cyanin pathway. The transferred *C1* may only be functional in *Arabidopsis* when it interacts with the transferred *R*.

The observed effect of *R* on both trichomes and anthocyanin suggests that the *ttg* locus in *Arabidopsis* encodes an *R* homologue. Alternatively, the *ttg* protein may activate an *R* homologue. The activator *R* is a useful visual marker of transformation in maize (*20*). Both *R* and *C1* may provide visual genetic markers in *Arabidopsis* and other flowering plants by controlling production of both pigment and trichomes.

REFERENCES AND NOTES

1. J. W. McClure, in *The Flavonoids*, J. B. Harborne, T. J. Mabry, H. Mabry, Eds. (Academic Press, New York, 1975), pp. 970–1055; M. C. Mehdy and C. J. Lamb, *EMBO J.* **6**, 1527 (1987); T. B. Ryder, *Mol. Gen. Genet.* **210**, 219 (1987); H. Habereder, G. Schröder, J. Ebel, *Planta* **177**, 58 (1989); A. M. Lloyd and R. W. Davis, unpublished data.
2. W. Kubasek and F. Ausubel, unpublished data.
3. R. L. Feinbaum and F. M. Ausubel, *Mol. Cell. Biol.* **8**, 1985 (1988).
4. E. H. Coe, Jr., M. G. Neuffer, D. A. Hoisington, in *Corn and Corn Improvement*, G. F. Sprague and J. W. Dudley, Eds. (American Society of Agronomy, Madison, WI, 1988), pp. 81–258; H. G. Dooner, T. P. Robbins, R. A. Jorgensen, *Annu. Rev. Genet.* **25**, 173 (1991).
5. G. Mendel, *Verh. Naturforsch. Ver. Brünn.* **4**, 3 (1865); W. Bateson, *J. R. Hortic. Soc.* **26** (no. 1), 1 (1901).
6. M. Koornneef, *Arabidopsis Inf. Serv.* **27**, 1 (1990).
7. ———, *ibid.* **18**, 45 (1981).
8. B. A. Roth, S. A. Goff, T. M. Klein, M. E. Fromm, *Plant Cell* **3**, 317 (1991); S. A. Goff, K. C. Cone, V. L. Chandler, *Genes Dev.* **6**, 864 (1992).
9. J. Paz-Ares, D. Ghosal, U. Wienand, P. A. Peterson, H. Saedler, *EMBO J.* **6**, 3553 (1987).
10. S. R. Ludwig, L. F. Habera, S. L. Dellaporta, S. R. Wessler, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7092 (1989).
11. Plasmid pKYLX71 was constructed by replacement of the polylinker sequence between the Hind III and Xba I sites in the CaMV 35*S* promoter–rbcs polyadenylation signal expression cassette of pKYLX7 with the following sequence: 5′-AAGCTTGGATCCTCGAGCTGCAGGAGCTCGAATTGATCCTCTAGA-3′ [C. L. Schardl *et al.*, *Gene* **61**, 1 (1987); A. G. Hunt, unpublished data]. This polylinker contains the following restriction sites: Hind III, Bam HI, Xho I, Pst I, Sac I, and Xba I, of which Hind III, Xho I, Sac I, and Xba I are unique in this vector. A 2.1-kb Eco RI fragment containing the *C1* cDNA cLC28 (*9*) was cloned into Eco RI–cut pBSKS+ (Stratagene) to create pAL70. The Hind III to Sac I fragment from pAL70 containing the *C1* cDNA was cloned into pKYLX71 that had been digested with Hind III and Sac I to create pAL71. Two versions of the *R* expression vector were made. A 2.4-kb Xba I fragment containing the entire cloned 5′ untranslated cDNA leader including three upstream AUG codons and the entire open reading frame (*10*) was cloned into Xba I–cut pKYLX71 to create pAL69. There is an internal Xba I site blocked by *dam* methylation in this fragment that is downstream of all three untranslated AUG codons but upstream of the AUG start codon. The plasmid pAL69 was transferred to a *dam–Escherichia coli* strain. The 2.2-kb Xba I fragment from pAL69 was cloned into Xba I–cut pKYLX71 to create pAL144. For plant transformations, the binary Ti vectors were placed in the *Agrobacterium* strain GV3101 that contained the helper plasmid pMP90 [C. Koncz, J. Schell, *Mol. Gen. Genet.* **204**, 383 (1986)]. In the R. W. Davis lab collection, pAL69 = pNN480, pAL70 = pNN481, pAL71 = pNN482, and pAL144 = pNN483.
12. P. N. Benfey, L. Ren, N.-H. Chua, *EMBO J.* **8**, 2195 (1989).
13. R. B. Horsch *et al.*, *Science* **227**, 1229 (1985).
14. D. Valvekens, M. Van Montagu, M. Van Lijsebettens, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5536 (1988).
15. The AUGs in the 5′ leader of the *R(Lc)* mRNA are part of an upstream open reading frame that represses *R(Lc)* expression about 30-fold in transient transfection assays in maize. R. D. Damiani, Jr., and S. R. Wessler, unpublished data.
16. K. H. Wolfe, M. Gouy, Y.-W. Yang, P. M. Sharp, W.-H. Li, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6201 (1989).
17. D. G. Oppenheimer, P. L. Herman, S. Sivakumaran, J. Esch, M. D. Marks, *Cell* **67**, 483 (1991).
18. J. Goodrich, R. Carpenter, E. S. Coen, *ibid.* **68**, 955 (1992).
19. B. W. Shirley, S. Hanley, H. M. Goodman, *Plant Cell* **4**, 333 (1992).
20. S. R. Ludwig, B. Bowen, L. Beach, S. R. Wessler, *Science* **247**, 449 (1990).
21. H. Drumm-Herrel and H. Mohr, *Photochem. Photobiol.* **33**, 391 (1981). Seedlings of *Arabidopsis* were grown at 22°C in Sunshine Mix 1 potting soil (Fisons Instruments) under constant illumination by Gro-Lux lamps (Sylvania) at 85 $\mu Es^{-1}\ m^{-2}$ The five seedlings used in Table 2 were grown for 7 weeks. Approximately equal amounts of rosette leaf tissue from each seedling were pooled and 0.5 g was extracted with 2 ml of acidic methanol. The extracts were diluted 100-fold before measurement of the absorption spectra.

