Inhibiting expression of specific genes in mammalian cells with 5′ end-mutated U1 small nuclear RNAs targeted to terminal exons of pre-mRNA

Puri Fortes*[†], Yolanda Cuevas*[‡], Fei Guan[‡§], Peng Liu[‡¶], Sara Pentlicky[§], Stephen P. Jung[§], Maria L. Martínez-Chantar*, Jesús Prieto*, David Rowe[¶], and Samuel I. Gunderson[†§]

*Department of Medicine, University of Navarra, Irunlarrea 1, Pamplona 31008, Spain; [§]Department of Molecular Biology and Biochemistry, Nelson Laboratory, Rutgers University, The State University of New Jersey, Piscataway, NJ 08854; and [¶]Department of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06030

Reducing or eliminating expression of a given gene is likely to require multiple methods to ensure coverage of all of the genes in a given mammalian cell. We and others [Furth, P. A., Choe, W. T., Rex, J. H., Byrne, J. C., and Baker, C. C. (1994) *Mol. Cell. Biol.* 14, 5278–5289] have previously shown that U1 small nuclear (sn) RNA, both natural or with 5′ end mutations, can specifically inhibit reporter gene expression in mammalian cells. This inhibition occurs when the U1 snRNA 5′ end base pairs near the polyadenylation signal of the reporter gene's pre-mRNA. This base pairing inhibits poly(A) tail addition, a key, nearly universal step in mRNA biosynthesis, resulting in degradation of the mRNA. Here we demonstrate that expression of endogenous mammalian genes can be efficiently inhibited by transiently or stably expressed 5′ end-mutated U1 snRNA. Also, we determine the inhibitory mechanism and establish a set of rules to use this technique and to improve the efficiency of inhibition. Two U1 snRNAs base paired to a single pre-mRNA act synergistically, resulting in up to 700-fold inhibition of the expression of specific reporter genes and 25-fold inhibition of endogenous genes. Surprisingly, distance from the U1 snRNA binding site to the poly(A) signal is not critical for inhibition, instead the U1 snRNA must be targeted to the terminal exon of the pre-mRNA. This could reflect a disruption by the 5′ end-mutated U1 snRNA of the definition of the terminal exon as described by the exon definition model.

mutant U1 snRNAs | gene expression inhibition | polyadenylation inhibition

Nearly all mature eukaryotic mRNAs, histone mRNAs being the only significant exception, contain a poly(A) tail that functions to control mRNA degradation (1, 2), translation efficiency (3, 4), and possibly mRNA export (5). The poly(A) tail is posttranscriptionally added to pre-mRNAs in the nucleus by a two-step reaction called cleavage and polyadenylation. In the first step, the pre-mRNA is endonucleolytically cleaved into two molecules at the cleavage site, also called the poly(A) site (6). In the second step, a poly(A) tail of ≈250 adenosine residues (70 in *Saccharomyces cerevisiae*) is added to the 3′ end of the upstream fragment by the enzyme poly(A) polymerase, whereas the downstream pre-mRNA fragment is degraded. Although much work remains to be done, the reaction mechanism is relatively well understood, and the major pre-mRNA sequence elements and nearly all of the polypeptides present in the cleavage/polyadenylation machinery have been identified in both *S. cerevisiae* and mammalian cells (6–8).

Regulation of poly(A) tail addition (7, 8) typically involves the choice between two or more poly(A) sites on a single pre-mRNA, resulting in mRNAs differing either in their coding regions or 3′ UTR sequences that may affect mRNA stability, localization, and translatability (1–8). A second, less common way to control poly(A) tail addition is by an "on–off" switch mechanism whereby a single poly(A) site is either active or

inactive. The U1A splicing factor represents the best understood example of this type of regulation, where excess U1A protein negatively autoregulates its own synthesis by inhibiting poly(A) tail addition to its own pre-mRNA (9). The mechanism involves dimerization of U1A on the U1A pre-mRNA, resulting in the formation of a binding pocket that interacts with and inhibits poly(A) polymerase (9, 10). Inhibition of poly(A) tail addition to the U1A pre-mRNA leads to reduced U1A mRNA and protein levels. Importantly, U1A selectively inhibits only the poly(A) addition step, which presumably allows the cleavage step to endonucleolytically remove downstream cryptic poly(A) sites that may become activated during the regulation.

A second example of such an on–off regulation is where the U1 small nuclear ribonucleoprotein (U1 snRNP) particle, normally involved in recognition of the 5′ splice site sequence (5′ss) during pre-mRNA splicing, binds upstream of a poly(A) site inhibiting its usage (11–14). Certain papillomaviruses use this type of regulation during early stages of infection to repress late gene expression via inhibition of the activity of the late poly(A) site by a U1 snRNP:5′ss complex (11, 14). The inhibitory mechanism is similar to U1A autoregulation and involves two regions of the U1 snRNP-specific U170K protein that bind to and inhibit poly(A) polymerase (14). An interesting and related example is in HIV, in which a poly(A) site in the 5′ LTR is maintained in an inactive state by the downstream binding of U1 snRNP, through a splicing independent mechanism that is still to be determined (15).

