The Stage-regulated Expression of *Leishmania mexicana* CPB Cysteine Proteases Is Mediated by an Intercistronic Sequence Element

Received for publication, September 4, 2001, and in revised form, October 9, 2001
Published, JBC Papers in Press, October 9, 2001, DOI 10.1074/jbc.M108498200

Darren R. Brooks‡, Hubert Denise, Gareth D. Westrop§, Graham H. Coombs§, and Jeremy C. Mottram¶

From the Wellcome Centre for Molecular Parasitology, University of Glasgow, the Anderson College, Glasgow G11 6NU and the §Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Joseph Black Building, Glasgow G12 8QQ, Scotland, United Kingdom

The tandemly arranged CPB genes of *Leishmania mexicana* are polycistronically transcribed and encode cysteine proteases that are differentially stage-specific; CPB1 and CPB2 are expressed predominantly in metacycles, whereas CPB3–CPB18 are expressed mainly in amastigotes. The mechanisms responsible for this differential expression have been studied via gene analysis and re-integration of individual CPB genes, and variants thereof, into a CPB-deficient parasite mutant. Comparison of the nucleotide sequences of the repeat units of CPB1 and CPB2 with CPB2.8 (typical of CPB3–CPB18) revealed two major regions of divergence as follows: one of 258 base pairs (bp) corresponding to the C-terminal extension of CPB2.8; another, designated InS, of 120 bp, with insertions totaling 57 bp, localized to the intercistronic region downstream of CPB1 and CPB2. Cell lines expressing CPB2.8 or CPB2 with the 3'-untranslated region and intercistronic sequence of CPB2.8 showed up-regulation in amastigotes. Conversely, metacyclic-specific expression occurred with CPB2 or CPB2.8 with the 3'-untranslated region and intercistronic sequence of CPB2. Moreover, the InS down-regulated expression in amastigotes of a reporter gene integrated into the CPB locus. It is proposed that the InS mediates metacyclic-specific stage-regulated expression of CPB by affecting the maturation of polycistronic pre-mRNA. This is the first well defined *cis*-regulatory element implicated in post-transcriptional stage-specific gene expression in *Leishmania*.

Protozoan parasites of the genus *Leishmania* are diploid eukaryotes that cause a range of cutaneous and visceral diseases that afflict ~12 million people worldwide (www.who.int/health-topics/leishmaniasis.htm). *Leishmania* species have a digenetic life cycle, passing between a sandfly vector and mammalian hosts. *Leishmania* exist as extracellular flagellated promastigotes within the alimentary canal of the insect, and these differentiate into the highly infectious metacyclic form that is found within the mouth parts of the insect. Parasites are transmitted to a mammal when the vector takes a blood meal, and following macrophage invasion they reside within phagolysosomes as aflagellated amastigotes (1). A stringently coordinated pattern of gene expression is crucial to life cycle progression.

The haploid *Leishmania* genome consists of 36 chromosomes with sizes that range between 0.3 and 2.5 megabases (2–4). Sequencing of chromosome 1 of *Leishmania major* revealed that it has an unusual distribution of predicted protein-coding genes. Remarkably, the first 29 genes are all encoded on one strand, whereas the remaining 50 genes are encoded in a head-to-head manner on the opposite strand (5). Several highly expressed protein-coding genes of *Leishmania*, such as those encoding the surface glycoproteins 63 (6) and 46 (7), α- and β-tubulin (8), A2 (9), and the cysteine proteases CPB10, occur in tandemly arranged clusters of identical or very similar genes that can show stage-specific expression (10–13). To date, there is no evidence for the existence of promoter elements for an α-amanitin-sensitive RNA polymerase II (5). Thus transcription of protein coding genes is considered to be constitutive in *Leishmania* and related trypanosomatids, and the regulation of gene expression is believed to occur predominantly at the post-transcriptional level (14, 15).

The polycistronic precursor mRNAs of trypanosomatids are converted into monocistronic mRNAs by two cleavage reactions that occur within the intercistronic regions. One cleavage is associated with *trans*-splicing of a capped 39-nucleotide spliced leader sequence to the 5'-end of the mature mRNA (16). The second cleavage occurs in a region several hundred bases upstream of the splice acceptor site and is necessary for polyadenylation (17). The *trans*-splicing and polyadenylation reactions are tightly coupled and regulated by polyuridymin tracts present within the intercistronic sequences (17). Numerous studies have shown that differential gene expression in *Leishmania* can be mediated by sequence elements in the 3′-UTRs that affect mRNA stability (13, 18–21). Intercistronic
regions have also been implicated in controlling gene expression (22), probably by mediating the events involved in pre-mRNA processing (trans-splicing and cleavage/polyadenylation), but no cis-regulatory elements have been defined (22). The precise mechanisms that govern leishmanial gene expression are still not well understood, although the recent identification and characterization of the L. major poly(A)-binding protein I (23) should shed some light on this area.

