Identification of a stem–loop structure important for polyadenylation at the murine IgM secretory poly(A) site

Cathy Phillips+, Christina B. Kyriakopoulou and Anders Virtanen*

Department of Genetics and Pathology, Uppsala University, Box 589, SE-751 23 Uppsala, Sweden

Received October 16, 1998; Revised and Accepted November 25, 1998

ABSTRACT
We have previously shown that a distal GU-rich downstream element of the mouse IgM secretory poly(A) site is important for polyadenylation in vivo and for polyadenylation specific complex formation in vitro. This element can be predicted to form a stem–loop structure with two asymmetric internal loops. As stem–loop structures commonly define protein RNA binding sites, we have probed the biological activity of the secondary structure of this element. We show that mutations affecting the stem of the structure abolish the biological activity of this element in vivo and in vitro at the level of cleavage and polyadenylation specificity factor/cleavage stimulation factor complex formation and that both internal loops contribute to the enhancing effect of the sequence in vivo. Lead (II) cleavage patterns and RNase H probing of the sequence element in vitro are consistent with the predicted secondary structure. Furthermore, mobility on native PAGE suggests a bent structure. We propose that the secondary structure of this downstream element optimizes its interaction with components of the polyadenylation complex.

INTRODUCTION
Polyadenylation sites of higher eukaryotic cells are characterized by a highly conserved AAUAAA hexanucleotide sequence located ~20 nt upstream of the site of cleavage and poly(A) addition (reviewed in 1). Downstream of the polyadenylation site, unconserved GU-rich regions are often present. The sequences and the locations of these GU-rich regions have been shown to play vital roles in determining the strength of the poly(A) site (reviewed in 1–3).

The hexanucleotide sequence and the GU-rich regions are recognized by a number of trans-acting factors which are necessary for the first steps of cleavage and polyadenylation. The hexanucleotide element interacts with the cleavage and polyadenylation specificity factor (CPSF) while the cleavage stimulation factor (CstF) interacts with the GU-rich regions (reviewed in 3,4). Neither factor binds RNA avidly by itself but the interaction between CPSF, CstF and the RNA substrate result in vitro in the formation of a stable complex known as the pre-initiation complex (5,6). In addition, a number of recently purified cleavage factors (7), as well as poly(A) polymerase, are necessary for the polyadenylation reaction to proceed to completion (reviewed in 4,8,9).

The usage of the IgM heavy chain (μ) secretory poly(A) site is regulated during B cell development (10–12). Regulation appears to occur by modulating polyadenylation activity at this site, making the secretory poly(A) site a good model to study the regulation of polyadenylation (13–21). It has been proposed that CstF is limiting for poly(A) site complex formation in B cells and may be involved in the regulation of the usage of the secretory poly(A) site during B cell differentiation (22). We have shown that the secretory poly(A) site contains dual upstream and downstream elements which all contribute to the formation of a stable CPSF/CstF pre-initiation complex on this poly(A) site in vitro (23). The distal downstream element is predicted to form part of a stem–loop structure with two asymmetric internal loops. It is also necessary for the binding of a 30 kDa polypeptide whose induction correlates with the upregulated usage of the secretory poly(A) site, suggesting that this element may play a role in the regulation of polyadenylation activity at this site (24).

We have, therefore, examined the sequences spanning the distal GU-rich region of the murine IgM secretory poly(A) site in detail to understand more clearly how these sequences function to enhance polyadenylation activity. We show that the downstream GU-rich region must be present in the form of a stem–loop structure in order to enhance the polyadenylation efficiency of the secretory poly(A) site and that this takes place at the level of CPSF/CstF complex formation. Furthermore, we show that the two internal asymmetric loops contribute to the enhancing effect of this element and that the predicted stem–loop structure is bent.