Mammalian U1 snRNP contains 10 polypeptides bound to the 164-nt U1 snRNA (Fig. 1*A*) and functions primarily to direct early steps in spliceosome formation by binding to the pre-mRNA exon–intron boundary (16, 17). Nucleotides 2–11 of the 5′ end of U1 snRNA base pair to the 5′ss of the pre-mRNA. There are a number of examples in which expression of 5′ end-mutated U1 snRNA leads to assembly of 5′ end-mutated U1 snRNP able to base pair to target sequences thereby changing the pattern of pre-mRNA processing of the targeted gene. Examples include suppressor U1 snRNAs (i.e., U1 snRNPs) (18, 19) that bind to and activate splicing at nonphysiological 5′ splice sites as well as 5′ end-mutated U1 snRNAs/U1 snRNPs that bind the 3′ UTR inhibiting expression of exogenous and stably expressed reporter genes (11–13). Here, for the first time, we use 5′ end-mutated U1 snRNAs/U1 snRNPs to inhibit expression of endogenous genes in mammalian cells (rat and human) and demonstrate that two U1 snRNAs base paired to a single target gene's pre-mRNA act synergistically, resulting in up to 700-fold
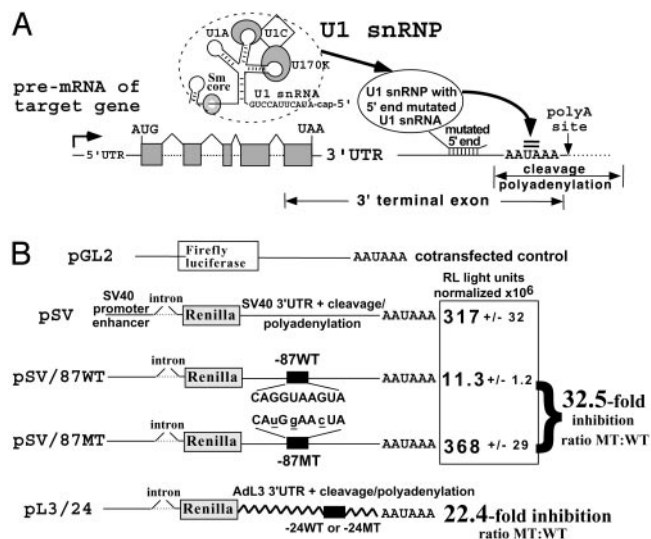

Fig. 1. Strategy and reporter system. (*A*) Schematic of strategy. Shown is the composition of endogenous U1 snRNP comprised of three U1 snRNP-specific proteins (U1A, U1C, and U170K), U1 snRNA (black line) with stemloops 1–4 and the Sm core consisting of seven Sm proteins bound to the U1 snRNA–Sm site. Also shown is a generic mammalian pre-mRNA containing in its 3′ UTR a nonphysiological U1 snRNP binding site bound to U1 snRNP containing a 5′ end-mutated U1 snRNA. Coding regions are shaded boxes, introns are dotted lines, the 5′ UTR and 3′ UTR are solid lines, and the double arrow indicates sequences that direct pre-mRNA cleavage and poly(A) tail addition. The U1 snRNP pre-mRNA complex inhibits, in a gene-specific manner, the addition of the poly(A) tail, thereby preventing maturation of the pre-mRNA leading to reduced mature mRNA levels. (*B*) Establishment of inhibition of reporter mRNA biosynthesis. pGL2 (Promega), expressing the firefly luciferase gene (*Photinus pyralis*), was cotransfected with all of the *Renilla* expression reporter plasmids. The *Renilla* light units were normalized to the firefly light units, thereby controlling for transfection efficiency, cell recovery, and lysis after transfection, a distinct advantage over other reporter assays. pSV (pRL-SV40, Promega) contains a SV40 promoter-enhancer, a β globin intron, the coding region for *Renilla* luciferase (RL), and an SV40-derived (SV) 3′ UTR and cleavage and polyadenylation signal. The pL3 construct is identical to pSV except the 3′ UTR and cleavage and polyadenylation signal was derived from adenovirus L3 (L3; wavy line). pSV/87WT and pSV/87MT are a pair of matched plasmids containing either a WT U1 snRNA binding site (solid box) or the MT (solid box) positioned 87 nt upstream of the AAUAAA sequence. Likewise, pL3/24 represents a pair of matched plasmids containing either a WT (pL3/24WT) or MT (pL3/24MT) site positioned 24 nt upstream. A total of 2.5 μg of each construct, along with 2.5 μg of pGL2 internal control, was transfected into 10⁶ HeLa cells on a 10-cm plate, and after 48 h the cells were harvested and luciferase levels were measured with the Promega Dual Luciferase assay. The *Renilla* light units were normalized to the cotransfected firefly light units, and the results from three or more independent experiments are shown. Standard deviations (not shown) were <15% in all cases. The fold inhibition is the ratio of normalized *Renilla* activity of the MT5′ss plasmid to the matching WT5′ss plasmid. Symbols are as in *A*.

inhibition of the expression of specific reporter genes and 25-fold inhibition of endogenous genes. We determine the inhibitory mechanism and establish a set of rules to use this technique, including the finding that the U1 snRNA must be targeted to the terminal exon of a pre-mRNA with a polyadenylation signal.

Materials and Methods

HeLa cells (14) and Ros 17/2.8 cells (13) were passaged and cultured under standard conditions. For details, see *Supporting Text*, which is published as supporting information on the PNAS web site, www.pnas.org.

Stable cell lines were produced as described (12, 13) with neomycin as the selection reagent. Luciferase assays were done according to the manufacturer's directions. Luciferase activity

was measured in a Berthold luminometer (Lumat LB 9507). Whole cell lysates used for Western blotting were prepared by lysis of cells in buffer containing 150 mM NaCl, 0.1% Triton, 50 mM Tris (pH 8), and 0.1 mM EDTA, followed by a brief sonication, centrifugation for 5 min at $20,000 \times g$, and collection of the supernatant. Chemiluminescence Western blots were performed and quantitated as described (14) with detection reagents from Amersham Pharmacia and Immobilon-P nylon membrane from Millipore. Antibody dilutions were 1:2,000 for the anti-aryl sulfatase A (ASA) antibody (kindly provided by P. Manowitz, University of Medicine and Dentistry, Piscataway, NJ), 1:10,000 for the anti-GAPDH antibody (Chemicon), and 1:5,000 for the anti-osteocalcin antibody (Diagnostic Systems Laboratories, Webster, TX). Northern blotting was done as described (12, 13) with probes generated by the random priming labeling method. The 900-nt Collagen 1a1 probe was derived from the C-terminal and 3′ noncoding region of the rat Collagen 1a1 gene, which is the least conserved part between collagen genes. The 500-nt osteocalcin probe was derived from a mouse osteocalcin cDNA, which is 89% identical to the rat gene.