Leishmania mexicana contains numerous lysosomal cysteine proteases, the majority of which are cathepsin-L-like and are expressed in increasing amounts during life cycle progression from promastigotes to amastigotes (24, 25). The type I cysteine proteases are encoded by the CPB genes (26), which map to one genomic locus as a tandem array of 19 copies (10). Targeted deletion of this array to generate the null mutant ΔCPB has shown that the genes encode virulence factors (27, 28). Analyses of enzyme substrate preferences have shown that some CPBs have differing substrate specificities (10). Moreover, the first two genes of the array, CPB1 and CPB2, are atypical because they encode enzymes that lack the C-terminal domain characteristic of trypanosomatid type I cysteine proteases (10, 29, 30). Furthermore, CPB1 and CPB2 are expressed almost exclusively in the infective metacyclic stage, whereas the remaining isogenes are expressed predominantly in amastigotes (10).

The present study was undertaken to investigate the mechanism controlling this differential stage regulation of CPBs. The strategy utilized was to restore by genetic manipulation CPB repeat units, and variants thereof, to the endogenous CPB-deleted locus of ΔCPB. Importantly, this allowed gene expression to be analyzed in its correct chromosomal context.

(i) The 635-bp HindIII-SalI fragment containing the 5′-flank of CPB was excised from plasmid pGL147 and used to replace the 436-bp HindIII-SalI unique 5′-flanking region of CPB present in pGL151. The resulting plasmid, pGL146, contains the original 436-bp unique 5′-flank of the CPB array and an additional 199 bp of non-unique 5′-flank that sequences the first intron for splicing 199 bp (20 times) 3′ 5′ 2°C for 5 min. The resulting 803-bp PCR product was cloned into pGEM T-vector (Promega) to generate plasmid pGL157.

(ii) The 168-bp SalI-SpeI fragment from pGL157 (which contains the remainder of the upstream sequence to the start codon that is required for expression of CPBs) was subcloned into pGL146 to generate pGL158. Plasmid pGL159 was subsequently used to integrate the native and chimeric CPBs by cloning into the SalI site.

(iii) The chimeric plasmids were first generated in pBS (Stratagene) prior to subcloning into pGL159. The 3684-bp EcoRV fragment from pGL28 (defined as plasmid lmcpgb2.8 in (27)) was used to replace the corresponding EcoRV fragment in pGL26 to generate chimeric CPB2 (pGL153). The 3741-bp EcoRV fragment from pGL26 was used to replace the corresponding fragment in pGL28 to generate chimeric CPB2.8 (pGL154).

(iv) The 2731-bp SalI fragment from pGL26 was subcloned into pGL159 to give the CPB2 re-integration plasmid pGL165.

(v) The 2875-bp SalI fragment from pGL28 was subcloned into pGL159 to give the CPB2.8 re-integration plasmid pGL166.

(vi) The 2125-bp SalI fragment of pGL153 was subcloned into the SalI site of pGL159 to generate the chimeric CPB2 re-integration plasmid pGL167.

(vii) The SalI fragment of pGL154 was subcloned into the SalI site of pGL159 to generate the chimeric CPB2.8 re-integration plasmid pGL168.

The bacterial chloramphenicol acetyltransferase (CAT) gene was amplified from pHB2 (34) (using primers OL221 (CTGAGATGCAGAAAAATTACTGTTATACCC) and OL222 (GATATCTTTCCCGCCTTGGACATC)) with Pfu Turbo (Stratagene) (20 cycles; 94°C for 30 s, 68°C for 30 s and 72°C for 1 min) to give plasmid pGL147. The 2731-bp EcoRV-SalI fragment of the non-coding region of pGL26 was subcloned into the corresponding sites of pGL294 to give pGL295. The 2731-bp EcoRV-SalI fragment that includes the corresponding sequence was subcloned into pGL298 as a SalI-SacI fragment derived from pGL157 (a pGEM T-vector clone of the 5′-flank region) to generate pGL299. The 5′-flank CAT-CPB2 non-coding region was excised from pGL299 as a SalI fragment and subcloned into the corresponding site of pGL159 to generate the construct used for re-integration (pGL390) into the ΔCPB cell line.

Further details on the sequence of the plasmids can be obtained from the authors upon request.

Transfection—Transfection of L. mexicana was as described previously (27). Briefly, 4 × 107 late-log phase ΔCPB promastigotes were subjected to electroporation with either 2 μg of pXCPB2 (pGL67) or 10 μg of the purified re-integration cassette obtained by restriction digestion (HindIII and BglII) from the appropriate plasmids (pGL165–pGL168 and pGL300). After overnight incubation in medium lacking antibiotics, the cell line transfected with the episome pGL67 was selected in 25 μg ml−1 G418 (Geneticin, Life Technologies, Inc.). Cell lines transfected for integration were spread onto modified Eagle’s medium plates supplemented with 10 μg ml−1 hygromycin (Cayla, France) and 25 μg ml−1 nourseothricin hydrolysulfate (Hans-Knoll Institute, Germany). Colonies, visible following 10–14 days incubation at 25°C, were picked and inoculated into complete liquid modified Eagle’s medium containing appropriate antibiotics.

