Transcription Initiation at the TATA-less Spliced Leader RNA Gene Promoter Requires at Least Two DNA-binding Proteins and a Tripartite Architecture That Includes an Initiator Element

(Received for publication, June 23, 1999, and in revised form, August 9, 1999)

Hua Luo, Gwen Gilinger, Devi Mukherjee, and Vivian Bellofatto

From the Department of Microbiology and Molecular Genetics, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, New Jersey 07103-2714

Eukaryotic transcriptional regulatory signals, defined as core and activator promoter elements, have yet to be identified in the earliest diverging group of eukaryotes, the primitive protozoans, which include the Trypanosomatidae family of parasites. The divergence within this family is highlighted by the apparent absence of the “universal” transcription factor TATA-binding protein. To understand gene expression in these protists, we have investigated spliced leader RNA gene transcription. The RNA product of this gene provides an m7G cap and a 39-nucleotide leader sequence to all cellular mRNAs via a trans-splicing reaction. Regulation of spliced leader RNA synthesis is controlled by a tripartite promoter located exclusively upstream from the transcription start site. Proteins PBP-1 and PBP-2 bind to two of the three promoter elements in the trypanosomatid Leptomonas seymouri. They represent the first trypanosome transcription factors with typical double-stranded DNA binding site recognition. These proteins ensure efficient transcription. However, accurate initiation is determined by an initiator element with a loose consensus of CYAC/AYR (+1), which differs from that found in metazoan initiator elements as well as from that identified in one of the earliest diverging protozoans, Trichomonas vaginalis. Trypanosomes may utilize initiator element-protein interactions, and not TATA sequence–TATA-binding protein interactions, to direct proper transcription initiation by RNA polymerase II.

Molecular studies of trypanosomatids, a ubiquitous and diverse family of protozoan pathogens, have revealed strikingly unusual mechanisms of mRNA synthesis. One central device is that two independent transcription events direct each mRNA produced in the trypanosome nucleus (for review, see Ref. 1). The protein-coding portion is transcribed as a single primary mRNA, often containing several open reading frames flanked by 5′- and 3′-untranslated regions. The capped 5′-end portion is transcribed as a short spliced leader (SL) RNA. The two parts are fused in a trans-splicing reaction that yields a functional mRNA. During fusion, the 39 nt present on the 5′-end of the SL RNA (and referred to as the SL) are transferred to a region upstream from the coding region on the primary mRNA (2). Addition of the SL provides each mRNA with an m7G cap as well as four extensively methylated nucleotides, at positions 1–4 within the 39-nt SL RNA (3).

The SL RNA is transcribed from a highly reiterated set of genes. In contrast to the long primary transcripts that form the bulk of the mature mRNA, each SL RNA has a discrete transcriptional start site. α-Amanitin studies show that it is very probable, though not proven, that the SL RNA gene is transcribed by RNA polymerase (pol) II. The primary SL RNA transcript and the transcript present in the trans-splicing spliceosome possess identical 5′- and 3′-ends, indicating that both transcription initiation and termination regulate the accumulation of SL RNA. SL RNA expression has been monitored using independent, tagged gene copies positioned on selectable shuttle vectors that are stably maintained in various trypanosomatids (4–6). In the simple trypanosomatid Leptomonas seymouri, a 95-bp region upstream of the SL RNA intragenic region followed by 70 bp of downstream sequence is sufficient to produce properly initiated and terminated SL RNA (7). These results have been recapitulated in vitro using homologous parasite nuclear extracts (8).

