Alterations in the pre-mRNA topology of the bovine growth hormone polyadenylation region decrease poly(A) site efficiency

Edward R.Gimmi[1,2+], Mitchell E.Reff[1,2+] and Ingrid C.Deckman[2]*

[1]Department of Microbiology and Immunology, Temple University School of Medicine, Philadelphia, PA 19140 and [2]Department of Molecular Genetics, Smith Kline and French Laboratories, 709 Swedeland Road, King of Prussia, PA 19406, USA

Received May 10, 1989; Revised and Accepted July 31, 1989

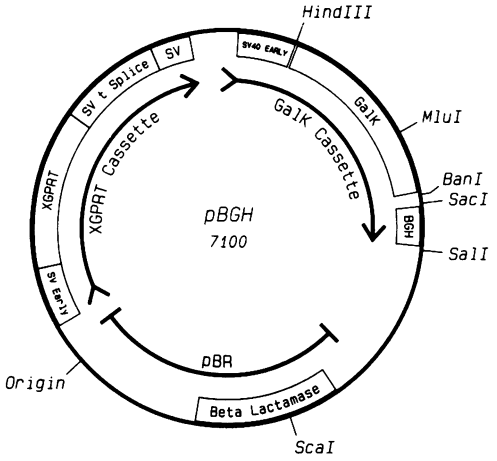
ABSTRACT

RNase mapping experiments show that the bovine growth hormone (bGH) poly(A) region forms an extensive hairpin loop. Mutants were prepared to change poly(A) region pre-mRNA structure and cleavage site efficiency without altering necessary sequences. An inverted repeat which includes the poly(A) cleavage site was created by insertion of a linker upstream of the poly(A) region to compete with any wild-type secondary structure. RNA mapping analyses show alterations in the nuclease accessibility of this mutant at the natural site of cleavage. This mutant shows a 75% drop in relative reporter gene expression at the steady-state protein and RNA levels. When the linker is inserted as a direct repeat, expression is equivalent to wildtype levels. To show that transcription was not terminated by the inverted repeat, the SV40 late poly(A) region was inserted downstream. These mutants show restored expression and processing at the downstream site. Our experiments reveal that the conformation of the poly(A) site pre-mRNA is important in mediating efficient cleavage-polyadenylation.

INTRODUCTION

The signals necessary for mediating poly(A) and processing at the 3'-end of RNA pol II transcripts remain unclear. Sequence and deletion analyses have revealed that a single specific necessary sequence, 5'-AAUAAA-3', or a close variant, is found in all poly(A) regions. Although sufficient for cleavage, this hexanucleotide is not sufficient for efficient 3'-end formation (1,2,3,4). Several studies have identified GU or U-rich motifs as downstream elements which function with the AAUAAA to mediate efficient processing. None of these downstream sequences are found in all poly(A) regions (5,6,7,8). In studying the sequences of several pre-messenger RNAs and deletion mutants of the SV40 late poly(A) region, we discovered a relationship between the efficiency of cleavage-polyadenylation and the thermodynamic free energy of formation (ΔG) for potential RNA secondary structures. A similar relationship was suggested for deletion mutatations in the bGH poly(A) region (8). Because there do not appear to be universal sequence signals other than the hexanucleotide, we decided to investigate the idea that structural features of the RNA provide information for efficient and accurate cleavage-polyadenylation (1,9,10,11,12).

The research presented in this paper focuses on mutations which alter pre-mRNA structure in the bGH poly(A) region and their effect on the efficiency of poly(A) site utilization. In order to study the effect of structural alterations in this pre-mRNA without changing the sequences near the poly(A) site, we designed a linker complementary to a region which includes the natural cleavage site of the bGH RNA. By positioning the linker upstream of the poly(A) signal, we were able to disrupt the wild type RNA structure, determined by *in vitro* mapping, as well as lower the efficiency with which the site is used. By reversing



BGH= Bovine Growth Hormone Polyadenylation Region
SV= SV40 Early Polyadenylation Region

Figure 1. Vector map of pBGH showing the transcriptional cassettes and control regions. Hind 3 and Sal 1 sites delineate the insert used in the antisense RNA 3'-end mapping vectors. Mlu 1 was used to linearize the end mapping vectors. Ban 1 and Sac 1 sites are to show the position of the deletion creating pΔ5. All of the vectors were linearized with Sca 1 prior to transfection.

the orientation of the linker, a mutant was created where the poly(A) site is as efficient as wild type. The secondary structure mapping of the *in vitro* synthesized RNAs, from the mutants that exhibited low expression *in vivo*, show changes in nuclease sensitivity in the poly(A) region. Our results show that pre-mRNA topology has an influence on the efficiency of bGH poly(A) site utilization.

MATERIALS AND METHODS

Cells, cell culture, transfection and enzyme analysis.

R1610 hamster lung fibroblasts (galactokinase minus)(13) were carried in DMEM with 5% fetal calf serum. 10μg of purified linearized vector DNA was used to transfect 10^6 cells by calcium phosphate coprecipitation (14). After 48 hours the cell lysates were prepared as described previously (15). Galactokinase and xgprt enzyme levels were analyzed by filter assays (15). For each data set two reactions were performed at two time points (15 and 30 min.) in order to demonstrate linearity of enzyme kinetics (15,16).

Vectors

The parental vector pBGH was constructed by D. Pfarr as a modified version of pDSP1BGH(16) (See figure 1). This vector contains a fragment of the bGH poly(A) region with 90 base pairs upstream of the AATAAA and 134 base pairs downstream. In this paper nucleotide 1 refers to the first nucleotide following the AAUAAA. Two reporter gene cassettes, galactokinase and xanthine-guanine phosphoribosyl transferase (xgprt) from *E. coli* are included in this vector. The galactokinase gene provides a way to measure expression mediated by mutants in the bGH poly(A) region introduced downstream of the gene. To control for transfection efficiency expression is measured from the xgprt gene in the vector. This vector system has been employed previously to quantitate relative