Southern Blot Analysis of Transfectants—Parasites that were found to be resistant to 10 μg ml−1 hygromycin and 25 μg ml−1 nourseothricin hydrolysulfate, but sensitive to 50 μg ml−1 hygromycin (Cayla, France) and 25 μg ml−1 nourseothricin hydrolysulfate (Hans-Knoll Institute, Germany), were picked and grown to stationary phase. DNA was isolated according to Medina-Acosta and Cross (35). The DNA (5 μg) was digested with restriction enzymes and processed as described previously (27). The probe consisted of a 1:1 ratio of a 203-bp KpnI-SalI fragment derived from pGL26 and pGL28 that composed part of the ORFs of CPB2 and CPB2.8.

Northern Blot Analysis—Total RNA was prepared from Leishmania cells using Tri-Reagent (Sigma) according to the manufacturer’s instructions. RNA samples from 2 × 107 cells were fractionated by
electrophoresis on 1.2% (w/v) agarose, 0.66 M formaldehyde gels and transferred to Hybond-N membranes (Amersham Pharmacia Biotech) as described previously (10). α-32P-Labeled DNA probes (spanning CPB2.8 or α-tubulin ORF) were prepared from agarose gel-purified restriction endonuclease fragments using a Prime-IT II kit (Stratagene) and purified on Micropip S-200 HR columns (Amersham Pharmacia Biotech). Filters were hybridized at 42 °C and washed at 55 °C as described previously (10). The labeled filters were exposed to storage phosphor screens and scanned using a Typhoon 9600 Imager (Amersham Pharmacia Biotech). Levels of mRNA were measured using ImageQuant software (Amersham Pharmacia Biotech) and normalized using α-tubulin mRNA.

RT-PCR Analysis and 3'-RACE—Total RNA was isolated from L. mexicana stationary phase promastigotes (containing a high proportion of metacyclics) as described previously (10). Firststrand cDNA was synthesized using 1 μg of total RNA, 0.5 μg of oligo(dT) adaptor primer (5'-TGCAGCTTGGAGCTGGC-3'), and 50 units of Moloney murine leukemia virus Superscript II reverse transcriptase as described by the manufacturer (Life Technologies, Inc.). Aliquots were subsequently analyzed by 2 rounds of PCR using Taq DNA polymerase (Applied Biosystems, Inc.). The initial PCR used OLS9 (5'-CGGGAGATGCGGCCCCCTG-3') in conjunction with the 3'-adapter primer (5'-TGCAGCTTGGAGCTGGC-3'). Re-amplification used CPB1–CPB2-specific primer DB1 or CPB3–CPB18-specific primer DB3 (10) and the 3'-adapter primer with the following cycling parameters: 94 °C for 1 min, 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s (30 cycles); 72 °C for 5 min. The annealing temperature was increased to 60 °C for 10 s for re-amplification with OL699 (5'-CGACCCGGCTATTCTGCTCTCG-3') and the 3'-adapter. PCR products were cloned into pGEM T-vector (Promega) and sequenced to localize the poly(A) addition sites.

RESULTS

Sequence Comparison of the Repeat Units of L. mexicana CPB1, CPB2, and CPB2.8—The 19 CPB genes of L. mexicana are arranged in a tandem array of 2.8-kb repeat units (Fig. 1A). In previous work (10, 27) we reported the sequence of CPB2.8 (a randomly cloned gene from the middle of the array) and CPB1 open reading frames, and we showed that expression of CPB1 and CPB2 is metacyclic specific, whereas CPB2.8 is predominantly expressed in amastigotes. In expectation that sequence elements regulating the stage-specific expression of the CPB genes would reside in the 3'-UTR or intercistronic regions of the genes, the complete repeat unit (defined as a 2.8-kb SalI fragment, Fig. 1) was sequenced for CPB1 and CPB2.8 (additional sequence deposited as updates to EMBL database accession numbers Z49962 and Z49963). As both CPB1 and CPB2 were found to be metacyclic specific, the SalI repeat unit containing the CPB2 gene was subcloned from a bacterial plasmid λ clone (13), see "Experimental Procedures." (10) and resequenced for comparative purposes.

A high level of nucleic acid sequence identity exists in the SalI repeat units of CPB1, CPB2, and CPB2.8, with the exception of two regions (Fig. 1B, the CPB1 sequence is identical to CPB2 over these two regions). First, there is significant sequence divergence in the region comprising the CTE of CPB2.8 (Fig. 1C and Ref. 10). Second, a region of 120 bp (termed the insertion sequence (InS), defined in Fig. 1D) is characterized by 4 short insertions, totaling 57 bp, in the CPB2 and CPB1 sequences. The final insertion of 32 bp is composed predominantly of GT and GC repeats (12 repeats in total). Additionally, CPB1 and CPB2 have a 31-bp region within the diverged sequence that is 90% AT-rich and includes 14 TA repeats. In contrast, comparison of CPB1 and CPB2 sequences from the stop codon to the downstream SalI site (a region comprising 1490 bp) revealed only 4 nucleotide changes (not shown). In addition, in the 1183 bp of sequence between the stop codon of CPB2.8 and the downstream SalI site (excluding the 120 bp InS), only 7 nucleotides differed from CPB1 and 3 differed from CPB2.