Results

Rationale. Even if we have successfully used 5′ end-mutated U1 snRNAs to inhibit the expression of reporter genes, we (and others) had two concerns about the application of the technique to endogenous genes: (*i*) the lack of explicit rules for choosing the target site on the pre-mRNA, and (*ii*) the significant potential for cytotoxicity of long-term, stable expression of 5′ end-mutated U1 snRNAs in cells. Cytotoxicity could arise by the 5′ end-mutated U1 snRNA, either inducing errors in pre-mRNA splicing or binding near and inhibiting the poly(A) signals of essential, nontargeted genes. Thus we first focused on establishing an explicit set of rules by determining (*i*) the effects of the location [distance to the poly(A) site and terminal exon versus upstream exons or introns] of the target site, (*ii*) the effects of the number of target sites on a given pre-mRNA, (*iii*) the effects of secondary structures in the target sites or between the target site and the polyadenylation site, and (*iv*) the importance of the presence of polyadenylation sequences to get inhibition. The results of this analysis are shown in Figs. 1–3.

Assay to Measure Inhibition. Fig. 1*A* is a schematic of the strategy where either natural or 5′ end-mutated U1 snRNA assembles into a U1 snRNP that base pairs to a target sequence in the terminal exon of the target gene, leading to inhibition of poly(A) tail addition. Advantages of this strategy include that (*i*) U1 snRNP proteins are highly abundant, and (*ii*) the natural activity of U1 snRNA/U1 snRNP to base pair to pre-mRNA and to localize to the desired cellular compartment is exploited. To systematically measure the degree and specificity of inhibition, we performed transient transfections with the dual luciferase reporter assay system (Promega), where the *Renilla* luciferase gene would be targeted for inhibition and a cotransfected firefly luciferase gene would not be targeted. Because both luciferase activities are measured in the same transfected cell lysate, this assay provides a more accurate internal control and normalization of data than assays based on normalization to cotransfected β-galactosidase or CAT control genes. The pRL-simian virus 40 (pSV) (Promega) *Renilla* reporter plasmid was modified by introducing a 10-nt-long, consensus, wild-type (WT) or mutant (MT) 5′ss 87 nt upstream of the poly(A) signal sequence (AAUAAA) to produce the pSV/87WT and pSV/87MT plasmids, respectively (Fig. 1*B*). The particular 3-nt mutation in the MT5′ss inhibits splicing *in vitro* and *in vivo* and does not bind purified U1 snRNP *in vitro* (14, 17). pSV/87MT and pSV had similar *Renilla* expression levels, whereas pSV/87WT had a 32.5-fold reduction in *Renilla* activity (Fig. 1*B*). Because the pSV/87WT and pSV/87MT plasmids differ by only 3 nt, we

conclude that the presence of the WT5′ss, i.e., a U1 snRNA/U1 snRNP binding site, confers specific inhibition of the expression of the *Renilla* gene.

Parameters Affecting Inhibition. Systematic increases in the length of the U1 snRNA binding site from 10 nt up to 16 nt had only slight reductions in the efficiency of inhibition, meaning potent (>10-fold in all cases) specific inhibition was still observed. However, when U1 snRNA binding sites were ≤7 nt or when the 5′ end of U1 snRNA was extended to interact with 25 or 50 nt of *Renilla* mRNA, no inhibition was observed (data not shown). Therefore, we kept the binding site at 10 nt throughout this work. We found inhibition does not depend on the adjacent simian virus 40 (SV40)-derived flanking sequences, because their replacement with corresponding sequences from adenovirus L3 (pL3/24 in Fig. 1*B*) or by a synthetic cleavage and polyadenylation signal (pSPA/20; see Fig. 3) gave similar levels of inhibition. The generality and robust nature of the U1 snRNA inhibitory strategy is underscored by the fact that similar levels of inhibition were observed when (*i*) these constructs were transfected into 25 different mammalian cell lines (e.g., HeLa, COS-1, BHK, K562, 293F, Jurkat, 3T3, etc.), (*ii*) the U1 snRNA binding site was inserted in the firefly luciferase reporter instead of the *Renilla* reporter, (*iii*) the time of transfection was varied from 24 to 72 h, and (*iv*) the amount of transfected reporter gene was varied across a 600-fold range (data not shown).

Effects of Increased Distance, Multiple Binding Sites, and Introns. Even if a 20- to 30-fold inhibition (reducing protein levels down to 3–5%) was respectable compared with either antisense techniques (20) or RNAi (21, 22), we decided to test whether additional U1 snRNA binding sites or their relative distance to the polyadenylation signal would affect their inhibitory activity. With increasing distance, only a very gradual reduction in inhibitory activity was observed even when the binding sites were 1,190 nt upstream of the polyadenylation signal in the 5′ UTR (Fig. 2*A*). This surprising "long-distance" inhibition suggests to us that even if poly(A) signals are being targeted, the inhibition could be mediated by the disruption of the definition of the terminal exon (see *Discussion*). In all cases, two or more binding sites were better than one, and their inhibitory activities were synergistic instead of being additive (Fig. 2*A*). This synergistic inhibition was observed regardless of whether the binding sites were placed in tandem (6 nt apart) or 171 nt apart (the pL3/195/24 pair of constructs). Increasing from two binding sites to three [Fig. 2*A* gives the example of pSV/(145)₃] further increased the level of inhibition; however, four binding sites was not better than three (data not shown). We note here that this remarkable increase in inhibition could also be explained as a stochastic process in which the two binding sites act independently and the resulting fold inhibition is the fold inhibition of one site multiplied with that of the other site. Also, insertion of an intron, but not a mutated intron, downstream of the binding site negated the inhibition [pL3/(1190)₂/wtInt is shown in Fig. 2*A* as an example]. Because the intron had no effect when inserted upstream, we conclude that U1 snRNA inhibits poly(A) tail addition only when base paired in the 3′ terminal exon of the pre-mRNA, an idea consistent with the exon definition model (23) and already supported by our earlier work (12, 13) and by data in Fig. 4*C*.