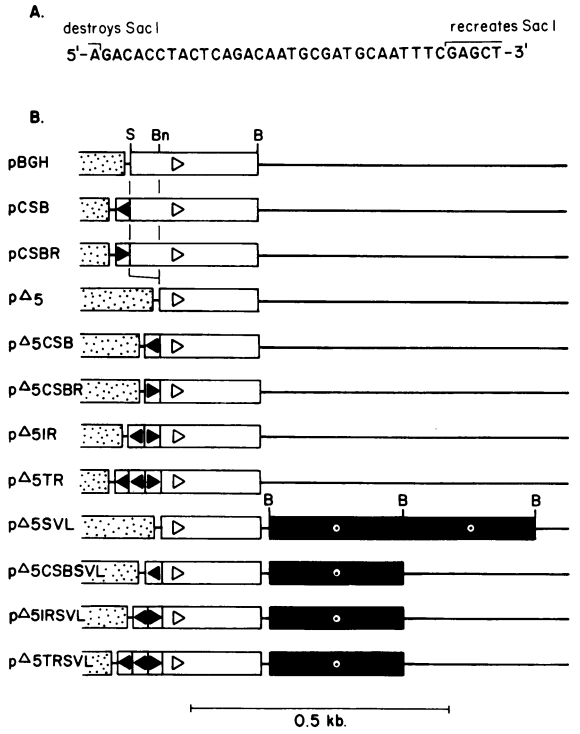
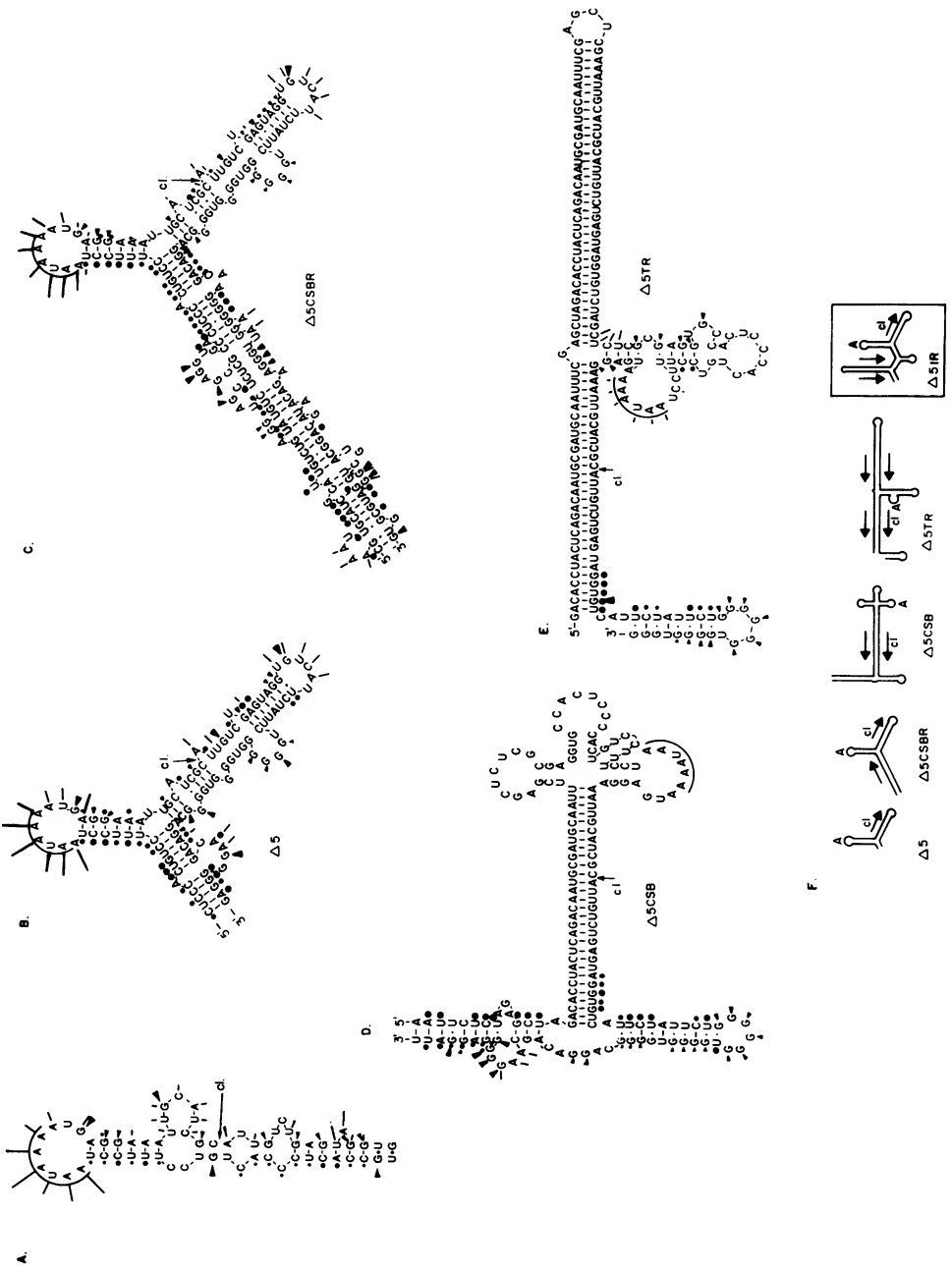


Figure 2. A)Sequence of the CSB (cleavage site blocker) synthetic oligonucleotide. Upon insertion this linker adds 33 base pairs to the vector. B)Bar diagrams of the linker mutants constructed in the bovine growth hormone polyadenylation region. Open triangles represent the wild type poly(A) addition site. Solid triangles represent the CSB linker and its orientation with respect to the natural cleavage site. Open circles containing dots represent the natural SV40 late poly(A) addition site. S=Sac 1, Bn=Ban 1 (not unique in the vector) and B=Bam H1. The brackets shown above the pΔ5 schematic indicate the position of the upstream Sac 1-Ban 1 deletion.

expression levels (15,16,17). In the experiments reported here, vectors were linearized with Sac 1 prior to transfection. These vectors remain linear for the duration of the transient transfection analyses (17). Linearization was done to prevent galactokinase transcription from interfering with the xgprt transcription, since alterations were made at the 3′ end of the galactokinase gene. Data not presented in this publication indicates that the relative expression from supercoiled vectors is identical to the relative expression from linear vectors (17). The vector pΔ5 has a 74 bp Sac 1-Ban 1 deletion into which a 9 bp Sac 1-Ban 1 linker was inserted 31 nucleotides upstream of the AATAAA (figure 2). Into the Sac 1 site of pBGH and pΔ5 a 33 bp linker (CSB) was inserted to create pBGHCSB, pBGHCSBR, pΔ5CSB,pΔ5CSBR, pΔ5IR and pΔ5TR (figure 2). The SV40 late poly(A) region, including 133 bp upstream of the AATAAA and 108 bp downstream, was inserted into the Bam Hl site of pΔ5 and linker mutants to create pΔ5SVL, pΔ5CSBSVL, pΔ5IRSVL and pΔ5TRSVL (figure 2). pGem3Δgalk and pGem3Δxgprt were described previously (17). A EcoR 1-Pvu 2 minor fragment was isolated from pBGH, pΔ5, pΔ5CSB, pΔ5CSBR and pΔ5TR and ligated into pGem4 digested with EcoR 1 and Hinc 2 to create vectors for production of poly(A) region RNAs for structural mapping. A Hind 3-Sal 1 minor


A.
B.
Δ5
C.
ΔSCSBR
D.
ΔSCSB
E.
ΔSTR
F.
Δ5
ΔSCSBR
ΔSCSB
ΔSTR
ΔSIR

fragment from pΔ5SVL, pΔ5CSBSVL and pΔ5TRSVL was inserted into pGem3Z to create pGem3ZΔ5SVL, pGem3ZΔ5CSBSVL and pGem3ZΔ5TRSVL which were used to probe the position of the mRNA 3′ ends. All recombinant DNA techniques were done using standard protocols (18).

Sequencing

Sequencing was carried out using the Sequenase kit from United States Biochemicals.