CPB2 Encodes a Major, High Mobility Gelatinase Activity Present in L. mexicana Metacyclics—cpb2 is predicted to encode a protein, CPB2, of 359 amino acids with a high degree of sequence identity to CPB1. In the pre-pro region, CPB2 shows absolute identity to that of CPB2.8, whereas there are 4 amino acid changes relative to the pre-pro-region of CPB1 (Table I). The primary amino acid sequence of the mature domain of CPB2 is the most divergent of the isoenzymes of the CPB array examined so far and encodes four amino acids that are unique to date: Glu21, Ser30, Ser115, and Glu160. Nevertheless, the mature domain of CPB2 is 95% identical to that of CPB2.8 and 98% identical to that of CPB1. The CPB2 coding region, like CPB1 (10), has a single base pair deletion that results in a frameshift and premature termination relative to the CTE of CPB2.8 (Fig. 1C). Thus, both of the metacyclic-specific genes encode CPB enzymes with truncated CTEs, whereas those encoded at a high level in amastigotes (CPB2.8, CPB2) are predicted to encode a full-length C-terminal extension.

Populations of L. mexicana stationary phase promastigotes that contain a high proportion of metacyclics express two predominant cysteine proteases that show activity toward gelatin in a non-denaturing activity gel (36) (Fig. 2, lane 1). When CPB1 was expressed in the L. mexicana CPB null mutant (ΔCPB), the activity of the encoded protein was found to correspond to the slower mobility gelatinase activity present in wild type metacyclic extracts (Fig. 2, lane 2). To investigate if CPB2 encoded the faster mobility activity in wild type metacyclic promastigotes, CPB2 was transfected into ΔCPB on an episomal shuttle vector. The cysteine protease activity of the resultant mutant was analyzed toward gelatin. CPB2 was found to encode an active protease (Fig. 2, lane 3) that corresponded in size to the faster mobility activity detected in wild type metacyclic promastigote cell extracts (Fig. 2, lane 1). Notably, fewer lower mobility cysteine protease activities, labeled activated precursors in the Fig. 2, were observed with CPB1 or CPB2 (10, 27). The equivalent lower mobility activity bands associated with CPB2.8 (~34 kDa) have been confirmed recently by protein sequence analysis of a recombinant CPB2.8 as being activated zymogen (37). This differences between the wild type parasites and the mutants re-expressing CPB1 and CPB2 are likely to reflect the different number of isogenes being expressed. Alternatively, they could relate to the presence or absence of the C-terminal extension domain in different CPB isoenzymes.

Generation of L. mexicana Mutants with CPB Genes Reintegrated into the Endogenous Locus—We postulated that the InS identified from comparison of the non-coding sequences of CPB2 and CPB2.8 contained the control elements necessary to modulate the expression of the different CPB genes. Thus, the presence of the InS would result in metacyclic-specific expression, as was shown previously by Northern blotting with a probe to detect CPB1 and CPB2 transcripts (10). Conversely, amastigote-specific expression is predicted in the absence of the InS. This was also suggested previously by Northern blot-
ting with a CTE-specific probe (10). To test this hypothesis, we utilized an *L. mexicana* cell line that has all CPB genes deleted (*H9004*) (27) to generate a series of mutants in which single CPB genes were restored to the CPB locus. The following four constructs used to target the CPB locus encoded: (i) the full CPB2 repeat unit, including the CPB2 ORF, the natural 3'-UTR, and downstream intercistronic region (giving the *H9004* CPB::CPB2 cell line); (ii) the full CPB2.8 repeat unit, including the CPB2.8 ORF, the natural 3'-UTR, and downstream intercistronic region (giving the *H9004* CPB::CPB2.8 cell line); (iii) a chimeric gene containing the CPB2 ORF fused to the 3'-UTR and the intercistronic region of CPB2.8 (giving the *H9004* CPB::CPB2A cell line), and (iv) a chimeric gene containing the CPB2.8 ORF fused to the 3'-UTR and the intercistronic region of CPB2 (giving the ΔCPB::CPB2.8M cell line), see Fig. 3A and “Experimental Procedures” for details.

Transfection of *H9004* with these constructs followed by double antibiotic selection with phleomycin and nourseothricin resulted in the replacement of the *HYG* gene with SAT. Confirmation that the CPB constructs had been correctly targeted to the natural CPB locus of *H9004* was obtained by Southern blotting (Fig. 3B). The probe, derived from the ORF of CPB2 and CPB2.8, the additional nucleotide of CPB2.8 is boxed. D, comparison of CPB2 and CPB2.8 nucleic acid sequences across the InS. Dashes indicate gaps in the CPB2.8 sequence introduced to allow maximal alignment with CPB2. Shading indicates differences between CPB2 and CPB2.8. The four insertion domains are labeled with *roman numerals*. The domains of TA repeats and G-pyrimidine repeats are overlined.

