Sequences Adjacent to the 5′ Splice Site Control U1A Binding Upstream of the IgM Heavy Chain Secretory Poly(A) Site*

Received for publication, February 6, 2003, and in revised form, March 18, 2003
Published, JBC Papers in Press, March 31, 2003, DOI 10.1074/jbc.M301349200

Catherine Phillips† and Samuel Gunderson
From the Nelson Laboratories, Rutgers University, Piscataway, New Jersey 08854

We have recently shown that the stability of the alternatively expressed immunoglobulin M heavy chain secretory mRNA is developmentally regulated by U1A. U1A binds novel non-consensus sites upstream of the secretory poly(A) site and inhibits poly(A) tail addition in undifferentiated cells. U1A’s dependence for binding and function upon a stem-loop structure has been extensively characterized for the consensus sites. We therefore probed the structure surrounding the novel U1A binding sites. We show that two of the three novel binding sites represent the major single-stranded regions upstream of the secretory poly(A) site, consistent with a major role at this site. The strength of binding and ability of U1A to inhibit poly(A) polymerase correlate with the accessibility of the novel sites. However, long range interactions are responsible for maintaining them in an open configuration. Mutation of an RNase V1-sensitive site 102 nucleotides upstream, directly adjacent to the competing 5′ splice site, changes the structure of one of the U1A binding sites and thus abolishes the binding of the second U1A molecule and the ability of U1A ability to inhibit poly(A) polymerase activity at this site. These sites bind U1A via its N-terminal domain but with a 10-fold lower affinity than U1 small nuclear RNA. This lower binding affinity is more conducive to U1A’s regulation of poly(A) tail addition to heterologous mRNA.

The major source of diversity in metazoans is the ability to process pre-mRNA into alternative mRNAs via the selection of alternative splice sites and alternative 3′-ends, giving rise to a range of proteins encoded by the same gene that are differentially expressed during growth and development (for review see Ref. 1). An important pattern of alternative processing involves a competition between splicing on of extra exons and 3′/H11032

22102

This paper is available on line at http://www.jbc.org

† The work was supported by New Jersey Commission on Cancer Research Grant 98-0717, American Heart Foundation Grant 98S0071T, and National Institutes of Health Grant GM57286. 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: Nelson Laboratories, Rutgers University, Rm. 322, 604 Allison Rd., Piscataway, NJ 08854. Tel.: 732-445-1017; Fax: 732-445-4231; E-mail: catphil8@rci.rutgers.edu.
high affinity binding (26), the stem-loop structure of the binding site has also been shown to be essential for maintaining the binding site in an accessible configuration and for making contact with certain U1A residues (14, 27).

We therefore wanted to determine the secondary structure surrounding the novel U1A binding sites upstream of the \(\mu\)-secretory poly(A) site. We show that the U1A binding sites are the major accessible regions upstream of the secretory poly(A) site, consistent with a major role at this site. These structures are dependent on the flanking sequences, which control the accessibility and function of the U1A binding sites. We identify the sequences in the 5’ flanking sequence necessary for the structure of the hexanucleotide proximal (8s) U1A binding site and find that they are directly adjacent to the 5’ splice site involved in the competing splicing reaction.

**EXPERIMENTAL PROCEDURES**

**Plasmid Constructs—**For *in vitro* transcription the PCR products from the secretory poly(A) site and surrounding sequences, containing mutations and a 5’ EcoRI site and a 3’ XbaI site, introduced as part of the synthetic primers, were cloned into pGem 3Zf (containing a T7 promoter in the forward direction) between the EcoRI and XbaI sites. The forward primers were as follows: for V1S1 and 2, CCGAATTCGGACCGTGGACAAGTGTTGAGGTTGAACCGTCTCTGGTACATGCTCTG; for V1S1, CCGAATTCGGACCGTGGACAAGTGTTGAGGTTGAACCGTCTCTGGTACATGCTCTG. The reverse primer was: GCGTCTAGATAGG GGGACCAAGTACGTCGACCCGACCGTGGACAAGTGTTGAGGTTGAACCGTCTCTGGTACATGCTCTG as previously described (12).

**RNA Substrates—**Templates for *in vitro* binding assays (EMSAs) were obtained by restriction of the pGem 3Zf constructs with XbaI or DdeI. Uniformly labeled RNA substrates were synthesized by *in vitro* transcription using T7 RNA polymerase as previously described (28). Transcripts that were 5’-end-labeled were obtained by the same protocol, substituting \(\gamma\)-\(^{32}\)P-GTP as the radiolabeled nucleotide. RNA substrates that were 5’-end labeled were synthesized by the same protocol omitting \(\alpha\)-\(^{32}\)P-UTP. The cold RNA was subsequently labeled at the 3’-end with \(\text{[}^3\text{P]}\text{UTP} \) using T4 RNA ligase as previously described (29).

All transcripts were purified by extraction after electrophoresis on 8% polyacrylamide gels containing 7 M urea.

**Recombinant Proteins—**Untagged recombinant wild type U1A and U1A\(_{11-117}\) were purified from *Escherichia coli* as described previously (30, 31). Recombinant bovine poly(A) polymerase, tagged at the C terminus with 6 histidines, was purified from *E. coli* on Ni\(_2\)NTA. The His-tagged C terminus ensured that all of the C-terminal residues were present after purification (31).