22. We thank F. Ausubel, M. Campbell, E. Dement, M.-I. Benito, W. Kubasek, K. Luehrsen, B. Osborne, D. Preuss, M. Schena, and M. Thomas for helpful discussion; A. Hunt, C. Koncz, M. Koornneef, C. Schardl, S. Wessler, and U. Wienand for donations of plasmids and strains; and M. Campbell, G. Hwang, A. B. Lloyd, and G. Yen for reading the manuscript. Supported by NIH grants GM 32422 (V.W.) and R37-H600198 (R.W.D.) and NSF grant DMB-910601 (R.W.D.).

21 July 1992; accepted 8 October 1992

Interaction of U6 snRNA with a Sequence Required for Function of the Nematode SL RNA in Trans-Splicing

Gregory J. Hannon,* Patricia A. Maroney, Yi-Tao Yu, Gretchen E. Hannon,* Timothy W. Nilsen†

Nematode trans-spliced leader (SL) RNAs are composed of two domains, an exon [the 22-nucleotide spliced leader] and a small nuclear RNA (snRNA)–like sequence. Participation in vitro of the spliced leader RNA in trans-splicing reactions is independent of the exon sequence or size and instead depends on features contained in the snRNA-like domain of the molecule. Chemical modification interference analysis has revealed that two short sequence elements in the snRNA-like domain are necessary for SL RNA activity. These elements are sufficient for such activity because when added to a 72-nucleotide fragment of a nematode U1 snRNA, this hybrid RNA could participate in trans-splicing reactions in vitro. One of the critical sequence elements may function by base-pairing with U6 snRNA, an essential U snRNA for both cis- and trans-splicing.

In nematodes, a subset of mRNAs contains a common 5′ terminal 22-nucleotide SL sequence. The SL sequence is acquired from a small (~100 nucleotides) SL RNA by trans-splicing. The SL RNA represents an unusual type of U snRNA that contains an exon, the 22-nucleotide SL, and an snRNA-like domain (*1*). Availability of ahomologous trans-splicing system in which synthetic SL RNA can serve as the trans-splice donor in vitro has facilitated a mutational analysis to identify sequences in the SL RNA that are important for its function (*2*). Such an analysis has shown that SL RNA participation in trans-splicing in vitro depends on assembly into an Sm small nuclear ribonucleoprotein (snRNP) and does not require specific exon sequences (*2*, *3*). These observations indicate that determinants of SL RNA function reside in the snRNA-like domain of the molecule.

We used chemical modification interference analysis as a probe to identify functionally important sequence elements in the snRNA-like domain of SL RNA. SL RNA that was 3′ end-labeled was treated with diethylpyrocarbonate (which modifies purines) under conditions in which less than one modification per molecule is introduced. The population of the SL RNAs carrying single modifications at random sites served as the substrate for trans-splicing (Fig. 1B). Because the SL RNA was 3′ end-labeled, only the Y-intron + exon intermediate (produced by the first step of trans-splicing) and the Y-intron product (produced by the second step of trans-splicing) were visualized by autoradiography of gel-fractionated trans-splicing reactions (Fig. 1, A and B). After recovery from preparative gels, the two Y-branched species, as well as unreacted SL RNA, were

Department of Molecular Biology and Microbiology, Case Western Reserve University School of Medicine, Cleveland, OH 44106.

*Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.
†To whom correspondence should be addressed.

treated with aniline to induce strand breakage at sites of modification (Fig. 1C). Comparison of the cleavage pattern of unreacted SL RNA with that of the Y intron + exon intermediate revealed that modification of any of nine purines prevented SL RNA participation in the first step of trans-splicing (Fig. 1C); no additional important purines were identified in the same analysis of the Y-intron product (Fig. 1C). Thus, no specific purines were uniquely required for the second step of trans-splicing. The purine residues that were shown by modification interference to be essential for trans-splicing were confined to two regions of the SL RNA: a short stretch on the 3′ side of stem II and the single-stranded region between stems II and III (Fig. 1D). This single-stranded region contains the Sm-binding sequence (AAUUUUGG) flanked by three upstream nucleotides (AUA) and three downstream nucleotides (AAC). The Sm-binding sequence in U snRNAs directs assembly of the RNAs into RNP particles that contain proteins with epitopes recognized by antibodies with Sm specificity (*4*). Modification of any of the four purines in the Sm-binding sequence or either of the two adjacent downstream adenosines prevented use of the SL RNA in trans-splicing.

A functional Sm-binding sequence is necessary in vitro for SL RNA participation in trans-splicing (*2*). For this reason, it seemed likely that Sm-binding site purines were identified by the modification interference analysis because of an effect on SL RNP assembly. However, when we assessed the effect of modification on SL RNP assembly (as measured by mobility shift on native gels) (*3*), no individual purine modification was sufficient to inhibit SL RNP assembly (*5*). Thus, the Sm-binding site purines apparently contribute to SL RNA function in a manner distinct from their role in directing RNP formation.

In contrast to the divergence found in the primary sequence of the remainder of the snRNA-like domain (the *Ascaris* SL RNA and *Caenorhabditis elegans* SL1 RNA are only ~40% identical in this domain), the elements identified by modification interference have been conserved among evolutionarily distant nematodes. The identical Sm-binding sequence (AAUUUUGG) is present in all nematode SL RNAs characterized to date. The sequence immediately 3′ of the Sm site (AAC) is found in the *C. elegans* SL RNAs (SL1 and SL2) and in the SL RNAs of *Brugia malayi*, *Angiostrongylus cantonensis*, and *Oncocera volvulus*; the sequence AAA is found in the analogous position in the SL RNA of *Dirofilaria imitis* (*6*). The element on the 3′ side of stem II has also been conserved both in primary sequence and its position relative to the Sm-binding site. The sequence GUGGC (which contains guanosines 61, 63, and 64 identified by modification interference) is found nine nucleotides upstream of the first adenosine of the Sm-binding site in *Ascaris*, *B. malayi*, *O. volvulus*, and *D. imitis*. A similar sequence UUGGC is found ten nucleotides upstream of the *C. elegans* SL1 Sm-binding site. The *C. elegans* SL1 RNA and *Ascaris* SL RNA function equally in trans-splicing reactions with *Ascaris* extracts (*7*), indicating that a guanosine per se at base 61 is not required for SL RNA function.