Effect of RNA Secondary Structure. In the context of splicing, there are many examples where pre-mRNA secondary structures (usually stem structures) inhibit the splicing activity of a WT5′ss (19, 24). Fig. 2*B* summarizes results where the placement of a U1 snRNA binding site in a secondary stem structure caused complete loss of inhibitory activity. Stem structures placed adjacent to the U1 snRNA binding sites had either a minor (Fig.
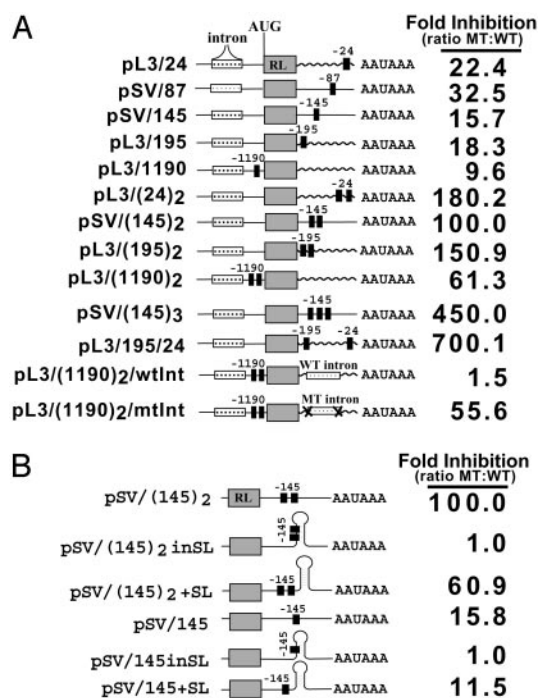

Fig. 2. Parameters affecting inhibition. (*A*) Effects of distance and number of binding sites on the level of inhibition. In every case, a pair of matched plasmids was constructed where 1, 2, or 3 binding sites (solid boxes) were positioned in various parts of the pre-mRNA as indicated. For each plasmid, the U1 snRNA binding sites were all WT or all MT. Transfection assays were done as described in Fig. 1*B*. Plasmids beginning with the letters "pSV" or "pL3" contain SV40 or adenovirus L3 3′ UTR sequences, respectively, as in Fig. 1*B*. For the pL3/(1190)2/wtInt and pL3/(1190)2/mtInt matching plasmids, a WT or a 5′ss/3′ss mutated intron was inserted in the 3′ UTR of the *Renilla* gene. For both intron-containing constructs, alternative splicing was avoided by designing the U1 snRNA binding sites located in the 5′ UTR to be poor matches to U6 snRNA. Ribonuclease protection experiments demonstrate that the WT intron was spliced as expected (data not shown). Note that a value of 1.0 represents, by definition, no inhibition. Note that standard deviations were <15% in all cases. (*B*) The effect of target accessibility and RNA secondary structure on inhibition is shown. In each case, a pair of matched plasmids was constructed where one or two binding sites (solid boxes) were positioned adjacent to or within stem loop structures. For each plasmid, the U1 snRNA binding sites were all WT or all MT. Transfection assays and standard deviations (<15% in each case) were done as described in Fig. 1*B*. The binding sites in pSV/145inSL and pSV/(145)2inSL are able to form a stem loop structure with adjoining pre-mRNA sequences, whereas the binding sites in pSV/145+SL and pSV/(145)2+SL are adjacent to (6 nt apart) but not within the stem loop structure.

2*B*) or no effect (Fig. 3, the pHist/SPA/20 pair of plasmids) on inhibition. The control set of plasmids, containing the mutant U1 snRNA binding site placed within the same type of stem structures, showed little change in expression, indicating such structures do not intrinsically affect *Renilla* expression.

Only "AAUAAA-Type" of Poly(A) Signals Are Inhibited. Although it is thought that U1 snRNA binding to the target mRNA inhibits the polyadenylation machinery that utilizes the AAUAAA sequence (11–14), a polyadenylation-unrelated mechanism could also be involved. It is possible that U1 snRNP binding could (*i*) retain pre-mRNA in the nucleus, (*ii*) inhibit mRNA-related processes by antisense interference, (*iii*) remove the cleavage and polyadenylation sequences by activation of splicing in the 3′ UTR by the 3′ UTR-bound U1 snRNP, or (*iv*) destabilize mRNA, which could occur by any of a number of processes. Also, it was important to directly test whether classical polyadenylation sequences are required for inhibition. There are many viral
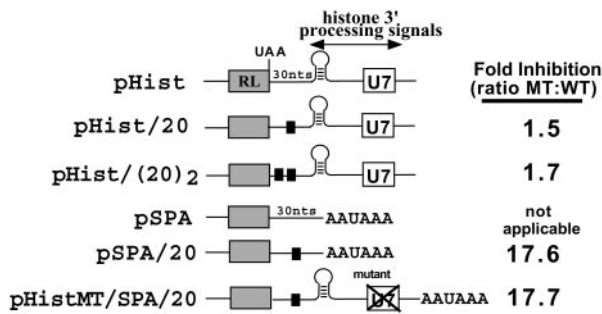
Fortes *et al.*