Unusual promoter architecture in this group of primitive eukaryotes, compared with typical metazoans, appears to be the rule. The U6 small nuclear (sn) RNA gene promoter contains three elements: one located within the 5′-portion of the intragenic region and two located within an upstream, but inversely oriented, tRNA gene. The two intragenic tRNA promoter elements, called A and B boxes, cofunction in both U6 and tRNA expression (9). Two abundant cell surface proteins in the African trypanosome Trypanosoma brucei are encoded by genes with promoter elements that resemble RNA pol I promoters in both structure and α-amanitin resistance. Aside from these two protein coding genes in T. brucei, all other trypanosomatid mRNAs are α-amanitin-sensitive and thus transcribed by RNA pol II (for review, see Refs. 10 and 11). Transcriptional start sites for primary mRNAs have been extremely difficult to detect. Two putative promoter regions were tentatively defined as transcriptionally void regions upstream from the highly transcribed actin and HSP 70 genes (12, 13). However, placement of these sequences upstream from a luciferase coding region did not yield even modest levels of reporter gene activity (14). Moreover, in the absence of any putative trypanosome promoter regions, Escherichia coli pBR 322-derived sequences drive expression of reporter genes, such as the chloramphenicol trypanosome initiator; wt, wild type; PCR, polymerase chain reaction; PBS, Bluescript SK II.
acetyltransferase gene. Models to explain these findings suggest that RNA pol II may not be recruited to specific promoter sites to initiate mRNA synthesis. Addition of an SL to these tagged mRNA 5'-ends would polish as RNAs mature into translatable units.

We present a detailed transcriptional analysis of the SL RNA gene promoter using an in vitro transcription system that faithfully recapitulates in vivo transcription. In a deoxy of any previously described trypanosome transcription factors, PBP-1 and PBP-2 (15), which are sequence-specific DNA-binding proteins initially identified in our laboratory, emerge as the first proteins that function to promote efficient SL RNA transcription. These studies also reveal that correct 5'-end formation of the SL RNA, which is crucial to proper capping of the SL, is dependent on the presence of a 5-bp element, the trypanosome initiator (Inr). Finally, we provide evidence for a trans-acting factor necessary for transcription which binds the Inr.

**EXPERIMENTAL PROCEDURES**

**Protein—** *T. brucei* (ATCC 30220) was grown in Trypansoma macro medium (16) at 28°C to log phase. Cells were harvested by centrifugation (10 min, 4°C) and washed twice with buffer 1 (20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 3 mM MgCl₂) (7). All of the following steps were performed at 4°C. The cell pellet was resuspended in 1 packed cell volume of buffer 2 (10 mM HEPES-KOH, pH 7.9, 150 mM sucrose, 2.5 mM MgCl₂, 1 mM EDTA, 2.5 mM diithiothreitol, 1 mM leupeptin) containing 20 mM potassium glutamate. After swelling cells on ice (10 min), they were Dounce homogenized (size A pestle) until disrupted. After centrifugation, the nuclear pellet was resuspended in 5 packed cell volumes of buffer 2 at 20 mM potassium glutamate, and 1 packed cell volume of (NH₄)₂SO₄ (4.1 M) was added dropwise. The extract was cleared by ultracentrifugation (100,000 × g, 35 min, 4°C). The supernatant was fractionated using solid (NH₄)₂SO₄ (0.33 g/ml of solution). 1 nL NaOH (0.1 M/10 g solid (NH₄)₂SO₄) was added to maintain a 34-nt (from the Ambion RPAII kit) with 5' to poooling of tRNA. The RNAs tagged U6 snRNA gene described above was included in each SL RNA transcription reaction. Because this gene contained the same tag as the SL RNA gene templates, the U6 snRNA and SL RNA transcripts were compared in RPA reactions and hybridized by the same probe. The tag was inserted in the U6 snRNA gene in a position such that the primer extension product from the U6 snRNA gene was 22 nt longer than that from the SL RNA gene. Each product was quantitated by true differences in transcription efficiencies and not experimental variation, an internal control was added to each reaction. Specifically, the tagged U6 snRNA gene described above was included in each SL RNA transcription reaction. Because this gene contained the same tag as the SL RNA gene templates, the U6 snRNA and SL RNA transcripts were compared in RPA reactions and hybridized by the same probe. The tag was inserted in the U6 snRNA gene in a position such that the primer extension product from the U6 snRNA gene was 22 nt longer than that from the SL RNA gene. Each product was quantitated by PhosphorImager analysis of dried gels. The ratio of the primer extension products from the U6 snRNA and wt SL RNA product was set as 1. The amount of primer extension product from each variant SL RNA gene template was compared with the U6-derived product within each reaction and expressed as a number relative to wt SL RNA transcription.

