Upstream Elements Present in the 3′-Untranslated Region of Collagen Genes Influence the Processing Efficiency of Overlapping Polyadenylation Signals*

Barbara J. Natalizio, Luis C. Muñiz, George K. Arhin‡, Jeffrey Wilusz‡, and Carol S. Lutz§

From the Department of Biochemistry and Molecular Biology and §Department of Microbiology and Molecular Genetics, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, New Jersey 07103

Poly(A) tails are found on the 3′ end of nearly every fully processed eukaryotic mRNA. The poly(A) tail has been suggested to influence mRNA stability, translation, and transport (for review, see Refs. 1–4). Polyadenylation is a two-step process that first involves specific endonucleolytic cleavage at a site determined by binding of polyadenylation factors (for review, see Refs. 5–10). The second step involves polymerization of an adenosine tail to an average length of ~200 residues. These steps are tightly coupled processes since reaction intermediates are not detectable under normal conditions.

The vast majority of eukaryotic polyadenylation signals contain the consensus sequence AAUAAA between 10 and 35 nucleotides upstream of the actual cleavage and polyadenylation site. In addition, sequences 10–30 nucleotides downstream of the cleavage site are known to be involved in directing polyadenylation (Refs. 11–13 and references therein). These downstream elements (DSEs) can be characterized as a block containing 4 of 5 uracil (U) residues. These two sequence elements recruit cleavage and polyadenylation specificity factor (CPSF) and cleavage stimulatory factor (CstF), respectively, to define the cleavage site; therefore, mutations within these sequences abolish polyadenylation.

The intricate nature of this process implies that polyadenylation might be a useful mechanism to regulate gene expression. The efficiency of 3′ end processing is a level at which regulation can occur. Because most pre-mRNAs in the cell are not efficiently processed, even small changes in the overall processing efficiency of a particular pre-mRNA may have a substantial effect on gene expression. Experimental evidence has demonstrated that poly(A) signal strength directly influences the amount of mature, exported mRNA (14, 15). Poly(A) signal strength is also directly correlated with the rapid assembly of the polyadenylation machinery on the nascent transcript (16) and with transcription termination efficiency (17). Detailed mechanistic studies on regulation of polyadenylation are now emerging, revealing both cis- and trans-acting factors (for review, see Refs. 5, 9, and 18).

In addition to the sequence elements previously described, elements upstream of the AAUAAA sequence and downstream of the DSE (see Refs. 19 and 20 and references therein) have been identified as auxiliary cis-acting polyadenylation efficiency elements. Such elements may play an important role in modulating the overall processing efficiency. Upstream elements (USEs) have been characterized in viral and cellular systems, including SV40 (21), human immunodeficiency virus (22–27), adenovirus major late region (28, 29), cauliflower mosaic virus (30), ground squirrel hepatitis virus (31, 32), and human C2 complement (33, 34). Spacing between the AAUAAA and the USEs plays a significant role in that the USEs closest to the AAUAAA are most important (21). Studies on human immunodeficiency virus suggest that definition of the polyadenylation site involves the recognition of multiple sequence elements, including the USE, in the context of the AAUAAA (27).

Comparisons between the polyadenylation signals of the SV40 late mRNA and other cellular mRNAs revealed that three human collagen genes, COL1A1, COL1A2, and COL2A1, each possess elements similar in sequence (USE consensus UAUnGUAs and position relative to the AAUAAA to the SV40 late USEs (21)). Collagens are extracellular proteins

* This work was funded by American Cancer Society Grant RPG-00-265-01-GMC, Arthritis Investigator Award 2AI-LUT-A-5, Arthritis Foundation, New Jersey Chapter Grant 3AI-LUT-A (to C. S. L.), and National Institutes of Health Grants CA80062 and GM63832 (to J. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: MSB E671, 185 S. Orange Ave., UMDNJ-NJMS, Newark, NJ 07103. Tel.: 973-972-0899; Fax: 973-972-5594; E-mail: lutzcs@umdnj.edu.

‡ To whom correspondence should be addressed: MSB E671, 185 S. Orange Ave., UMDNJ-NJMS, Newark, NJ 07103. Tel.: 973-972-0899; Fax: 973-972-5594; E-mail: lutzcs@umdnj.edu.

§ To whom correspondence should be addressed: MSB E671, 185 S. Orange Ave., UMDNJ-NJMS, Newark, NJ 07103. Tel.: 973-972-0899; Fax: 973-972-5594; E-mail: lutzcs@umdnj.edu.

The abbreviations used are: DSE, downstream element; CPSF, cleavage and polyadenylation specificity factor; CstF, cleavage stimulatory factor; USE, upstream sequence element; BGH, bovine growth hormone.

1 The abbreviations used are: DSE, downstream element; CPSF, cleavage and polyadenylation specificity factor; CstF, cleavage stimulatory factor; USE, upstream sequence element; BGH, bovine growth hormone.