The chemical modification interference experiment identified elements necessary for function of the SL RNA's snRNA-like domain in trans-splicing reactions. If these elements were sufficient to confer function, their introduction into a heterologous RNA in the appropriate structural context would be predicted to create a functional artificial SL RNA. As a recipient RNA for such a substitution experiment, we chose a 72-nucleotide fragment of an *Ascaris* U1 snRNA (bases 91 to 162). This RNA was selected for the following reasons. (i) Computer-generated secondary structure analysis predicted that deletion of the first 90 nucleotides of the *Ascaris* U1 snRNA would result in an RNA that would adopt an

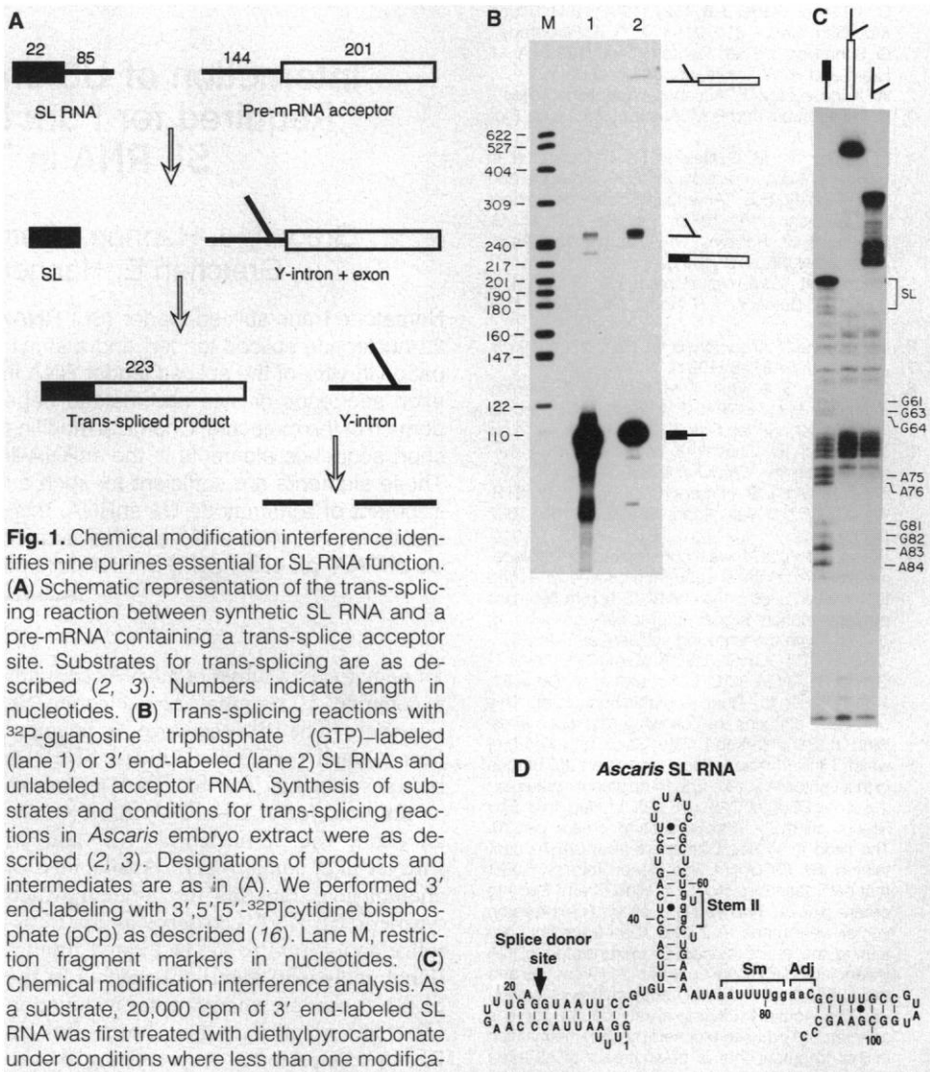

Fig. 1. Chemical modification interference identifies nine purines essential for SL RNA function. (**A**) Schematic representation of the trans-splicing reaction between synthetic SL RNA and a pre-mRNA containing a trans-splice acceptor site. Substrates for trans-splicing are as described (*2*, *3*). Numbers indicate length in nucleotides. (**B**) Trans-splicing reactions with ^{32}P-guanosine triphosphate (GTP)–labeled (lane 1) or 3′ end-labeled (lane 2) SL RNAs and unlabeled acceptor RNA. Synthesis of substrates and conditions for trans-splicing reactions in *Ascaris* embryo extract were as described (*2*, *3*). Designations of products and intermediates are as in (A). We performed 3′ end-labeling with 3′,5′[5′-^{32}P]cytidine bisphosphate (pCp) as described (*16*). Lane M, restriction fragment markers in nucleotides. (**C**) Chemical modification interference analysis. As a substrate, 20,000 cpm of 3′ end-labeled SL RNA was first treated with diethylpyrocarbonate under conditions where less than one modification per molecule was introduced (*17*) and then used for trans-splicing. After gel electrophoresis, Y-intron + exon, and Y-intron species were purified from preparative gels such as the one shown in (B), lane 2, and cleaved with aniline (*17*). Indicated bases are those whose modification prevented SL RNA use in trans-splicing. Bases were identified by comparison to a chemical sequencing ladder of unreacted SL RNA (*5*) and are numbered according to (D). Purines contained within the 22-nucleotide SL (indicated) are not present in the Y-intron + exon, or Y-intron molecules. (**D**) Schematic representation of chemical modification interference results. The sequence and computer-generated secondary structure of the *Ascaris* SL RNA in which purines identified in (C) are shown in lowercase letters and the 3′ side of stem II, the Sm-binding site, and the 3′ adjacent (Adj) nucleotides are indicated by brackets.

overall structure similar to that of the SL RNA's snRNA-like domain (two stem loops separated by a single-stranded region containing the U1 snRNA Sm-binding site). (ii) The U1 snRNA has the sequence UUGGC ten nucleotides upstream of the first nucleotide of its Sm-binding site. (iii) This fragment was predicted to lack the binding sites for U1-specific proteins (*8*). After introduction of an exon and a 5′ splice site (for convenience, in initial constructs we used the 22 nucleotide SL sequence and the accompanying 5′-splice site), the resultant chimeric RNA was predicted to have an arrangement of structural domains resembling those of the authentic SL RNA (Fig. 2A).