Isolation of RNA and RNase protection assays

Cytoplasmic RNA was isolated 48 hours post transfection. RNA was extracted by the lithium chloride precipitation method as described previously (19). RNase protection assays were performed by the method of Melton as modified by Toscani (20,21). Probes were synthesized from galactokinase and xgprt gene constructions using the bacteriophage T7 promotor in pGem3. These probes are designed to protect transcript coding sequence and are distinguishable by size on acrylamide denaturing gels. 5−10 μg of cytoplasmic RNA was hybridized with both, or a single probe, at 55°C for 12 hours. The RNA was digested with RNases T1 and A, run on a 10% acrylamide denaturing gel and dried onto Whatman 3MM paper for autoradiography. Films were scanned by densitometry with an LKB laser densitometer at various exposures to ensure the linearity of the film response. Peak areas were determined by integration and final tabulations show the relative ratio of the galactokinase to xgprt peak area (17). 3′-end mapping was performed using antisense RNA probes synthesized from the T7 promoter in pGem3Z.

RNA secondary structure mapping of in vitro synthesized bGH polyadenylation region mutants

RNAs to be mapped were synthesized in pGem4-based vectors using SP6 bacteriophage polymerase basically as described by Melton (20). Structure mapping experiments were performed as described by Deckman and Draper (22). RNAs were 5′-end labelled using T4 kinase and ^{32}[P]-γATP. Several of the RNAs (CSB and TR) were not readily 5′end labelled (probably due to the inaccessibility of the 5′end to the kinase because of secondary structure). These RNAs were 3′-end labelled with ^{32}[P]-pCp and T4 RNA ligase. Labelled RNAs were gel purified and renatured in a buffer containing 10 mM Tris pH 7.6, 0.2 M NaCl, 5mM $MgCl_2$ at 42°C for 10 minutes. Reactions with V1, T1 and T2 RNases were performed in this buffer with several dilutions of enzyme at both 37° and 0°C, and run on a 85 cm gel at 40 volts for 16 to 20 hours. Gels were dried and exposed to film. The extent of the sequence labelled is as follows: BGH, −127 to 134; Δ5, −56 to 134; ΔCSB, −90 to 134; ΔCSBR, −90 to 134 and ΔTR, −158 to 134. All of these RNAs contain a 26 nucleotide stretch from the pGem4 polylinker upstream of the polyadenylation region-specific sequence.

Figure 3. A−E)Stem loop structures predicted for the synthetic RNAs based on RNase digestion and computer modelling. Structures determined from the nuclease accessibility data in Table 1 used in tandem with computer algorithm prediction. The structures of pBGH, pΔ5, pΔ5CSBR, pΔ5CSB and pΔ5TR are shown. Solid dots represent double-strand specific nuclease V1 cuts. Triangles represent single-strand nuclease T1 cuts. Bars represent single-strand specific T2 cuts. The relative intensity of the cuts is indicated by the size of the symbol on the structure. An arrow and the letters 'cl.' denote the position of the natural cleavage site within 3 bases (8). The AAUAAA is indicated by a curved line. E)Schematic showing the position of the natural cleavage site (indicated by 'cl.' and an arrow indicating its orientation) and the orientation of the CSB linker(s) denoted by arrows. The structure of Δ5IR, as predicted by the computer algorithm is shown in this diagram for comparison, however, this RNA structure was not mapped *in vitro*.

TABLE 1. Nuclease accessibility data from RNA structure mapping experiments in the hexanucleotide and cleavage site region.

		-10	1	10	20
		CUGUCCUUUCCUAAUAAAAUGAGGAAAUUGCAUCGCAUUGUCUGAG			
V1	BGH	2 232322	3		
	Δ5	333 22232332	2 2	2	2 2 33
	Δ5CSBR	222 22233332	11 1	2	1 1
	Δ5TR	221			
T1	BGH	2	2 11	3 1	1 1 1
	Δ5	0	3 22	1 1	3 2 0
	Δ5CSBR	0	2 22	1 1	2 1 1
	Δ5TR	0	0 22	0 0	0 0 0
T2	BGH	4324322	1	1111111	2 1 1
	Δ5	44344323 4			2 2
	Δ5CSBR	1443342111		1	111 1
	Δ5TR	1 2112211 1	1		

Nuclease accessibilities are graded on a relative scale of 1-4, with 4 representing the most intense cutting. The relative intensity of cutting of single-strand (T1, T2) and double-strand specific (V1) ribonucleases is indicated for the region including the AAUAAA and the natural cleavage site (both underlined). Zeros indicate guanylate residues which are totally T1-insensitive. Δ5CSB was not included in this table since no cleavage was detected in the region shown.

Computer Analysis

To help fit the nuclease data to the molecular structures, the algorithm in the FOLD program was used (23). This algorithm predicts the structure with the lowest free energy of formation based on stacking and loop destabilizing energy values (24). In all cases where nuclease data is available the data correlated very well with the predictions.

RESULTS

Structural Features of an Efficient Poly(A) Region

The bGH poly(A) region was chosen for this study because the computer predicted structure was the longest and most stable stem-loop of any poly(A) region analyzed. Furthermore, previous results show this polyadenylation region mediated the most efficient relative expression of any region analyzed (16).

In order to examine RNA structure, labelled transcripts of the bGH poly(A) region were produced *in vitro* using SP6 polymerase and probed with single- and double-strand specific nucleases. Nuclease V1, which cleaves double-stranded RNA, was used to probe helices within the *in vitro* folded structure. RNases T1 and T2, were used to probe for single stranded regions in the RNA. T1 cuts 3′ to guanylate residues and T2 attacks on all phosphodiester bonds with a preference for adenylate residues. Nuclease data were used in tandem with a computer program to determine the RNA secondary structure (23,24). These data are presented in figure 3 and summarized in Table 1. In the computer structure of the bGH region the AAUAAA is predicted to be in a hairpin loop and the cleavage

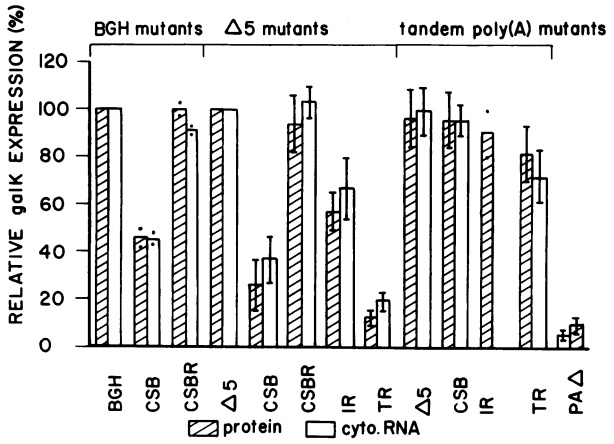


Figure 4. Bar graph showing the relationship between the relative galactokinase enzyme levels and the levels of steady-state cytoplasmic RNA for the bGH mutants. Cross hatched bars represent a compilation of the relative galactokinase enzyme levels. Empty bars represent a compilation of the relative densitometric data from RNase protection assays. All assay data has been normalized to either the pBGH or pΔ5 vector data, which are equivalent and considered 100% for these experiments, therefore no error bars are shown for these vectors. Mutant vector data are shown with error bars which represent 4 or more enzyme assays or 3 or more RNase protection assays from independent transfections. Data shown with 2 points indicates an average of two independent transfection assays. Superscripts over the graph field indicate which mutant series is illustrated; BGH mutants are insertions in pBGH, Δ5 mutants are insertions in pΔ5 and tandem poly(A) mutants are pΔ5 mutants containing the SV40 late polyadenylation region downstream of the bGH polyadenylation region.