2 C. S. Lutz, unpublished data.
that are responsible for the strength and flexibility of connective tissue. They account for 25–30% of all proteins present in animals and are the major fibrous element of skin, bone, tendon, cartilage, blood vessels, and teeth (see Ref. 35 and references therein). In addition to their structural role, collagens have a directive role in tissue development. The basic structural unit of collagen consists of three polypeptide chains that are extensively covalently cross-linked to each other. The composition of the chains depends on the type of collagen. Type I collagen consists of two COL1A1 chains and one COL1A2 chain, whereas type II collagen consists of three COL2A1 chains.

Interestingly, COL1A1, COL1A2, and COL2A1 each possess 3'-UTRs that are extremely conserved between human and other vertebrates, including mice, cows, chickens, and puffer fish (Refs. 36–43, see also GenBank™ and Table I). For example, human COL1A1 has a 3'-UTR ~1.5 kilobases in length with 2 polyadenylation signals. The first ~500 bases, containing the first polyadenylation signal, are 86.5% identical in mouse, followed by a ~600-base block of little conservation, and the final ~400 bases, including the second polyadenylation signal, are 71.1% identical (44). This high degree of evolutionary conservation suggests important regulatory functions of the collagen 3'-UTRs. It is important to note that the use of one polyadenylation signal over another will shorten the 3'-UTR, and this shortening could potentially remove important regulatory elements. When the sequences surrounding the polyadenylation signal are carefully examined, the percent identity is even higher, especially in the case of the final polyadenylation signal (see Table I). Mechanisms for alternative polyadenylation have been extensively studied for the calcitonin and IgM genes (45–47); however, it is not mechanistically understood how one collagen polyadenylation signal is chosen from among several.

Up-regulation of collagen gene expression takes place in a variety of diseases, including osteoarthritis and scleroderma, but it is unclear how this regulation of expression is accomplished (48–51). Some studies suggest in scleroderma that collagen production may be up-regulated by increased mRNA transcription (47, 52), but the altered expression may not be fully explainable by changes in transcription rates and may additionally be accomplished by regulated post-transcriptional mechanisms, such as polyadenylation.

This study examines the regulation of 3' end formation in human collagen genes. The strong evolutionary conservation of the 3'-UTRs, particularly around polyadenylation signals, led us to believe that these regions contained key regulatory elements. We asked whether cis-acting USEs function as auxiliary elements to influence polyadenylation of the collagen mRNAs, whether these elements acted in a similar fashion to other defined USEs, and how these elements affect utilization of overlapping polyadenylation signals. We determined that the USEs present in these collagen genes do influence 3' end formation efficiency in these genes. Furthermore, the organization of alternative poly(A) signals in the COL1A2 gene suggests

### Table I

| Gene   | Size of 3'-UTR | No. polyadenylation signals | % Sequence identity to final PA signal |
|--------|---------------|----------------------------|--------------------------------------|
| COL1A1 | ~1.5 kb       | 2                          | 92% cow, 98% macaque, 92% mouse      |
| COL1A2 | ~900 bases    | 5 or 6                     | 79% puffer fish, 84% chicken         |
| COL2A1 | ~400 bases    | 2                          | 95% dog, 86% cow, 89% rat            |

FIG. 1. Inclusion of a USE stimulates in vitro processing of a weak polyadenylation signal. Top, schematic of constructs used, including relative positions of the AAUAAA and USE. An arrow marks the cleavage site. Bottom, in vitro polyadenylation reactions. 0 min, unreacted substrate RNA; 30 min, reaction products. Polyadenylated products are noted as poly(A)⁺ on the right. Quantitation of percent polyadenylation is noted at the bottom. Percent polyadenylation was calculated as the quantitation of the polyadenylated product divided by the total quantitated RNA in the lane.

that assembly of 3' end processing factors on the distal signal inhibits assembly of the processing complex on the proximal signal. This suggests that protein-RNA interactions between core polyadenylation signal elements may represent a novel method of down-regulating polyadenylation signal usage.

### MATERIALS AND METHODS

**In Vitro Transcription of RNA Substrates**—RNA transcripts for in vitro polyadenylation and cleavage reactions were synthesized by use of SP6 RNA polymerase according to the supplier (Promega) in the presence of 50 μCi of [³²P]UTP (Amersham Biosciences or PerkinElmer Life Sciences). Transcription of COL1A2 yielded a 311-base RNA, transcription of COL2A1 yielded a 323-base RNA, and transcription of COL1A1 yielded a 374-base RNA. RNAs were gel-purified from 5% polyacrylamide, 8M urea gels by overnight crush elution in high salt buffer (0.4 M NaCl, 50 mM Tris at pH 8.0, 0.1% SDS) before use in reactions. Eluted RNAs were ethanol-precipitated and resuspended in water.