This chimeric RNA was synthesized, and it efficiently assembled into an Sm RNP, as assessed by native gel electrophoresis and cap trimethylation assays (*2*, *3*); however, it did not function in trans-splicing (*5*). The chimeric RNA was further modified by replacement of the U1 sequence AAUUUUUGC-GUUU (the Sm-binding site and adjacent four 3′ nucleotides) with the corresponding region of the SL RNA (AAUUUUGG-AACG). This altered chimeric RNA appeared to function in trans-splicing (Fig. 2B, lane 4) because of the appearance of RNAs with the expected mobilities of the trans-spliced product and branched molecules (Y intron and Y intron + exon). To confirm the authenticity of the trans-splicing, we subjected the two putative products (ligated exons and Y-intron) to two-dimensional ribonuclease (RNase) T1 fingerprinting. The 22-nucleotide SL sequence was accurately transferred from the chimeric SL-U1 snRNA to the acceptor molecule, and the branch point adenosines (*2*) used to generate the Y-branched product were identical to those used in trans-splicing with authentic SL RNA (*5*). The artificial SL-U1 RNA was used as a trans-splice donor in spite of an excess of endogenous SL RNA present in the extract (*2*, *3*).

Because the functional chimeric RNA contained a substitution that replaced both the U1 Sm-binding site and the adjacent 3′ nucleotides, we examined whether each of these substitutions was necessary. Therefore, two additional chimeric RNAs were constructed. The first contained the U1 Sm-binding region with the adjacent nucleotides from the SL RNA (AAUUUUUGC-AACG). This RNA did not function in trans-splicing (Fig. 2B, lane 2). Similarly, a

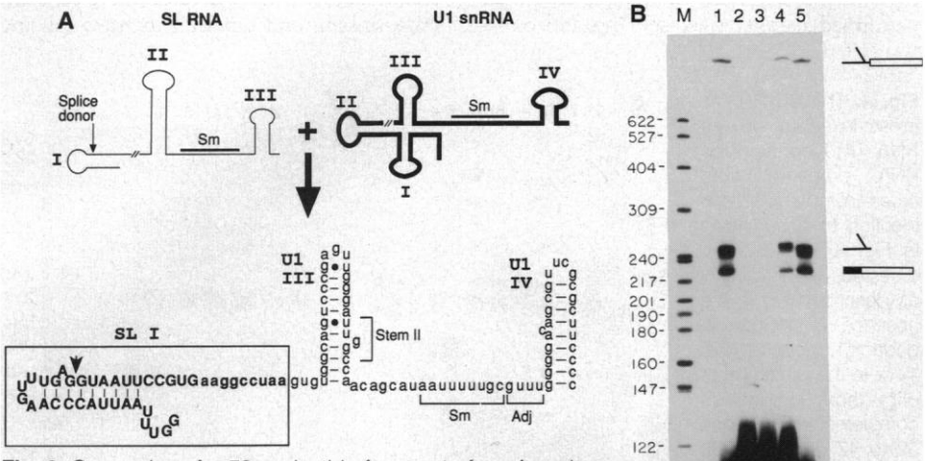

Fig. 2. Conversion of a 72-nucleotide fragment of an *Ascaris* U1 snRNA into a functional SL RNA by site-directed mutagenesis. (**A**) Schematic representation of chimeric synthetic SL RNAs. (Top) Both the SL RNA and U1 snRNA are shown in predicted secondary structures with stem loops numbered. (Bottom) Nucleotide sequences of chimeric SL-U1 SL RNAs. The first 33 nucleotides of the SL RNA (bold uppercase and boxed) were fused to a 72-nucleotide fragment of *Ascaris* U1 snRNA (sequence shown in lowercase). Nucleotides indicated by bold lowercase and boxed are not derived from either the U1 snRNA or the SL RNA and were introduced as a consequence of the cloning strategy we used to generate the chimeric molecule. (**B**) Trans-splicing of chimeric SL-U1 SL RNAs. Lane 1, wild-type SL RNA; lane 2, SL-U1 chimeric RNA with U1 Sm-binding site AAUUUUUGC and SL RNA adjacent bases AACG; lane 3, SL-U1 chimeric RNA containing the SL RNA Sm-binding site AAUUUUGG and U1 3′ adjacent bases GUUU; lane 4, SL-U1 chimeric RNA with SL RNA Sm-binding site and adjacent bases; and lane 5, wild-type SL RNA. We used 150,000 cpm (5 ng) of each RNA in trans-splicing reactions identical to those shown in Fig. 1B. After incubation, reactions were analyzed as described (*3*). Schematic designations of products and intermediates are as in Fig. 1A. Lane M, restriction fragment markers in nucleotides.