MATERIALS AND METHODS
Plasmid constructs
PCR products from the secretory poly(A) site and surrounding sequences, containing mutations and a 5′ BglII site introduced as part of the synthetic primers, were cloned into pGem T vector. Inserts were subcloned into pPKLT55 (kindly provided by Walter Dietrich-Goetz) (24) between the BglII and NotI sites. The forward primer, 5′-GACAGATCTGATCATGTTCTGACACAGGCG-3′, was used in combination with either of the following reverse primers (nucleotide positions are numbered according to the mouse IgM sequence with accession no. V00818). For deletions

*To whom correspondence should be addressed. Tel: +46 18 4714908; Fax: +46 18 526849; Email: anders.virtanen@medgen.uu.se

*Present address: Department of Molecular Biology and Biochemistry, Rutgers University, Nelson Laboratories, Room A322, Piscataway, NJ 08855, USA
Figure 1. (A) Schematic diagram of the sequences surrounding the IgM heavy chain secretory poly(A) site and the deletion mutant constructs used. The multiple hexanucleotide and GU-rich regions have been indicated and are shown as follows: open box, upstream adjacent AU-rich region; black box, consensus hexanucleotide sequence; hatched boxes, proximal and distal GU-rich regions. The nucleotide sequences of the AU-rich region, the consensus hexanucleotide element and the GU-rich regions are shown. The location of the polyadenylation cleavage site is indicated with a vertical arrow. The 55 nt element is indicated with a horizontal arrow. Names of the mutant constructs and diagram indicating the position of their 3′ ends are depicted at the bottom. (B) 3′ deletion analysis of the 55 nt downstream element.

Sequences were cloned into pPKLT55 and transfected into J558L cells, luciferase activity was adjusted for CAT activity and expressed as percentage of LuµspA2 activity. Bars represent the means of luciferase activity of three transfection experiments ± standard error. Each triplicate experiment was performed three times with similar results. Samples are as indicated.

(A): µspA1: 5′-GGTCTAGAGCAGGCAGCATGCATGTTATAT-3′; µspA2: 5′-GGTCTAGAGTGGGCAAGCAGTCATCGATTG-3′; Igsdel(2071): 5′-GCAAGCAAACACAACTGTCAGCAG-3′; Igsdel(2058): 5′-TGCAAGGAAGACAAACAACTGTCAGCTGAGCAGAAGC-3′; Igsdel(2052): 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igsdel(2052): 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igsdel(2064): 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igsdel(2071): 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′.

For mutations in the stem (introduced point mutations are indicated with bold letters) (see also Fig. 2A): Igs(G2056U,C2061A), 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igs(C2048A,G2042U), 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igs(C2059CA), 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igs(C2064AA): 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′; Igsdel(2071): 5′-TGCAAGGAAGAACAACAACTGTCAGCTGAGCAGAAGC-3′.

To generate plasmids belonging to the pBsIgs category the forward primer 5′-GGTCTAGAGCGTTGCAATTTTAATATAT-3′ was combined with the reverse primers used for Igs(G2056U,C2061A), Igs(C2048A,G2042U), Igs(C2056U,C2061A,C2048A,G2042U) and Igsdel(2071) respectively, during PCR amplification. The resulting PCR products were cloned into Bluescript II KS between the BamHI and Smal sites. The plasmids obtained were named pBsIgs(G2056U,C2061A), pBsIgs(C2048A,G2042U), pBsIgs(G2056U,C2061A,C2048A,G2042U) and pBsIgsdel(2071) respectively. An additional plasmid named pBslgs(U1962,1968,1969,1975G) mutated in the hexanucleotide and the extended hexanucleotide sequence elements was generated similarly using primers 5′-GACAGATCTGTCATTTAATAAAAATAGAATATATATAT-3′ and the reverse primer Igsdel(2071) during PCR amplification.
Figure 2. An intact stem and the two asymmetrically paired internal loops are necessary for the enhancement of polyadenylation activity in vivo. (A) Schematic diagram of the mutations in the stem–loop structure used for assessment of in vivo activity. (I) Those used for the compensatory mutations. (II) Those used for the mutations in the hairpin and asymmetric internal loops. (B) Compensatory mutations which destroy and reconstruct the stem–loop structure. Sequences were cloned into pPKL T55 and transfected into J558L cells and luciferase activity was adjusted for CAT activity and expressed as percentage of λµspA2 activity. Bars represent the means of luciferase activity of three transfection experiments ± standard error. Each triplicate experiment was performed three times with similar results. Samples are as indicated. (C) Hairpin and internal loops were mutated. Procedures were the same as for (B). Samples are as indicated.