**Nuclear Extracts and in Vitro Polyadenylation and Cleavage Reactions**—HeLa cell nuclear extracts were prepared as described previously (53). In vitro polyadenylation reactions contained a final concentration of 58% v/v HeLa nuclear extract, 16 mm phosphocholine (Sigma), 0.8 mM ATP (Amersham Biosciences), 2% polyvinyl alcohol, and 1 × 10⁶ cpm of [³²P]labeled substrate RNA (~50 fmol) in a reaction volume of 12.5 μl (54, 55). Reactions were allowed to incubate at 30 °C for 1 h. Reaction products were then extracted with phenol/chloroform/isoamyl alcohol, precipitated with ethanol, and analyzed on (19:1) 5% polyacrylamide gels containing 8 μ Urea gels by overnight wash in high salt buffer (0.4 M NaCl, 50 mM Tris at pH 8.0, 0.1% SDS) before use in reactions. Eluted RNAs were ethanol-precipitated and resuspended in water.
2.6% polyvinyl alcohol, and 1 × 10^6 cpm of 32P-labeled substrate RNA in a reaction volume of 12.5 µl (56). Cleavage reactions were allowed to incubate at 30 °C for 1 h. Products were then processed and analyzed as described for polyadenylation products. Competition reactions were performed by adding increased concentrations of specific or nonspecific oligoribonucleotides as indicated into the typical polyadenylation reactions using COL1A2 as substrate RNA. COL1A2 specific or nonspecific competitor oligoribonucleotides (see "Materials and Methods") were added to the reaction in the amounts indicated at the top of the lane. Percent polyadenylation was calculated as the quantitation of the polyadenylated product divided by the total quantitated RNA in the lane.

**FIG. 2.** Competition studies suggest USE binding factors influence the processing efficiency of collagen polyadenylation signals. A, sequence of SV40 late polyadenylation signal compared with three human collagen genes, COL1A1, COL1A2, and COL2A1. Canonical AUUAAA or AUUAA elements are shown in bold, and putative auxiliary upstream elements are underlined and italicized. B, in vitro polyadenylation reactions using COL1A2 as substrate RNA. COL1A2 specific or nonspecific competitor oligonucleotides were added to the reaction in the amounts indicated at the top of the lane. Percent polyadenylation was calculated as the quantitation of the polyadenylated product divided by the total quantitated RNA in the lane.

| Primers for cloning | COL1A2 | COL1A1 | COL1A2 | COL2A1 |
|---------------------|--------|--------|--------|--------|
| Forward primer: TAGGATCTGGTGTTGAGAGAAAACATTC | CUGUAACUAAAGGUGUGAGAGAAACATACTTAC | CUGUAACUAAAGGUGUGAGAGAAACATACTTAC | CUGUAACUAAAGGUGUGAGAGAAACATACTTAC | CUGUAACUAAAGGUGUGAGAGAAACATACTTAC |
| Reverse primer: TATATAAGCTTTCCTTAGGACTGCTATTTG | TATATAAGCTTTCCTTAGGACTGCTATTTG | TATATAAGCTTTCCTTAGGACTGCTATTTG | TATATAAGCTTTCCTTAGGACTGCTATTTG | TATATAAGCTTTCCTTAGGACTGCTATTTG |

**TABLE II**

| Primers used in construct preparation (written 5' to 3') |
|---------------------------------------------------------|
| COL1A2 mutant A: TCAGCATTTGTTCTTTGCCAGATCTCATTTTCAGTTCTTAC CTCTTCAG | COL1A2 mutant 1: TGGGGGCAAGCAGAAAAACTAAAGATCTCATTACCTTCAAG | COL1A2 mutant 2: GCAGAAAAACTAAATTGTACCTAGATCTATATGTGAGATG | COL1A2 mutant 3: ATATGTGAGATGTTTAAATAAAGATCTAAAAAAATGAAATAAAGCAT |
| COL1A2 double mutant complement 1: 500 TAGGATCCATGTGAGATGTTTAAATAAATTG | COL1A2 double mutant 1: TTGGGGCAAGCAGAAAAACTAAAGATCTCATTACCTTCAAG | COL1A2 double mutant 2: GCAGAAAAACTAAATTGTACCTAGATCTATATGTGAGATG | COL1A2 double mutant 3: ATATGTGAGATGTTTAAATAAAGATCTAAAAAAATGAAATAAAGCAT |

**Transfection and RNase Protection**—Human 293T and HeLa cells were maintained in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% fetal bovine serum (Sigma) and 1% penicillin-streptomycin (Invitrogen). Cells were seeded in 100-mm plates ~12 h before transfection. When cells reached 80% confluency, they were transfected using the Quantum Prep Cytofectene reagent (Bio-Rad). Plasmid DNA (8.4 µg) was diluted in 700 µl of serum-free medium to which 40 µl of Cytofectene was added, and the mixture was incubated at room temperature for 20 min. After the addition of 6.3 ml of medium (plus fetal bovine serum) to the mixture, the medium on the cells was removed and replaced with the entire transfection mixture. The Cytofectene-DNA mixture was removed 6 h after transfection and replaced with fresh medium. After 24 h, cells were washed once with phosphate-buffered saline. Cells were scraped and collected into 1 ml of phosphate-buffered saline and transferred into microcentrifuge tubes. Cells were then centrifuged at 1000 rpm for 5 min. The phosphate-buffered saline was aspirated, and pellets were stored at −80 °C for no more than 2 days.