**RNase Protection Assays—**The Ambion RPAII kit was used for the RNase protection assay (Ambion). To obtain a DNA template for the production of the specific riboprobe, a PCR was performed using the baculovirus DNA from DE1S, which contains the SL RNA gene, and two SL RNA-specific primers. The sense primer, VB 331, starts at n = 38 of the SL RNA gene; the antisense primer, VB 330, starts at nt + 196 and has the T7 promoter at its 5'-end to allow for labeling in vitro transcription with T7 polymerase. T7 RNA polymerase-dependent transcription produced a [α-32P]UTP-labeled 176-nt riboprobe that could protect 81 nt of the in vitro transcribed tagged SL RNA and 51 nt of the endogenous SL RNA. Trypanosome in vitro transcriptions were performed as described (8), and RNA was precipitated and resuspended in 20 µl of solution A (from the Ambion RPAII kit) with 5 × 10⁸ cpm of riboprobe. The RNAs were hybridized overnight at 42°C, single-stranded regions were digested using RNase T1 and A, and the resultant RNA was precipitated with ethanol and resuspended in formamide loading buffer. Samples were electrophoresed on a 10% polyacrylamide-urea gel in 1 × TBE, and results were visualized using PhosphorImager analysis.

**RESULTS**

**In Vivo Transcription Assays Identify Promoter Element Function in SL RNA Transcription—**Our investigation of a
tractable gene promoter in these ancient eukaryotic protists have established that the SL RNA gene promoter lies within the proximal −100 nt upstream of the transcription start site (see Fig. 1) (6, 7, 17, 18). In vivo promoter analysis has revealed a tripartite promoter architecture (7, 17). However, the limiting component of in vivo analyses is that promoter mutations that produced inaccurately initiated SL RNA, which would turn over rapidly, could not be distinguished from mutations that down-regulated transcription. To investigate the role each promoter element contributes to the transcriptional process, homologous nuclear extracts were produced which could initiate transcription accurately on DNA templates containing the upstream proximal 100 nt adjacent to a guanosine-less coding sequence (G-less cassette) (8). The L. seymouri in vitro transcription extracts contain few RNA-processing enzymes and thus directly assay transcription independently of other nuclear activities. In a recent modification of these assays, shown here, the exogenous SL RNA gene template possesses a 19-nt tag that is transcribed as part of the SL RNA (see “Experimental Procedures” and Ref. 8). RNAs are detected by primer extension reactions using the complement of the tag sequence as the radiolabeled primer. Fig. 2A, lane 1, demonstrates that the SL RNA gene was transcribed in vitro to produce an accurately initiated SL RNA. Detailed transcriptional analysis of mutated templates, drawn schematically below the data, revealed that the two upstream elements (PBP-1E and PBP-2E) of the core SL RNA gene promoter were necessary for efficient transcription. However, these mutations did not effect RNA start site selection (Fig. 2A, lanes 4, 5, 7, and 8). This was surprising because an analogy between snRNA gene promoters in higher eukaryotes and the SL RNA gene promoter would have predicted that PBP-1 and or PBP-2 would be directly responsible for determining transcriptional start sites (19). Unexpectedly, a 10-nt mutation of the third core element, which resides at −1/10 nt, completely abolished proper transcription initiation (lane 2). The in vivo phenotype had been a loss of detectable SL RNA. Clearly, the effect of improperly initiated SL RNAs resulted in their rapid turnover to produce a null phenotype for the −1/10 mutant in vivo.