**RNA Secondary Structural Probing and Footprinting—**Probing of the secondary structure of RNA was carried out as previously described (29). Briefly, all RNA substrates were first refolded by heating in reaction buffer at 70 °C for 2 min followed by cooling to 37 °C over a 30-min period. For RNase T1, RNase V1 and lead (II) cleavage, 10,000 cpm of 3’- or 5’-end-labeled RNA substrate was incubated at 37 °C in a total volume of 5 \(\mu\)l with either 0.4-0.001 unit of RNase T1 (Roche Applied Science) or 0.072-0.00072 unit of RNase V1 (originally Amer sham Biosciences, kind gift of Jean Patterson) in 50 mM PIPEs, pH 7.0, 100 mM NaCl, 10 mM MgCl\(_2\) for 5 min or 0.1-0.5 mM Pb(OAc)\(_2\) in 50 mM Tris HCl, pH 7.5, 100 mM NaCl, and 10 mM MgCl\(_2\) for 10 min. The reactions were stopped by the addition of 10 \(\mu\)l of 5X EDTA. Samples were analyzed on 15% PAGE containing 7 M urea. A reference nucleotide ladder was made by boiling of substrates with 50 mM NaOH, 10 mM EDTA for 10 s. Footprints were obtained by adding U1A before T1 digestion. These reactions were stopped by the addition of 180 \(\mu\)l of proteinase K buffer (50 mM Tris, pH 7.9, 10 mM EDTA, 230 mM NaCl, 0.2% SDS) and 50 \(\mu\)g of proteinase K and incubated for 10 min at 30 °C. This was followed by phenol/chloroform extraction and ethanol precipitation before loading on PAGE as above.

**Electrophoretic Mobility Shift and Poly(A) Addition Assays—**A U1A binding assay were performed as previously described (14) using 10,000 cpm of \(\text{[}^3\text{P]}\text{UTP} \) to label RNA substrate and recombinant U1A as indicated. Nonspecific poly(A) addition assays were performed as previously described (32) with 50 ng of recombinant poly(A) polymerase, and incubations were carried out for 30 min at 30 °C.

**RESULTS**

**U1A binding sites are in the major accessible regions upstream of the secretory poly(A) site—**Fig. 1A is a schematic diagram of the alternative processing of the \(\mu\)-heavy chain pre-mRNA, showing that the \(\mu\)-secretory poly(A) site is located within the intron used to splice on the membrane exons. Fig. 1B shows the precise positions of the various elements known to be important for \(\mu\)-secretory mRNA expression mentioned in the introduction, including the location of the U1A binding motifs, as well as the location of the 5’- and 3’-ends of the RNA substrate used in this report. It also shows the proximity between the 5’ splice site of this intron and the \(\mu\)-secretory poly(A) site used for the \(\mu\)-secretory mRNA. Positions are numbered according to the mouse IgM sequence with accession number V00818. Fig. 1C shows the sequence between 1838 and 1950, which we have previously scanned with mutation (mut1 through mut10), and the position of the 2s, 4s, and 8s mutations of the three novel U1A binding sites.

Because U1A binding to the consensus sequence on U1 snRNA is dependent upon a particular RNA hairpin stem-loop structure, we wanted to probe the structure surrounding the novel U1A binding sites that we have recently identified upstream of the \(\mu\)-secretory poly(A) site. For this we used partial RNase T1 digestion. Substrates were digested with limiting amounts of RNase T1 for 5 min and then immediately run on 15% denaturing PAGE. RNase T1 cuts 3’ to Gs with a preference for single-stranded RNA. Thus Gs that give a strong pattern of cuts will be situated in the single-stranded region of hairpin loops or bulges (33). At first we examined the pattern of RNase T1 digestion of 3’-end-labeled IgM (1970-2085) substrate, which spans all the

![Fig. 1. Schematic model of the immunoglobulin secretory poly(A) site.](https://example.com/fig1.png)
known important elements involved in regulation of expression of \( \mu \)-secretory mRNA (Fig. 2A). These include the upstream 5' splice site, the novel U1A binding sites, the hexanucleotide sequence, and both downstream GU-rich regions. In this way the RNA structure surrounding the U1A binding sites will be preserved in a larger context and more accurately represents the physiological structure. A single-nucleotide ladder was generated by hydroxy radical cleavage and served as a marker for the RNA structure surrounding the U1A binding sites and more accurately represents the RNA structure surrounding the U1A binding sites will be preserved in a larger context and more accurately represents the physiological structure. A single-nucleotide ladder was generated by hydroxy radical cleavage and served as a marker for the RNA structure surrounding the U1A binding motifs (Fig. 2A, lane 1). The deepest RNase T1 cuts occur around the hexanucleotide sequence and the cleavage/polyadenylation site (1971–2007), which is predicted to form a loop containing these sequences in a form accessible to components of the polyadenylation and cleavage complex (and RNase T1) (Fig. 2A, lanes 3–8). The proximal and distal GU-rich regions are also accessible (labeled on Fig. 2A). The region of RNA surrounding the U1A binding motifs (Gs at 1854, 1877, and 1925) is indicated by a bracket. It can be seen from Fig. 2A that these are the major accessible regions upstream of the hexanucleotide sequence.

We labeled the IgM-(1790–2085) RNA substrate either at the 3' end or at the 5' end and ran the gels long enough to be able to accurately determine the position of the T1 cuts in the region surrounding the U1A binding motifs (1854–1930) and were able to identify a reproducible pattern. The RNase T1 partial digestion of the 5'-end-labeled wild type substrate is shown in Fig. 2B (lane 2). A distinct reproducible pattern of cuts can be seen around U1A binding motifs 4s (1875–1882) and 8s (1923–1930) (Fig. 2B, lane 2). The 8s motif consists of one deep cut (1925) and two lighter cuts (1919 and 1921) that are closer to a mirror image set of cuts between the two U1A binding motifs (1903). The 4s motif consists of a pattern of two deep cuts (1871 and 1925) followed by two further cuts (1881 and 1883). The U1A binding motifs are labeled individually in Fig. 2B (8s, 4s, and 2s). The 2s motif was not RNase T1-sensitive.