Total RNA was extracted from the cell pellet using the RNeasy mini kit (Qiagen) according to the manufacturer's spin protocol for isolation of total RNA from animal cells. Probe RNA was prepared as described above using T7 polymerase and generating the antisense of the pCpG-COL1A2 construct. The reporter RNA levels were determined by RNase protection using the RPAIII kit (Ambion, Austin, TX) for 1 h at 37 °C. DNA templates were then removed by DNase I digestion, and the RNA was phenol-extracted, ethanol-precipitated, and analyzed on 5% polyacrylamide 8 M urea gels as described above.

**Plasmids**—COL1A2, COL2A1, and COL1A1 inserts were generated and used for the construction of plasmids for mammalian expression. COL1A2 and COL2A1 inserts were generated.
by PCR using complementary primers (Table II). The COL2A1 and COL1A1 primers contained BamHI (forward) and HindIII (reverse) recognition sites to allow insertion into appropriately digested vectors, whereas the COL1A2 primers contained BamHI (forward) and PstI (reverse) recognition sites. Gel-purified and appropriately digested PCR fragments were ligated into appropriately digested pGEM4 with T4 DNA ligase (Invitrogen) at 17°C overnight. The constructs were then transformed into Escherichia coli XL1-Blue cells. Positive clones were both sequenced and assayed for expression of appropriately sized clones. Sequencing was completed by the New Jersey Medical School Molecular Resource Facility using the Applied Biosystems 373 DNA Sequencer and the resulting sequences were analyzed by BLAST computer programs for accuracy. Amplification reactions were performed using platinum Taq polymerase (Invitrogen) in a total volume of 50 μl using standard reaction mixtures for 35 cycles of 95°C (1 min), 55°C (30 s), and 72°C (45 s) with an initial denaturation step of 95°C (5 min). PCR products were purified on a 1% agarose gel before use. Primers used in individual PCR reactions are referred to in Table II.

The COL1A2 double mutant 1,2 and the triple mutant 1,2,3 were prepared via the Stratagene QuikChange kit. Amplification reactions using Pfu polymerase were for 20 cycles of 95°C (30 s), 55°C (1 min), and 68°C (8 min) with a 95°C (5 min) initial denaturation step. To remove the remaining template after amplification, a DpnI digestion was performed for 90 min at 37°C. The PCR product was then transformed into E. coli XL-1 Blue cells. Positive clones were isolated and sequenced to determine correct expression before use in polyadenylation reactions.

The PA2-G mutant COL1A2, the single USE mutants, and the double mutant 1,3 were created by the use of the megaprimer method of PCR mutagenesis as described previously (57). Briefly, a mutant PCR product was generated using an internal upstream primer containing the mutant, such as the PA2-G mutant COL1A2, and the wild-type COL1A2 downstream primer. This product containing the mutant sequence was then gel-purified and used as the downstream megaprimer for the second PCR. The COL1A2 wild type upstream primer was used for the second PCR. A third PCR was then performed using the gel-purified second PCR product as the template and COL1A2 wild type upstream and downstream primers. This final PCR amplified the inefficient product resulting from the second PCR. This final product was gel-purified, cloned into pGEM4, and transformed into E. coli XL-1 Blue cells.

The construction of pIVA2 was described in Wilusz et al. (58). Briefly, the 155-base BamHI to PvuII fragment of pAd5-E1B (which contains adenovirus type 5 sequences from 3943 to 4122) was cloned into pGEM4 at the BamHI and PvuII sites. Linearization with BglII yields a 155-base RNA containing the IVA2 poly(A) signal. The DNA template to generate IVA2-USE RNA, which contains a USE 5′ to the AAUAAA, was constructed by a two-step PCR reaction using the megaprimer approach (57). The first PCR reaction used a standard SP6 primer and 5′-TTTTAGGGGTTTGGCGGGTAAATTTACTAA (forward) and upstream of the AAUAAA element. The second PCR reaction contained the megaprimer and the primer 5′-AGCTTGCAATGGCTG-CAGGTCGAC. The product of this PCR reaction was then cut with BglII and used as a template to generate IVA2-USE RNA using SP6 RNA polymerase.

For transfection assays, all COL1A2 constructs were cloned into the BamHI-PstI sites of vector pCS8 (a gift of David Fritz, UMDNJ) downstream of a cytomegalovirus promoter and upstream of a bovine growth hormone polyadenylation (BGH) signal. This vector also includes intron 1 of the rabbit β-globin gene accompanied by the splice donor and acceptor sites. Constructs were verified by sequencing.

**RESULTS**

**USEs Can Stimulate In Vitro 3′ End Processing of a Weak Polyadenylation Signal**—Previously, auxiliary polyadenylation elements known as upstream efficiency elements (USEs) have been described and characterized by use of the SV40 late polyadenylation signal (21). It was determined that these elements function as efficiency elements in the SV40 system since their disruption resulted in reduced polyadenylation function (21). It has also been shown that a USE from the SV40 signal can replace the human immunodeficiency virus type I (HIV) 3′ end formation in transient transfection assays (23). Both SV40 and human immunodeficiency virus have very strong polyadenylation signals. However, it has not previously been determined if USEs could be added to a weak polyadenylation signal, such as the adenovirus IVA2 polyadenylation signal, to enhance 3′ processing efficiency. A USE motif from SV40 having the sequence GCUUUAAUUUGUAACC was inserted upstream of the AAUAAA in the IVA2 polyadenylation signal to create IVA2-USE, substrate RNAs were prepared from both pIVA2 and pIVA2-USE, and the RNAs were added to in vitro polyadenylation reactions. Fig. 1 shows that the presence of a USE in IVA2-USE enhanced polyadenylation efficiency—4-fold as compared with IVA2 alone. These data indicate that insertion of a USE can increase polyadenylation efficiency of a weak processing signal, suggest that USEs can modulate poly(A) site definition, and suggest that USEs may be commonly found in cellular, not only viral, genes. It is important, therefore, to evaluate mammalian polyadenylation signals for USE elements.