**Fig. 1.** Comparison of the repeat units containing the CPB1, CPB2, and CPB2.8 genes. A, map of the CPB locus: E, EcoRV; S, *Sal* I, and X, *Xho* I. CPB1 and CPB2 are metacyclic-specific and CPB19 is a pseudogene. B, similarity plot highlighting the differences between the nucleic acid sequences of the *Sal* I repeat units for CPB2 and CPB2.8. SL, spliced leader addition site; CTE, C-terminal extension; InS, insertion sequence; PT, polypyrimidine tract. C, comparison of CPB2 and CPB2.8 nucleic acid sequences across the CTE. Coding sequence is shown in *bold*, and the stop codons are *underlined*. *Shading* indicates differences between CPB2 and CPB2.8. The additional nucleotide of CPB2.8 is boxed. D, comparison of CPB2 and CPB2.8 nucleic acid sequences across the InS. Dashes indicate gaps in the CPB2.8 sequence introduced to allow maximal alignment with CPB2. *Shading* indicates differences between CPB2 and CPB2.8. The four insertion domains are labeled with *roman numerals*. The domains of TA repeats and G-pyrimidine repeats are overlined.
confirmed by the absence of hybridization to ΔCPB DNA restricted with EcoRV (lane 5). The sizes of all fragments detected are as predicted from the map of the CPB locus (Fig. 1A).

The InS Regulates Expression of L. mexicana CPBs—The expression of CPB in the four cell lines was analyzed by Northern blotting to determine levels of steady-state mRNA and by gelatin SDS-PAGE to determine the proteolytic activity (Fig. 4). Two life cycle stages of the parasite were analyzed, stationary phase promastigotes that contain a high proportion of the metacyclic form and relative to the metacyclic form and relative to the mature mRNA in the mutant lines was the same as wild type (data not shown) providing evidence for correct processing of precursor RNA.

In the metacytic form the chimeric gene CPB2Δ (lane 3) was expressed at a similar level to CPB2 (lane 1) and CPB2.8 (lane 2), whereas CPB2.8Δ was 2-fold higher (lane 4). In the amastigote form CPB2Δ (lane 7) was expressed at a level a third lower than CPB2.8 (lane 6), whereas the chimeric CPB2.8 gene containing the InS (CPB2.8Δ, lane 8) was down-regulated 30-fold compared with CPB2.8. These data implicate the InS in down-regulation of CPB gene expression in the amastigote form and hence tightly regulated expression of CPB1 and CPB2 in the metacytic form.

The stage-regulated CPB activity of the mutant cell lines was analyzed by gelatin SDS-PAGE (Fig. 4B). The ΔCPB::CPB2 cell line was found to have substantial gelatinase activity in the metacytic form (lane 1) but little activity in the amastigote form (lane 5), confirming the expression profile shown for mRNA levels by Northern blotting (Fig. 4A). Likewise, the ΔCPB::CPB2.8 cell line had CPB2.8 activity in the amastigote form (lane 6), but little was detected in the metacytic form (lane 2). This stage-regulated difference parallels the data from the Northern blots that show that expression of CPB2.8 is up-regulated in the amastigote form.
Stage Regulation of Leishmania Cysteine Proteases

The InS Is Located in the Intercistronic Region between Poly(A) Addition Sites and the Site of Trans-splicing—The above data confirm that CPB1 and CPB2 are expressed in the metacyclic form, whereas CPB2.8 is up-regulated in the amastigote form and that the InS is likely to be of importance in controlling the differential stage-regulated expression. To determine whether the InS is located in the 3′-UTR of CPB2 or in the intercistronic region between the polyadenylation site and the site of addition of the spliced leader, the poly(A) addition sites were determined for CPB in wild type L. mexicana metacyclics by RT-PCR. Fig. 5 shows the localization of poly(A) addition sites using a primer combination to detect CPB2 and CPB1 mRNAs, a primer combination to detect CPB3-CPB18 (including CPB2.8) mRNAs, and a generic primer combination that detects all CPB mRNAs (see “Experimental Procedures”). A number of poly(A) addition sites were utilized, and these mapped to two regions, designated S1 and S2. Region S1 was between 825 and 810 bp 5′ of the AG dinucleotide previously identified as the splice acceptor site (26). The second region, S2, was located between 520 and 486 bp 5′ of the splice acceptor site. Further experiments using the generic primer combination showed that both sites were also utilized in axenic amastigotes. With the exception of one PCR product, generated with a 5′-oligonucleotide (DB1) present in both CPB1 and CPB2, showing a poly(A) addition site to exist within the InS, all other identified sites were upstream of the InS.