To confirm the identities of the cuts ascertained by comparison with the hydroxy ladder, we employed mutants that replace the AUGC (or AGGC) motifs with AAAA (see Fig. 1C). These mutations abolish U1A binding (13). As the G is replaced by an A, these mutants lose the ability to be cut by RNase T1 at that site. It can be seen in Fig. 2B that mutation of either the 4s or 8s site results in the loss of the deep cuts at 1877 or 1925, respectively (see asterisks in lane 3 and 4). This confirms that these accessible sites are at the position of the novel U1A binding sites. To further confirm that open structures form around the 4s and 8s U1A binding sites, we performed lead cleavage analysis, which we have successfully used previously to identify bulges in RNA structure (29). Lead (II) cleavage patterning is a technique to probe tertiary structures of RNA (34). Furthermore, lead (II) preferably cleaves at bulges of RNA structure independent of nucleotide sequence (35). It can be seen that lead cuts around the hexanucleotide sequence, and two patterns of cuts form around the 4s and 8s U1A binding sites (lane 6). Taken together these data show that the 4s and 8s novel U1A binding sites are the major accessible regions upstream of the hexanucleotide sequence.

We noted that the pattern of RNase T1 cuts surrounding the 8s binding motif changes only in the region of the 8s mutation, in that the deep cut at 1925 is lost (Fig. 2B, compare lanes 2 and 4). In contrast, for the 4s mutation, the pattern of RNase T1 cuts surrounding the (1875–1882) motif changes more extensively (compare lanes 2 and 3), suggesting a remodeling of the surrounding area upon 4s mutation.

**Is the Double-stranded 2s Motif Remodeled upon U1A Binding?**—To determine the extent of remodeling with the 4s mutation and to further probe the structure surrounding the 2s binding site, we performed RNase V1 partial digestion. RNase V1 preferentially cuts double-stranded RNA (33). It can be seen in Fig. 3A that the 2s motif lies within a double-stranded region.
In fact it appears to be within a symmetrical structure consisting of four RNase V1-sensitive sites. The 8s mutation once again appears not to be remodeled as probed by RNase V1 (compare lane 2 and 6). In contrast, the 4s mutation leads to remodeling of the RNase V1-sensitive sites from 1911 to 1858 (compare lanes 2 and 4, labeled with a bar in lane 4). The RNase V1-sensitive site at 1855 (the 2s motif) is still visible but diminished in the 4s mutation (compare lanes 2 and 4).

We wondered whether the susceptibility this region to remodeling upon mutation of the AUGC motif might reflect its ability to be remodeled upon U1A binding this motif. We therefore attempted to footprint the RNA with U1A to look for changes in RNA structure brought about by the addition of the flanking sequences. The results are presented in Fig. 4 (A and B), and the pattern of cuts for each substrate is summarized in the schematic diagram in C. A longer exposure of the gel allows visualization of lighter cuts for better comparison.

Deletion of the 5' flanking region alone (IgM-(1838–2085)), Fig. 4A, lanes 2–7) results in the loss of the deep cut at 1925 and only a slight diminution of the cut at 1877. On the other hand, deletion of the 3' flanking region alone, which spans the downstream stem-loop (IgM-(1790–2030), Fig. 4B, lanes 9–14), results in the diminution of cuts surrounding 1877, whereas those at 1903 and 1925 are maintained. The diminution of the cut at 4s is dependent upon the presence of the 5' flanking region (1790–1838), because there is almost no difference in its

---

**Fig. 3. Remodeling of the structure surrounding the 2s and 4s motifs.** A, RNase T1 and V1 structural probing of the structure surrounding the U1A binding motifs. Samples in lane pairs 1 and 2, 3 and 4, and 5 and 6 contain digested 5'-end-labeled wild type, 4s, and 8s mutated IgM-(1790–2085), respectively. Samples in lanes 1, 3, and 5 contained 0.1 unit of RNase T1. Samples in lanes 2, 4, and 6 contained 0.0072 unit of RNase V1. B, U1A footprinting: 80 ng of U1A was allowed to bind the RNA before partial RNase T1 digestion. Samples were treated with proteinase K before PAGE. Footprints are marked with an asterisk.

**Structure of the U1A Binding Sites**—As a first step to determine the secondary structure surrounding the single-stranded U1A binding motifs, we were interested in knowing the minimal RNA sequence that could maintain the U1A binding sites in accessible regions. For this we tested the effect of deleting the 5' and 3' flanking regions. The substrates were 3'-end-labeled and run for 5 h on 15% gels so that the pattern surrounding the U1A binding sites was well separated and that the pattern surrounding the hexanucleotide sequence was retained within the gel for convenient reference. The pattern of RNase T1 digestion in the whole substrate, which can be used as a reference point, can be seen in lanes 9–14 of Fig. 4A. We performed a partial RNase T1 digest of 3'-end-labeled substrates, with or without the flanking regions to look for changes in RNA structure brought about by the addition of the flanking sequences. The results are presented in Fig. 4 (A and B), and the pattern of cuts for each substrate is summarized in the schematic diagram in C. A longer exposure of the gel allows visualization of lighter cuts for better comparison.

Deletion of the 5' flanking region alone (IgM-(1838–2085), Fig. 4A, lanes 2–7) results in the loss of the deep cut at 1925 and only a slight diminution of the cut at 1877. On the other hand, deletion of the 3' flanking region alone, which spans the downstream stem-loop (IgM-(1790–2030), Fig. 4B, lanes 9–14), results in the diminution of cuts surrounding 1877, whereas those at 1903 and 1925 are maintained. The diminution of the cut at 4s is dependent upon the presence of the 5' flanking region (1790–1838), because there is almost no difference in its
U1A Binding at the IgM Secretory Poly(A) Site

A comparison of the effects of the 3' and 5' flanking regions on the RNase T1 partial digestion pattern. A, substrates with the 3' flanking region with and without the 5' flanking region. Samples in lanes 1–7 and 8–14 contain 3'-end-labeled IgM-(1838–2085) and IgM-(1790–2085), respectively. B, substrates without the 3' flanking region with and without the 5' flanking region. Samples in lanes 1–7 and 8–14 contained 3'-end-labeled IgM-(1838–2030) and IgM-(1790–2030), respectively. Samples in each set of seven lanes (1–7 and 8–14 in A and B) contained 0, 0.4, 0.2, 0.1, 0.04, 0.02, and 0.01 unit of RNase T1, respectively. Samples were run on 15% denaturing PAGE and visualized by phosphorimaging. C, schematic diagram of the RNase T1 patterns obtained in A and B. The sizes of the solid circles indicate the intensity of the cut. Boxes mark the positions of the U1A binding sites. Substrates are labeled.

intensity between IgM-(1838–2085) and IgM-(1838–2030), which are both greater than when the 5' flanking region is present. The 4s cut at 1877 is only 50 nucleotides from the end of the substrate when the 5' flanking sequence is missing and may be more accessible for this reason. We presume that a comparison of the pattern of sites affected by the deletion of the 3' flanking sequence has more physiological relevance in the longer flanking region, which includes the 5' flanking regions, but that sequences in the 5' flanking region have a small effect on the 4s site.