**USEs Can Be Found in Collagen Genes**—A survey of numerous cellular polyadenylation signals revealed that elements resembling the USE motifs present in SV40 can also be found in many 3′-UTRs; that is, similar to the consensus AAUAAA_GUNA and within 75 bases of the AAUAAA (2). We chose to focus on three collagen genes since their 3′-UTRs are highly conserved through evolution, suggesting regulatory function. Fig. 2 shows the comparison of the SV40 late polyadenylation signal to three human collagen genes for type I (COL1A1 and COL1A2) and type II (COL2A1) collagens. The polyadenylation signals AAUAAA/AAUAAA and the putative USE elements are underlined. We next wanted to determine whether these...
putative USEs present in the collagen 3'-UTRs could stimulate 3'-end processing efficiency like the SV40 USEs. Previously it was shown that polyadenylation reactions containing an SV40 substrate RNA could be inhibited specifically by oligoribonucleotides representing the USE motifs (54). Plasmids encoding a portion of each collagen gene 3'-UTR containing the polyadenylation signals were created by PCR of human genomic DNA. Substrate RNAs for the polyadenylation reactions were prepared by *in vitro* transcription using SP6 polymerase in the presence of [32P]UTP. *In vitro* coupled cleavage and polyadenylation reactions were performed using HeLa nuclear extract and COL1A2 (Fig. 2B) substrate RNA. Oligoribonucleotides representing a putative USE corresponding to COL1A2 or a nonspecific oligoribonucleotide were also added to the reactions. The results for COL1A2 are shown as an autoradiogram of a typical *in vitro* polyadenylation reaction. The second lane in Fig. 2B, marked 0, represents a reaction performed in the absence of competitor oligoribonucleotides and demonstrates that the COL1A2 substrate RNA was efficiently polyadenylated in our *in vitro* system. Similar results were found with COL1A1 and COL2A1 substrate RNAs and their specific oligoribonucleotides (data not shown). In each case, the specific USE oligoribonucleotide inhibited polyadenylation, whereas the nonspecific had no effect on polyadenylation (see Fig. 2B and data not shown). No effect was also noted when a different nonspecific oligoribonucleotide was used for each substrate RNA (data not shown). Quantitation of the percent polyadenylated product is indicated below the autoradiogram of the gel in Fig. 2B. Additionally, oligoribonucleotides representing the collagen USEs can cross-compete in this assay (*i.e.* a COL1A1 oligo can compete with a COL1A2 substrate RNA; data not shown). Taken together, these data suggest that the oligoribonucleotides specifically bind and sequester a common factor(s) important for polyadenylation and suggest that the similarity to the SV40 motifs identified in the collagen 3'-UTR is functionally significant.

**COL1A2 USEs Act as Auxiliary Polyadenylation Elements in Vitro**—Because of the strong processing efficiency observed
A competing polyadenylation signals. A, in vitro cleavage assay reveals poly(A) signal 2 is the predominantly used polyadenylation signal, whereas poly(A) signal 1 is a minor signal. Mutation of poly(A) signal 2 from AAUAAA to AAGAAA switches this predominance. Marker lane, transcript prepared from COL1A2 construct that was linearized with NspI (cuts between the two cleavage sites). B, in vivo polyadenylation assays reveal that USE mutants plus mutation of poly(A) signal 2 results in a marked decrease in polyadenylation efficiency. Lighter gray bars, 293T cell transfections; darker gray bars, HeLa cell transfections. Percent polyadenylation (% PA) was measured as the ratio of COL1A2 polyadenylation site utilization to the downstream bovine growth hormone polyadenylation site utilization as quantitated by RNase protection assays.

Using the COL1A2 substrate RNA, we chose to focus our attention on that polyadenylation signal (diagrammed in Fig. 3A). We were also intrigued by the high degree of sequence conservation of this signal from diverse organisms (see Fig. 3A). We next made a series of substitutions replacing the COL1A2 USEs with BglII linkers to assess the contribution of the USEs to in vitro polyadenylation (diagrammed in Fig. 3B). USEs were replaced individually, as well as two at a time and three at a time. A nonspecific mutation was created by introducing a mutation outside the USE region (NS mut) had no effect on polyadenylation. We conclude that none of the USEs is absolutely required for COL1A2 polyadenylation but that mutation of USE 2 in conjunction with at least one other USE led to the most dramatic decreases in polyadenylation.