To confirm that the same polyadenylation sites were used in the mutant cell lines, 3′-RACE was performed on mRNA isolated from metacyclic and axenic amastigote forms of ΔCPB:CPB2, ΔCPB:CPB2.8, ΔCPB:CPB2Δ, and ΔCPB:CPB2.8Δ using the generic primer combination. Two major sites of polyadenylation, correlating with the data obtained from wild type metacyclics and axenic amastigotes, were found in all four cell lines. This shows that correct polyadenylation is taking place in the mutant cell lines and that the stage-regulated expression mimics the wild type state.

InS Directs Stage-regulated Expression of the CAT Reporter Gene—The 3′-UTR and downstream intercistronic region of CPB2.8 have been shown recently to direct expression of reporter genes in L. mexicana amastigotes (38). To test the ability of the InS to direct stage-regulated expression of a reporter gene, the EcoRV-SalI non-coding region downstream of CPB2 was fused to the CAT gene and the cassette to target the CPB locus of ΔCPB. Clones resistant to 10 μg ml−1 phleomycin and 25 μg ml−1 nourseothricin hydrosulfate, and sensitive to 50 μg ml−1 hygromycin B, were isolated. Southern blotting confirmed the HYG cassette in ΔCPB had been replaced by homologous recombination with the CAT cassette to generate ΔCPB::CAT (data not shown). Assays performed with 5 × 105 cell equivalents showed that CAT activity was down-regulated by ∼20-fold in axenic amastigotes relative to stationary phase (metacyclic) promastigotes. This provides evidence that the InS element can down-regulate in amastigotes of Leishmania not only CPB genes but also a heterologous gene.

DISCUSSION

This study addressed the mechanism mediating differential stage-regulated expression of the L. mexicana CPB genes. We have identified a 120-nucleotide sequence (designated the InS) containing 4 insertions totaling 57 nucleotides that is downstream of the metacyclic-specific genes CPB1 and CPB2 and absent from the downstream region of the amastigote-specific gene CPB2.8. By analyzing expression of CPB genes re-integrated into their native genomic environment, we have shown that the InS element can modulate expression of cysteine

![Figure 4](http://www.jbc.org/)

**Fig. 4.** Stage-specific expression of CPB is dependent upon the presence of the InS. A, Northern blot analysis. 20 μg of total RNA was separated by denaturing formaldehyde gel electrophoresis, blotted, and probed with a 32P-labeled CPB-specific probe. The blot was scanned on a PhosphorImager. Filters were re-probed with an α-tubulin-specific probe to allow normalization of mRNA levels. Normalized CPB mRNA levels were plotted relative to ΔCPB:CPB2 (metacyclic) and ΔCPB:CPB2.8 (amastigote), taking into account the 2.8 times higher levels of CPB2.8 mRNA in amastigotes. Lanes 1 and 5, ΔCPB:CPB2; lanes 2 and 6, ΔCPB:CPB2.8; lanes 3 and 7, ΔCPB:CPB2Δ; and lanes 4 and 8, ΔCPB:CPB2.8Δ. The results represent the means ± S.D. from three experiments. B, gelatin SDS-PAGE analysis. Activity gel using extracts from 106 metacyclic promastigotes (lanes 1–4) and axenic amastigotes (lanes 5–8); lanes as for A. The low mobility activities (>46 kDa) are due to protease other than CPB.

(Fig. 4A, lane 6). In contrast, the mutants generated with chimeric constructs show a complete reversal of their respective wild type CPB activity profiles. Thus, ΔCPB:CPB2Δ showed absence of CPB2 activity in the metacyclic form (lane 3) and presence of activity in the amastigote form (lane 7). Additionally, ΔCPB:CPB2.8Δ exhibited CPB2.8 activity in the metacyclic form (lane 4), and drastically reduced activity in the amastigote form (lane 8). These data therefore are consistent with the hypothesis that stage regulation of CPBs is mediated via sequence elements present in the downstream regions of the respective genes. More precisely, the stage regulation correlates with the presence or absence of the InS located within the EcoRV-SalI non-coding region of CPB2. The presence of these short insertions results in down-regulation of gene expression and hence CPB activity in the amastigote form (lanes 1 and 4), whereas amastigote-specific expression and CPB activity occur in the absence of the insertions (lanes 6 and 7). Of note is the finding that expression of CPB2.8 resulted in significant levels of zymogen (appearing as a slower mobility activity band of >30-kDa apparent molecular mass in the gels due to in situ activation) in both the metacyclic (lane 4) and amastigote (lane 6) forms, whereas less precursor was apparent with metacyclic or amastigote forms expressing CPB2 (lanes 1 and 7). This provides further evidence that the processing of the zymogen in vivo differs between isoenzymes with and without a C-terminal extension.
protease genes in vivo in a stage-specific manner. The InS is also capable of modulating the expression of a heterologous gene, as demonstrated using CAT that was expressed some 20-fold less in amastigotes than in metacyclics of ΔCPB::CAT mutants. This degree of down-regulation is consistent with the 30-fold reduction in CPB2.8 mRNA levels observed between ΔCPB::CPB2.8 and ΔCPB::CPB2.8Δ repeat units (Fig. 4A). Given the apparent lack of promoters within the intercistronic region of transcription units transcribed by a-amanitin-sensitive RNA polymerase II in Leishmania and other trypanosomatids (5, 14, 15), it is reasonable to propose that the CPB genes are transcribed polycistronically and that differential stage-regulated gene expression is controlled post-transcriptionally. For many genes of Leishmania (13, 17, 18, 22, 39–41) and trypanosomes (42–45) it is thought that mRNA stability is the major mechanism of post-transcriptional control. By demonstrating that the InS is located in the intercistronic region between the CPB genes and is therefore present in the polycistronic pre-mRNA but not in the mature mRNA, we have provided evidence that pre-mRNA processing is a major mechanism for control of stage-regulated CPB gene expression. Polycistronic pre-mRNA processing events in trypanosomatids include the tightly coupled processes of trans-splicing and polyadenylation (17, 46–48). Thus, we propose a model whereby a trans-acting RNA-binding factor (or factors) associates with sequences in the InS and interacts with the trans-splicing and polyadenylation machinery to modulate polycistronic pre-mRNA processing. This modulation leads to increased expression of CPB1 and CPB2 in metacyclics but conversely down-regulation of expression of CPB1 and CPB2 in amastigotes.