Fig. 3. Inhibitory mechanism targets poly(A) signals. The U1 snRNA binding site inhibits poly(A) signal-directed 3′ end processing but not histone 3′ end processing. Transfection assays and quantitation were done as described in Fig. 1B. pHist replaces the 3′ UTR and cleavage polyadenylation signals of the pSV plasmid with the histone 3′ end processing signals consisting of a histone stem loop and U7 snRNP binding site. As in Fig. 1, pHist/20 represents a pair of matched plasmids containing one WT or MT U1 snRNA binding site placed 20 nt upstream of the histone stem loop. pHist/(20)₂ represents a pair of matched plasmids containing two WT or MT U1 snRNA binding sites. pSPA replaces the histone 3′ end processing signals with artificially strong cleavage and polyadenylation sequences. pSPA/20 represents a pair of matched plasmids containing one WT or MT U1 binding site. Likewise, pHistMT/SPA/20 is a matching pair of WT/MT plasmids containing both types of processing signals except the histone signal is inactivated by deletion of the U7 binding site. Note that standard deviations were <15% in all cases.

mRNAs that are polyadenylated by a molecular mechanism that is distinct from the classical "AAUAAA" type of polyadenylation (25). There are several mRNAs whose 3′ end processing does not depend on the polyadenylation machinery and that do not have a poly(A) tail, with histone mRNAs being the best characterized of this type (26). Furthermore, it would be of interest to know whether U1 snRNA can inhibit the expression of polymerase I and polymerase III transcripts.

In Fig. 3 we analyzed the involvement of the polyadenylation machinery in inhibition. First we note that dot-blot and RNase protection assays showed a decrease in the amount of poly(A)+ *Renilla* mRNA levels only for those constructs containing a U1 snRNP binding site (see Fig. 5, which is published as supporting information on the PNAS web site). This decrease in mRNA levels correlates with the *Renilla* luciferase activity lending support to a polyadenylation-inhibition-based mechanism. To check whether polyadenylation sequences are important, we replaced the 3′ UTR and cleavage and polyadenylation sequences of the pSV plasmid with sequences containing the histone 3′ end processing signals to produce the pHist plasmid (Fig. 3). Most histone mRNAs lack a poly(A) tail and form their 3′ ends by an entirely different reaction mechanism that requires a histone stem loop sequence and a binding site for U7 snRNP (26). Transfection of the pHist plasmid resulted in efficient *Renilla* expression, and as expected, the mRNA produced lacked a poly(A) tail as measured by dot-blot analysis of poly(A)+ and poly(A)− fractions (see Fig. 5). Insertion of one or two U1 snRNA binding sites into pHist did not affect *Renilla* expression (Fig. 3). Inhibitory activity was restored, however, when the histone processing signals were replaced by an artificially strong, synthetic cleavage and polyadenylation sequence (SPA; ref. 27) as found in the pSPA and pSPA/20 plasmids. We excluded the possibility that the histone stem loop structure itself was masking the inhibitory activity of the U1 snRNA binding site by observing inhibition in the pHistMT/SPA/20 plasmids where the U7 binding site was inactivated by deletion. This result is all the more remarkable because it strongly supports the idea that U1 snRNA specifically inhibits poly(A) sites, but not having a significant effect on mRNA export, stability, or translatability because it is present in the 3′ UTRs of the normally expressed, mature

mRNAs transcribed from the pHist/20WT and pHist/(20WT)₂ plasmids. We also determined that U1 snRNA binding sites inserted into the 5s rRNA transcript (Pol 1) or the U6 snRNA transcript (Pol 3) had no effect on their relative expression (data not shown). Thus, we conclude that U1 snRNA specifically inhibits AAUAAA type of polyadenylation and will not be useful for inhibition of other types of RNA transcripts.

Inhibition of Expression of Endogenous Genes. As diagrammed in Fig. 4A, we targeted the mRNA of endogenous genes with 5′ end-mutated U1 snRNA transgenes designed to recognize a 10-nt-long sequence in the 3′ terminal exon. Even if the binding site sequence is not long enough to be completely unique, the inhibition only works in 3′ terminal exons and where the RNA is unstructured. Because inhibition levels are limited by transfection efficiencies (a 90% transfection rate can at most give a 10-fold inhibition), we used the 293F human fibroblast cell line where 95–98% of transfection efficiencies are routinely obtained as confirmed by fluorescence-activated cell sorter analysis of a GFP reporter plasmid. Note that we transfected different amounts of the U1 snRNA transgene because increased amounts of 5′ end-mutated U1 snRNA results in increased inhibition. Transfection of 14 μg of each anti-ASA U1 snRNA plasmid individually resulted in 7- to 8-fold levels of inhibition of the ASA protein as measured by Western blotting (Fig. 4B, lanes 4 and 5). Note that transfection of 7 μg of each anti-ASA U1 snRNA plasmid individually gave 3-fold inhibition (data not shown). In contrast, cotransfection of 4 or 7 μg of each transgene together resulted in synergistic inhibition levels of 15- or 24-fold, respectively (Fig. 4B, lanes 6 and 7), far better than inhibition levels seen with the corresponding micrograms of each transgene alone. Use of >14 μg of plasmid did not further increase the inhibition, and no inhibition was observed when control U1 snRNA transgenes were expressed either alone (Fig. 4B, lanes 8 and 9) or together (data not shown). Inhibition was specific, as reprobing the same Western blot showed no changes in the level of either GAPDH or α-tubulin, and long exposures of the Western blot failed to detect truncated ASA protein, as can happen with antisense oligodeoxynucleotides (data not shown) (28). It was also possible that the anti-ASA U1 snRNAs were inhibiting expression by inducing aberrant splicing. Indeed, Niwa *et al.* (29) observed loss of polyadenylation and exon skipping when U1 snRNA sites were placed in the terminal exon (see *Discussion* for more details). To look for aberrant splicing we performed a battery of RT-PCR and length measurement poly(A) tail assays and found none (see Fig. 6, which is published as supporting information on the PNAS web site). Because aberrant splicing could remove the primer binding sites used for these PCR-based assays, we cannot formally rule out the possibility that U1 snRNA binding sites activate aberrant splicing.