RNA substrates

Templates for in vitro complex experiments were obtained by restriction of the ‘Bs’ plasmid constructs with EcoRI. For the structural mapping, templates for in vitro transcription were obtained by PCR amplification of plasmid G1ms359 (25,26) using the oligonucleotide 5′-GAAATTAACAAGACTCACATAAGCTGCTGGAGACAGTTGTTG-3′ containing the T7 RNA polymerase promoter at its 5′ end (27) and nucleotides 2025–2076 of the IgM secretory locus at its 3′ end combined with either of the following reverse primers: wild-type stem–loop: 5′-GGCAAGTATGCGGGTTGTTTTGTTGAAAGCAAGCAAAACA-3′; closed stem–loop: 5′-GGGAAGTTGTTTGGCAGGCAAGCAAAACA-3′.

Capped uniformly labelled RNA substrates were synthesized by in vitro transcription using T7 RNA polymerase (Stratagene) according to the manufacturer’s instructions in the presence of 50 µCi [α-32P]UTP (40 Ci/mmol in the transcription mixture). Transcripts were purified by extraction after electrophoresis on 3.5% polyacrylamide gels containing 7 M urea. Capped and 3′ end labelled RNA substrates were synthesized by the same protocol omitting [α-32P]UTP. Obtained RNA was purified by Sephadex G50 spin column chromatography, followed by phenol/chloroform extraction and ethanol precipitation. Subsequently the purified RNA was labelled by 5′-32pCp at the 3′ end using T4 RNA ligase. 3′ End labelled RNA was purified as above.

Cell culture and transfection

J558L (28) was obtained from the European Collection of Animal Cell Cultures (Salisbury, UK). Plasmids were transfected into J558L cells in log phase using DEAE-dextran according to Grosschedl and Baltimore (29) as previously described (24). Transfection efficiency (24) was measured by co-transfection of SVCA Tb (30).

Gel retardation assay

Reactions (10 µl) were assembled on ice having final concentrations: 25 mM Tris–HCl buffer system pH 8.3, 10% (v/v) glycerol, 1% (w/v) polyvinyl alcohol, 50 mM KCl, 2 mM MgCl2, 0.1 mM EDTA, 0.2 mg/ml BSA, 0.01% NP-40, 2 mM DTT, 1 mM ATP and 5 U ribonuclease inhibitor. In each reaction 32P-radiolabelled RNA substrate (400 000 c.p.m.) was added. Partially purified CPSF (31) was added to each reaction followed by HeLa cell nuclear extract as a source of CstF. Polyadenylation complexes were formed for 20 min at 30°C. Five microlitres of the reaction mixture were loaded on a native 3% polyacrylamide/agarose composite gel as described (7).
boiling of RNA substrates with 50 mM NaOH, 10 mM EDTA for 10 s. For RNase H probing, 100,000 c.p.m. of uniformly radiolabelled RNA substrate was refolded in 0.2 M NaCl, 10 mM MgCl₂, 25 mM Tris–HCl, pH 7.5, 1 mM DTT. 0–50 pM of oligo (5′-GAGCAA-3′) or 0–800 pM of oligo (5′-AACACA-3′) complementary to the RNA at positions 2049–2054 or 2039–2044, respectively, were added and allowed to hybridize for 10 min before the addition of 3 U of RNase H. Reaction mixtures were incubated for a further 30 min at 37°C and products were analyzed on 15% PAGE containing 7 M urea. For mobility on native gels, 100 000 c.p.m. of uniformly radiolabelled substrate were refolded in 30 mM Tris–HCl pH 7.8, 10 mM MgCl₂ and 270 mM KCl and immediately loaded onto 15% native polyacrylamide gels (80:1).