It is interesting to note that the AT-rich nature of the sequence within the InS resembles motifs known to bind trans-acting factors involved in 3′-end processing of pre-mRNA, although the spatial context with respect to the poly(A) addition sites is altered in Leishmania. For example, the AAUAAA motif binds the multisubunit cleavage and polyadenylation specificity factor that is required for cleavage and polyadenylation of mammalian pre-mRNA (49, 50). There is also an absolute requirement for this motif in the nematode Caenorhabditis elegans, an organism that like Leishmania has a significant number of genes within operons and a requirement for trans-splicing (51, 52). Another major component of mammalian pre-mRNA 3′-end processing machinery is the cleavage-stimulatory factor, which binds GU-rich elements downstream of poly(A) addition sites (49, 50). Interestingly a series of 12 GT repeats are found in insertion IV of the InS. As pre-mRNA, GU repeats have the potential to form stable hairpin-loop structures that could interact with regulatory factors involved in 3′-end processing and are themselves developmentally regulated (53). Furthermore, a conserved U-rich element localized in the intercistronic region of several C. elegans and Caenorhabditis briggsae operons has been recently identified (54). Mutational analysis has shown the U-rich element is essential for pre-mRNA processing of genes within a C. elegans operon (54). A U-rich RNA-binding protein, TcUBP-1, has also recently been identified in Trypanosoma cruzi. It binds a 44-nucleotide AU-rich RNA instability element, located in the 3′-UTR of mucin SMUG mRNAs (55). TcUBP-1 is stage-regulated and is thought to control the stability of mRNAs containing the AU-rich instability sequence. As a family of genes encoding similar U-rich-binding proteins has been identified in the Leishmania genome (55), it is possible that a member of this RNA-binding protein family associates with the InS and regulates stage-specific expression of CPB1 and CPB2 in L. mexicana.

3′-RACE analysis of transcripts from wild type and mutant L. mexicana metacyclic promastigotes indicates that a number of poly(A) sites were used within two closely defined regions (designated S1 and S2, Fig. 5). The lack of an identifiable recognition sequence for polyadenylation in Leishmania has been reported previously for several genes (see Ref. 15 for Fig. 5. Localization of poly(A) addition sites relative to InS. A, comparison of the CPB2 and CPB2.8 repeat units. The annotation shows the relative positions of the 5′-oligonucleotides used for RT-PCR analysis and the position of poly(A) addition sites S1 and S2. Other labeling as in Fig. 1. B and C, sequences of the S1 and S2 poly(A) addition sites. Nucleotides in bold represent poly(A) addition sites determined by RT-PCR. Arrows above the poly(A) addition sites indicate the origin of the clones sequenced and the primer combinations used for RT-PCR. ▼, DB1 or DB3; ▼, OL699; ↓, the poly(A) addition site of the amastigote expressed CPB cDNA (26). The InS sequence is double underlined.
Stage Regulation of Leishmania Cysteine Proteases

details). All the poly(A) sites identified were located >400 bp upstream of the trans-splicing site and were therefore consistent with the trypanosomatid model of the poly(A) polymerase scanning the pre-mRNA for suitable addition sites (17). Most important for this study, the sites of polyadenylation in the mutant cell lines where the CPB genes have been reintroduced into the CPB locus were very similar to those of wild type parasites. Indeed it is a notable aspect of this investigation that the experimental strategy allowed gene regulation to be studied in its correct chromosomal context, rather than from episomal expression vectors as used in previous studies on the expression of other Leishmania genes (13, 17, 18, 22, 39–41, 56). Integration allowed the study of clones and therefore precluded the problems associated with variations in copy number and rates of transcription that occur with episomal elements. The validity of the approach was confirmed by an analysis of CPB2, which in wild type L. mexicana is metacyclic-specific, and CPB2.8, which is up-regulated in the amastigote form (10, 27). Integration of the complete repeat units for CPB2 and CPB2.8 resulted in the expected stage-regulated expression (Fig. 4). The resulting mutants showed stage-regulated CPB activity that is dependent upon the context of the non-coding sequences downstream of the CPB ORFs.