Because the sub −1/10 mutation produced transcripts that initiated at multiple sites, we deemed it interesting to assess if correct 3′-end formation had been affected similarly. An RNase protection assay was performed on in vitro transcripts using an antisense riboprobe that would recognize SL RNA sequences from nt +38 to nt +196 (Fig. 2B, schematic). The 3′-end of in vivo synthesized RNAs maps immediately upstream of the T stretch beginning at nt +99. Accordingly, the endogenous SL RNAs present within the nuclear extract protected a fragment of 51 nt (Fig. 2B). Correctly terminated and/or 3′-end processing in vitro transcribed SL RNAs would protect a larger, 81-nt fragment because of the presence of the internal 19 nt tag. The 81-nt RNA shown in Fig. 2B demonstrated that the 3′-end of the SL RNA, transcribed from a wt SL RNA gene template, was properly generated in vitro. Interestingly, the 3′-end of the SL RNA transcribed from the sub −1/10 DNA template was also correctly formed. Previous in vivo analysis of SL RNAs by Northern blot analysis is consistent with this finding (7). The significant reduction of the 81-nt product (sub −1/10) compared with the left lane (wt) was consistent with the overall reduction in transcription levels from the mutant template. The 51-nt internal control band (which represents the endogenous RNA) demonstrated that equal amounts of total RNA were included in each RNase protection assay. These results demonstrate the requirement for the −1/10 nt region exclusively in transcription start site selection and not in transcription termination of the SL RNA gene.
1) NT/AYY in which it is crucial for the difference. In metazoans, a loose consensus exists of shown in Fig. 2 directing proper initiation (21–25). As a result of the data stream-most element within the tripartite promoter of the SL trypanosome gene promoters, we tested whether that down-
YR(1) Sequence—
referred to as a trypanosome Inr, or Inrt. However, a sequence comparison between the metazoan Inr and Inrt shows a distinct this distinguishing difference between trypanosome and metazoan Inr, or Inrt. However, a sequence comparison between the metazoan Inr and Inrt shows a distinct difference. In metazoans, a loose consensus exists of YYA(+1)NTAYY in which it is crucial for the +3 position to be an A or T for optimal Inr activity (26). A database survey of the sequence that flanks SL RNA transcription start sites in related Trypanosomatidae reveals a consensus YHYBYA(+1)ACT in which the C (+3) is invariant. Hence, the A/T (+3) found in metazoans is absent in trypanosomatids. Because of this distinguishing difference between trypanosome and metazoan Inr, it is appropriate to refer the trypanosome Inr as Inrt.

To delineate the boundaries of the Inr, element, mutations were introduced in and around the −1/10 nt region (5′-AGACCTTACCA (+1)ACT-3′) of the SL RNA promoter. Initially, each half of the 10 nt region was mutated separately, and mutant templates were used to program nuclear extracts. Fig. 3, lanes 2–4, illustrates the transcription results. Mutation of the −6/10 nt region caused 36% of the transcripts to initiate at the −3 position (lane 4). A comparison of the sequence with the consensus YHYBYA(+1)ACT Inr, showed that a new Inr, had been generated by the substitution of the −6/10 nt region (see Fig. 5). Specifically, the wt sequence from −1/10, which is AGACCTTACCA(+1), had been replaced with TACGGTTACCA(+1). In the mutant construct, a CGTCTA(+1) was recognized at high efficiency to initiate SL RNA transcription. Substitution of the −1/5 nt region (lanes 2 and 3) completely abolished synthesis of properly initiated SL RNAs. Clearly, replacement of the CTACCA(+1) sequence with GATGGA(+1) abolished Inr, function. In the absence of a wt Inr, in the −1/5 mutation, cryptic sites, partially generated by the replacement nucleotides, were recognized by the transcription machinery (see Fig. 5). Consequently, RNAs initiated at several sites, albeit with decreased efficiency. These initiation sites were as follows: CTACCA(+1), located at +8 (relative to the wt (+1) start site); ATGGA(+1), located at +2; CCGATG(+1), located at −2; AGACCC(+1), located at −5. In the case of the start site utilized with the highest efficiency (the −5 site; 50% of total RNAs produced) the initiating purine (G, in this case), was preceded by an ACC trinucleotide that is identical to that within the wt Inr. The consistently best utilized start site in the sub −1/10 mutation (the +8 site, 45% of total RNA produced; fastest migrating band in lane 2) also functioned with modest efficiency in the −1/5 mutation (20% of total RNA synthesized; lanes 2 and 3). This start site is preceded by a CTACCA(+1) sequence that is identical to the wt CTACCA(+1) region except that the −2 position is an A in place of a C, and the purine that is used as the initiation nucleotide is G in place of A. Thus, conservation of at least three nt within the five nt upstream of the start site is clearly important for Inr, activity. In del-Inr, 2, the nucleotides at positions −6 and −7 nt were altered from CC to GA. This alteration had no effect on start site selection (lane 5), nor did this dinucleotide substitution generate a new Inr, sequence. Taken together, these data show that the Inr, is restricted to the five nt adjacent to the initiating purine.

