THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 36, Issue of September 7, pp. 33833–33839, 2001

Printed in U.S.A.

Strong RNA Splicing Enhancers Identified by a Modified Method of Cycled Selection Interact with SR Protein*

Received for publication, April 3, 2001, and in revised form, June 27, 2001
Published, JBC Papers in Press, July 13, 2001, DOI 10.1074/jbc.M102957200

Huicheng Tian‡§ and Ryszard Kole¶

From the ‡Department of Genetics, Center for Genetic and Cellular Therapies, Duke University, Durham, North Carolina 27710 and the ¶Lineberger Comprehensive Cancer Center, Department of Pharmacology and Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, North Carolina 27599

A modified method of cycled selection was used to characterize splicing enhancers for exon inclusion from a pool of β-globin-based three exon/two intron pre-mRNAs with a variable number of random nucleotides incorporated in the internal exon. The pre-mRNAs generated by this method contained random sequences ranging from 0 to 18 nucleotides in length. This method was used to isolate particular splicing enhancer motifs from a previously enriched pool of extremely diverse enhancers. After four cycles of selection for mRNA containing the internal exon, a distinct enhancer motif (GACGAC···CAGCAG) was highly enriched. This motif served as strong splicing enhancers in a heterogeneous exon. We have shown here that the selected enhancer motif promotes exon inclusion through specific interaction with SRp30. We have also shown that although present in many of our selected splicing enhancers conforming to this motif, a typical purine-rich enhancer sequence is dispensable for either enhancer activity or binding with SRp30.

Numerous studies have identified very diverse exonic splicing enhancers that either stimulate constitutive splicing or regulate alternative splicing (1–10). Transacting splicing factors, namely the serine/arginine-rich (SR) proteins, have been shown to interact with exonic splicing enhancers and influence splicing or exon inclusion (Refs. 3, 6, 7, 11, 12, 13, 14 and reviewed in Refs. 15, 16). The SR proteins contain a unique RS domain at the C-terminal region and one or two RNA recognition motifs at the N-terminal region (reviewed in Ref. 15). The SR proteins appear to be involved in the selection of both 3′- and 5′-splice sites. It has been shown that SR proteins recruit U2AF65 to the upstream 3′-splice site through direct interaction with U2AF35 (17–19). SR proteins also mediate the recognition of a downstream 5′-splice site by U1 snRNP in certain pre-mRNAs (12, 20, 21). SR proteins may play an important role in bridging the 5′- and 3′-splice sites during splicing (18, 22, 23, 24). It has been shown that SR proteins, such as SC35 and SF2/ASF, can simultaneously interact with the U170k and U2AF35 protein subunits to stabilize the association of the 5′- and 3′-splice sites in the early spliceosome (18).

In most studies, the exonic splicing enhancers are identified as purine-rich sequences present in cellular pre-mRNAs (*e.g.* Refs. 3, 6, 10), whereas in recent years certain non-purine-rich splicing enhancers are also found through cycled selection (SELEX) from pre-mRNAs containing random sequences (9, 25, 26). Specific RNA sequences with high affinity for purified SR proteins have been identified as a short consensus selected from pools of random sequences and shown to serve as splicing enhancers (7, 27, 28). In these studies, the RNA consensus for binding each SR protein ranges from 6 to 10 nucleotides in length (reviewed in Ref. 16). However, when assayed for splicing enhancer activity, both the selected consensus and its flanking sequences or multiple copies of the consensus must be inserted in pre-mRNAs. This raises the question of whether each selected consensus alone is the complete binding motif for the SR protein and is sufficient as a splicing enhancer. Our previous study has obtained very efficient yet extremely diverse splicing enhancers that stimulate exon inclusion in a two intron/three exon pre-mRNA, DUP184 (9). Such diversity was likely because of the fact that no consensus sequences could be derived from limited sequencing data of highly complex pool of enhancers enriched through various mechanisms of action. We have attempted to specifically isolate individual motifs through selectively enriching subpopulations of the whole pool. We hope that such subpopulations with greatly reduced sequence complexity would yield one or more enhancer motifs. In the present investigation, we have developed a unique method to directly isolate and identify minimum enhancer motifs from particular subpopulations of enhancers containing variable number of random nucleotides. We have identified RNA motifs that have strong enhancer activity and specifically interact with SRp30.

