polycistronic nature of the mRNAs and the fact that very few reported sequences have precisely defined UTR could be hindering its identification. Besides, in this analysis, the distinction between 3′ or 5′ flanking regions for sequences that contain more than one ORF is arbitrary. As trivially expected, these repeated sequences are more frequently found in flanking sequences compared with coding sequences. The largest $[dT\text{-}dG]_n$ elements (n ≥ 7) are only found in CDS-flanking regions, and when n ≥ 5 more than one tract is usually found. No relationship between the presence of the poly[dT-dG] in the CDS-flanking sequences and characteristics as expression level, codon usage and properties of the encoded protein could be detected. It is worth to mention the existence of a clear bias in the database to sequences encoding antigens or proteins from the noninfective stage. Nevertheless, the absence of a correlation with particular characteristics could favor the interpretation that these elements are involved in general processes of gene expression.

In the SI data set, the magnitudes of the frequency deviations between the complementary poly-dinucleotides are similar with only one exception: the poly[dT-dG]·poly[dC-dA] (Table 1 and Fig. 1). The poly[dT-dG]·poly[dC-dA] turn up as the only pair of complementary poly-dinucleotides that displays a significant deviation of frequency proportion in the SI data set ($Z = 3.40$ $P < 0.0003$). This peculiarity is also observed when comparing the theoretical sense strands in the contig ($Z = 2.0$ $P < 0.02$). The location, high quantity and asymmetrical distribution of poly[dT-

TABLE 2

GenBank Accession No. and Definition of the Sequences Presenting Poly[dT-dG] in the CDS-Flanking Regions

Accession No.	Definition	Length (n) and location
Z47798	cystatin	11_{-86}
L76077	elongation factor 1-alpha	4_{-310}, 7_{-127}, 6_{+193}
M65021	insect stage-specific antigen (GP72)	5_{-575}, 7_{-485}, 6_{-95}, 6_{+193}, 5_{+416}
AH008347	KMP11	10_{-82}, 9_{+178}
U70620	Ca2+-ATPase gene	4_{-428}, 9_{-408}, 4_{+10}
U24190	Tc40 antigen	8_{+130}, 5_{+151}
M25364	kinetoplast-associated protein (KAP)	8_{+63}, 6_{+177}
AF047023/AF004380	paraflagellar rod component Par1h	6_{+147}, 5_{+208}
X62144	P0 ribosomal protein	5_{+46}
U11272/U06070/AF044733	poly(A) binding protein	4_{+10}, 7_{+65}
U31282	dehydrogenase gene	6_{+410}, 4_{+390}, 7_{+377}
M61732/X57235/M21582	neuraminidase (TCNA), acute phase antigen	5_{+68}, 6_{+549}
M97956	beta tubulin and alpha tubulin	16_{+330}
Y09115	histone H2A	5_{+264}
U04340	TCA33 amastin	5_{+242}, 4_{+642}
X02838	1F8 protein	9_{-162}
AF099099	TolT1, TolT2 and TolT3	6_{-111}, 16_{+649}, 7_{+849}
U25030	tuzin gene	5_{-633}, 5_{-308}
L01584	calcium-binding protein (CUB2.8)	4_{-399}, 5_{-388}
Z49222	P-glycoprotein	4_{-46}, 4_{+284}, 4_{+416}, 5_{+438}, 14_{+798}
Y11262	dihydrolipoamide dehydrogenase	4_{+435}
AF051696	sialidase homolog (P85.2)	4_{+80}
L27659	histone H3	4_{+354}

Note. The length of the perfect poly[dT-dG] repeat is expressed as n; the subindex indicates the respective location relative to the CDS.