A recent study has shown that sequences present in the 3′-UTR of some amastigote-specific genes stabilize the mRNA leading to higher levels of expression (2). Comparative sequence analysis of the 3′-UTR of genes expressed at elevated levels in amastigotes revealed a 150-nucleotide region that is capable of inducing reporter gene expression specifically in amastigotes. This sequence element, however, does not appear to be present in the 3′-UTR of the L. mexicana CPB cDNA. The results of this study, coupled with previous work that showed that the 3′-UTR and downstream intercistronic sequence of CPB2.8 can be used to generate high level expression of a reporter gene in amastigotes from constructs targeted to the rRNA locus (38), indicate that CPB2.8 mRNA also contains amastigote-specific stabilization sequences.

So far the complete CPB locus has not been characterized in detail in any other species of Leishmania, so it is not known if the InS is present downstream of the first two genes in the corresponding CPB loci. Also, we have been unable to identify an InS sequence downstream of any other multicopy genes of Leishmania. However, the small size and the repetitive nature of these insertions make other functional elements difficult to identify in DNA sequence data bases. It is interesting to note that the complete tandem gene array that encodes the Trypanosoma brucei cysteine proteases homologous to CPB of Leishmania has been sequenced recently as part of the African trypanosome genome project (GenBank accession number AC073906). There are no insertions in the intercistronic sequences downstream of the first two copies of the array. Thus the tight expression of individual genes within a tandem array, which is a feature of the CPBs of L. mexicana, appears to be absent from African trypanosomes such as T. brucei. Unlike the L. mexicana CPBs, which vary significantly in sequence between different encoding genes of the array, especially in the C-terminal extension, the 11 T. brucei cysteine proteases encoded on the tandem array are very similar (only 5 of 450 amino acids have variation between some of the 11 isozymes). Thus the enzymes of T. brucei may all be functionally very similar such that different expression profiles would not be advantageous.

ΔCPB has reduced infectivity to macrophages and animals compared with the wild type parasites, implicating the CPB enzymes as virulence factors (27). The mutants in which individual CPB genes have been chromosomally re-integrated are powerful tools to study the roles of individual CPBs in parasite virulence. To this end, detailed phenotype/immunological studies are underway to shed light on why L. mexicana requires such an exquisitely stage-regulated array of CPB genes.
Stage Regulation of Leishmania Cysteine Proteases

45. Di Noia, J. M., D’Orso, I., Sánchez, D. O., and Frasch, A. C. C. (2000) J. Biol. Chem. 275, 10218–10227
46. Ullu, E., Matthews, K. R., and Tschudi, C. (1993) Mol. Cell. Biol. 13, 720–725
47. Matthews, K. R., Tschudi, C., and Ullu, E. (1994) Genes Dev. 8, 491–501
48. Vassella, E., Braun, R., and Roditi, I. (1994) Nucleic Acids Res. 22, 1359–1364
49. Keller, W. (1995) Cell 81, 829–832
50. Wahle, E., and Ruegsegger, U. (1999) FEMS Microbiol. Lett. 23, 277–285
51. Blumenthal, T. (1998) BioEssays 20, 480–487
52. Liu, Y., Huang, T., MacMorris, M., and Blumenthal, T. (2001) RNA (New York) 7, 176–181
53. Bai, C., and Tolias, P. P. (1998) Nucleic Acids Res. 26, 1597–1604
54. Huang, T., Kuersten, S., Deshpande, A. M., Speith, J., MacMorris, M., and Blumenthal, T. (2001) Mol. Cell. Biol. 21, 1111–1120
55. D’Orso, I., and Frasch, A. C. C. (2001) J. Biol. Chem. 276, 34801–34809
56. Kelly, B. L., Nelson, T. N., and McMaster, W. R. (2001) Mol. Biochem. Parasitol. 116, 101–104
The Stage-regulated Expression of *Leishmania mexicana* CPB Cysteine Proteases Is Mediated by an Intercistronic Sequence Element

Darren R. Brooks, Hubert Denise, Gareth D. Westrop, Graham H. Coombs and Jeremy C. Mottram

*J. Biol. Chem. 2001, 276:47061-47069.*
doi: 10.1074/jbc.M108498200 originally published online October 9, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M108498200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 55 references, 26 of which can be accessed free at http://www.jbc.org/content/276/50/47061.full.html#ref-list-1