Transcriptional analysis of SL RNA genes in three related trypanosomatids have shown that alterations within the SL RNA sequence had minor effects on both transcription efficiency and start site selection, although these effects were not studied in detail (5, 17, 27). As an important component of our Inr studies, we determined directly if the SL RNA sequence must follow the Inr, immediately. Moreover, by the insertion of four nt, we altered the helical face of the SL RNA sequence relative to the upstream promoter. Any protein-DNA recognition that straddled the upstream and intragenic regions would be disrupted in this mutated template. The insertion mutation, add-C4, did not alter start site selection (Fig. 3, lane 6). In addition, substitution of the entire coding region with pBS sequences did not change the start site, although efficiency was decreased (data not shown). These results indicate that the downstream component of the Inr, includes only the A(+1) nucleotide. In both the pBS substitution and add-C4 mutations, the five nt (CTACC) upstream from the start site were followed by an A(+1)TC trinucleotide sequence, which is similar but not identical to the wt A(+1)AC sequence. The alteration of the A to T at +2 demonstrates that this invariant A, present in SL RNA sequences across all trypanosomatid species, is not essential for Inr function.

The Critical Region of the Inr, Is a Consensus CYA C/A YR(+1) Sequence—In the absence of a TATA box to direct start site selection in any of the known, albeit few, characterized trypanosome gene promoters, we tested whether that downstream-most element within the tripartite promoter of the SL RNA gene contributed directly to accurate RNA initiation (20). A growing collection of higher eukaryotic and yeast genes relies on an Inr element, often without a nearby TATA sequence, for directing proper initiation (21–25). As a result of the data shown in Fig. 2A, lane 2, the downstream-most element maintains functional homology to the metazoan Inr and is now referred to as a trypanosome Inr, or Inrt. However, a sequence comparison between the metazoan Inr and Inrt shows a distinct difference. In metazoans, a loose consensus exists of YYA(+1)NTAYY in which it is crucial for the +3 position to be an A or T for optimal Inr activity (26). A database survey of the sequence that flanks SL RNA transcription start sites in related Trypanosomatidae reveals a consensus YHYBYA(+1)ACT in which the C (+3) is invariant. Hence, the A/T (+3) found in metazoans is absent in trypanosomatids. Because of this distinguishing difference between trypanosome and metazoan Inr, it is appropriate to refer the trypanosome Inr as Inrt.

To delineate the boundaries of the Inr, element, mutations were introduced in and around the −1/10 nt region (5′-AGACCTTACCA (+1)ACT-3′) of the SL RNA promoter. Initially, each half of the 10 nt region was mutated separately, and mutant templates were used to program nuclear extracts. Fig. 3, lanes 2–4, illustrates the transcription results. Mutation of the −6/10 nt region caused 36% of the transcripts to initiate at the −3 position (lane 4). A comparison of the sequence with the consensus YHYBYA(+1)ACT Inr, showed that a new Inr, had been generated by the substitution of the −6/10 nt region (see Fig. 5). Specifically, the wt sequence from −1/10, which is AGACCTTACCA(+1), had been replaced with TACGGTTACCA(+1). In the mutant construct, a CGTCTA(+1) was recognized at high efficiency to initiate SL RNA transcription. Substitution of the −1/5 nt region (lanes 2 and 3) completely abolished synthesis of properly initiated SL RNAs. Clearly, replacement of the CTACCA(+1) sequence with GATGGA(+1) abolished Inr, function. In the absence of a wt Inr, in the −1/5 mutation, cryptic sites, partially generated by the replacement nucleotides, were recognized by the transcription machinery (see Fig. 5). Consequently, RNAs initiated at several sites, albeit with decreased efficiency. These initiation sites were as follows: CTACCA(+1), located at +8 (relative to the wt (+1) start site); ATGGA(+1), located at +2; CCGATG(+1), located at −2; AGACCC(+1), located at −5. In the case of the start site utilized with the highest efficiency (the −5 site; 50% of total RNAs produced) the initiating purine (G, in this case), was preceded by an ACC trinucleotide that is identical to that within the wt Inr. The consistently best utilized start site in the sub −1/10 mutation (the +8 site, 45% of total RNA produced; fastest migrating band in lane 2) also functioned with modest efficiency in the −1/5 mutation (20% of total RNA synthesized; lanes 2 and 3). This start site is preceded by a CTACCA(+1) sequence that is identical to the wt CTACCA(+1) region except that the −2 position is an A in place of a C, and the purine that is used as the initiation nucleotide is G in place of A. Thus, conservation of at least three nt within the five nt upstream of the start site is clearly important for Inr, activity. In del-Inr, 2, the nucleotides at positions −6 and −7 nt were altered from CC to GA. This alteration had no effect on start site selection (lane 5), nor did this dinucleotide substitution generate a new Inr, sequence. Taken together, these data show that the Inr, is restricted to the five nt adjacent to the initiating purine.