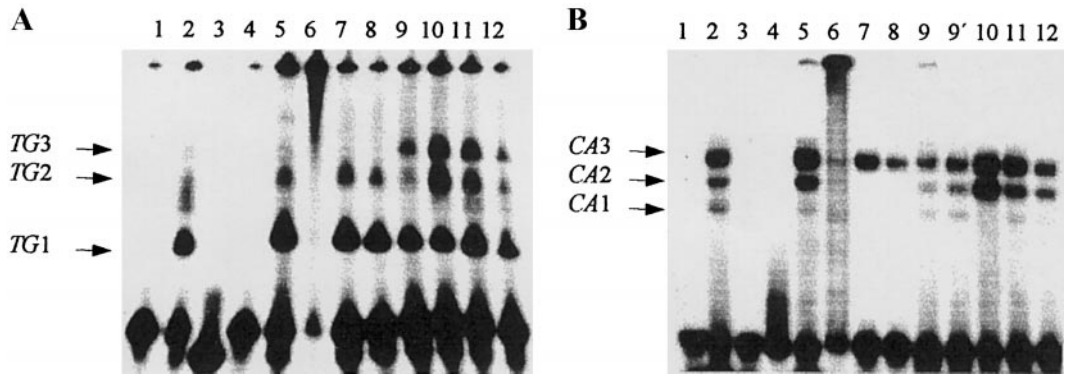

FIG. 2. EMSA specificity analysis of the complexes formed by nuclear extracts from *T. cruzi* epimastigotes and the repetitive dinucleotide probes: (A) $[dT\text{-}dG]_{15}$ and (B) $[dC\text{-}dA]_{15}$. Lane 1, labeled probe. Lane 2, binding reactions with 5 μg of nuclear extract and 1 ng of labeled probe (30,000 cpm approx.). Lanes 3–12, binding reactions in the presence of different competitors (100-fold molar excess). Lane 3, $[dC\text{-}dA]_{15}$; 4, $[dT\text{-}dG]_{15}$; 5, 300-bp double stranded *Nco*I fragment from the intergenic region *tcpgp2-ptr1*, which presents a 28-bp poly[dT-dG]) stretch; 6, the former competitor previously heated at 90°C for 3 min and snap-cooled on ice for 5 min; 7, tRNA; 8, tRNA previously heat denatured; 9, ssM13 DNA; 9′, ssM13 DNA previously heat denatured; 10, $[dT]_{15}$; 11, CR4 oligodeoxynucleotide (TAATCCGCATGCGTTGCACAGGTAGT); and 12, MP1 oligodeoxynucleotide (CGCCGGGCGTGTGCTGCCAAACCGCTGACC).

dG]·poly[dC-dA] suggest a functional role for these elements.

Specific Complexes with $[dT\text{-}dG]_{15}$ and $[dC\text{-}dA]_{15}$ Oligodeoxynucleotides and Nuclear Extracts of T. cruzi Epimastigotes

Poly[dT-dG]·poly[dC-dA] repeats can adopt particular conformations (21–23) that are specifically recognized by proteins (21, 24) involved in different processes such as recombination and transcriptional events. In order to determine whether the highly frequent and asymmetrical strand distributed poly[dT-dG] repeats constitute specific target for gene regulatory proteins, we attempted EMSA using nuclear extracts from *T. cruzi* epimastigotes. Due to the evidence that binding of specific proteins in trypanosomatids mainly requires the target DNA to be single-stranded, we tested not only the double-stranded ($[dT\text{-}dG]_{15}$·$[dC\text{-}dA]_{15}$) but also the $[dT\text{-}dG]_{15}$ and $[dC\text{-}dA]_{15}$ oligodeoxynucleotides as probes. While in the assayed conditions the double stranded probe did not produce any specific complex, a variety of band-shifts were seen for both oligodeoxynucleotides (Fig. 2). Structural and sequence components of specificity were analyzed including different competitors in the binding reactions. RNA and M13 DNA were added to distinguish between binding due to RNA or single strand DNA non-sequence specific binding proteins. The intergenic region *tcpgp2-ptr1*, which has a 28-bp poly[dT-dG], was used both as single or double stranded to study the putative effect of the target context. Finally, poly[dT] as a polypyrimidine sequence and two random oligodeoxynucleotides were assayed to distinguish the sequence component of specificity. Titration experiments varying the amount of competitors were done (data not shown) and illustrative results are shown at 100-fold molar excess (Fig. 2).