EXPERIMENTAL PROCEDURES

Primers, Templates, and Plasmids—The materials and methods used for selection of exon inclusion enhancers were essentially as previously described (9), except that a newly designed primer (primer 7, 5′-GCCATTCGTAATGTCTGCCCATTCTGCT-3′) replaced the previously used primer 6. The positions of both primers are shown in Fig. 1*B*. For this study, the pool of sequences obtained by five cycles of *in vitro* splicing and reverse transcriptase-PCR[1] from the previous study was initially used as the starting material for PCR with primers 7 and 5 to generate fragment R'. The full-length DNA template was obtained by overlapping extension with two overlapping halves of the template obtained by PCR with primers 1 and 2, fragment R′, and primer 4, respectively. After an additional four cycles of selection the reconstructed DNA template was ligated into pSP64 vector through *Hin*dIII-*Bam*HI restriction sites (Promega). The plasmid DNA was used to transform DH5α *Escherichia coli* cells, and individual clones were sequenced.

Templates of LP4.19R, 4.7R, 4.7RA, 4.7RB, 4.7RC, and 4.7RD for transcription were constructed via site-directed mutagenesis from two halves. To construct the 5′-halves of the templates, DUP184REV plas-

* This work was supported in part by National Institutes of Health Grant GM32994 (to R. K.) and a National Institutes of Health viral oncology grant (to H. T.). 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: Box 2601, DUMC, Durham, NC 27710. Tel.: 919-684-6787; Fax: 919-681-7970; E-mail: h.tian@cgct.duke.edu.

[1] The abbreviations used are: PCR, polymerase chain reaction; nt, nucleotide(s).

mid[2] was subjected to PCR with primer 1 (Fig. 1*A* and Ref. 9) and the primers 5′ 4.19 (CTGCTGCTCCTCGTCGTCTTATCCGCCAAGTACAA), 5′ 4.7 (CAGTTCTGCTGAGTCGTCTTATCCGCCAAGTACAA), 5′ 4.7A (CAGTTCTGCTGACTGCTGTTATCCGCCAAGTACAA), 5′ 4.7B (CAGTTCTGGTGAGTGGTCTTATCCGCCAAGTACAA), 5′ 4.7C (CAGTTCTGGTGACTGGTGTTATCCGCCAAGTACAA), 5′ 4.7D (GGAGGGCACAGTTAGTCGTCTTATCCGCCAAGTAC), respectively. The 3′-halves of the templates containing an overlapping sequence with the 5′-half were obtained by PCR with primer 4 and the primers 3′ 4.19 (GACGACGAGGAGCAGCAGACTGTGCCCTCCATGC), 3′ 4.7 (GACGACTCAGCAGAACTGTGCCCTCCATGC), 3′ 4.7A (CAGCAGTCAGCAGAACTGTGCCCTCCATGC), 3′ 4.7B (GACCACTCACCAGAACTGTGCCCTCCATGC), 3′ 4.7C (CACCAGTCACCAGAACTGTGCCCTCCATGC), 3′ 4.7D (GACGACTAACTGTGCCCTCCATGC). The full-length templates were obtained by overlapping extension of the two halves through PCR, digested with *Hin*dIII and *Bam*HI and were cloned into the pSP64 vector.