**Sequence comparisons**

Sequence searches were carried out by connection to the NCBI BLAST Email Server at ‘NCBI BLAST E-Mail Server’ (blast@ncbi.nlm.nih.gov) (32). Comparisons were carried out using the COMPARE and DOTPLOT programs of the ‘gcg’ program (33).

**RESULTS**

**Sequences 3′ of the distal GU-rich region are necessary for full polyadenylation activity in vivo**

We have previously shown that a 55 nt element containing a distal GU-rich region enhances polyadenylation activity at the secretory poly(A) site (23,24) using transfection of luciferase reporter constructs. A similar reporter system has been previously used to probe the function of a ‘consensus’ downstream sequence (34). In order to more precisely define the sequences necessary for the enhancing effect of the element spanning the distal GU-rich region of the IgM secretory poly(A) site, we generated a series of 3′ deletion mutants (Fig. 1). These constructs, as well as plasmids Lu pspA1 and Lu psfpA2, were transfected into J558L cells and the luciferase activities were subsequently determined. Figure 1B shows that a sequence located 20 nt beyond the GU-rich sequence of this 55 nt element was necessary to retain the enhancing activity of the whole element, as deletions more proximal than position 2071 abolished the enhancing activity (Fig. 1B, compare samples 3 with 1 and 2).

**An intact stem–loop structure is necessary for the enhancing activity**

RNA folding predictions suggested that the distal GU-rich region could form a stem–loop structure with the nucleotides located downstream of it (Fig. 2A). We therefore examined whether the putative secondary structure might have a role in the biological function of this element since these nucleotides together with the GU-rich region were necessary for full luciferase activity. For this we generated mutants predicted to destroy and restore the stability of the stem–loop (Fig. 2). Mutations which destroyed the stem–loop structure on either side of the stem reduced the enhancing effect of the element to the level of Lu psfpA1 lacking the whole 55 nt element (Fig. 2B, compare samples 1 and 2 with samples 4 and 5). However, when these mutations were combined to reconstruct the predicted stem–loop structure the enhancing effect was restored (Fig. 2B, sample 3). Thus, we conclude that the presence of a stem–loop structure is important for the enhancing activity of the element. Furthermore, the compensatory mutational analysis suggested that nucleotides G2042 and C2048 interacted with nucleotides C2061 and G2056, respectively.

Two asymmetric internal loops contribute to in vivo polyadenylation efficiency

As hairpin and internal loops have been previously shown to be the main sites of protein binding (reviewed in 35–37), we examined how these structures might contribute to the activity of the putative stem–loop in vivo. We designed mutated substrates to either exchange the entire loop for the stable tetraloop sequence GCAA, or closed each internal loop in turn by pairing across the stem (Fig. 2A). The mutant constructs were transfected into J558L cells and luciferase activities were measured as before. Whereas changing the loop had little or no effect on the activity of the stem–loop structure (Fig. 2C, compare samples 1 and 2), partial or full closing of either internal loop reduce the enhancing activity of this element (Fig. 2C, samples 3, 4 and 5). These data suggest that the putative internal loops are important for full activity while the primary sequence of the hairpin loop is relatively less important.

The stem–loop structure contributes to polyadenylation complex formation in vitro