**COL1A2 USE Mutations Are More Deleterious in in Vivo Assays**—Because all of our experiments so far have been performed in vitro, we found it important to verify our results in in vivo assays. We next cloned our COL1A2 wild type and mutant constructs into pCMV downstream of a cytomegalovirus promoter and upstream of a BGH poly(A) signal. Tandem polyadenylation signals have been used previously to examine requirements for a different type of auxiliary sequences in the lamin B2 gene (59). A T7 promoter on the opposite strand downstream of the BGH poly(A) signal was also present for ease in making antisense probes. The constructs were then transfected into HeLa or 293T cells, and after 24 h, total RNA was harvested. This RNA was added to RNase protection assays (using an antisense transcript from the T7 promoter as a probe). Representative RNase protection assays are shown in Fig. 4A, and the results of all our experiments were quantitated and are shown in Fig. 4B. The results show that use of the COL1A2 polyadenylation signal prevailed over use of the BGH polyadenylation site when the COL1A2 signal was wild type (lane 2) or had a nonspecific mutation (lane 7), but the USE mutations altered this ratio (lanes 4–6 and 8–10). The quantitated results were analyzed as the ratio of the protected RNA fragment corresponding to RNA polyadenylated at the COL1A2 site relative to those polyadenylated at the BGH poly(A) site (Fig. 4B). A large number means that the COL1A2 polyadenylation signals were preferentially used rather than the BGH polyadenylation signal, whereas a small number means that the BGH polyadenylation signal was preferentially used. The overall trends in the in vitro data correlate with the in vitro data; however, the USE mutants are more deleterious in varying degrees in vivo as compared with in vitro. USE 1 and 3 mutations alone had little effect on use of the COL1A2 polyadenylation signal, whereas USE 2 mutation decreased polyadenylation efficiency to approximately half of wild type. Mutation of the USEs in duplicate or triplicate also reduced analyzed on 5% polyacrylamide, 8 M urea gels. The data were quantitated from multiple in vitro reactions and are presented in Fig. 3. Mutation of either USE 1, 2, or 3 alone had little effect on polyadenylation. Mutation of both USEs 1 and 2 simultaneously also had only a slight effect, but co-mutation of USEs 1 and 2 or USEs 1, 2, and 3 diminished polyadenylation efficiency to approximately half of wild type levels. A nonspecific mutation outside the USE region (NS mut) had no effect on polyadenylation.

FIG. 6. CPSF binding between the core elements of the proximal polyadenylation signal of COL1A2 may inhibit complex assembly on the distal polyadenylation signal. Polyadenylation machinery can successfully assemble on the proximal signal (poly(A) signal 2) but may not support assembly of processing factors on the distal signal (poly(A) signal 1) due to spacing or steric constraints.
polyadenylation efficiency to approximately half of wild type. **COL1A2 Has Unusual, Overlapping, Competing Polyadenylation Sequences**—Close examination of the COL1A2 mRNA sequence revealed an unusual feature, that there are in fact two polyadenylation signals within 15 bases of each other (see Fig. 3A). Based upon the composition and spacing of the downstream CstF binding site (also known as the DSE) relative to the AUAUUAA, it might seem that poly(A) signal 1 would be preferentially used instead of poly(A) signal 2. To formally investigate the question of which poly(A) signal was the major site of polyadenylation, whereas poly(A) signal 1 is the minor site (Fig. 5A). When a non-usable mutant of poly(A) signal 2 was created (AAUAAA to AAGAAA; PA-2 G), polyadenylation now switched to poly(A) signal 1 (Fig. 5A). This suggested that perhaps something more than USEs and sequence spacing of the AUAUUAA relative to the DSE influences poly(A) signal choice in this system.

We then wanted to know how mutation of the USEs in combination with the poly(A) signal 2 mutant (PA-2 G) affected 3′ end processing. In our in vitro assays, mutation of the poly(A) signal 2 consensus hexamer from AUAUUAA to AA-GAAA did not affect the overall polyadenylation efficiency (see Fig. 3B, PA-2 G mut). In our in vivo RNase protection assays, mutation of poly(A) site 2 had no effect on overall polyadenylation, but that mutation in conjunction with the double and triple USE mutations decreased polyadenylation to approximately one-fifth of wild type (see Fig. 4A, lanes 2–3 and 11–12, and Fig. 5B). Taken together, these data demonstrate that USEs influence 3′ end formation efficiency in the COL1A2 gene.

**DISCUSSION**

In this study we have identified auxiliary 3′ end processing elements in highly conserved regions of the 3′-UTRs of human collagen genes. These elements promote efficient polyadenylation in vitro and in vivo. In addition, COL1A2 has the unusual feature of overlapping polyadenylation signals, one of which predominates, and suggests a novel mechanism for poly(A) signal 1 utilization since mutation of the USEs affects polyadenylation at that site more than when both poly(A) signals are intact. Our in vitro data also support this, although the effects are not as dramatic (data not shown).

As has been appreciated more completely in recent years, 3′ end formation is interconnected both to the other major RNA processing events, splicing and capping, and also to mRNA transcription (for review, see Refs. 6, 9, and 60). This interconnection likely results in most efficient utilization of cis- and trans-acting signals. Thus, it is reasonable to expect that the in vivo data most closely mimic regulation at the cellular level and reflect the co-transcriptional nature of these processes.