site is included in a bulge of an extensive helix (figure 3A). A similar overall structure for the bGH poly(A) region pre-mRNA has been suggested previously (8). Nuclease mapping reveals that the AAUAAA is extremely sensitive to nuclease T2 at both 0°C and 37°C (Table 1). In this paper nucleotide 1 refers to the first nucleotide following the AAUAAA. A strong helix is suggested by extensive V1 cutting between bases 112 to 130. These bases form the stem of the stem loop. A pattern of weak T1 and T2 cuts is displayed at 0° between bases 10−19. These cuts intensify at 37°C, indicative of unpaired bases. This region correspondes to a bulge region in the extensive helix (figure 3A) and contains the cleavage site of polyadenylation.

In order to measure relative expression mediated by different poly(A) regions the galactokinase (galk) reporter gene vector system was used. The galk system has been used to measure the effects on transient gene expression in mammalian cells when polyadenylation regions are ligated downstream from the reporter gene (15,16,17). We believe that the alterations in the relative expression reflect the efficiency of cleavage-polyadenylation. The relative level of either steady-state galactokinase messenger RNA or galactokinase enzyme activity is calculated by normalizing the expression to the second reporter gene, xgprt, which provides an internal control for the efficiency of independent transfections. Figure 4 is a compilation of relative galactokinase activities (galk/xgprt) observed for the various constructions. In order to measure the level of steady-state messenger RNA in this system, RNase protection assays were performed using probes to the coding sequence (figure 5). Previous studies on poly(A) region mutants using this vector system showed that the relative enzyme data correlates with the relative steady-state messenger RNA data in every case tested (15,16,17).

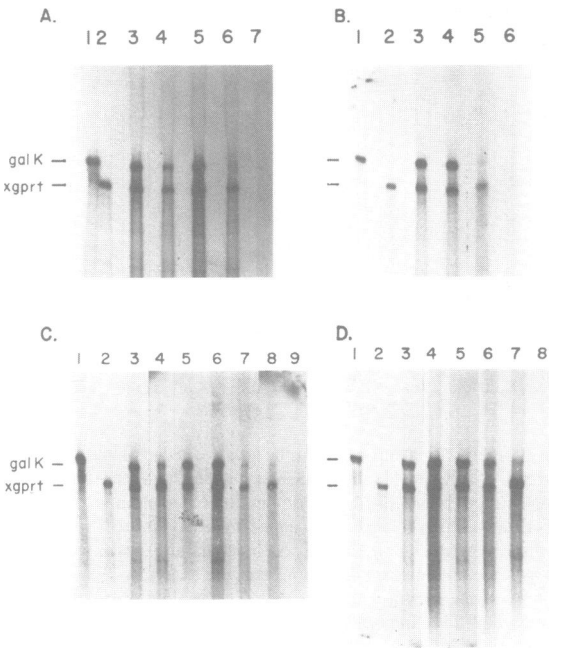


Figure 5. RNase protection assay of galactokinase and xgprt steady-state cytoplasmic RNA levels from linker insertion mutants of the bovine growth hormone polyadenylation region in R1610 hamster fibroblasts. 10μg. of cytoplasmic RNA isolated from R1610 cells at 48 hours post-transfection was either hybridized to an antisense galactokinase, an xgrpt RNA probe separately, or these probes were mixed. The hybridized RNAs were subjected to RNase A and T1 digestion, denatured and a run on a 10% acrylamide-8 M urea denaturing gel. These fragments represent regions which protect message coding sequence. The positions of the galactokinase and xgprt protected fragments are indicated. A)Introduction of a linker complementary to the sequences surrounding the cleavage site. Lane 1:pBGH, galactokinase probe alone. 2:pBGH, xgprt probe alone. 3:pBGH, mixed probes. 4:pCSB, mixed probes. 5:pCSBR, mixed probes. 6: pPAΔ, mixed probes. 7:R1610 RNA, mixed probes. B)Deletion upstream of the AAUAAA in pBGH to create pΔ5. Lane:pΔ5, galactokinase probe alone. 2:pΔ5, xgjprt probe alone. 3:pBGH, mixed probes. 4:pΔ5, mixed probes. 5:pPAΔ, mixed probes. 6:R1610 RNA, mixed probes. C)Lane 1:pΔ5, galactokinase probe only. 2:pΔ5, xgprt probe only. 3:pΔ5, mixed probes. 4:pΔ5CSB, mixed probes. 5:pΔ5CSBR, mixed probes. 6:pΔ5IR, mixed probes. 7:pΔ5TR, mixed probes. 8:pPAΔ, mixed probes. 9:R1610 RNA, mixed probes. D)Lane 1:pΔ5, galactokinase probe alone. 2:pΔ5 xgprt probe alone. 3:pΔ5, mixed probes. 4:pΔ5SVL, mixed probes. 5:pΔ5CSBSVL, mixed probes. 6:pΔTRSVL, mixed probes. 7:pPAΔ, mixed probes. 8:R1610 RNA, mixed probes.

By deleting 74 nucleotides upstream of the AAUAAA in the bGH poly(A) region, a smaller construct, pΔ5 was prepared (figure 2). This deletion does not remove any of the downstream sequences reported to be essential for efficient polyadenylation. RNase mapping revealed that important sequences for polyadenylation, the hexanucleotide AAUAAA and the cleavage site, had similar structural features, in both pΔ5 and pBGH (figures 3A and 3B and table 1). Relative galactokinase expression data shows that pΔ5 is equivalent to pBGH at both the level of steady state RNA and protein (figures 4 and 5B).