Using the $[dT\text{-}dG]_{15}$ probe (Fig. 2A), at least three band shifts with different specificity were observed (noted as *TG*1, *TG*2, and *TG*3). Both unlabeled $[dT\text{-}dG]_{15}$ and $[dC\text{-}dA]_{15}$ were able to compete the three complexes. When the complementary oligodeoxynucleotide was used, a variation in the migration pattern was observed probably due to the formation of the dsDNA. *TG*1 and *TG*2 complexes correspond to specific DNA-protein interactions, as they were not displaced by the three different oligodeoxynucleotides used. An RNA-binding protein may be involved in the *TG*3 complex since it is competed by yeast tRNA. Reasons for the enhanced formation of complex *TG*3 in the presence of the oligodeoxynucleotides remains to be understood. Interestingly, the double stranded intergenic sequence could not displace any of the complexes, and when heat denatured produced an aggregation in the well. *TG*1 and *TG*2 complexes were Proteinase K sensitive in different extent, and resistant to RNase and DNase activity (data not shown). A different mobility pattern was obtained using the $[dC\text{-}dA]_{15}$ probe. At least three band shifts, a faster minor one, an intermediate and a sharp upper complex were observed (noted as *CA*1, *CA*2, and *CA*3 in Fig. 2B). Both unlabeled $[dC\text{-}dA]_{15}$ and $[dT\text{-}dG]_{15}$ were able to compete the three complexes. *CA*1 and *CA*2 complexes were completely displaced by the yeast tRNA competitor. *CA*3 showed high specificity, as it was not competed by the employed competitors. *CA*3 complex was Proteinase K sensitive, and resistant to RNase and DNase activity (data not shown).

It is interesting to mention that the specific complexes *TG*1, *TG*2 and *CA*3 were formed even in the presence of high ionic strength (1 M KCl). Our ap-

proach cannot discard that a single protein recognizing the repetitive poly(Pu-Py) motif could be involved. As expected, single-stranded DNA binding protein (SSBP) from *E. coli* (Promega) was able to interact with $[dT\text{-}dG]_{15}$ and $[dC\text{-}dA]_{15}$ probes but the band shift was eliminated when any of the competitors were used. In addition, similar EMSA patterns were observed for different *T. cruzi* strains (Tulahuen and Y strains) (data not shown).

It is not possible to rule out that these proteins could be involved in RNA maturation. In trypanosomatids, a set of proteins participating in splicing and RNA processing (25, 26 and references therein) has been identified. RNA-binding proteins are of crucial relevance in these organisms since regulation of gene expression occurs primarily at the posttranscriptional level.

The high frequency and the distribution of the polydinucleotide sequences described here points out a functional role of these elements in genome processes. Particularly, the abundance, location, asymmetrical distribution as well as the *in vitro* behavior of poly[dT-dG]·poly[dC-dA] as a specific target for single stranded nuclear binding protein(s) strongly suggest an important role in *T. cruzi* genome physiology and constitute a novel contribution to the understanding of genome regulation in these organisms.

ACKNOWLEDGMENTS

We thank Dr. R. Ehrlich for constant encouragement and helpful discussions and Dr. H. Musto for critical reading of this manuscript. We also thank Dr. N. Williams and Dr. W. Ruyechan for valuable suggestions and comments. We are indebted to H. Romero for his assistance in sequence analysis. This work was financially supported by CSIC (Uruguay). M. A. Duhagon received a PEDECIBA fellowship.