The overlapping polyadenylation signals present in the COL1A2 3′-UTR are unusual. Our data demonstrate that poly(A) signal 2 is the major site of polyadenylation, whereas poly(A) signal 1 is used, but to a lesser extent (see Fig. 5A). These data suggest a model, shown in Fig. 6. The configuration of the overlapping signals sets up a competition between CstF binding to poly(A) signal 1 versus CPSF binding to poly(A) signal 2. Mutation of the AUAUUAA in poly(A) signal 2 activates usage of poly(A) signal 1 (see Figs. 5A and 6). These data suggest that the two polyadenylation signals are in competition with each other. Steric hindrance may not allow the interaction between CPSF and CstF bound at the corresponding sites for poly(A) signals 1 and 2 simultaneously, or it may suggest that CPSF-RNA interactions are dominant over CstF interactions at the DSE for poly(A) signal 1. These data demonstrate a principle that protein-RNA interactions can interfere with scaffold assembly, suggesting a novel mechanism for repressing poly(A) signal usage. It remains to be seen whether this arrangement could be used to decrease polyadenylation efficiency at selected signals.

**Acknowledgments**—We thank the members of the Lutz, Wilusz, and O’Connor laboratories for helpful experimental suggestions and comments on the paper.

**REFERENCES**

1. Lewis, J., Gunderson, S., and Mattaj, I. W. (1995) *J. Cell Sci. Suppl.* 19, 13–19
2. Jacobson, A., and Peltz, S. W. (1996) *Annu. Rev. Biochem.* 65, 693–739
3. Sachs, A. B., Sarnow, P., and Hentze, M. W. (1997) *Cell* 90, 831–838
4. Wickens, M., Anderson, P., and Jackson, R. J. (1997) *Curr. Opin. Genet. Dev.* 7, 220–232
5. Proudfout, N. (1996) *Cell* 87, 779–781
6. Proudfout, N. (2000) *Trends Biochem. Sci.* 25, 290–293
7. Keller, W., and Minvielle-Sebastia, L. (1997) *Curr. Opin. Cell Biol.* 9, 329–336
8. Celgan, D. F., and Manley, J. L. (1997) *Genes Dev.* 11, 2755–2766
9. Zhao, J., Hyman, L., and Moore, C. (1999) *Microbiol. Mol. Biol. Rev.* 63, 405–445
10. Shatkin, A. J., and Manley, J. L. (2000) *Nat. Struct. Biol.* 7, 838–842
11. Connor, F., MacDonald, C. C., and Wilusz, J. (1996) *Nucleic Acids Res.* 24, 2614–2620
12. Chou, Z. F., Chen, F., and Wilusz, J. (1994) *Nucleic Acids Res.* 22, 2525–2531
13. Graber, J. H., Cantor, C. R., Mohr, S. C., and Smith, T. F. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 14055–14060
14. Denome, R. M., and Cole, C. N. (1988) *Mol. Cell. Biol.* 8, 4829–4839
15. Edwards-Gilbert, G., Prescott, J., and Falck-Pedersen, E. (1993) *Mol. Cell. Biol.* 13, 3472–3480
16. Chao, L. C., Jamil, A., Kim, S. J., Huang, L., and Martinson, H. G. (1999) *Mol. Cell. Biol.* 19, 5588–5600
17. Osheim, Y. N., Proudfout, N. J., and Beyer, A. L. (1999) *Mol. Cell* 3, 379–387
18. Edwards-Gilbert, G., Vernaldi, K. L., and Milcarek, C. (1997) *Nucleic Acids Res.* 25, 2547–2551
19. Chen F., and Wilusz, J. (1998) *Nucleic Acids Res.* 26, 2891–2898
20. Arbizu, C., Sarnow, P., and Hentze, M. W. (1991) *Nature* 350, 10–15
21. Holley, N., and Anderson, P. (1996) *Curr. Opin. Genet. Dev.* 6, 695–701
22. Hentze, M. W., and Zavolan, M. (1999) *Curr. Opin. Genet. Dev.* 9, 482–489
23. Varrone, M. J., and Gravely, B. R. (1995) *J. Cell Sci.* Suppl. 18, 5–12
24. Osheim, Y. N., and Beyer, A. L. (1999) *Mol. Cell. Biol.* 19, 5588–5600
25. DeZazzo, J. D., Kilpatrick, J. E., and Imperiale, M. J. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 14055–14060
26. DeZazzo, J. D., Kilpatrick, J. E., and Imperiale, M. J. (1991) *Mol. Cell. Biol.* 11, 1842–1850
27. Connor, F., MacDonald, C. C., and Wilusz, J. (1996) *Nucleic Acids Res.* 24, 2614–2620
28. Arbizu, C., Sarnow, P., and Hentze, M. W. (1991) *Nature* 350, 10–15
29. Holley, N., and Anderson, P. (1996) *Curr. Opin. Genet. Dev.* 6, 695–701
30. Sanfacon, H., Hromack, P., and Hohn, T. (1991) *Genes Dev.* 5, 141–149
31. Russnak, R., and Ganem, D. (1996) *Genes Dev.* 4, 764–776
Regulation of Human Collagen Pre-mRNA Polyadenylation