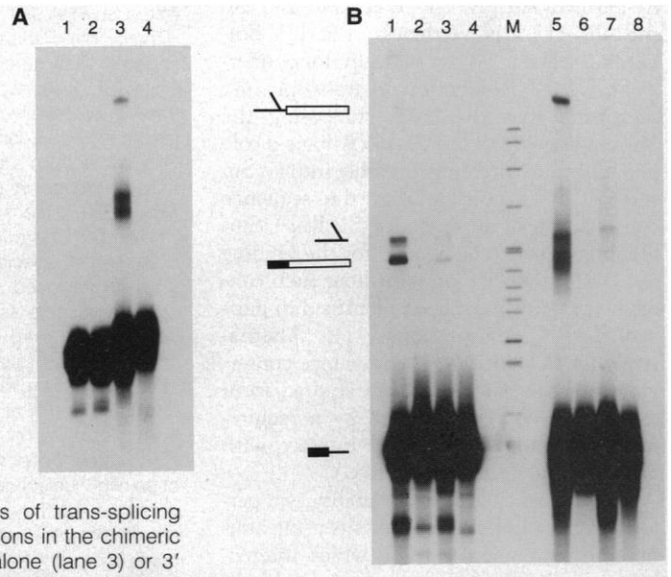

Fig. 3. Function of the SL RNA in trans-splicing correlates with cross-linking efficiency. (**A**) AMT-dependent UV cross-linking of wild-type SL RNA. We incubated 150,000 cpm (5 ng) of [^{32}P]GTP (guanosine triphosphate)–labeled SL RNA for 30 min at 30°C as described for trans-splicing reactions (Fig. 1B) without addition of adenosine triphosphate (ATP) or acceptor pre-mRNA (lanes 1 to 3). The reaction in lane 4 did not contain extract. After incubation, AMT (40 μg/ml) was added to the reactions shown in lanes 2 to 4, and all reactions were incubated an additional 20 min at 4°C in the presence (lanes 1, 3, and 4) or absence (lane 2) of irradiation with 365-nm light as described (*12*). After incubations, reactions were analyzed as described in Fig. 2B. (**B**) Function of trans-splicing correlates with cross-linking. Lanes 1 and 5, wild-type SL RNA; lanes 2 and 6, SL RNA in which the Sm-binding region (AAUUUUGGAACG) was replaced with the analogous region of U1 snRNA (AAUUUUUGCGUUU); lanes 3 and 7, SL RNA in which the Sm-binding site (AAUUUUGG) was replaced with the U1 snRNA Sm-binding site (AAUUUUUGC); and lanes 4 and 8, SL RNA in which the four nucleotides 3′ of the Sm-binding site (AACG) were replaced with the analogous nucleotides of U1 snRNA (GUUU). Lanes 1 to 4, trans-splicing reactions performed as in Fig. 1B; lanes 5 to 8, cross-linking reactions performed as in (A, lane 3). Lane M, restriction fragment markers identical to those in Fig. 2B (nucleotide sizes not indicated). Schematic designations of trans-splicing intermediates and products are as in Fig. 1A. Unlike similar substitution mutations in the chimeric SL-U1 RNA (Fig. 2B), replacement of either the SL RNA's Sm-binding site alone (lane 3) or 3′ adjacent sequences alone (lane 4) with analogous U1 snRNA sequences does not completely eliminate trans-splicing function. Replacement of the entire region (as it does with the chimeric SL-U1 RNA) abolishes function (lane 2).

chimeric RNA containing the SL RNA Sm-binding site and U1 adjacent nucleotides (AAUUUUGG-UUUG) was not active (Fig. 2B, lane 3). Thus, the primary sequences of both the Sm-binding site itself and the adjacent 3′ nucleotides were required for chimeric SL RNA function. We investigated the possibility that the 22-nucleotide SL sequence itself might be important in conferring SL RNA function to the chimeric molecule, even though exon sequence is irrelevant in the context of the native SL RNA (*3*). Chimeric SL-U1 RNAs were constructed with a variety of artificial exons derived from pBR 322 (*3*), and in each case these RNAs functioned in trans-splicing with an efficiency comparable to that seen when the same exons were affixed to the wild-type SL RNA (*5*).

We have identified two sequence elements in the snRNA-like domain of the nematode SL RNA that are necessary for its participation in trans-splicing and have shown that these elements can confer SL RNA function to a fragment of a nematode U1 snRNA. The first element, the sequence on the 3′ side of stem II, may serve as the recognition sequence for an SL RNP–specific protein or might interact with other trans-spliceosomal components. The presence of a nearly identical sequence element upstream of the U1 snRNA Sm-binding site may be coincidental because this element is not conserved either in the C. *elegans* U1 snRNA or in the U1 snRNAs of higher eukaryotes. The second element includes the Sm-binding site of the SL RNA. In other eukaryotes, the Sm-binding sequences of U snRNAs, although subject to variation, appear to direct the binding of identical core proteins (*9*). Similar variation in Sm-binding sequences is tolerated in nematode U snRNAs (*10*), and a similar set of core proteins appears to be associated with affinity-purified *Ascaris* U1, U2, U4, and SL snRNPs (*11*). Nevertheless, the U1 Sm-binding sequence is not sufficient for chimeric SL-U1 RNA function in trans-splicing, suggesting that the bases composing the Sm-binding site of the SL RNA have a role in addition to directing assembly into an Sm snRNP. It is possible that the sequence conservation of an extended SL RNA Sm-binding sequence is required for the binding of SL RNP–specific proteins [four such candidate proteins have been identified in purified SL RNP preparations (*11*)]. Alternatively (or in addition), the sequence conservation of the Sm-binding site and adjacent sequences could be dictated by a requirement for the SL RNA to interact with another trans-spliceosomal RNA.

To address the latter possibility, we performed cross-linking experiments using aminomethyltrioxsalen (AMT), which intercalates into double-stranded nucleic acid helices and forms covalent adducts after irradiation with long-wavelength (365 nm) ultraviolet (UV) light (*12*). Using labeled SL RNA, we observed three cross-linked species (Fig. 3A, lane 3). The appearance of the cross-linked species was dependent on incubation in extract, the presence of AMT, and irradiation with 365-nm light (Fig. 3A). To determine if the capacity of the SL RNA to form cross-linked species correlated with its ability to function in trans-splicing, we performed parallel trans-splicing and cross-linking assays using either wild-type SL RNA or mutant SL RNAs altered in the Sm-binding site region (Fig. 3B). Although the mutant SL RNAs were efficiently assembled into Sm snRNPs (*5*), they either did not function (Fig. 3B, lane 2) or functioned poorly (Fig. 3B, lanes 3 and 4) as trans-splice donors. These same RNAs cross-linked inefficiently (Fig. 3B, lanes 6 to 8). These results demonstrated a correlation between cross-linking and function in trans-splicing