The importance of the stem–loop structure for CPSF/CstF polyadenylation complex formation was investigated by gel retardation assays. In order to ensure that the formation of the CPSF/CstF complex was dependent on the downstream binding factor, CstF, and not the hexanucleotide interacting factor, CPSF, we provided CPSF in excess. To do this we first established the conditions in which a complex between the RNA substrate and CPSF could be formed, in the absence of CstF: HeLa cell nuclear extract was then added in increasing amounts to provide CstF necessary for stable complex formation. Under these conditions complex formation is dependent on the CstF factor binding to the downstream elements and differences among the mutated substrates could be detected. Figure 3 shows that complexes formed easily on the wild-type RNA substrate (BsIgswt), while the RNA substrate containing mutations in the descending arm of the stem–loop BsIgs(G2056U,C2061A) formed complexes poorly (Fig. 3, lanes 10 and 1–3 respectively). Furthermore, the effect of the two point mutations in BsIgs(G2056U,C2061A) could be suppressed by the compensatory mutations in RNA substrate BsIgs(G2056U,C2061A,C2048A,G2042U) restoring the stem–loop structure (Fig. 3, lanes 7–9). We conclude from this that the stem–loop structure of this element is important for complex formation. However, the RNA substrate BsIgs(C2048A,G2042U) in which the ascending arm of the stem–loop was mutated, formed complexes which were only slightly reduced in comparison with the wild-type substrate (Fig. 3, lanes 4–6 and 10). The formation of a complex on the hexanucleotide mutated RNA substrate, BsIgs(U1962,1968,1969,1975G) (Fig. 3, lane 11) was most likely due to the presence of excess CPSF in these reactions, since this complex was not formed unless CPSF was added in excess together with CstF (data not shown).
Independent of nucleotide sequence (42). Thus, lead cleavage by lead (II) preferably cleaves at bulges of RNA structures. This technique can be used to probe tertiary structures of RNA (41). Furthermore, lead (II) cleavage patterns have been used to monitor changes in RNA secondary structure induced by mutations (38–40) and as a technique to locate the loop, as the loop is predicted to be the sole open structure in this structure. These RNA constructs, where the stem is predicted to be destroyed by two point mutations; and (iii) a closed stem–loop structure in which the two predicted internal loops are paired across the stem (Fig. 4 C). The RNase T1 digest of the wild-type stem–loop shows a stronger cut at this G (Fig. 4 A, lane 14). Thus, the RNase T1 cleavage pattern is consistent with the conclusion that this G is located within the loop of the wild-type stem–loop structure.

Comparing the 5’ section of the wild-type stem–loop with that of the open stem–loop construct after 15% PAGE (Fig. 4 A, compare lanes 1–4 with 9–12) shows very similar lead (II) cleavage patterns, although the cuts in the open stem–loop construct are stronger. However, comparing their 3’ sections after 20% PAGE (Fig. 4 B, compare lanes 7–9 with lanes 1–3), the patterns differ. Nucleotides A2060 and C2061A of the open stem–loop are inaccessible to lead (II) even at 0.5 mM. Moreover, the 5’ adjacent GCAC sequence (position 2056–2059) is more accessible to lead (II) cleavage in the open stem–loop construct than in the wild-type stem–loop structure. Finally, the 5’ section of the wild-type stem–loop construct in general is more susceptible to lead (II) cleavage than the 3’ section, while this is not true for either of the open or closed stem–loop constructs.

**Lead (II) cleavage of the stem–loop structure**

Lead (II) cleavage patterns have been used to monitor changes in RNA secondary structure induced by mutations (38–40) and as a technique to probe tertiary structures of RNA (41). Furthermore, lead (II) preferably cleaves at bulges of RNA structures independent of nucleotide sequence (42). Thus, lead cleavage patterns can provide more structural information than RNase T1, which cuts only after guanosine. To monitor the changes in folding brought about by the mutations designed to disrupt the putative secondary structure formed between the GU-rich region and the following 20 nt, we prepared three RNA constructs representing: (i) the wild-type stem–loop; (ii) an open stem–loop construct where the stem is predicted to be destroyed by two point mutations; and (iii) a closed stem–loop structure in which the two predicted internal loops are paired across the stem (Fig. 4 C). The latter construct helps to locate the loop, as the loop is predicted to be the sole open structure in this structure. These RNA constructs were labelled at their 3’ ends and probed by lead (II) cleavage. The cleavage products were analyzed by PAGE and the obtained cleavage patterns of the three RNA constructs were compared. A summary of cleavage sites and putative secondary structures are pictorially represented in Figure 4 C. Lead (II) can be seen to cleave at positions in the predicted hairpin and internal loops.