Deletion of both 5' and 3' flanking regions results in the complete disruption of the RNA structure surrounding the hexanucleotide sequence and a new deep cut appears in the 1938–54 region (Fig. 4B, lanes 2–7). The region between 1925 and 1903 is remodeled, and the deep cuts at 1925 and 1903 are lost. In both substrates missing the 5' flanking region, a cut appears at 1855 in the region surrounding the 2s (1854–1860) hexanucleotide sequence, which spans the U1A binding motif (Fig. 4B, lanes 2–7). From this we conclude that sequences 5' of 1838 are necessary to maintain the 2s binding site in an inaccessible form.

From consideration of the pattern of sites affected by the respective flanking regions, a "criss-cross" pattern emerges: Deletion of the 3' flanking sequence has the greatest effect on the cut at 1877 (compare substrate IgM-(1790–2085) with IgM-(1790–2030), Fig. 4A, lanes 8–14 with Fig. 4B, lanes 8–14, respectively). In contrast, deletion of the 5' flanking region has the most effect on the 25 cut and results in the appearance of the cut at 1855 (compare substrate IgM-(1790–2085) with IgM-(1838–2085), Fig. 4A, lanes 8–14 with Fig. 4A, lanes 1–7, respectively). Thus the 3' flanking region affects the 4s U1A binding site, whereas the 5' flanking sequence affects the 8s U1A binding site in a criss-cross effect. However, the 5' flanking region also affects both the 4s and the 2s site and appears to have a greater effect on the 8s motif than does the 3' flanking sequence on the 4s motif. These effects are summarized in Fig. 4C.

The Accessibility of the U1A Binding Sites Correlates with the Stronger Binding of U1A and Its Ability to Inhibit Poly(A) Addition in Vitro—We next wanted to test if the accessibility of the U1A binding sites had an effect on U1A binding. We performed electrophoretic mobility shift assays comparing the various substrates using uniformly labeled in vitro transcribed RNA substrates and recombinant non-tagged full-length wild type U1A. U1A binds the substrate IgM-(1838–2030) that spans the U1A binding motifs, the hexanucleotide sequence of the poly(A) site, and the proximal GU-rich region (Fig. 5A, lanes 1–4) (see Fig. 1 for location of sequences). Inclusion of the sequence 2030–2085 enhances the binding (lanes 5–8). This sequence spans a stem-loop structure that includes the distal GU-rich region and is necessary for full polyadenylation activity. However, this sequence does not bind U1A on its own (data not shown). We next tested the effect of 5' flanking sequences. Adding 5' flanking sequences (1790–1838) greatly increased U1A binding to both the IgM-(1838–2030) and IgM-(1838–2085) substrate (Fig. 5A, compare lanes 1–4 with 9–12 and lanes 5–8 with 13–16, respectively). Thus both the 3' and 5' flanking sequences separately enhance the binding of U1A to
the substrate, although inclusion of the 5' flanking sequence has a greater effect than the 3' flanking sequence. Furthermore, 5' and 3' flanking sequences in combination enhance this binding further (Fig. 5A, lanes 13–16). Thus the strength of binding correlates with the accessibility of the binding sites determined by the RNase T1 partial digestion (compare Figs. 4A and 5A). Because the extra sequences at either end of the U1A binding sites do not bind U1A themselves, we conclude that the flanking sequences contribute to the enhanced binding by helping to maintain the binding sequences in an accessible configuration.

We note that, on all substrates with one or more of the flanking regions, three bands form, the slowest mobility band representing a complex of RNA with three molecules of U1A. This is in contrast to two accessible regions on the naked RNA (see Fig. 4C). This raises the possibility that two U1A molecules binding the accessible site recruit the third molecule to the in accessible 2s site.

**Accessibility of U1A Motifs Correlates with the Ability to Inhibit Poly(A) Tail Formation**—We next examined the functional effect of the changes in secondary structure brought about by inclusion of the 5' flanking sequence on the ability of U1A to inhibit poly(A) polymerase activity. For this we used a reconstituted system containing only recombiant PAP and U1A and in vitro transcribed RNA (Fig. 5B). We compared the two cleaved substrates IgM-(1838–2030) and IgM-(1790–2030) for the ability of U1A to inhibit poly(A) polymerase activity (Fig. 5B, lanes 5–8 and 9–12, respectively). Both substrates contain the U1A binding motifs but, as can be seen from the structural analysis, IgM-(1838–2030) does not present both U1A binding motifs in single-stranded, RNase T1-accessible regions (Fig. 4A, compare lanes 1–7 with 8–14, position 1925 (8s)). In addition, it can be seen in Fig. 5A that the second molecule of U1A necessary for co-operative inhibition of PAP binds this substrate poorly (Fig. 5A, lanes 1–4). These substrates were compared with the minimal RNA substrate containing only the hexanucleotide sequence and no upstream U1A binding motifs (Fig. 5B, lanes 1–4). All three substrates produced strong poly(A) tails in the absence of U1A (lanes 1, 5, and 9). However, U1A was unable to inhibit poly(A) polymerase activity on the IgM-(1838–2030) substrate, although it contains all the U1A binding motifs. A mild inhibition is seen with 100 ng of U1A (lane 8), which is comparable to that seen with the minimal substrate that does not contain the binding sites (lane 4) and is therefore a nonspecific effect. In contrast, on IgM-(1790–2030), U1A could begin to inhibit PAP activity at 10 ng of U1A and completely abolish it with 30 ng. Thus we conclude that the ability to inhibit PAP activity at the μ-secretory poly(A) site correlates with the secondary structure of RNA that holds both U1A binding motifs in a single-stranded configuration.