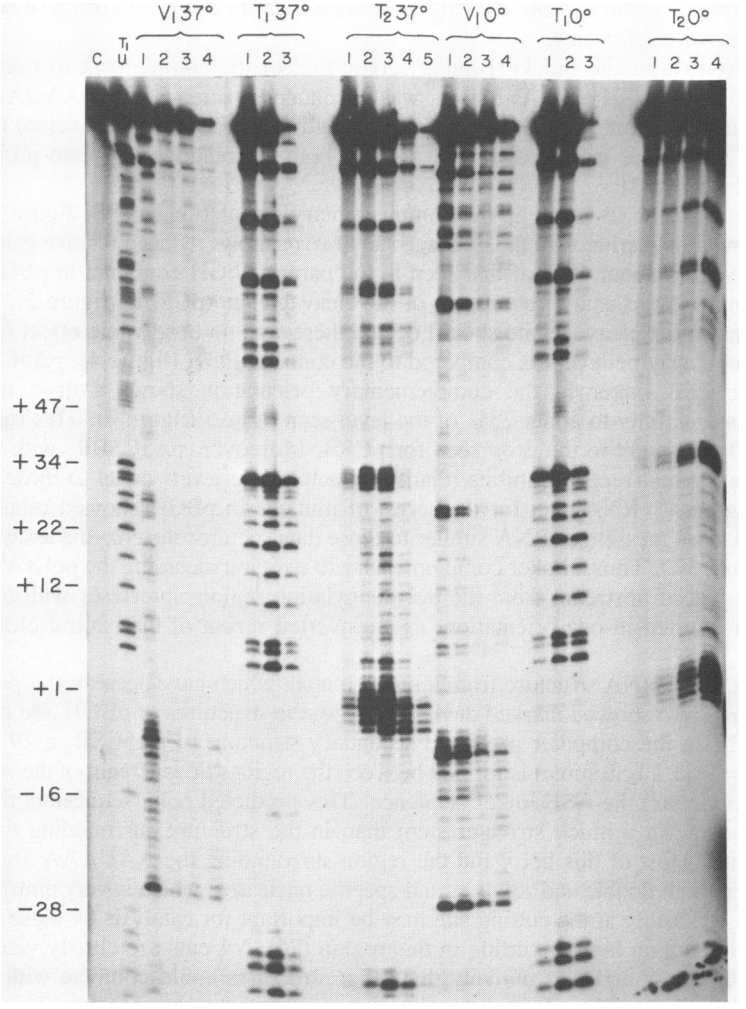


Figure 6. *Example of a Structure Mapping Gel.*
Autoradiograph of partial V1, T1 and T2 digestion of CSBR RNA run on an 8% denaturing acrylamide gel. Digestion was carried out for 5 minutes at 0° or 37°C as indicated, using various dilutions of ribonuclease in each case. Reactions were run in two loadings in order to get clear data throughout the sequence. The structure determination for each RNA was performed at least three times. Cleavage intensities were assigned by eye, and are reported in Table 1. Cleavage intensities assigned by eye have been shown to be as consistent as densitometry scanning data for these experiments (Kean and Draper (1985) Biochemistry 124, 5052−5070). Results for each RNA were quite consistent, especially in areas of very strong cleavage intensity (i.e. AAUAAA in this gel) at position −6 to −1. C is the control lane with no ribonuclease added. T_1U is digestion with RNase T1 in a denaturing 8M urea buffer which produces a cut after each guanidine residue in order to locate the position of bands in a sequence. Also run on each gel was an alkaline hydrolysis lane (not shown) which cleaves after each base to produce a ladder.

A linker complementary to the cleavage site region alters the RNA structure and lowers the cleavage site efficiency.

The CSB (cleavage site blocker) linker, composed of the complement to nucleotides 6 through 36 of the bGH poly(A) region, was introduced upstream of the AAUAAA signal in both orientations into pBGH to create pCSB and pCSBR (linker in reverse) (figure 2). Similarly, this linker was inserted into pΔ5 in both orientations to create pΔ5CSB and pΔ5CSBR (figure 2).

Insertion of the CSB linker in the complementary orientation (pCSB, figure 2), which should allow base pairing with the cleavage site *in vivo*, shows reduced relative galactokinase enzyme levels to about 40% of that seen in the parental bGH construct in pBGH (figure 4). When introduced as a direct repeat of the cleavage site (pCSBR, figure 2), where no base pairing to the cleavage site should occur, there was no observable effect on relative galactokinase enzyme levels as compared to the control pBGH (figure 4). pΔ5CSB which contains a CSB linker in the complementary orientation shows a drop in relative galactokinase activity to about 25% of the level seen in pΔ5 (figure 4). This result is not significantly different to the drop seen for pCSB. Moreover, pΔ5CSBR, with the linker inserted as a direct repeat, exhibits relative galactokinase levels equal to those observed in pΔ5 (figure 4). RNA data for the series of mutants in pBGH showed relative levels of galactokinase messenger RNA similar to those data seen for the enzyme assays (figures 4 and 5A and 5C). Thus a linker complementary to a region including the poly(A) cleavage site, when placed upstream from the polyadenylation region, interferes with expression only when inserted in one orientation, as an inverted repeat of the natural cleavage site region.

Analysis of the RNA structure from a linker mutant which gave decreased reporter gene expression *in vivo* showed marked deviations from the structures of pBGH and pΔ5 (table 1, figure 3). In the computer predicted secondary structure of pΔ5CSB, a 29 base pair helix ($\Delta G = -53.1$ kcal/mole) is formed between the regions downstream of the AAUAAA hexanucleotide and the CSB linker sequence. This predicted helix sequesters the natural site of cleavage in a much stronger stem than in the structure surrounding the natural cleavage site. Most of this helix and the region surrounding the AAUAAA are nuclease insensitive to both double and single strand-specific nucleases, which is very unusual (figure 3D). RNA flexibility at the cutting site may be important for catalysis by these nucleases which could explain lack of cutting in this region (26). V1 cuts are clearly visible at the 3′ end of the helix and it is unlikely that other structures could compete with the helix in this region (figure 3D). Therefore, this stem structure probably forms. This topology cannot be verified, but the decreased intensity of this RNA to T1 cleavage in the presence of the denaturant urea (8 M) suggests the presence of extensive base pairing (data not shown).

The construct pΔ5CSBR, which showed reporter gene expression levels equal to that of the parental bGH poly(A) region yielded RNA structural data similar to that seen for the parental region. The hexanucleotide is predicted to be in a hairpin loop and the cleavage site is predicted to be in the bulge of an extensive helix (figure 3C and figure 6). Therefore, these structural elements were shared by all of the mutants which mediated high reporter gene expression (pBGH, pΔ5, pΔ5CSBR).

Linker multimers also characterized were an inverted repeat, pΔ5IR, and a triplet, pΔ5TR (figure 2). The mutant containing an inverted repeat of the linker (pΔ5IR) showed a relative galactokinase activity of 60% of the control pΔ5 (figure 4). This value was also observed

for the relative steady-state cytoplasmic RNA levels and is intermediate between the level seen for the parental vector and the single linker mutant in pΔ5CSB (figures 4 and 5C). The mutant pΔ5TR contains three copies of the CSB linker. Two form a direct repeat and the other forms an inverted repeat. It resembles the structure of pΔ5IR with an extra linker (figure 2). pΔ5TR, however, shows a dramatic drop in relative enzyme activity to 13% of the activity observed for the controls and similar relative RNA levels (figures 4 and 5C). In summary, expression after the addition of a second linker is higher than the single linker mutant (pΔ5IR vs pΔ5CSB) and addition of a third linker shows activity which is lower (pΔ5TR vs pΔ5CSB).

The RNA structure of pΔ5TR differs from the parental RNA structure. This transcript is computer predicted to have two extensive RNA helices of 30 and 37 base pairs (figure 3E). The 30 base pair helix sequesters the natural cleavage site as in pΔ5CSB (figure 3D). These predicted helices are totally resistant to RNAse digestion, but a V1-sensitive area again appears at the 3′ end of the helix formed by complementary linker sequences, suggesting that these stem structures are present (figure 3E and table 1). Computer predicted structures show that the AAUAAA is in a single stranded bulged region of a hairpin loop stem (figure 3E). Nuclease data confirm that the AAUAAA is slightly sensitive to the single strand-specific nuclease T2 with weak double strand-sensitive cuts on the 5′ side in the predicted stem loop (figure 3E and table 1). However, most of the guanylate residues in this region are sensitive to T1 which suggests that any structure which is forming is weak and unstable.