Lead (II) was only able to cleave the closed stem–loop construct in the region of the hairpin loop (Fig. 4 A, lanes 5–8; B, lanes 4–6; and C). This construct is 2 nt longer, resulting in a different mobility on PAGE. Furthermore, the cuts in the hairpin loop of the closed stem–loop construct are stronger on the 3’ side while the cuts in the hairpin loop of the wild-type stem–loop construct are stronger on both the 3’ and the 5’ sides of the loop and weaker in the center at UGC (Fig. 4 A, lanes 1–4; B, lanes 7–9; and C). The RNase T1 digest of the wild-type stem–loop shows a stronger cut at this G (Fig. 4 A, lane 14). Thus, the RNase T1 cleavage pattern is consistent with the conclusion that this G is located within the loop of the wild-type stem–loop structure.

In order to examine the relative openness of the wild-type stem–loop, the open stem–loop and the closed stem–loop RNA constructs, we performed RNase H cleavage using two hexamer DNA oligomers, complementary to nucleotides 2049–2054 of the hairpin loop, and nucleotides 2039–2044 of the stem between the two internal loops, as the probes (Fig. 4 C). Figure 5 A shows that the accessibility of the oligonucleotide complementary to the hairpin loop sequence, as detected by RNase H cleavage, was ordered thus: the open stem–loop (lanes 9–12) being most accessible, the wild-type stem–loop (lanes 1–4) intermediary and the closed stem–loop (lanes 5–8) least accessible.

Figure 5 B shows that the oligonucleotide complementary to the paired section of the stem in the wild-type stem–loop construct could access the stem at high concentrations (Fig. 5 B, lanes 1–4) while the stem of the open stem–loop was accessible already at lower concentrations of oligonucleotide (Fig. 5 B, lanes 9–12). It appears that this oligonucleotide can bind to other GU-rich sequences of the open stem–loop structure, as a number of cleavage products appeared above and below the expected cleavage products (Fig. 5 B, lanes 9–12). No prominent cleavage products were seen, even at high concentration of oligonucleotide, when the stem of the closed stem–loop structure was probed (Fig. 5 B, lanes 5–7).

Taken together, results from Figure 5 A and B are consistent with the prediction that the wild-type stem–loop construct contains a relatively accessible hairpin loop between positions 2049 and 2055 and a partially basepaired stem structure below this hairpin loop. This is based on the finding that the open stem–loop construct was more accessible to oligonucleotide binding both in the region of the hairpin loop and the stem compared to the wild-type stem–loop construct, and that the stem of the closed stem–loop construct was much less accessible to oligonucleotide hybridization compared with the wild-type stem–loop construct.

**RNase H probing of the stem–loop structure**
Figure 4. Lead (II) and RNase T1 cleavage of RNA constructs spanning the putative stem–loop structure. (A) 3' end labelled wild-type, closed and open stem–loop RNA constructs, were incubated with increasing concentrations of lead (II) and run on 15% PAGE containing 7 M urea (lanes 1–12). A reference ladder was constructed by boiling of the wild-type RNA construct in 50 mM NaOH for 10 s (lane 13). The wild-type RNA construct was partially digested with 0.001 U of RNase T1 (sample 14). The positions of G residues have been indicated to the right. (B) 3' end labelled RNA substrates were incubated with increasing concentration of lead (II) and run on 20% PAGE containing 7 M urea. The positions of three representative nucleotides have been indicated to the right. (C) A pictorial representation of RNase T1 (open arrow) and lead (II) cleavage (black arrows). Size of arrows represent strong and weak cuts. The putative secondary structures of the three indicated RNA constructs are shown.
Mobility on native gels suggest that the stem–loop structure is bent

Asymmetric internal loops can often cause bending of the structure depending upon how the nucleotides within the loop are stacked (43). Different relative electrophoretic mobility of bent structures under denaturing or native conditions provide a simple assay for this possibility (44). We therefore compared the mobilities of the three RNA stem–loop constructs on denaturing and native PAGE. On denaturing PAGE the wild-type stem–loop construct ran with the same mobility as the open stem–loop construct and faster than the closed stem–loop structure which is 2 nt longer (Fig. 6A, compare lanes 1, 2 and 3). However, on native PAGE, the wild-type stem–loop construct ran with the slowest mobility (Fig. 6B, compare 1, 2 and 3). Remarkably, the wild-type stem–loop, in spite of being 2 nt shorter, ran with a slower mobility than the closed stem–loop structure. This suggests that the wild-type stem–loop structure is bent compared to the open or closed stem–loop constructs.