**Identifying Potential Hybridizing Sequences in the 5'-Flanking Region**—As a first approach to identifying sequences in the 5' flanking regions that could interact with the 8s U1A binding site, we probed the RNA structure of this region. For this, we examined the pattern of partial RNase T1 and V1 of the 5'-end-labeled (1790–2085) substrate run on 15% gels for shorter times so that the smallest fragments at the 5'-end of the substrate are retained on the gel (Fig. 6A). We were able to discern two possible structures. One of these lies between 1836 and 1865; 1838, 1843, 1855, and 1863 are the centers of four patterns of RNase V1 cuts, suggesting a “strong double-stranded structure” (Fig. 6A, lanes 6 and 7), which contained the 2s motif as noted in Fig. 3A. The second, between 1803 and 1824, spans the 5' splice site. The Gs at 1803 and 1824 are more accessible than the two Gs that flank the 5' splice site at 1810 and 1811 (Fig. 6A, lanes 3 and 4, labeled "structure surrounding the 5' splice site"). This is surrounded on either side by a pattern of RNase V1 cuts, suggesting double-stranded regions, the most prominent occurring at 1808 and 1820, which are both Cs (see Fig. 6B, where Cs are boxed). From an examination of the sequences on either side of the 5' splice site there is no obvious pattern of hybridization that would be consistent with a hairpin stem-loop structure (see Fig. 6B, sequences indicated by lines and labeled VIS1 and VIS2, where VIS = RNase V1-sensitive). We therefore wondered if these would hybridize with sequences in the vicinity of the 8s U1A binding site. Thus we mutated these sequences into their opposite and investigated the effect on the structure of the 8s U1A binding site.

As can be seen in Fig. 6C mutating both RNase V1-sensitive sites (VIS1 plus 2), alters the structure of the 8s binding site (Fig. 6C, compare lanes 5 and 6 with lanes 2 and 3 at position 1925 (8s)). The cut at 4s was diminished, but the pattern was maintained. Furthermore, the accessibility of the region between 1971 and 1925 is altered, demonstrating that mutation these two sequences have multiple effects on the structure. Nevertheless, the inaccessibility of 2s was unaffected. To isolate the effect of the mutations on the 8s binding site, we mutated the two sequences separately. The VIS1 mutation yielded a pattern of T1 partial digest very similar to that of the wild type (compare lanes 8 and 9 with lanes 2 and 3). In contrast, the VIS2 mutation produced a pattern around the 8s...
site that resembled that of the double-mutated RNA, whereas the 4s site and the region between 1971 and 1925 are comparable with the wild type. We therefore conclude that the sequence downstream of the 5'/H11032 splice site (V1S2) is responsible for the structure of the 8s U1A binding site. Inclusion of the V1S1 mutation with the V1S2 mutation appears to have an effect on the overall structure but the effect of the V1S1 alone was minor.

Sequences Adjacent to the 5' Splice Site Are Responsible for the Ability of the Second U1A Molecule to Bind and Function

We next investigated whether the change in structure at the 8s U1A binding site affects U1A binding and function at this site. We first compared the wild type and the double mutant (V1S1 plus 2) in EMSAs (Fig. 7A). It can be seen that, although two molecules of U1A can bind the wild type substrate (lanes 1–6), only one molecule can bind the mutated substrate (lanes 7–12). We next compared the effect of the V1S1 or V1S2 mutant on U1A binding (Fig. 7B). Whereas both the wild type and V1S1 mutant could bind two molecules of U1A (lanes 1–4 and 5–8), the V1S2 mutant could only bind one molecule and the second could no longer bind.

To determine the functional significance of these changes, we tested the consequences of the mutations in reconstituted polyadenylation assays. Because it has been shown in the auto-regulatory system that two molecules are necessary to inhibit poly(A) polymerase activity, any reduction in the number of molecules of U1A bound would be predicted to result in loss of polyadenylation inhibitory activity at the secretory poly(A) site. It can be seen in Fig. 7C that in the case of the wild type (lanes 1–6) and the V1S1 mutation (lanes 11–15) in which two molecules of U1A can bind, increasing concentrations of U1A result

Fig. 6. RNase T1 and V1 structural probing (A) Analysis of the 5' flanking region. Lanes 1 and 2 are the hydroxy ladders. All samples contain the 5'-end-labeled wild type IgM-(1790–2085). Samples in lanes 3 and 4 contained 0.04 and 0.1 unit of RNase T1, respectively. Samples in lanes 6 and 7 contained 0.0072 and 0.0036 unit of RNase V1, respectively. Sample in lane 5 contained only the RNA substrate. Samples were run on 15% denaturing PAGE and visualized by phosphorimaging. B, model showing sequences in the region of the 5' splice site 1803–1827. The most RNase V1-sensitive nucleotides are boxed (1808 and 1820), and those showing some sensitivity are marked with a line. The 5' splice site is indicated with an arrow. C, RNase T1 analysis of the U1A binding sites with or without mutations in the RNase V1-sensitive sites surrounding the 5'/H11032 splice site: positions are indicated on the left side. The hexanucleotide sequence (A2UA3) is marked by brackets, and the cleavage site is indicated by a line. U1A binding sites 8s, 4s, and 2s are indicated. Lanes 2, 5, 8, and 11 contained 0.1 unit of RNase T1, and lanes 3, 6, 9, and 12 contained 0.04 unit of RNase T1.