We next determined whether stable expression of 5′ end-mutated U1 snRNA would lead to stable reduction in two genes (Fig. 4A), osteocalcin (OC), and collagen 1a1 (COL), found in Ros 17/2.8 cells (rat bone). We consider this a rigorous test of the method because these genes are highly expressed and any potential deleterious effects of the 5′ end-mutated U1 snRNA would become apparent, as the cells are stably expressing the 5′ end-mutated U1 snRNPs. Such toxicity had been observed previously by another group, though in that case the stably expressed, 5′ end-mutated U1 snRNAs were participating in splicing, not polyadenylation inhibition (30). As shown in the Northern and Western blots of Fig. 4C, stable expression of anti-OC/U1 snRNA (targeting the OC pre-mRNA) resulted in a specific 15-fold decrease in OC mRNA levels and a corresponding 15-fold decrease in the OC polypeptide with no changes in the levels of control mRNAs or polypeptides. Likewise, the stable expression of anti-COL/U1 snRNA (targeting the COL pre-mRNA) resulted in a specific 30-fold decrease in
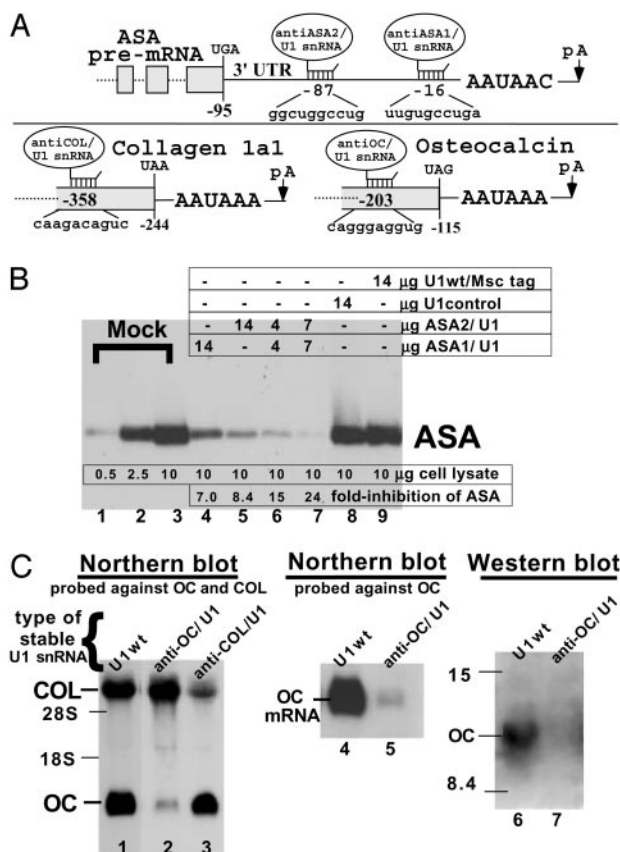
Fig. 4. U1 snRNA transgenes target inhibition of endogenous genes. (*A*) The schematic is of the 3′ end of the pre-mRNA for three target genes: human ASA, rat osteocalcin, and rat collagen 1a1 with the symbols as described in Fig. 1*A*. Shown are the binding sites for two different anti-ASA U1 snRNA transgenes (designated antiASA1/U1 and antiASA2/U1) each directed to bind 10 nt in different parts of the 3′ UTR of the ASA pre-mRNA. Also shown are the binding sites for the anti-osteocalcin U1 snRNA transgene (anti-OC/U1) and the anti-collagen 1a1 U1 snRNA transgene (anti-COL/U1). The numbers are the positions relative to the polyadenylation signal. (*B*) Synergistic inhibition of ASA expression by two anti-ASA-specific U1 snRNA transgenes. Western blot is of whole-cell lysates from cells transiently transfected with the indicated U1 snRNA transgenes or mock-transfected cells (lanes 1–3). Plasmid abbreviations are as in Fig. 4*A*. Lanes 3–9 each contain 10 μg of transfected cell lysate where either one U1 snRNA transgene (lanes 4, 5, 8, and 9) or two different U1 snRNA transgenes (lanes 6 and 7) were transfected as indicated. Lanes 1–3 each contain 0.5, 2.5, and 10 μg of lysate, respectively, from mock-transfected cells to gauge the sensitivity of the ECL-based Western blot. The Western blot was probed with an anti-ASA antibody and then stripped and reprobed for GAPDH or alpha-tubulin as a control (data not shown). Indicated in lanes 4–7 are the "fold inhibition of ASA" (from three independent transfections) as compared with lane 3, where each ASA signal was first normalized to the levels of GAPDH. Lanes 8 and 9 had no inhibition of ASA as compared with lane 3. The U1 control plasmid in lane 8 is a 5′ end-mutated U1 snRNA designed to recognize an internal region of the ASA gene, whereas the "U1wt/Msc tag" plasmid in lane 9 is an additional control that expresses U1 snRNA with a WT 5′ end. All anti-ASAU1 snRNA plasmids have a mutation in stem loop 3 allowing us to monitor their level of expression by reverse transcription. (C) Stable expression of single U1 snRNA transgenes specifically inhibits expression of osteocalcin and collagen 1a1 in Ros 17/2.8 cells, a rat osteosarcoma osteoblast cell line. Three neomycin-resistant stable cell lines were produced from Ros 17/2.8 cells that stably express different U1 snRNA genes. Anti-COL/U1 snRNA is designed to target the rat collagen 1a1 gene, anti-OC/U1 snRNA is designed to target the rat osteocalcin gene, and U1wt is a control that expresses WT U1 snRNA. Lanes 1–3 and 4 and 5 are two separate Northern blots of total RNA taken from each stable cell line as indicated and were probed to detect both the osteocalcin and collagen 1a1 mRNAs (lanes 1–3) or just the osteocalcin mRNA (lanes 4 and 5). The arrows indicate the positions of the mRNAs and rRNAs. Lanes 6 and 7 are a Western blot probed with an anti-osteocalcin antibody where each lane contains 20 μg of total cellular extract from stable cells expressing tagged U1wt snRNA (lane 6) or tagged anti-OC/U1 snRNA (lane 7).