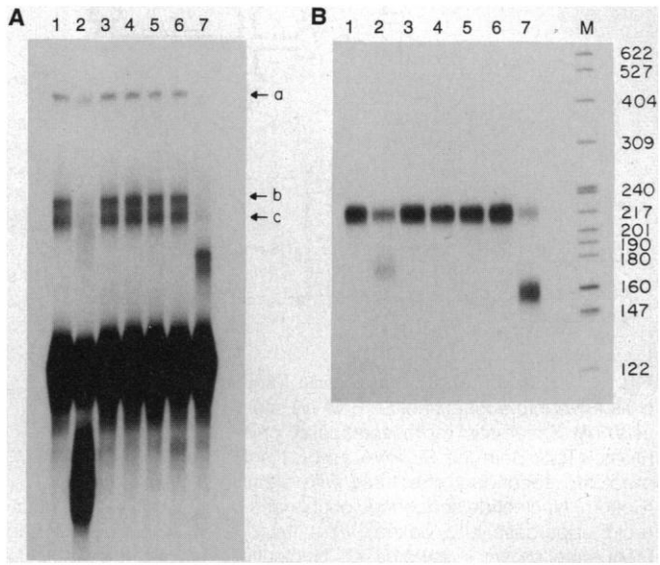

Fig. 4. The SL RNA is cross-linked to U6 snRNA. (**A**) Deproteinized RNA prepared from cross-linking reactions identical to those shown in Fig. 3, lane 3 were digested with RNase H as described (*2*) in the absence of added oligodeoxynucleotide (lane 1) or in the presence of oligodeoxynucleotides complementary to bases 22 to 42 of *Ascaris* SL RNA (lane 2), bases 1 to 13 of *Ascaris* U1 snRNA (lane 3), bases 29 to 45 of *Ascaris* U2 snRNA (lane 4), bases 64 to 83 of *Ascaris* U4 snRNA (lane 5), bases 37 to 61 of *Ascaris* U5 snRNA (lane 6), or bases 37 to 55 of *Ascaris* U6 snRNA (lane 7). Somewhat less efficient but similar digestion to that seen in lane 7 was observed when digestion was carried out with an oligodeoxynucleotide complementary to bases 73 to 90 of U6 snRNA (*5*). After digestion, RNAs were analyzed as described (*2*). The three cross-linked species are designated a, b, and c. Sequences of oligodeoxynucleotides are given in (*18*). (**B**) Cross-linked species c in (A) was excised from preparative gels and digested with the same panel of oligodeoxynucleotides described in (A). Lane designations are identical to those in (A). Lane M, restriction fragment markers in nucleotides.

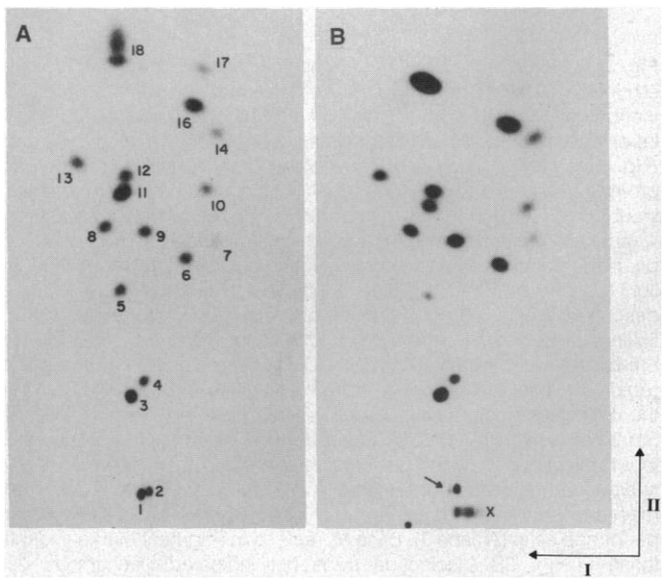

Fig. 5. RNase T1 fingerprint analysis of cross-linked SL RNA. (**A**) Two-dimensional RNase T1 fingerprint of uncross-linked SL RNA excised from a preparative gel similar to Fig. 3, lane 3. Oligonucleotide designations are as in (*19*). (**B**) Two-dimensional RNase T1 fingerprint of cross-linked species c (Fig. 4A) excised from preparative gels as described in Fig. 4B. The small arrow indicates an oligonucleotide [oligonucleotide 1 in (A)] absent from this fingerprint. The X indicates the position of an oligonucleotide not present in the fingerprint shown in (A). Fingerprints of cross-linked species a and b (Fig. 4A) appear identical to the fingerprint of cross-linked species c (*5*). Fingerprinting was as described (*19*) and arrows I and II indicate the first (gel electrophoresis) and second (homochromatography) dimensions, respectively.

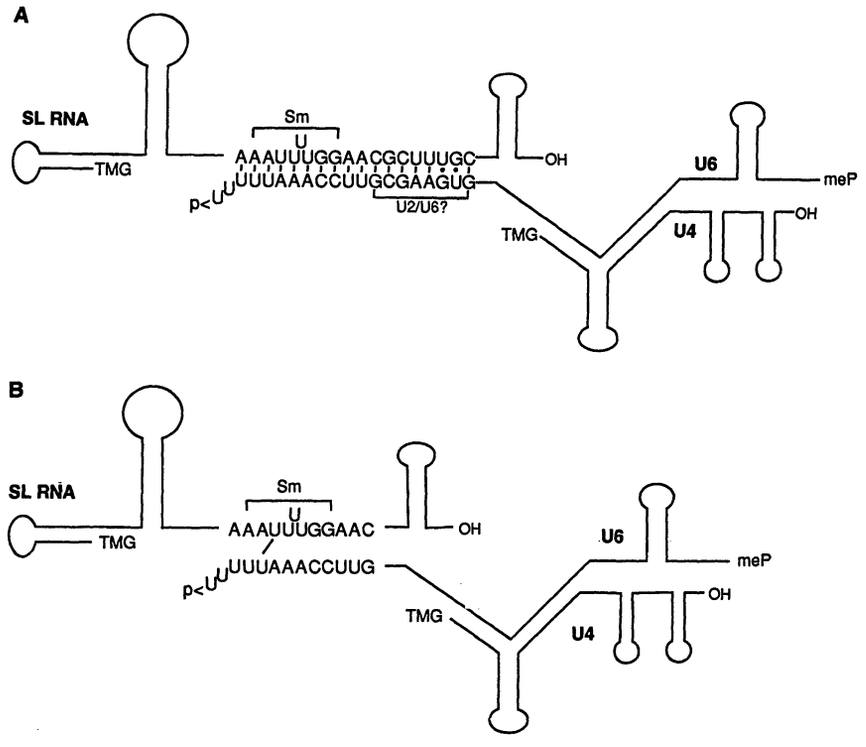

Fig. 6. A potential base-pairing interaction between the SL RNA and U6 snRNA. (**A**) The sequences of a possible site of base pairing between the *Ascaris* SL RNA and U6 snRNA are shown. The potential interaction is drawn to maximize base pairing, and the specific bulged uridine in the SL RNA's Sm-binding site was arbitrarily chosen. The region of U6 snRNA known to interact with U2 snRNA (*12, 14*) in other eukaryotes is indicated. (**B**) The same interaction shown in (A), indicating the position of an AMT cross-link between the SL RNA and U6 snRNA (*13*). The presumptive cap structures of the snRNAs are indicated to provide orientation: TMG, trimethylguanosine (U4 snRNA and SL RNA); meP, gamma monomethyl phosphate (U6 snRNA).

and suggested that the Sm-binding region of the SL RNA could be interacting with another RNA by base pairing.