Transcriptional analysis of SL RNA genes in three related trypanosomatids have shown that alterations within the SL RNA sequence had minor effects on both transcription efficiency and start site selection, although these effects were not studied in detail (5, 17, 27). As an important component of our Inr studies, we determined directly if the SL RNA sequence must follow the Inr, immediately. Moreover, by the insertion of four nt, we altered the helical face of the SL RNA sequence relative to the upstream promoter. Any protein-DNA recognition that straddled the upstream and intragenic regions would be disrupted in this mutated template. The insertion mutation, add-C4, did not alter start site selection (Fig. 3, lane 6). In addition, substitution of the entire coding region with pBS sequences did not change the start site, although efficiency was decreased (data not shown). These results indicate that the downstream component of the Inr, includes only the A(+1) nucleotide. In both the pBS substitution and add-C4 mutations, the five nt (CTACC) upstream from the start site were followed by an A(+1)TC trinucleotide sequence, which is similar but not identical to the wt A(+1)AC sequence. The alteration of the A to T at +2 demonstrates that this invariant A, present in SL RNA sequences across all trypanosomatid species, is not essential for Inr function.
The choice of alternative sites when the wt CTACCA(1) 1 sequence was destroyed, cryptic sites functioned as retention of either the CTA trinucleotide or ACC trinucleotide initiator activity. Conservation of pyrimidine richness, as well as retention of either the CTA trinucleotide or ACC trinucleotide, followed by a purine, was necessary for initiator function. The choice of alternative sites when the wt CTACCA(1) was mutated or separated from the PBP-2E region provides us with important new information. First, the utilization of sites that insert an AG at positions 5/4 argues that a CY dinucleotide is not essential for Inr function. However, the presence of three out of five nt followed by a purine (ACCG(+1) at −3/+1) in the transcriptional machinery. Instead, cryptic Inr1 were used for transcription initiation (see Fig. 5). Because mutations between any two of the PBP-1E, PBP-2E, and Inr elements (lanes 5–7) were assayed for transcription in vitro. Properly initiated RNA is indicated by the arrow. A control U6 snRNA in vitro transcription was also performed; these data are not included because the U6 and improperly initiated SL RNAs comigrate. The amount of wt SL RNA transcription relative to U6 synthesis was normalized to 1. The transcription efficiency of variant SL RNA templates was determined relative to WT. These values are as follows: add-5 mutant (1.1), add −10 mutant (0.69), del-4 mutant (0.39) add-Inr1, 2 (0.12), and add-Inr1, 5 and 10 variants (0.02). Panel B, transcription assays were performed in the absence (−) or presence (+) of 200 μg/ml α-amanitin. The control U6 snRNA and SL DNA contained the identical tag, therefore a single primer extension reaction identified both the SL RNA (69 nt) and the U6 snRNA (92 nt). Lanes 4 and 5 are duplicates as are lanes 7 and 8.