COL mRNA (Fig. 4*C*) and a corresponding 10-fold decrease in the COL polypeptide (data not shown). The fact that no aberrant signals were visible in the Northern blots suggests that aberrant splicing or activation of alternative poly(A) sites was either not occuring or occuring at a relatively low level. No changes in the expression of either OC or COL were observed in the control stable cell line (U1wt), and all three stable cell lines exhibited normal growth phenotypes. Thus, the stable expression of 5′ end-mutated U1 snRNA leads to permanent, inheritable reduction in the expression of target endogenous genes. One of our future goals is to stably express multiple U1 snRNAs targeted to a single endogenous gene, which will require usage of two different selection markers and two different tagged U1 snRNAs so as to simultaneously monitor expression of the different 5′ end-mutated U1 snRNAs.

Discussion

A protocol along with information on how to obtain the plasmids to use the U1 snRNA method can be found at www.rci. rutgers.edu/~gundersn. Essentially, a plasmid harboring the 5′ end-mutated U1 snRNA gene is constructed through a routine cloning step and used for either transient expression or production of stable cell lines. Although the data show that this technique is gene-specific, the overall specificity of this method and of other gene-inhibitory techniques remains an unresolved issue that will best be addressed with microarray-based approaches. Ideally, we could tune the inhibition of a given gene by varying the number and RNA binding capacities of the 5′ end-mutated U1 snRNAs or by developing an inducible U1 snRNA promoter. Note that inhibition of endogenous genes by transient transfection is less efficient than with stable cell lines because of (*i*) limits on transfection efficiency (e.g., a 90% transfection efficiency can at best lead to a 10-fold inhibition) and (*ii*) the time required for the transiently transfected U1 snRNA gene to be transcribed, assembled into U1 snRNP, and imported into the nucleus.

U1 snRNA Inhibition and the Exon Definition Model. It was surprising that U1 snRNA inhibited gene expression even when placed far upstream (up to 1,190 nt) of the AAUAAA poly(A) signal. This inhibition could reflect a direct, but "long-distance," interaction between 5′ end-mutated U1 snRNP and the poly(A) signals. Alternatively, this long-distance inhibition could result from disruption of the definition of the terminal exon as described in the exon definition model (23, 29). This model as applied in higher eukaryotes, proposes that exons, as opposed to much longer introns, are recognized by the splicing machinery through interactions bridging from the 5′ss back to the upstream 3′ss of the same exon (23). This model holds true for internal exons, but for the first and last exons it has to be modified in that the first exon is defined by interactions between the 5′ CAP and the first 5′ss, while the last, terminal exon is defined by interactions between the last 3′ss and the downstream poly(A) signal. It is with the definition of this terminal exon that we focus on. Experimental data supporting terminal exon definition includes the observation that mutations in the poly(A) signal (AAUAAA) reduce splicing of the adjacent terminal intron but not other internal upstream introns, and likewise, mutations in the terminal 3′ss reduce the activity of the poly(A) signal (reviewed in ref. 23). We propose that assembly of an artificial 5′ end mutated U1 snRNA/U1 snRNP bound anywhere within the terminal exon disrupts the terminal 3′ss/poly(A) signal interactions. Inhibition of terminal exon definition would lead to an incompletely processed, immature pre-mRNA that is eventually degraded. This explanation, as opposed to one invoking a simple direct inhibition of the poly(A) signal, predicts that distance from the binding site of 5′ end-mutated U1 snRNA to

the poly(A) signal will have at most a minor role in inhibition, a prediction that matches our observations in Fig. 2.

Do Mutant U1 snRNAs Induce Aberrant Splicing? It is possible that the 5′ end-mutated U1 snRNAs inhibit expression by inducing aberrant splicing. However, when we target endogenous genes, we chose the mutant U1 snRNA binding sites so that they cannot splice, or at least would not be predicted to splice. In all cases shown in Fig. 4, the U1 snRNA binding sites would be "bad" for splicing because (*i*) they lack the canonical GU at the exon–intron boundary and (*ii*) they lack a good binding site for U6 snRNA and U5 snRNA. Of course it is possible that the binding could shift up- or downstream by a few bases and activate aberrant splicing. Thus, we cannot rule out by design only that a mutant U1 snRNA activates aberrant splicing. Indeed, there is precedent for this, as Niwa *et al.* (29) observed loss of polyadenylation and exon skipping when U1 snRNA sites were placed in the terminal exon. Although we performed extensive RT-PCR and length measurement poly(A) tail analysis (see Fig. 6) and failed to detect aberrant splicing, we cannot rule out that it occurs. We observe inhibition even when the the U1 snRNP binds >1,000 nt away from the 3′ss (Figs. 1 and 2), whereas Niwa *et al.* (29) reported that inhibition and exon skipping were lost when the last 3′ss and the downstream U1 snRNA site were >300 nt apart. So why is it that we do not observe a 300-nt-long distance constraint in the data presented here? We note that Niwa *et al.* produced the different spacing mutants by inserting plasmid sequences into their terminal exon near the U1 binding site. Thus, they cannot rule out that their results reflect changes in the secondary structure of the pre-mRNA that mask the U1 binding site.