Therefore, both mutants exhibiting lower levels of expression, pΔ5CSB and pΔ5TR probably contain natural cleavage sites sequestered in helices which are more stable than the stem loop of the parental transcript. Moreover, the AAUAAA, which is extremely T1 sensitive and present in a hairpin-loop in all of the mutants showing high level reporter gene expression, is either completely nuclease insensitive or only partially sensitive in those linker mutants exhibiting reduced expression.

Blockage of Processing in the bGH Poly(A) Region allows Processing at a Heterologous downstream Poly(A) Site.

A series of polyadenylation region chimerae were constructed to show that the lowered expression observed from the linker mutants was not due to transcription termination but was mediated by inefficient cleavage-polyadenylation and presumed degradation of the unprocessed RNAs. An intact SV40 late poly(A) region was ligated downstream from mutants of the bGH poly(A) region which showed drops in reporter gene expression. If the linkers inhibited expression by blocking cleavage-polyadenylation then readthrough messenger RNA would be processed at the SV40 late 'catcher' region downstream, elevating the observed relative expression. This processing in the SV40 late poly(A) region would not occur if the inserted linkers in the bGH poly(A) region mediated a transcription termination event.

pΔ5SVL, which contains an SV40 late poly(A) region downstream from the bGH region mutant exhibits relative galactokinase enzyme levels equal to those of its parental vector pΔ5 (figure 4). The mutant pΔ5CSBSVL, derived from pΔ5CSB by the introduction of a downstream SV40 poly(A) region, has relative galactokinase levels restored to that of pΔ5 (pΔ5, 100%; pΔ5CSB, 25%; pΔ5CSBSVL, 100%) (figure 4). In every case the relative levels of galactokinase messenger RNA were identical to the relative enzyme values (figures 4 and 5D). Therefore, placement of a second heterologous polyadenylation region downstream from the linker mutants showing lowered expression exhibit relative expression values equivalent to the parental vector.

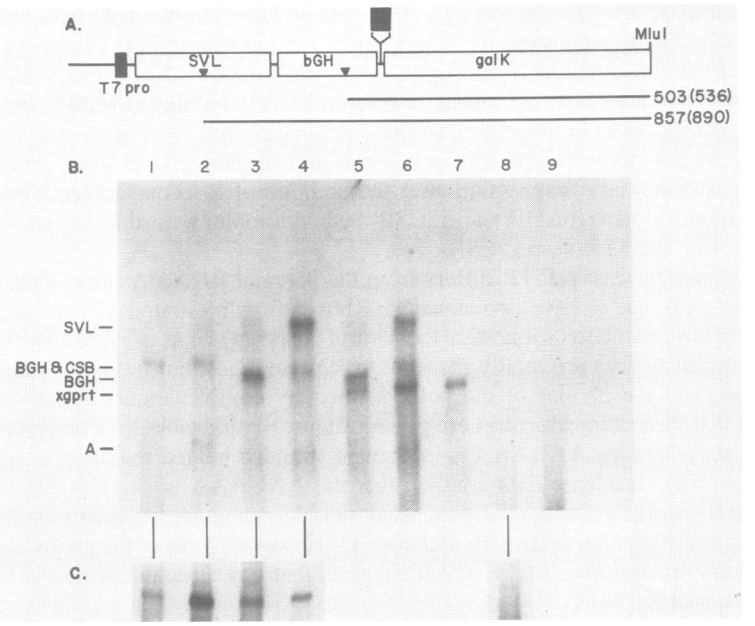


Figure 7. RNase protection mapping of the 3'ends of the bGH-SV40 late polyadenylation region chimerae in R1610 hamstser lung fibroblasts. 10μg. of cytoplasmic RNA isolated at 48 hours post-transfection was hybridized to labelled anti-sense RNA probes complementary to a region spanning the terminal portion of the galactokinase coding sequence and the polyadenylation region(s). The assay is essentially as described in the legend of figure 4. A)Bar diagram of the vector used to generate labelled transcripts for termini mapping. These vectors were linearized with Mlu 1 and transcripts are produced off of the T7 promoter. The black box indicates the CSB linker insertion in pGem3zCSBSVL. bGH denotes the Δ5 bovine growth hormone deletion polyadenylation region. SVL denotes the SV40 late polyadenylation region (there are two copies flanking pGem3ZΔ5SVL and a single copy downstream in pGem3ZCSBSVL). Solid triangles denote the poly(A) addition sites. Protected fragment sizes are shown below for pΔ5SVL and in parentheses for pΔ5CSBSVL, ending in bGH pol(A), SV40 late poly(A). B)Lane 1:pΔ5, pGem3ZΔ5SVl probe. 2:pΔ5CSB, pGem3ZCSBSVL probe. 3:pΔ5SVL, pGem3ZSVL probe. 4:pΔ5CSBSVL, pGem3ZCSBSVL probe. 5:pΔ5SVL, pGem3ZΔ5SVL and xgprt probes. 6:pΔ5CSBSVL, pGem3ZCSBSVL and xgprt probes. 7:pΔ5CSBSVL, xgprt probe. 8:R1610 RNA, pGem3ZΔ5SVL, pGem3ZCSBSVL and xgprt probes. 9:mixed probes, no RNA. Protected fragment sizes are indicated. The xgprt probe protects a 46S bp fragment and is used to indicate size. The 'A' indicates a reproducible minor hybridization product which is believed to be due to intramolecular hybridization of the CSB and cleavage site sequences in the probe, the cytoplasmic RNA or both. This would generate a small protected galactokinase fragment. Since this product is a protected fragment from a competing hybridization reaction and represents a constituent of the cytoplasmic RNA population being measured, densitometric data from this band are included in the quantitation. C)Xgprt probe mixed with the RNAs analyzed to normalize for RNA amount. Lane 1:pΔ5. 2:pΔ5CSB. 3:pΔ5SVL. 4:pΔ5CSBSVL. 5:R1610 RNA.

RNase protection was performed to determine the location of the 3′ ends of the steady-state messenger RNA. The structure of these probes is shown in figure 7A. These probes contain a fragment of the galactokinase cassette which allowed for more efficient hybridization (data not shown), and the bGH and SV40 late polyadenylation regions. Figure 7B shows that both pΔ5 and pΔ5CSB contain 3′ ends within the bGH polyadenylation region generating protected fragments of 503 nucleotides for pΔ5 (lane 1), and 536 nucleotides for pΔ5CSB (lane 2), which is slightly larger due to the CSB linker insert.

A lower protected pΔ5CSB band (lane 2, band 'A') represents a protected fragment and was used in the quantitation (refer to figure legend). The relative level of 3′ ends seen for pΔ5CSB (figure 7B, lane 2) is 20% of that seen for the parental vector (pΔ5)(figure 7B, lane 1) when added band intensities are normalized to the internal control xgprt transcript (figure 7C). This is equivalent to the earlier relative RNase protection data (figure 5C) and the relative enzyme data (figure 4).