To confirm the requirement for PBP-1 and PBP-2 in SL RNA transcription, we used a phosphocellulose-fractionated nuclear extract that was transcriptionally competent but differed from the original unfractonated extract in that it contained limiting amounts of PBP-1 and PBP-2. Removal of the remaining PBP-1 and PBP-2 protein using competitor DNA almost completely blocked transcription (Fig. 6B, lane 2). Transcriptional activity was increased 3–4-fold when highly purified PBP-1 and PBP-2 were added to the reaction (lanes 3 and 4). Equivalent amounts of bovine serum albumin did not restore activity (lanes 5 and 6). Restoriation of activity was not expected to be greater than severalfold observed because preinitiation complex assembly, using highly enriched protein fractions, is often inefficient in eukaryotic systems. Because PBP-1 and PBP-2 bind to the PBP-1E and PBP-2E elements in gel shift assays and these two proteins stimulate transcription in in vitro transcription assays, it is probable that PBP-1 and PBP-2 function as bona fide transcription factors in regulating SL RNA expression. In bacterial systems, the holc enzyme of RNA polymerases often recognize promoters in the absence of additional proteins in vitro; this is in contrast to yeast and metazoan RNA polymerases. Because trypanosomatids are of an ancient eukaryotic lineage, these data suggest that the recognition of gene promoters by transcription factors in contrast to RNA polymerase itself is an early evolutionary adaptation.

**DISCUSSION**

An examination of the SL RNA gene promoters from the trypanosomatids in which the data are available has revealed a highly conserved YTHBYA(+1) motif at the transcriptional start site. In the work presented here, in vitro transcription analysis demonstrated that when the L. seymouri wt CTACCA(+)1 sequence was destroyed, cryptic sites functioned as Inr elements. Analysis of these new initiation sites served to highlight the critical regions of the CTACCA(+)1 necessary for initiator activity. Conservation of pyrimidine richness, as well as retention of either the CTA trinucleotide or ACC trinucleotide, followed by a purine, was necessary for initiator function. The choice of alternative sites when the wt CTACCA(+)1 was mutated or separated from the PBP-2E region provides us with important new information. First, the utilization of sites that insert an AG at positions −5/4 argues that a CY dinucleotide is not essential for Inr function. However, the presence of three out of five nt followed by a purine (ACCG(+1) at −3/+1) in the
case where the CY at −5/4 is absent may be thermodynamically important for DNA unwinding and negate the need for the upstream pyrimidines (see Major sites in Fig. 5). Second, the −3 nt position is a T, A, or C in the Inr of all the known trypanosomatid SL RNA gene sequences. Interestingly, our mutagenesis revealed that an A at −3 nt contributed to each major start site selected. This argues for potential protein-DNA interactions that are stabilized when an A is present at position −3 nt. Third, the −2 nt position was less well conserved in both the Inr elements identified by mutagenesis and the Inr elements compiled in the trypanosomatid survey. Finally, the −1 and +1 nt positions were always YR(+1) in the major start sites and tended toward this dinucleotide motif in the minor start sites. This trend is identical to that found in metazoan Inr elements.

The trypanosome Inr sequence is functionally analogous to a metazoan Inr element because it serves to direct correct transcription initiation. The metazoan consensus sequence is loosely defined as YYA(+1)NWYY, which is pyrimidine-rich but otherwise distinct from the trypanosome sequence. Recent analysis of the Inr region from a parasitic protozoan even more anciently diverged than trypanosomes, *Trichomonas vaginalis*, has uncovered an Inr element that contains a highly conserved TCA(+1)YT/A motif (33). Although the *T. vaginalis* and trypanosome Inr elements share a pyrimidine richness with the metazoan Inr, they are distinct from each other.

It is also remarkable that start sites in each case clustered between 22 and 38 nt from the upstream PBP-2E element. The DNA scanning model put forth by Giardina and Lis (34) to explain start site selection by RNA pol II transcription of TATA-containing mRNA coding genes in yeast may be directly relevant to this observation. Specifically, promoter melting, associated with RNA polymerase entry into the preinitiation complex, occurs at a fixed distance downstream from the TATA...