Fig. 7. The RNase V1-sensitive sites adjacent to the 5' splice site affect U1A binding and function. A, effect of mutation of both RNase V1-sensitive sites on U1A binding. EMSAs with recombinant U1A and uniformly labeled in vitro transcribed substrates. B, effect of single RNase V1-sensitive sites on binding. EMSAs are as in A. C, effect of double- and single-RNase V1-sensitive sites on the ability of U1A to inhibit poly(A) polymerase activity. Nonspecific polyadenylation assays using in vitro transcribed uniformly labeled substrates, recombinant poly(A) polymerase, and U1A at the indicated amounts.
Fig. 8. Affinity of the novel U1A binding sites in the context of IgM-(1790–2085) relative to U1 snRNA. Shown are sample EMSAs used for binding curve for (A) full-length wild type U1A and (B) U1A 1–117 containing only the N-terminal RRM. EMSAs were done with uniformly labeled RNA substrates and the indicated amounts of recombinant U1A. Samples were run on 5% native PAGE, and the amount bound versus input was quantitated by phosphorimaging.

in the inhibition of poly(A) polymerase activity at this site. However, in the double mutant (V1S1 plus 2) (lanes 6–10) and the single V1S2 mutant (lanes 16–20) in which only one molecule can bind, increasing concentrations of U1A can no longer inhibit poly(A) polymerase activity. From this we conclude that the mutation in the sequence directly downstream of the 5’ splice (V1S2), which affects the structure of the 8s U1A binding site, also affects the ability of the second U1A molecule to bind and thus the ability of U1A to inhibit poly(A) polymerase activity at this site.

Estimation of Relative Binding Affinity of Full-length U1A and the Ability of the N-terminal RRM of U1A to Bind Upstream of the Secretory Poly(A) Site—Having characterized the secondary structure necessary for forming the U1A binding sites, we wanted to look at U1A binding to these novel motifs more closely to relate it to what is known for U1A binding to the consensus sequence in U1snRNP. U1A binds U1 snRNA via RRM1 its N-terminal domain (27) with a Kd value of 2 \times 10^{-11} \text{ M} (19).

We first compared the ability of IgM-(1790–2085) versus U1 snRNA to bind full-length U1A in EMSAs (Fig. 8A). U1 snRNA binds one molecule of U1A, whereas multiple U1As bind upstream of the secretory poly(A) site. We therefore compared low U1A concentrations when only one U1A bound per RNA substrate (Fig. 8A). This would presumably represent the U1A binding site with the highest affinity for U1A. We adjusted the relative input of uniformly [32P]UTP-radiolabeled U1 snRNA and IgM-(1790–2085) for the relative numbers of Us in the substrates, because IgM-(1790–2085) is almost twice as long as the U1 snRNA substrate (295 nucleotides versus 164 nucleotides, compare input in Fig. 8A, lane 1 with lane 7). We quantitated the amount of RNA bound to U1A relative to input for each substrate by phosphorimaging and constructed binding curves for four separate assays. We estimated the affinity of the IgM-(1790–2085) substrate for the wild type full-length U1A to be 10 ± 4-fold lower than that of U1 snRNA. We then constructed binding curves for U1A binding to the substrates containing single or double mutations in the U1A binding sites to gauge the relative strength of each of the binding sites (data not shown). We established the order of strength of the novel U1A binding sites as 4s > 8s > 2s (data not shown).

It has been previously shown that the N-terminal RRM is solely responsible for the binding of U1A to U1 snRNA (19, 21). To test whether this is also the case for IgM-(1790–2085), we examined its affinity, relative to U1 snRNA, for the truncated version of U1A containing only amino acids 1–117, which span RRM1. We employed two different EMSA assays for the two lengths of U1A. The shift due to the very small truncated U1A (1–117) was not visible above the long IgM-(1790–2085) substrate unless we ran the gel for 5 h (Fig. 8B). This was not necessary for the U1A wild type, which is larger, and the shift was visible after 1.5 h (Fig. 8A). Quantitation and analysis showed that the truncated U1A bound the substrate IgM-(1790–2085) with a similar relative affinity to that of the wild type (Fig. 8, compare A and B). We overexposed the results for U1 snRNA relative to IgM-(1790–2085) for easier visual comparison of RNA bound versus input (Fig. 8B, compare lanes 1, 2, and 8). Taken together, we conclude that U1A binds upstream of the μ-secretory poly(A) site via RRM1 with a 10-fold lower affinity than to U1 snRNA.

Discussion

We have recently shown that U1A binds novel sites upstream of the secretory poly(A) site and regulates poly(A) tail addition in a developmental manner (13). Here we further characterize these novel binding sites in terms of what is known about U1A binding to consensus sites. The U1A binding to consensus sites requires a stem loop structure in which both an accessible binding site and a specific interaction with the closing pair of the stem play essential roles. We therefore wanted to determine how far these novel sites satisfy these requirements. The major findings of this work are that two of the three novel sites occur at the major accessible regions upstream of the secretory poly(A) site, that long range interactions maintain these in accessible configuration, and that U1A binding and function correlate with the accessibility of these sites. Furthermore, we have identified sequences responsible for the structure of the 8s sites, and these are directly adjacent to the upstream 5’ splice site involved in the splicing out of the intron containing the secretory poly(A) site and the U1A binding sites. Thus these novel sites satisfy the requirements determined for the consensus sites but at the same time extend the definition of a U1A binding site. U1A binds these novel sites via its N-terminal RRM1, which is the same domain that binds the consensus sites, but at a 10-fold lower affinity. We suggest that the lower binding affinity is more consistent with a regulatory role of U1A on a heterologous mRNA. These findings have consequences when searching for similar binding sites on other genes that undergo the same type of regulation.