Placement of the SV40 late poly(A) region downstream from the parental bGH poly(A) region showed less than 5% of the protected fragments mapping to the downstream poly(A) addition site (857 nucleotides), suggesting little processing at the downstream site (pΔ5SVL) (figure 7B, lane 3). A second 'catcher' polyadenylation region was added to the vector to quantitate the level of unprocessed transcript which readsthrough the two upstream polyadenylation regions. No protected fragments could be detected mapping to this distal SV40 late poly(A) site, which would be expected to be about 1100 nucleotides, (figure 7B, lane 3). Conversely, the CSB chimera showed that the majority of the 3′ ends mapped to the downstream site (890 nucleotides) (pΔ5CSBSVL) (figure 7B, lane 4). Total relative levels of 3′ ends for the two tandem polyadenylation region mutants were equal (pΔ5SVL=pΔ5CSBSVL).

These data suggest that the CSB linker alters the efficiency of the bGH poly(A) region function preventing processing in the 'blocked' region. Introduction of the SV40 late poly(A) region downstream served to 'catch' and process these readthrough transcripts. In the parental vector, the efficient bGH poly(A) region allowed little unprocessed transcript to readthrough.

Insertion of a second poly(A) region behind the BGH poly(A) in the mutants pΔ5IR and pΔ5TR (pΔ5IRSVL and pΔ5TRSVL) also restored relative galactokinase expression (figures 4 and 5D) to the parental pΔ5 level.

DISCUSSION

There has been much speculation on the role of poly(A) region RNA structures in 3′-end processing. Most of the recent studies are directed at discovering the essential sequence elements needed for this processing event. Indeed, both sequence elements and RNA structure may be needed for efficient processing. Our work has focused on determining whether altering the pre-mRNA structure of the bGH poly(A) region would have an effect on the efficiency of the processing reaction. Preliminary results from studies with the SV40 late poly(A) region indicated a correlation between relative gene expression and the stability of the stem loop formed by that region.

Constructs exhibiting levels of relative expression equal to that of the control vector pBGH all contained similar structural features in their synthetic pre-mRNA transcripts (figure 3). Similarities in RNA conformation between these constructs may indicate some of the necessary structures in efficient 3′-end processing. Each of the mutants mediating efficient expression show an AAUAAA which is extremely sensitive to single strand specific nucleases and is predicted to be at the end of a hairpin-loop. Furthermore, the natural site of cleavage in these mutants, exists as a single stranded region downstream from the stem loop containing the AAUAAA. There had been speculation that the cleavage site exists in a single stranded conformation (3,8,27,28).

To examine if RNA features are needed for efficient 3′-end processing, a linker designed to hybridize with the region surrounding the cleavage site and change the wild-type structure, was used to construct various mutants. We attempted to make the cleavage site inaccessible

to the processing machinery by base pairing. Also, we performed our study without modifying any of the sequence in the poly(A) region itself, attempting to exert effects on 3′-end processing in *cis* from upstream of the region. Introduction of this linker in the complementary orientation, to sequester the cleavage site in a stem, lowered the level of relative galactokinase enzyme activity and messenger RNA. To show that this effect was not due to transcription termination an SV40 late poly(A) region was inserted downstream from the bGH poly(A) region. It was reasoned that if the linker was hybridizing to the cleavage site region, affecting the interaction of this sequence with the 3′-end processing machinery, then transcription should readthrough the site without being processed. The introduction of a downstream poly(A) site would serve to 'catch' this transcript and process it, thereby increasing both the steady-state messenger RNA and enzyme activity. Examination of these tandem poly(A) region mutants showed relative galactokinase enzyme and messenger RNA levels equal to those seen for the control vectors. Furthermore, 3′-end mapping indicated that the 'blocked' site was utilized inefficiently, as in pΔ5CSB, and the downstream 'catcher' served as a processing site for the transcript. Since the total relative amount of 3′ends of pΔ5SVL and pΔ5CSBSVL is equal, and correlates to both the relative protein and messenger RNA levels, this suggests that all of the RNA species mapped are translated. Since there is no increase in the amount of stable messenger RNA produced from the mutants containing the 'catcher' poly(A) region(s), this suggests that the utilization of each site is not additive and implies a competition between them. A similar effect was also shown for tandem homologous poly(A) regions in transient assays (29).

Computer prediction indicates that in the poly(A) region of pΔ5CSB, the CSB linker hybridizes to the cleavage site region, embedding it in a stable helix (calculated $\Delta G = -53.1$ kcal/mole and $T_m = 109.6°C$) (24,30). This taken together with the observation that the AAUAAA is inaccessible to the RNases *in vitro*, suggests that the effects seen on the level of steady-state messenger RNA may be due to a reduced ability of the 3′ end-processing enzymes to recognize the pre mRNA region *in vivo*, leading to inefficient processing. The readthrough transcripts which are not processed may be unstable and do not constitute steady-state messenger RNA.

pΔ5IR was constructed in order to relieve the effect of the CSB linker on its supposed site of action. Two copies of the linker are present as an inverted repeat inserted upstream of the bGH poly(A) region. We reasoned that the close proximity of these two complementary sequences would favor their hybridization over the more distant interaction of the CSB linker and the natural site of cleavage. Moreover, the appearance of these sequences in the transcript at an earlier time than the cleavage site region would allow for this hairpin to form first. These linker sequences are in the 'window' suggested for naked RNA available for folding (31). The formation of this upstream hairpin loop would presumably leave the cleavage site in its normal conformation and allow processing. The relative steady-state messenger RNA levels seen from this mutant were higher than those seen for the cleavage site blocker mutant, but lower than the parental vector. So, the inhibitory effect of the CSB linker insertion is partially overcome. The intermediate effect of this mutant suggests that there may be a mixed population of RNA, some having the cleavage site bound to CSB sequence and others with the upstream stem-loop formed by the inverted linker repeat.

Insertion of three CSB linkers mediated a dramatic drop in reporter gene expression (figures 2 and 4). Both relative galactokinase activity and relative steady-state messenger RNA levels were reduced to about 10% of the controls (pBGH and pΔ5) (figures 4 and

5C). It was predicted by computer that the third linker interacts with the accessible cleavage site altering the expression in a fashion similar to pΔ5CSB (figure 3E). This analysis showed the presence of two very stable stem loops which may sequester the cleavage site more tightly by base stacking effects (figure 3E). Alternatively, two linkers complementary to the cleavage site could increase the number of pre-messenger RNAs which form a stem loop containing the cleavage site region. Such extensive stem loops competing for the wild-type structure of the region may be responsible for the very low levels of expression from this mutant, presumably by being unrecognizable to the 3'-end processing enzymes. The AAUAAA, in this mutant is predicted to exist in a bulge loop, different from the hairpin-loop observed in the structure of the parental bGH poly(A) region. Introduction of a heterologous 'catcher' poly(A) region downstream from this construct showed relative expression levels equal to that seen in the control vector. Again, this indicates that the new structures presumably formed *in vivo* do not interfere with transcription.