To determine if AMT cross-linking had identified an interaction between the SL RNA and a characterized U snRNA, we digested the cross-linked species with RNase H in the presence of oligodeoxynucleotides complementary to SL, U1, U2, U4, U5, or U6 snRNAs (Fig. 4). As expected, the cross-linked species were digested with RNase H in the presence of an oligodeoxynucleotide complementary to SL RNA (Fig. 4, A and B, lane 2). The cross-linked species were not digested in the presence of oligodeoxynucleotides complementary to U1, U2, U4, or U5 snRNAs (Fig. 4, A and B, lanes 3 to 6) but were digested in the presence of either of two oligodeoxynucleotides complementary to U6 snRNA [(Fig. 4, A and B, lane 7) and (*5*)]. These results indicated that synthetic SL RNA could base pair with U6 snRNA present in the extract. The same interaction occurs between endogenous SL RNA and U6 snRNA because cross-linked species of identical mobility to those seen with synthetic SL RNA are detected by Northern blot analysis of RNAs prepared from AMT-UV–treated extracts (*5*).

To determine the site of cross-linking with respect to the SL RNA, we performed two-dimensional RNase T1 fingerprint analysis (Fig. 5). Comparison of fingerprints generated from uncross-linked (Fig. 5A) or cross-linked SL RNA (Fig. 5B) revealed a single difference. In the fingerprint of the cross-linked species, oligonucleotide 1 (OH-AAAAUAAAUUUUG-P) spanning the SL RNA's Sm-binding site was absent and was replaced by a heterogeneous slowly migrating species (X in Fig. 5B). Photoreversal of the cross-linked species with shortwave (254 nm) UV light showed that oligonucleotide X was a cross-linked derivative of oligonucleotide 1 of the SL RNA (*5*). These results indicated that the AMT cross-link between the SL RNA and U6 snRNA was confined to a 13-nucleotide region of the SL RNA that contained the SL RNA's Sm-binding sequence. Further mapping by partial alkaline hydrolysis placed the major site of cross-linking at uridine 77 of the SL RNA, the 5′-most uridine of the SL RNA's Sm-binding site (*13*). In the U6 snRNA, we have mapped the major site of cross-linking to uridine 98, five nucleotides from the 3′ end (*13*). Inspection of the sequences of the SL RNA and U6 snRNA indicated that these RNAs are complementary over a stretch of 18 nucleotides (with one bulged nucleotide) that includes the cross-link site (Fig. 6). In the SL RNA, the region of complementarity includes the Sm-binding site and adjacent critical nucleotides. In U6 snRNA, the region of complementarity adjoins and overlaps the portion of U6 snRNA that in other organisms interacts with U2 snRNA by base pairing (*12, 14*).

Although the complementarity between the SL RNA and U6 snRNA extends into stem III of SL RNA, we speculate that the base pairing interaction (if functionally significant) is probably limited to the region depicted in Fig. 6B for the following reasons. (i) No critical purines were detected in stem III by modification interference. (ii) No cross-links were detected 3′ of the Sm-binding region (*13*). (iii) The more limited interaction would permit simultaneous base pairing of U6 snRNA with both the SL RNA and U2 snRNA. These potential interactions, together with the well-established [in other systems (*15*)] base-pairing interaction of U2 snRNA with the pre-mRNA branch site, suggest a plausible model for trans-spliceosome assembly in which U6 and U2 snRNAs provide a connecting bridge between the SL RNA and the acceptor pre-mRNA. If this model is experimentally verified, it would answer, at least in part, one of the major questions in trans-splicing, that is, how the two substrates for trans-splicing (the SL RNA and acceptor pre-mRNA) efficiently associate in the absence of significant sequence complementarity to each other.

REFERENCES AND NOTES

1. P. A. Sharp, *Cell* **50**, 147 (1987); J. D. Thomas, R. C. Conrad, T. Blumenthal, *ibid.* **54**, 533 (1988); J. P. Bruzik, K. van Doren, D. Hirsh, J. A. Steitz, *Nature* **335**, 559 (1988); K. van Doren and D. Hirsh, *ibid.*, p. 556.
2. G. J. Hannon, P. A. Maroney, J. A. Denker, T. W. Nilsen, *Cell* **61**, 1247 (1990); P. A. Maroney, G. J. Hannon, J. A. Denker, T. W. Nilsen, *EMBO J.* **9**, 3667 (1990).
3. P. A. Maroney, G. J. Hannon, J. D. Shambaugh, T. W. Nilsen, *EMBO J.* **10**, 3869 (1991).
4. I. W. Mattaj, in *Small Nuclear Ribonucleoprotein Particles*, M. L. Birnstiel, Ed. (Springer-Verlag, New York, 1988), pp. 100–114.
5. P. A. Maroney, Y.-T. Yu, G. J. Hannon, G. E. Hannon, T. W. Nilsen, unpublished data.
6. A. M. Takacs *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7932 (1988); T. W. Nilsen *et al.*, *Mol. Cell. Biol.* **9**, 3543 (1989); W. Zeng, C. M. Alarcon, J. E. Donelson, *ibid.* **10**, 2765 (1990); G. W. P. Joshua *et al.*, *Mol. Biochem. Parasitol.* **460**, 209 (1991).
7. P. A. Maroney, G. J. Hannon, T. W. Nilsen, unpublished observation.
8. C. S. Surowy *et al.*, *Mol. Cell. Biol.* **9**, 4179 (1989); M. Bach, A. Krol, R. Lührmann, *Nucleic Acids Res.* **18**, 449 (1990); C. C. Query, R. C. Bentley, J. D. Keene, *Mol. Cell. Biol.* **9**, 4872 (1989); D. Scherly *et al.*, *EMBO J.* **8**, 4163 (1989); J. R. Patton *et al.*, *Mol. Cell. Biol.* **9**, 3360 (1989).
9. R. Lührmann, in *Small Nuclear Ribonucleoprotein Particles*, M. L. Birnstiel, Ed. (Springer-Verlag, New York, 1988), pp. 71–99.
10. J. Thomas, K. Lea, E. Zuker-Aprison, T. Blumenthal, *Nucleic Acids Res.* **18**, 2633 (1990); J. Sham-

baugh, P. A. Maroney, Y.-T. Yu, T. W. Nilsen, in preparation.
11. D. LaSala and T. W. Nilsen, in preparation.
12. T. P. Hausner, L. M. Giglio, A. M. Weiner, *Genes Dev.* **4**, 2146 (1990).
13. Y.-T. Yu, P. A. Maroney, T. W. Nilsen, in preparation.
14. B. Datta and A. M. Weiner, *Nature* **352**, 821 (1991); J. Wu and J. L. Manley, *ibid.*, p. 818; K. P. Watkins and N. Agabian, *Genes Dev.* **5**, 1859 (1991).
15. C. Guthrie, *Science* **253**, 157 (1991).
16. T. E. England and O. C. Uhlenbeck, *Nature* **275**, 560 (1978).
17. D. A. Peattie, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1760 (1979).
18. G. J. Hannon, P. A. Maroney, T. W. Nilsen, *J. Biol. Chem.* **266**, 22792 (1991).
19. P. A. Maroney, G. J. Hannon, T. W. Nilsen, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 709 (1990).
20. Supported by PHS grants GM31528 and AI28799, the John D. and Catherine T. MacArthur Foundation, and an award from the Burroughs Wellcome Fund (to T.W.N.). We thank J. Shambaugh for cloning and sequencing of the *Ascaris* U6 RNA.