The IgM Secretory Poly(A) Site Exhibits a Criss-cross Structural Pattern—RNA structural probing of the whole μ-secretory poly(A) site and regions 5’ of this that span the 5’ splice site used for the competing splicing reaction show that two of the U1A binding sites are the major single-stranded regions upstream of the hexanucleotide sequence, consistent with a major role in the control of the secretory poly(A) site. These U1A binding sites (4s and 8s) consist of open structures, which are maintained in an open configuration by surrounding sequences. The sequences that maintain the most 3’ U1A binding site (8s) in an open configuration are located 102 nucleotides upstream in the 5’ flanking region, directly adjacent to the competing 5’ splice site. Furthermore, the 3’ flanking sequences affect the upstream 4s U1A binding site. Taken together, the flanking sequences affect the U1A binding site most distal to them in a criss-cross fashion, suggesting a more com-
plicated secondary structure than a simple stem loop. Because U1A binding is dependent on the secondary structure of RNA, this raises the possibility that the effect of U1A could be regulated by structural re-modeling caused by developmentally induced factors binding quite distant from the U1A binding site.

**The Novel Sites Compared with the Crystallographic and NMR Structures of Consensus Sites**—We have shown that the binding strength and ability of U1A to inhibit poly(A) polymerase at the secretory poly(A) site correlate with the accessibility of two of its binding sites (4s and 8s), which is consistent with what is known to be important for consensus U1A binding sites (14, 27). We have also shown that U1A binds this site via its N-terminal domain, the same as is used for binding to U1snRNP. Based on crystallographic and NMR studies of the interaction of the N-terminal domain of U1A with the consensus binding sites on U1snRNP and PIE RNA (19, 21, 22, 24), it is possible to envision how U1A binds the novel sites. The bases of the consensus binding site, in particular the UGC of the AUUGACG of the unpaired loop, stack with amino acid side chains within the β sheet of the U1A protein via electrostatic interactions (see Protein data bank pdb accession number 1DZ5). One of the loops connecting the β sheet of the U1A protein (loop 3) protrudes through the RNA loop in a “button in a button-hole” effect, and another (loop 1) interacts with the 5’ stem (22). Via these loops, U1A makes specific contacts with the 3’ C of the AUUGACG and the C-G pair at the base of the RNA hairpin, respectively (22). Both loop 3 of U1A and the RNA hairpin to which it binds becomes more structured in the protein-RNA complex in an “induced fit” effect (16, 18, 22). Furthermore, loop 3 contacts U1 snRNA and the PIE elements differently, suggesting flexibility in this interaction (15, 22). The UGC sequence is common to the AUGC of the novel sequences and could be predicted to interact in the same way. There is a C three nucleotides downstream of the AUGC motif for both motifs (see Fig. 1). The distance between the electrostatic interacting portion and the C is increased by two nucleotides, which could be accommodated by the flexibility of the interaction. There is no C in close proximity 5’ of either the novel motifs that could form a C-G pair for interaction with loop 1. However, the interaction between the 5’-adjacent sequence and those adjacent to the 8s U1A binding site may provide the “stem” for interaction with loop 1. In this case the most likely pairing would be G-C, which has been shown to bind with lower affinity (22). We suggest that these adjustments for binding are responsible for the 10-fold lower binding affinity of U1A to the novel motifs.

**Juxtaposition of Two U1A Molecules to Form the PAP Inhibitory Pocket**—In the autoregulatory system the two dimerizing U1A molecules are juxtaposed by the complicated stem loop structure of PIE. In this way they form a bi-partite binding pocket for PAP that is necessary to inhibit its activity (14, 25, 37). We show here that U1A behaves in the same way in that failure to bind a second U1A molecule abolishes the ability of U1A to inhibit PAP activity at the secretory poly(A) site. Thus two U1A molecules can dimerize to form the PAP inhibitor binding pocket on this RNA despite the distance of at least 40 nucleotides between them. Therefore, their juxtaposition is imposed either by a rigid RNA structure or by the dimerization of two U1A molecules across a distance. The latter would require a flexible RNA structure, which we favor because of the remodeling of the RNA surrounding the 4s U1A binding sites when we introduced mutations into this site (see Fig. 2B). Furthermore, U1A is known to bind RNA with an induced fit as discussed above.

*Is the 2s Site Remodeled?*—We found that, although two of the novel motifs (4s and 8s) were accessible to RNaseT1 on the naked RNA, the third (2s) was within a double-stranded RNase V1-sensitive site. We know that 2s is a U1A binding site, because we have previously shown that mutation of 2s alone causes a release in the inhibition of poly(A) site expression in vivo and that mutation of all three motives is necessary for the complete loss of binding to this substrate in vitro. In addition, we have shown previously (13) and it can be seen in Figs. 6A and 7A that three bands bind these substrates. We noted that mutation of the 4s motif caused a remodeling of the surrounding regions to at least within three nucleotides of the 2s site. This finding, combined with what is known about U1A dependence upon an unpaired region for binding, raises the possibility that U1A binding at the 4s site remodels this region, increasing the accessibility of the 2s site and at the same time recruiting U1A to the 2s site via the dimerization interaction. The question then arises that, if a U1A at 4s alone is sufficient to recruit a second molecule to the 2s site, then these two U1A molecules would presumably be able to dimerize and create an inhibitory binding pocket for PAP even when the 8s site is lost. However, this is not the case (see Fig. 1). A scenario consistent with these findings is that U1A binding at both the 4s and 8s site is necessary to achieve sufficient binding strength to recruit a third U1A to 2s. However, this is a subject for future investigation.

*Is This a General Mechanism?*—We reasoned that other non-immunoglobulin genes that produce alternative forms by activation of a poly(A) site within an internal exon might be regulated by U1A binding these novel motifs. Using the findings from this study, we identified eight candidate genes that contained clusters of potential novel U1A binding sites near the regulated poly(A) site. These are predominantly receptors but also represented are transcription factors and biologically important enzymes. For instance, the oncogenes epidermal growth factor receptor and HER2 receptor are both receptors with truncated forms (38, 39). In both cases the truncated versions are festooned with novel U1A binding motifs that may play a role in silencing the truncated forms. An investigation of these genes will determine whether this is a more widely used regulatory mechanism.