32. Russnak, R. (1991) *Nucleic Acids Res.* 19, 6449–6456
33. Moreira, A., Wollerton, M., Monks, J., and Proudfoot, N. J. (1995) *EMBO J.* 14, 3809–3819
34. Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J. L., and Proudfoot, N. J. (1998) *Genes Dev.* 12, 2522–2534
35. Persikov, A. V., and Brodsky, B. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 1101–1103
36. de Wet, W., Bernard, M., Benson-Chanda, V., Chu, M. L., Dickson, L., Weil, D., and Ramirez, F. (1987) *J. Biol. Chem.* 262, 16032–16036
37. Sangiorgi, F. O., Benson-Chanda, V., de Wet, W. J., Sobel, M. E., and Ramirez, F. (1985) *Nucleic Acids Res.* 13, 2815–2826
38. Sandell, L. J., Prentice, H. L., Kravis, D., and Upholt, W. B. (1984) *J. Biol. Chem.* 259, 7826–7834
39. Ninomiya, Y., Showalter, A. M., van der Rest, M., Seidah, N. G., Chretien, M., and Olsen, B. R. (1984) *Biochemistry* 23, 617–624
40. Upholt, W. B., Strom, C. M., and Sandell, L. J. (1985) *Ann. N. Y. Acad. Sci.* 460, 130–140
41. Metsaranta, M., Toman, D., de Crombrugghe, B., and Vuorio, E. (1991) *J. Biol. Chem.* 266, 16862–16869
42. Myers, J. C., Dickson, L. A., de Wet, W., Bernard, M. P., Chu, M. L., Di Liberto, M., Pepe, G., Sangiorgi, F. O., and Ramirez, F. (1983) *J. Biol. Chem.* 258, 10128–10135
43. Elima, K., Vuorio, T., and Vuorio, E. (1987) *Nucleic Acids Res.* 15, 9499–9504
44. Maatta, A., Bernstein, P., and Penttinen, R. P. K. (1991) *FEBS Lett.* 279, 9–13
45. Lou H., Neugebauer, K. M., Galg, R. F., and Berget, S. M. (1998) *Mol. Cell. Biol.* 18, 4977–4985
46. Takagaki, Y., Seipelt, R. L., Peterson, M. L., and Manley, J. L. (1996) *Cell* 87, 941–952
47. Martinicek, K., Campbell, R., Edwalsd-Gilbert, G., Souan, L., Lotze, M. T., and Milsarek, C. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 11095–11100
48. Jimenez, S. A., Hitraya, E., and Varga, J. (1996) *Rheum. Dis. Clin. N. Am.* 22, 647–674
49. Fehr, J. E., Trotter, G. W., Oxford, J. T., and Hart, D. A. (2000) *Am. J. Vet. Res.* 61, 960–965
50. Martin, J., Jakub, M., Schaefer, D., Dick, W., Spagnoli, G., and Heberer, M. (2001) *Osteoarthritis Cartilage* 9, 112–118
51. LeGraverand, M. P. H., Eggerer, J., Vignon, E., Otterness, I. G., Barclay, L., and Hart, D. A. (2002) *J. Orthop. Res.* 20, 535–544
52. Tasanen, K., Palatsi, R., and Oikarinen, A. (1998) *Br. J. Dermatol.* 139, 23–26
53. Moore, C. L. (1998) *Methods Enzymol.* 255, 1–15
54. Lutz C. S., and Alwine, J. C. (1994) *Genes Dev.* 8, 576–586
55. Lutz C. S., Murthy, K. G. K., Schek, N., O'Connor, J. P., Manley, J. L., and Alwine, J. C. (1998) *Nucleic Acids Res.* 26, 325–337
56. Lutz, C. S., Cooke, C., O'Connor, J. P., Kobayashi, R., and Alwine, J. C. (1998) *RNA (N. Y.)* 4, 1493–1499
57. Aiyar, A., and Leis, J. (1993) *Biotechniques* 14, 366–369
58. Wilusz, J., Feig, D. I., and Shenk, T. (1988) *Mol. Cell. Biol.* 8, 4477–4483
59. Brackenridge, S., Ashe, H. L., Giacca, M., and Proudfoot, N. J. (1997) *Nucleic Acids Res.* 25, 2326–2335
60. Proudfoot, N. J., Forger, A., and Dye, M. J. (2002) *Cell* 108, 501–512
Upstream Elements Present in the 3′-Untranslated Region of Collagen Genes Influence the Processing Efficiency of Overlapping Polyadenylation Signals
Barbara J. Natalizio, Luis C. Muñiz, George K. Arhin, Jeffrey Wilusz and Carol S. Lutz

J. Biol. Chem. 2002, 277:42733-42740.
doi: 10.1074/jbc.M208070200 originally published online August 27, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M208070200

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 60 references, 27 of which can be accessed free at http://www.jbc.org/content/277/45/42733.full.html#ref-list-1