Concluding Remarks. Although we have not examined this experimentally, the U1 snRNA-based approach compares favorably with antisense-based inhibition (20) or RNAi (21, 22, 31, 32). The 700-fold inhibition is the highest level yet published for a reporter gene. For endogenous genes, the 15-fold inhibition with one stably expressed U1 snRNA or the 25-fold inhibition with two transiently expressed U1 snRNA genes, are also respectable. Although this method targets only the 3′ terminal exon, that limitation has the benefit that restoration of the protein or mutant versions thereof is relatively straightforward with a transgene with a different 3′ UTR, which allows for control of the phenotypes observed with the inhibition. Finally, we note that stably expressed, 5′ end-mutated U1 snRNA can target a reporter gene while not inhibiting a nearly identical reporter gene differing by a single nucleotide (3), thus allowing targeted inhibition of a dominant negative gene while not targeting the homologous functional copy of the gene. In conclusion, reprogrammed U1 snRNA transgenes are potent and specific inhibitors of gene expression in mammalian cells that provide researchers with a powerful tool to study mammalian gene function.

1. Carpousis, A. J., Vanzo, N. F. & Raynal, L. C. (1999) *Trends Genet.* **15,** 24–28.
2. Mitchell, P. & Tollervey, D. (2000) *Curr. Opin. Genet. Dev.* **10,** 193–198.
3. Macdonald, P. (2001) *Curr. Opin. Cell Biol.* **13,** 326–331.
4. Gingras, A. C., Raught, B. & Sonenberg, N. (1999) *Annu. Rev. Biochem.* **68,** 913–963.
5. Huang, Y. & Carmichael, G. C. (1996) *Mol. Cell. Biol.* **16,** 1534–1542.
6. Keller, W. & Minvielle-Sebastia, L. (1997) *Curr. Opin. Cell Biol.* **9,** 329–336.
7. Zhao, J., Hyman, L. & Moore, C. (1999) *Microbiol. Mol. Biol. Rev.* **63,** 405–445.
8. Edwalds-Gilbert, G., Veraldi, K. L. & Milcarek, C. (1997) *Nucleic Acids Res.* **25,** 2547–2561.
9. Boelens, W. C., Jansen, E. J., van Venrooij, W. J., Stripecke, R., Mattaj, I. W. & Gunderson, S. I. (1993) *Cell* **72,** 881–892.
10. Gunderson, S. I., Beyer, K., Martin, G., Keller, W., Boelens, W. C. & Mattaj, L. W. (1994) *Cell* **76,** 531–541.
11. Furth, P. A., Choe, W. T., Rex, J. H., Byrne, J. C. & Baker, C. C. (1994) *Mol. Cell. Biol.* **14,** 5278–5289.
12. Beckley, S. A., Liu, P., Stover, M. L., Gunderson, S. I., Lichtler, A. C. & Rowe, D. W. (2001) *Mol. Cell. Biol.* **21,** 2815–2825.
13. Liu P., Gucwa, A., Stover, M. L., Buck, E., Lichtler, A. & Rowe, D. (2002) *Nucleic Acids Res.* **30,** 2329–2339.
14. Gunderson, S. I., Polycarpou-Schwarz, M. & Mattaj, I. W. (1998) *Mol. Cell* **1,** 255–264.
15. Ashe, M. P., Furger, A. & Proudfoot, N. J. (2000) *RNA* **6,** 170–177.
16. Will, C. L. & Luhrmann, R. (1997) *Curr. Opin. Cell Biol.* **9,** 320–328.
17. Burge, C. B., Tuschl, T. & Sharp, P. A. (1999) in *The RNA World* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 525–560.
18. Zhuang, Y. & Weiner, A. M. (1986) *Cell* **46,** 827–835.
19. Cohen, J. B., Snow, J. E., Spencer, S. D. & Levinson, A. D. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10470–10474.
20. Probst, J. C. (2000) *Methods* **22,** 271–281.
21. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T. (2001) *Nature* **411,** 494–498.
22. Brummelkamp, T. R., Bernards, R. & Agami, R. (2002) *Science* **296,** 550–553.
23. Berget, S. M. (1995) *J. Biol. Chem.* **270,** 2411–2414.
24. Li, B., Vilardell, J. & Warner, J. R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 1596–1600.
25. Krug, R. M., Alonso-Caplen, F. V., Julkunen, I. & Katze, M. G. (1989) in *The Influenza Viruses* (Plenum, New York), pp. 89–152.
26. Dominski, Z. & Marzluff, W. F. (1999) *Gene* **239,** 1–14.
27. Levitt, N., Briggs, D., Gil, A. & Proudfoot, N. J. (1989) *Genes Dev.* **3,** 1019–1025.
28. Thoma, C., Hasselblatt, P., Kock, J., Chang, S. F., Hockenjos, B., Will, H., Hentze, M. W., Blum, H. E., von Weizsacker, F. & Offensperger, W. B. (2001) *Mol. Cell* **8,** 865–872.
29. Niwa, M., MacDonald, C. C. & Berget, S. M. (1992) *Nature* **360,** 277–280.
30. Cohen, J. B., Broz, S. D. & Levinson, A. D. (1993) *Mol. Cell. Biol.* **13,** 2666–2676.
31. Tusch, T. (2002) *Nat. Biotechnol.* **20,** 446–448.
32. Nishikura, K. (2001) *Cell* **107,** 415–418.

CELL BIOLOGY