**Acknowledgments**—We thank K. Nagai for donation of the U1A 1–117 and Jean Patterson, Youngtae Ro, and Ricardo Carion for their willingness to part with half of their last 40 μl of RNase V1 when it was no longer in production.

**REFERENCES**

1. Maniatis, T., and Tasic, B. (2002) Nature 418, 236–243
2. Edwalds-Gilbert, G., Veraldi, K. L., and Milcarek, C. (1997) Nucleic Acids Res. 25, 2547–2561
3. Early, P., Rogers, J., Davis, M., Calumane, K., Bond, M., Wall, R., and Hood, L. (1980) Cell 20, 313–319
4. Lamsen, G., and Koshland, M. E. (1984) J. Exp. Med. 160, 877–892
5. Cox, A., and Emtage, J. S. (1989) Nucleic Acids Res. 17, 10439–10454
6. Mason, J. O., Williams, G. T., and Neuberger, M. S. (1988) Genes Dev. 2, 1003–1010
7. Berberich, I., and Schimpl, A. (1990) Eur. J. Immunol. 20, 445–448
8. Zhao, J., Hyman, L., and Moore, C. (1999) Microbiol. Mol. Biol. Rev. 63, 405–445
9. Murthy, K. G., and Manley, J. L. (1995) Genes Dev. 9, 2672–2683
10. Ecker, R., Ellmeyer, W., and Biernat, M. L. (1991) EMBO J. 10, 3531–3532
11. Huang, Y., and Carmichael, G. C. (1996) Mol. Cell. Biol. 16, 1534–1542
12. Ellmore, P., McCarty, T., Rosbash, M., Parker, R., and Jensen, T. H. (2001) Nature 413, 538–542
13. Phillips, C., Jung, S., and Gunderson, S. I. (2001) EMBO J. 20, 6443–6452
14. van Zyl, C. W. G., Gunderson, S. I., Jansen, E. J. R., Boelens, W. C., Polycarpou-Schwarz, M., Mattaj, I. W., and Venrooij, W. J. v. (1993) EMBO J. 12, 5191–5200
15. Allain, F. H., Howe, P. W., Neuhaus, D., and Varani, G. (1997) EMBO J. 16, 5764–5772
16. Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K., and Neuhaus, D. (1998) J. Mol. Biol. 287, 398–411
17. Beck, D. L., Stump, W. T., and Hall, K. B. (1998) RNA (N. Y.) 4, 331–339

*C. Phillips, unpublished results.*
U1A Binding at the IgM Secretory Poly(A) Site

18. Gubser, C. C., and Varani, G. (1996) *Biochemistry* **35**, 2253–2267
19. Hall, K. B., and Stump, W. T. (1992) *Nucleic Acids Res.* **20**, 4283–4290
20. Hall, K. B. (1994) *Biochemistry* **33**, 10076–10088
21. Howe, P. W., Nagai, K., Neuhaus, D., and Varani, G. (1994) *EMBO J.* **13**, 3873–3881
22. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H., and Nagai, K. (1994) *Nature* **372**, 432–438
23. Rimmele, M. E., and Belasco, J. G. (1998) *RNA (N. Y.)* **4**, 1386–1396
24. Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J., and Mattaj, I. W. (1990) *Nature* **345**, 502–506
25. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D., and Varani, G. (2000) *Nat. Struct. Biol.* **7**, 329–335
26. Stump, W. T., and Hall, K. B. (1995) *RNA (N. Y.)* **1**, 55–63
27. Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamms, J., and Mattaj, I. W. (1989) *EMBO J.* **8**, 4163–4170
28. Phillips, C., and Virtanen, A. (1997) *Nucleic Acids Res.* **25**, 2344–2351
29. Phillips, C., Kyrnakopoulou, C. B., and Virtanen, A. (1999) *Nucleic Acids Res.* **27**, 429–438
30. Boelens, W. C., Jansen, E. J. R., Venrouj, W. J. v., Stripecke, R., Mattaj, I. W., and Gunderson, S. I. (1993) *Cell* **72**, 881–892
31. Gunderson, S. I., Vagner, S., Polycarpou-Schwarz, M., and Mattaj, I. W. (1997) *Genes Dev.* **11**, 761–773
32. Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C., and Mattaj, I. W. (1994) *Cell* **76**, 531–541
33. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) *Nucleic Acids Res.* **15**, 9109–9128
34. Hüttenhofer, A., Westhof, E., and Bock, A. (1996) *RNA (N. Y.)* **2**, 354–366
35. Pan, T., and Uhlenbeck, O. C. (1992) *Nature* **358**, 560–564
36. Teunissen, S. W. M., van Gerlier, C. W. G., and van Venrooij, W. J. (1997) *Biochem.* **36**, 1762–1789
37. Klein Gunnewiek, J. M., Hussein, R. I., van Aarssen, Y., Palacios, D., de Jong, R., van Venrooij, W. J., and Gunderson, S. I. (2000) *Mol. Cell. Biol.* **20**, 2209–2217
38. Scott, G. K., Robles, R., Park, J. W., Montgomery, P. A., Daniel, J., Holmes, W. E., Lee, J., Keller, G. A., Li, W. L., and Fendly, B. M. (1993) *Mol. Cell. Biol.* **13**, 2247–2257
39. Reiter, J. L., and Maihle, N. J. (1996) *Nucl. Acids Res.* **24**, 4050–4056
Sequences Adjacent to the 5′ Splice Site Control U1A Binding Upstream of the IgM Heavy Chain Secretory Poly(A) Site
Catherine Phillips and Samuel Gunderson

J. Biol. Chem. 2003, 278:22102-22111.
doi: 10.1074/jbc.M301349200 originally published online March 31, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M301349200

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 39 references, 13 of which can be accessed free at http://www.jbc.org/content/278/24/22102.full.html#ref-list-1