Structural probing of stem–loop structure within the context of the IgM secretory poly(A) site

We used partial RNase T1 digestion to examine the structure of the stem–loop within the context of the full IgM secretory poly(A) site from position 1838 to 2085 (Fig. 7). RNase T1 cleavage cannot provide the detailed information of secondary structure obtained by lead cleavage patterns of shorter structures. However, lead cleavage patterns becomes more difficult to interpret when longer substrates involving more than one structure are used. Partial RNase T1 digestion of the IgM secretory poly(A) site revealed that the hexanucleotide sequence, the cleavage site and the proximal GU-rich region were located in the most RNase T1 accessible region with the ascending GU-rich arm of the stem and the loop providing a further RNase T1 accessible region downstream of this (Fig. 7). Furthermore, the descending arm of the downstream stem–loop structure is relatively RNase T1 inaccessible compared with the ascending arm as was found to be the case for the isolated structure (compare Figs 4 and 7).

The downstream sequence and the stem–loop structure are conserved

We were interested to compare sequences downstream of the hexanucleotide sequence of the µ-secretory poly(A) site with other mammalian species to look for conservation of particular sequence elements. Unfortunately, although cDNAs are available for a number of species, mammalian genomic sequences are only available for hamster and human in addition to mouse. A summary of these comparisons is presented in Figure 8. Overall the mouse and hamster sequences are 80% similar, while the...
mouse and human sequences are 63% similar. A highly conserved 14 nt long sequence element, TGCAC(G, A or C)CACCCTGC, was found in all three species immediately downstream of the distal GU-rich region. This 14mer element corresponds to the descending arm of the mouse stem–loop structure which basepairs with the preceding GU-rich region. The region preceding the highly conserved 14mer element, which includes the GU-rich region, is conserved between mouse and hamster. However, in human the corresponding sequence appears to have been deleted and the preceding sequence is highly mutated relative to the hamster and mouse (Fig. 8A).

We next investigated whether a stem–loop structure similar to that in mouse could be found downstream of the hamster and human secretory poly(A) sites. As can be seen in Figure 8B, both the hamster and human sequences could form similar stem–loop structures to that of the mouse. In all cases the GU-rich regions are basepaired with the conserved 14mer element. The hamster stem–loop is almost identical to the mouse counterpart, while the human stem–loop structure differs. In the human structure the ascending arm is now formed by the mutated G-rich sequence. Thus, stem–loop structures can be folded downstream of the mouse, hamster and human IgM secretory poly(A) sites.

**DISCUSSION**

We have shown here that an intact stem–loop structure is necessary for the biological activity of the distal downstream element of the murine IgM secretory poly(A) site, which include a GU-rich element and a conserved element as ascending and descending arms of the stem–loop respectively. The stem–loop structure contains two asymmetric internal loops both of which contribute to activity. Furthermore, CPSF/CstF complex formation in vitro is disrupted by mutations in the descending arm of the stem–loop which destroy the structure and restored by compensatory mutations which reform the structure. Structural probing, using lead (II) cleavage, RNase H probing or RNase T1 digestion of the stem–loop structure, are consistent with the predicted folding. Native gel mobility analysis suggested that the stem–loop structure may bend, possibly due to the stacking of the nucleotides within the two asymmetric internal loops.

There are numerous examples where hairpin loops, internal loops and bulges represent targets for site-specific recognition of RNA by proteins (45,46). Such structures appear to function to open the major groove and expose unpaired nucleotides for recognition by proteins (36,47). The unpaired nucleotides in the internal asymmetric loops may function to increase the accessibility of these sequences to interact with components of the polyadenylation machinery. Furthermore, the distortability introduced by the bend in the stem–loop structure may make the protein–RNA interaction thermodynamically more favorable (36,37,48). In addition, the strong downstream stem–loop structure may help to maintain the area surrounding the hexanucleotide sequence in a
non-basepaired state, increasing its accessibility for CPSF binding (49).