8 May 1992; accepted 22 September 1992

Nucleosome Core Displacement in Vitro via a Metastable Transcription Factor–Nucleosome Complex

Jerry L. Workman*† and Robert E. Kingston

In order to function, transcription factors must compete for DNA binding with structural components of chromatin, including nucleosomes. Mechanisms that could be used in this competition have been characterized with the use of the DNA binding domain of the yeast GAL4 protein. The binding of GAL4 to a nucleosome core resulted in a ternary complex containing GAL4, the core histone proteins, and DNA. This ternary complex was unstable; upon the addition of nonspecific competitor DNA, it dissociated into either the original nucleosome core particle or GAL4 bound to naked DNA. Nucleosome core destabilization by GAL4 did not require a transcriptional activation domain. These data demonstrate the displacement of nucleosome cores as a direct result of binding by a regulatory factor. Similar mechanisms might affect the establishment of factor occupancy of promoters and enhancers in vivo.

An initial hurdle in activating gene expression involves the ability of regulatory factors to access their binding sites in chromatin. For numerous genes, activation results in the interruption of nucleosome arrays at enhancer and promoter sequences and the binding of specific regulatory factors to these sequences (*1*). For most genes, it is not clear whether binding by transcription factors causes the observed disruptions in chromatin structure or if chromatin structure must first be perturbed (by DNA replication, for example) to subsequently allow factor binding (*2*, *3*). However, an example of the former mechanism is provided by studies of the mouse mammary tumor virus (MMTV) promoter that have demonstrated that the glucocorticoid receptor can bind to a nucleosome associated with this promoter (*4*, *5*). It has been proposed that this binding alters the nucleosome so that the additional factors nuclear factor 1 (NF1) and TATA box binding factor (TFIID) can bind to this promoter and activate transcription (*5*). Similar mechanisms that involve replication-independent nucleosome displacement are illustrated in the activation of the PHO5 promoter in *Saccharomyces cerevisiae* (*6*), and indeed manipulation of nucleosome density in vivo alters the regulation of PHO5 and additional promoters (*7*). These studies and others have led to the proposal that transcription factors and nucleosomes are involved in a dynamic competition for occupancy of regulatory DNA sequences (*3*). We tested the hypothesis that binding by a regulatory factor can directly destabilize nucleosome cores. This mechanism would allow the initiation of occupancy of DNA sequences by regulatory factors in chromatin.

We have previously shown that derivatives of the yeast regulatory protein GAL4 can bind to nucleosomal DNA and transcriptionally activate nucleosomal templates in vitro (*8*, *9*). The binding of GAL4 to nucleosomes is facilitated by the presence of multiple GAL4 sites, which perhaps is relevant to the occurrence of multiple GAL4 sites in naturally found promoters (*9*). To determine the effect of this binding on nucleosome stability, we developed a protocol using purified components so that the protein composition of the resulting complexes could be determined directly. We used polymerase chain reaction (PCR) amplification to generate homogeneous DNA fragments of nucleosome core length (150 bp). The central portion of these fragments contained five GAL4 binding sites that covered 95 bp. These fragments were reconstituted into nucleosome core particles by salt-gradient dialysis with purified HeLa core histones and purified as 11S particles by sedimentation (*10*). Gradient-purified nucleosome core particles reconstituted by this protocol contain equal stoichiometries of the four core histones as illustrated by Coomassie blue–stained gels of the purified reconstituted particles (Fig. 1A) (*11*). Nucleosomes reconstituted by this method also demonstrate a kinetic barrier to micrococcal nuclease digestion at approximately 146 bp and demonstrate the 10-bp deoxyribonuclease (DNase I) digestion ladder characteristic of native nucleosome core particles (*11*).

We first verified that the GAL4 protein could bind to these purified nucleosome cores. We used GAL4-AH, a fusion protein that contains the DNA binding and dimerization domains of GAL4 and an artificial activation domain (*12*). This protein binds to the 150-bp PCR fragment that contains the five GAL4 sites and also binds to the same fragment reconstituted into nucleosome cores as measured by electrophoretic mobility-shift assay (EMSA) (Fig. 1B). The purified DNA fragment was shifted through a series of complexes, which represents the binding of an increasing number of GAL4-AH dimers (from one to five dimers). Similarly, a series of complexes was generated upon GAL4-AH binding to the reconstituted nucleosome cores. The specificity of this interaction was demonstrated by DNase I footprinting of templates under similar conditions (*9*) and by the lack of formation of these complexes when the same amounts of protein were incubated with PCR-raised templates that did not contain GAL4 binding sites. GAL4 binding to the purified nucleosome cores, relative to binding naked DNA, was inhibited even less than the tenfold reported previously (*9*) in this assay. This lowered inhibition was a result of the high concentration (above the dissociation constant) of nucleosome cores used to facilitate the protein analysis. The intermediate complexes with increasing numbers of GAL4 dimers bound to the nucleosome cores were decreased in mobility relative to the analogous complexes formed with the naked DNA probe. Although this "supershift" became more subtle with an increasing number of bound GAL4 dimers, it was still apparent when five dimers were bound (*9*). These data suggest that components of the original nucleosome core (core histones) could still be present in the complexes with bound GAL4 dimers.

To address this question, we took advantage of the fact that there were only five

J. L. Workman, Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.
R. E. Kingston, Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114 and Department of Genetics, Harvard Medical School, Boston, MA 02115.

*Present address: Department of Molecular and Cell Biology, The Pennsylvania State University, University Park, PA 16802.
†To whom correspondence should be addressed at The Pennsylvania State University.