REFERENCES

1. Teixeira, S. M. (1998) Control of gene expression in *Trypanosomatidae. Braz. J. Med. Biol. Res.* **31,** 1503–1516.
2. Lee, M. G., and Van der Ploeg, L. H. (1997) Transcription of protein-coding genes in trypanosomes by RNA polymerase I. *Annu. Rev. Microbiol.* **51,** 463–489.
3. Swindle, J., and Tait, A. (1996) Trypanosomatid genetics. *In* Molecular Biology of Parasitic Protozoa (Smith, D. F., and Parsons, M., Eds.), pp. 6–34, Oxford University Press, NY.
4. Lee, M. G. (1996) An RNA polymerase II promoter in the *hsp*70 *locus* of *Trypanosoma brucei. Mol. Cell. Biol.* **16,** 1220–1230.
5. Campbell, D. A., Sturm, N. R., and Yu, M. C. (2000) Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol. Today* **16,** 78–82.
6. Gillinger, G., and Bellofatto, V. (2001) Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. *Nucleic Acids Res.* **29,** 1556–1564.
7. Kelly, J. M., Ward, H. M., Miles, M. A., and Kendall, G. (1992) A shuttle vector which facilitates the expression of transfected genes in *Trypanosoma cruzi* and *Leishmania. Nucleic Acids Res.* **20,** 3963–3969.
8. Myler, P. J., Audleman, L., deVos, T., Hixson, G., Lemley, C., Magness, C., Rickel, E., Sisk, E., Sunkin, S., Swartzell, S., Westlake, T., Bastien, P., Guoliang, F., Ivens, A., and Stuart, K. (1999) *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc. Natl. Acad. Sci. USA* **96,** 2902–2906.
9. Andersson, B., Aslund, L., Tammi, M., Tran, A. N., Hoheisel, J. D., and Pettersson, U. (1998) Complete sequence of a 93.4-kb contig from chromosome 3 of *Trypanosoma cruzi* containing a strand-switch region. *Genome Res.* **8,** 809–816.
10. Mc Andrew, M., Graham, S., Hartmann, C., and Clayton, C. (1998) Testing promoter activity in the trypanosome genome: Isolation of a metacyclic-type VSG promoter, and unexpected insights into RNA polymerase II transcription. *Exp. Parasitol.* **90,** 65–76.
11. Mc Donagh, P. D., Myler, P. J., and Stuart, K. (2000) The unusual gene organization of *Leishmania major* chromosome 1 may reflect novel transcription processes. *Nucleic Acids Res.* **28,** 2800–2803.
12. Beberof, M., Vanhamme, L., Alexandre, S., Lips, S., Tebabi, P., and Pays, E. (2000) A single-stranded DNA binding protein shared by telomeric repeats, the variant surface glycoproteins transcription promoter and the procyclin transcription terminator of *Trypanosoma brucei. Nucleic Acids Res.* **28,** 597–604.
13. Pham, V. P., Rothman, P. B., and Gottesdiener, K. M. (1997) Binding of *trans*-acting factors to the double-stranded variant surface glycoprotein (VSG) expression site promoter of *Trypanosoma brucei. Mol. Biochem. Parasitol.* **89,** 11–23.
14. Espinosa, J., Martinetto, H., Portal, D., Dángelo, M., Torres, H. N., and Flawia, M. M. (1999) Factors from *Trypanosoma cruzi* with AP-1 sequences. *J. Eukaryot. Microbiol.* **46,** 516–521.
15. Benson, G. (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580.
16. Robello, C., Navarro, P., Castanys, S., and Gamarro, F. (1997) A pteridine reductase gene *ptr*1 contiguous to a P-glycoprotein confers resistance to antifolates in *Trypanosoma cruzi. Mol. Biochem. Parasitol.* **90,** 525–535.
17. Di Noia, J. M., D'Orso, I., Sanchez, D. O., and Frasch, A. C. (2000) AU-rich elements in the 3′-untranslated region of a new mucyn type family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *J. Biol. Chem.* **275,** 10218–10227.
18. Shimizu, M., Mori, T., Sakurai, T., and Shindo, H. (2000) Destabilization of nucleosomes by an unusual DNA conformation adopted by polydA)·polydT) tracts *in vivo. EMBO J.* **19,** 3358–3365.
19. Dallagiovanna, B., Gamarro, F., and Castanys, S. (1996) Molecular characterization of a p-glycoprotein-related tcpgp2 gene in *Trypanosoma cruzi. Mol. Biochem. Parasitol.* **75,** 145–157.
20. Robello, C., Dallagiovanna, B., Engel, J. C., Gamarro, F., and Castanys, S. (1998) A new member of YER057c family in *Trypanosoma cruzi* is adjacent to an ABC-transporter. *Gene* **220,** 1–12.
21. Gaillard, C., and Strauss, F. (2000) DNA loops and semicatenated DNA junctions. *Biochem. Struct. Biol.* **1,** 1, http://biomedcentral.com/1471-2237/1/1.
22. Ho, P. S. (1994) The non-B-DNA structure of d(CA/TG)n does not differ from that of Z-DNA. *Proc. Natl. Acad. Sci. USA* **91,** 9549–9553.
23. Kladde, M. P., Kohwi, Y., Kohwi-Shigematsu, T., and Gorski, J. (1994) The non-B-DNA structure of d(CA/TG)n differs from that of Z-DNA. *Proc. Natl. Acad. Sci. USA* **91,** 1898–1902.
24. Dutreix, M. (1997) (GT)n repetitive tracts affect several stages of RecA-promoted recombination. *J. Mol. Biol.* **273,** 105–113.
25. Manger, I. D., and Boothroyd, J. C. (1998) Identification of a nuclear protein in *Trypanosoma brucei* with homology to RNA-binding proteins from *cis*-splicing systems. *Mol. Biochem. Parasitol.* **97,** 1–11.
26. Zhang, J., Ruyechan, W. T., and Williams, N. (1998) Developmental regulation of two nuclear RNA binding proteins, p34 and p37, from *Trypanosoma brucei. Mol. Biochem. Parasitol.* **92,** 79–88.