The GU-rich sequence is surprisingly underconserved between mouse and human. A clue to this lack of conservation may be its sequence similarity to a part of the U2-IR2 domain of Epstein–Barr virus (EBV), a tumourigenic virus which invades human B cells. The sequence TTGTGTGTGCT in the EBV genome, which is identical to the distal GU-rich region of the murine µ-secretory poly(A) site, is located downstream of potential 3′ splice sites and immediately precedes the AUG codon in the open reading frame of the U2 protein which has been implicated in tumourigenesis (50). There is evidence that selective pressure by EBV has induced mutations in other human genes (reviewed in 51). Thus it is possible that it was advantageous to modify the distal GU-rich region of the human µ-secretory poly(A) site once EBV became endemic in humans.

The most conserved sequence between mouse, hamster and human within the 55 nt enhancing sequence is the sequence TGCAC(G, A or C)CACCTGC immediately following the GU-rich sequence in mouse and hamster and predicted to form the descending arm of the stem–loop structure. The conservation of the TGCAC(G, A or C)CACCTGC sequence between species including humans, in which the ascending arm of the stem–loop is not conserved, raises the possibility that the formation of a secondary structure may not be the only function of this conserved sequence. The two point mutations on the 5′ arm of the stem which disrupt the conserved sequence reduce polyadenylation activity in vivo as well as reduce complex formation in vitro, whereas the two point mutations on the 3′ side of the stem, while reducing in vivo activity, have only a minor effect on complex formation. This raises the possibility that the conserved sequence may play an extra role in complex formation. Alternatively, the stem–loop structure may form differently in vivo and in vitro under the conditions used here. In this case, it is possible that mutations on the 5′ arm may disrupt this structure less severely in vitro than in vivo than those in the 3′ arm. Furthermore, activity is restored both in vivo and in vitro when these mutations are combined even though this construct still contains mutations in the conserved sequence. This argues that it is the secondary structure which is important here. Furthermore, if the GU-rich region has been deleted in the human under selective pressure, it is relevant that the sequence preceding the GU-rich sequence deletion is heavily mutated relative to the surrounding sequences. These heavily mutated sequences can form a stem–loop structure with the conserved sequence resulting in a structure which bears a great deal of resemblance to the murine structure.

The IgM secretory poly(A) site contains dual upstream and downstream sequence elements, the latter consisting of two GU-rich elements which are both suboptimally placed and both required for the formation of the pre-initiation complex in vitro (23). That the distal GU-rich element is presented as the ascending arm of a bent stem–loop structure may be an important feature making the region more efficiently recognized by the polyadenylation machinery, thus compensating for its suboptimal location. We have proposed that the dual structure may provide a flexibility to accommodate regulatory events at this particular poly(A) site (23). The stem–loop structure may also provide novel possibilities for regulatory events as the target for structure modifying activities such as RNA helicases (52), single-strand RNA binding proteins, such as HnRNP proteins (53) or strand-annealing activities (54). It will be an important task for future work to identify the protein or protein factors that specifically interact with the stem–loop structure. Obvious candidates are the basal polyadenylation factor CstF, as the 55 nt element spanning this stem–loop structure is necessary for
CPSF/CstF complex formation in vitro (23), and the inducible 30 kDa polypeptide that we previously have identified (24).

ACKNOWLEDGEMENTS

We would like to thank Walter Dietrich Goetz for the donation of the pPKLT55 luciferase construct and Clinton MacDonald for the 3A7 antibodies. We would also like to thank Leif Kiresebom, Gerhart Wagner and Devapriya Choudhury for helpful discussions and suggestions throughout the project, and Hans Johansson for helpful comments on the manuscript. C.P. was supported by a Wellcome Travelling Fellowship and as a gästforskare by the Swedish Medical Research Council. C.B.K. was supported by an EU-TMR fellowship. The work was supported by the Wellcome Trust, the Swedish Cancer Society and the Swedish Natural Science Research Council.

REFERENCES