Our data suggest that pre mRNA topology has an influence on the efficiency of poly(A) site utilization. It has been shown that splice site use can be regulated by altering the RNA secondary structure (32). Perhaps alternative poly(A) site utilization is accomplished by changes in poly(A) site efficiency due to differences in pre mRNA structure. Further characterization of these structures, their function and interaction with the processing machinery *in vivo* and *in vitro* will be needed to more subtly elucidate the role of RNA structure in the efficiency of mammalian cleavage-polyadenylation.

ACKNOWLEDGEMENTS

We thank David Pfarr, John Trill, Steven Cosenza and Dr. Antonio Toscani for their excellent technical assistance and Dr. Ganesh Sathe for synthesizing the oligonucleotides utilized in this study. We also thank Dr. Robert Lyons and Dr. Kenneth Soprano for their helpful discussion and criticism. We are grateful to Dr. Fritz Rottman for providing us with the clone of the bovine growth hormone polyadenylation region.

*To whom correspondence should be addressed

+Present address: The Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, USA

REFERENCE

1. Fitzgerald, M. and Shenk, T. (1981). The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV 40 mRNAs. Cell 24, 251–260.
2. Proudfoot, N.J. and Brownlee, G.G. (1976). 3'-Non-coding region sequences in eukaryotic messenger RNA. Nature. 263,211–214.
3. Wickens, M. and Stephenson, P. (1984). Role of the converved AAUAAA sequence: Four point mutants prevent messenger RNA 3'end formation. Science 226, 1045–1051.
4. Wilusz, J. and Schenk, T. (1988). A 64 kd nuclear protein binds to RNA segments that include the AAUAAA polyadnylation motif. Cell. 52, 221–228.
5. Gil, A. and Proudfoot, N.J. (1984). A sequence downstream of the AAUAAA is required for rabbit β-globin mRNA 3'-end formation. Nature (London) 312, 473–474.
6. McDevitt, A.M., Imperiale, M.J., Ali, H. and Nevins, J.R.(1984). Requirement of a downstream sequence for generation of a poly(A) addition site. Cell. 37, 993–999.
7. Sadofsky, M., Connelly, S., Manley, J. and Alwins, J.C.(1985). Identification of a sequence element on the 3'side of AAUAAA which is necessary for simian virus 40 mRNA 3'-end processing. Molec. Cell. Biol. 5(10), 2713–2719.
8. Woychick, R.P., Lyons, R.H., Plst, L. and Rottman, F.(1984). Requirement for the 3' flanking region of the bovine growth hormone gene for accurate polyadenylation. Proc. Natl. Acad. Sci. 81, 3944–3948.

9. Goins, W.F. and Stinski, M.F.(1986). Expression of a human cytomegalovirus late gene is postranscriptionally regulated by a 3'-end-processing event occurring exclusively late after infection. Molec. Cell. Biol. 6(12), 4202–4213.
10. Jenh, C., Deng, T., Li, D., Dewille, J. and Johnson, L.F.(1986). Mouse thymidilate synthase messenger RNA lacks a 3' untranslated region. Proc. Natl. Acad. Sci. USA. 83, 8482–8486.
11. Ryner, L. and Manley, J.C.(1987). Requirements for accurate and efficient mRNA 3' end cleavage and polyadenylation of a simian virus 40 early pre-mRNA in vitro. Molec. Cell. Biol. 7(1), 495–503..
12. Seiki, M., Hattori, S., HIrayama, Y. and Yoshida, M.(1983). Human T-cell leukemia virus: complete nucleotode sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. 80, 3618–3622.
13. Thirion, J.P., Banville, D. and Noel, H.(1976). Galactokinase mutants of Chinese hamster ovary somatic cells resistant to 2-deoxygalactose. Genetics. 83, 137–147.
14. Wigler, M., Sweet, R., Sim, G.K., Wold, B., Pellicer and Axel, R.(1979). Transformation of mammalian cells with genes from procaryotes and eucaryotes. Cell. 16, 777–785.
15. Pfarr, D.S., Sathe, G. and Reff, M.E.(1985). A highly modular vector for the analysis of eukaryotic genes and gene regulatory elements. DNA. 4(6), 461–467.
16. Pfarr, D.S., Rieser, L.A., Woychick, R.P., Roffman, F.M., Rosenberg, M. and Reff, M.E.(1986). Differential effects of polyadenylation regions on gene expression in mammalian cells. DNA. 5(2), 115–122.
17. Gimmi, E.R., Soprano, K.J., Rosenberg, M. and Reff, M.E.(1988). Deletions in the SV40 late polyadenylation region downstream of the AAUAAA mediate similar effects on expression in various mammalian cell lines. Nucl. Acids Res. 16(18), 8977–8997.
18. Maniatis, T., Frisch, E.F. and Sambrook, J.(1982). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).
19. Tushinski, R.J., Sussman, P.M., Yu, L.Y. and Bancroft, F.L.(1977). Prgrowth hormone messenger RNA; Glucocorticord induction and identification in rat pituitary cells. Proc. Natl. Acad. Sci. USA. 74, 2357–2361.
20. Melton, D.A., Krieg, P.A., Regagliati, M., Maniatis, T., Zinn, K. and Green, M.R.(1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucl. Acids Res. 12, 7035–7056..
21. Toscani, A., Soprano, D.R., Cosenza, C.S., Owen, T.A. and Soprano, K.J.(1987). Normalization of multiple RNA samples using an *in vitro* synthesized external standard RNA. Anal. Biol. 165, 309–319.
22. Deckman, I.C. and Draper, D.E.(1987). S-4α mRNA translation regulation complex II. Secondary structures of the RNA regulatory site in the presence and absence of S4. J. Mol. Biol. 196, 323–332.
23. Zucker, M. and Steigler, P.(1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucl. Acids Res. 9, 133–148.
24. Freier, S.M., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H.(1986). Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA., 83, 9373–9377.
25. Gimmi, E.R., Polyadenylation of mammalian mRNAs.(1989). Ph.D. Dissertation. Temple University School of Medicine.
26. Lowman, H.B. and Draper, D.E.(1986). On the recognition of helical RNA by Cobra Venom V_1 Nuclease. J. Biol. Chem. 261(12), 5396–5403.
27. Berget, S.M.(1984). Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature. 309, 179–182.
28. Cole, C.N. and Stacy, T.P.(1985). Identification of sequences in herpes simplex virus thymidine kinase gene required for efficient processing and polyadenylation. Molec. Cell. Biol. 5(8), 2104–2113.
29. Denome, R.M. and Cole, C.N.(1988). Patterns of polyadenylation site selection in gene constructs containing multiple polyadenylation signals. Molec. Cell. Biol. 8(11), 4829–4839.
30. Marky, L.A. and Bresslauer, K.J.(1987). Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curver. Biopolymers. 26, 1601–1620.
31. Eperon, L.P., Graham, I.R., Griffiths, A.D. and Eperon, I.C.(1988). Effects of RNA secondary structure on alternative splicing of pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? Cell. 54, 393–401.
32. Solnick, D.(1985). Alternative splicing caused by RNA secondary structure. Cell. 43, 667–676.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.