![FIG. 5. Positioning of the start site is determined by the Inr element. PBP-2E is boxed. Arrows indicate the position and relative strength of the start sites that were identified in the transcription assays. The 5 nt that precede each start site are underlined. A broader arrow signifies a stronger start site. The lower case letters indicate nt mutations. The wt transcribed sequence is in large bold letters, and the sequences are aligned based on the wt start site.](image)

![FIG. 6. The three promoter elements bind proteins required for SL RNA transcription. Panel A, competitor DNAs, which contain one or two of the three promoter elements, were added to transcription reactions in 2-, 4-, and 8-fold molar excess over template DNA. The ramp indicates increased amounts of competitor. Properly initiated RNA is marked by the arrow. The diagram depicts the competitor DNAs. The dashed line in the Inr competitor refers to pBS sequences. NS refers to pBS sequences. Panel B, add-back experiments show that PBP-1 and PBP-2 increase transcription. Specific DNA affinity-purified PBP-1 and PBP-2 proteins or bovine serum albumin (10 and 20 μg) was added to transcription reactions in which PBP-1 and 2 had been sequestered by the addition of a 5-fold molar excess of the PBP-1E/2E region. Transcription assay results were quantitated by PhosphorImager analysis. Lanes 2–6 correspond to those in the top of the figure. The filled box shows activity from the PBP-1 and 2-depleted extract. The open boxes refer to the addition of PBP-1 and PBP-2; the striped boxes refer to the addition of bovine serum albumin. Activity is in arbitrary units.](image)
box. However, in the two genes compared in the study, transcription initiation occurs within a 40-nt window from the TATA boxes in the GAL 10 and GAL 1 gene promoters. To account for this finding, RNA pol II is hypothesized to melt the DNA and locate the start site by downstream scanning of the DNA. This scanning would then detect an Inr sequence within the scanning boundaries and appropriately initiate transcription. The trypansomine data presented here are consistent with an active RNA polymerase-scanning mechanism in these early divergent eukaryotes. Specifically, the introduction of a consensus start sites between the PBP-2E region and the wt Inr (mutants add-Inr, 2, 5, and 10) caused the RNA polymerase to initiate transcripts at nucleotides between the PBP-2E region and the wt Inr. Moreover, scanning appears to be restricted to a set window distance of 22–38 nt downstream from the PBP-2E region. In the case of Inr, substitution mutations, the loss of the wt Inr, caused initiation to occur only at similar sequences that were within a limited distance from the PBP-2E region, notably between 22 and 38 nt downstream. Strikingly, consensus Inr sequences, including the wt Inr, located outside of this ~16-nt window failed to be recognized. In the case of the sub–1/10 template, the wt site was completely nonfunctional, suggesting that it was outside the scanning window of the polymerase complex. Consequently, RNA polymerase located the CTAGCG (+1) element, with limited homology to the CYAC/AYR (+1) consensus but proximal to the PBP-2E region, during its scanning activity. Recent data in a DNA replication system in a prokaryotic organism suggest that the threading of DNA through a protein complex can occur in cells (35). These data support the mechanism proposed for RNA polymerase scanning for initiation sites within transcriptional promoters.

The SL RNA gene promoter is tripartite, containing three closely spaced elements that all reside within the upstream 85 bp of the intergenic region (5, 7, 17). Two of these elements interact with sequence-specific DNA-binding proteins. In the case of the two upstream-most elements, two transcription factors have been identified and shown to be necessary for transcription in vitro. Evidence for a third transcription factor, which interacts with the Inr element, is provided by the sequestration assay (Fig. 6). This complex promoter structure and the identification of two (possibly three) cognate protein interactions are significantly different from the much simpler snRNA basal promoters found in other eukaryotic organisms. In the case of snRNA genes in yeast and metazoans, a single proximal sequence element (PSE) element is recognized by a cognate SNAPc protein to nucleate the complex that directs scanning for initiation sites within transcriptional promoters.

Acknowledgments—We thank Jeffrey Wilusz, Michael Hempsey, and members of the Bellofatto laboratory for many helpful comments on this work.

REFERENCES

Transcription Initiation at the TATA-less Spliced Leader RNA Gene Promoter Requires at Least Two DNA-binding Proteins and a Tripartite Architecture That Includes an Initiator Element

Hua Luo, Gwen Gilinger, Devi Mukherjee and Vivian Bellofatto

doi: 10.1074/jbc.274.45.31947

Access the most updated version of this article at http://www.jbc.org/content/274/45/31947

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 36 references, 14 of which can be accessed free at http://www.jbc.org/content/274/45/31947.full.html#ref-list-1