In Vitro Transcription, Splicing, and Electrophoresis—To prepare pre-mRNAs for *in vitro* splicing, the plasmids were linearized with *Bam*HI and transcribed by SP6 RNA polymerase (9, 29). The HeLa nuclear extract was prepared as described (30). RNA splicing was carried out in a mixture of around 10 μg/μl HeLa nuclear extract, 50,000 cpm of ^{32}P-labeled pre-mRNA and splicing buffer containing 1.3% polyvinyl alcohol, 1.5 mM ATP, 10 mM creatine phosphate, and 3 mM $MgCl_2$ (31). The splicing products were extracted with phenol and precipitated in ethanol and resolved on an 8% polyacrylamide, 7 M urea gel in 1× Tris-borate-EDTA. The ^{32}P-labeled RNA probes used for filter binding and UV cross-linking assays were transcribed *in vitro* by T7 RNA polymerase (Promega) from synthesized DNA templates and four different [α-^{32}P]NTP following the manufacturer's instructions and purified on a 10% polyacrylamide, 7 M urea gel.

Double Filter Binding, UV Cross-linking, and Immunoprecipitation—For the binding reaction, 3 μg/μl of HeLa nuclear extract, 0.2 μg/μl of tRNA, and splicing buffer were incubated on ice for 20 min with 30,000 cpm of ^{32}P-labeled RNA probes in a total volume of 10 μl. The mixtures were added with 40 μl of splicing buffer before loaded onto a set of stacking nitrocellulose and Nylon membranes (Schleicher & Schuell) pre-equilibrated with splicing buffer and mounted on a 96-well manifold (Schleicher & Schuell). After passed through the filters by vacuum, the membranes were washed once with 50 μl of splicing buffer. The membranes were separated and autoradiographed on a phosphorimager. For UV cross-linking, protein/RNA mixtures in a volume of 10 μl as above were irradiated with UV (short wave range) at about 7 milliwatts/cm^2 on ice with a UV transilluminator (Fotodyne) for 10 min. For immunoprecipitation, the samples from UV cross-linking were diluted with 40 μl of IP buffer (20 mM pH 7.9 Hepes, 150 mM KCl, 20% glycerol, 1 mM dithiothreitol, 1% Triton X-100, 0.1% Nonidet P-40, and 0.2 mM EDTA) and precipitated with 4.8 μg of monoclonal antibody 1H4 specific for SR proteins (BAbCO) and 4 μl of flurry protein G-Sepharose beads (Amersham Pharmacia Biotech) at 4 °C for overnight. The beads were washed twice with 50 μl of IP buffer. The precipitated proteins were treated with 5 μg of RNase A for 15 min at 37 °C, dissolved in SDS loading buffer, and resolved on a 10% polyacrylamide/SDS gel. The gel was dried before autoradiography.

Depletion of SRp30 through RNA-based Affinity Chromatography—Equal amounts of HeLa nuclear extract were loaded onto columns containing streptavidin-agarose beads (Amersham Pharmacia Biotech) without RNA or tagged with RNA containing the enhancer LP4.7 (5′-GACGACUCAGCAGCCAGAGGUCGAUGUACU-3′), LP4.19 (5′-GACGACGAGGAGCAGCAGCCAGAGGUCGAUGUACU-3′), or the equivalent DUP184 sequence (5′-GUAUACAGAAUAGCAGAAUGCCAGAGGUCGAUGUACU-3′). The RNA tags were appended to the beads by hybridizing them to a biotinylated 2′-*O*-methyl-RNA oligonucleotide (5′-biotin-AGUACAUCGACCUCUGG-3′) that is complementary to the extended 3′ common sequence of the RNA tags. The flow-through extract samples were collected and adjusted to equal volumes with Buffer D (30) and tested for splicing activities.

Autoradiography and Imaging—Autoradiography was either processed with Fuji x-ray film or by a Storm 840 phosphorimager (Molecular Dynamics). The efficiency of internal exon inclusion was quantitated by densitometry of the autoradiograms using the NIH Image 1.47 software or directly quantitated on a phosphorimager. The percentage of products with an included internal exon relative to the sum of the total mRNA products was calculated.

[2] Z. Dominski and R. Kole, unpublished data.

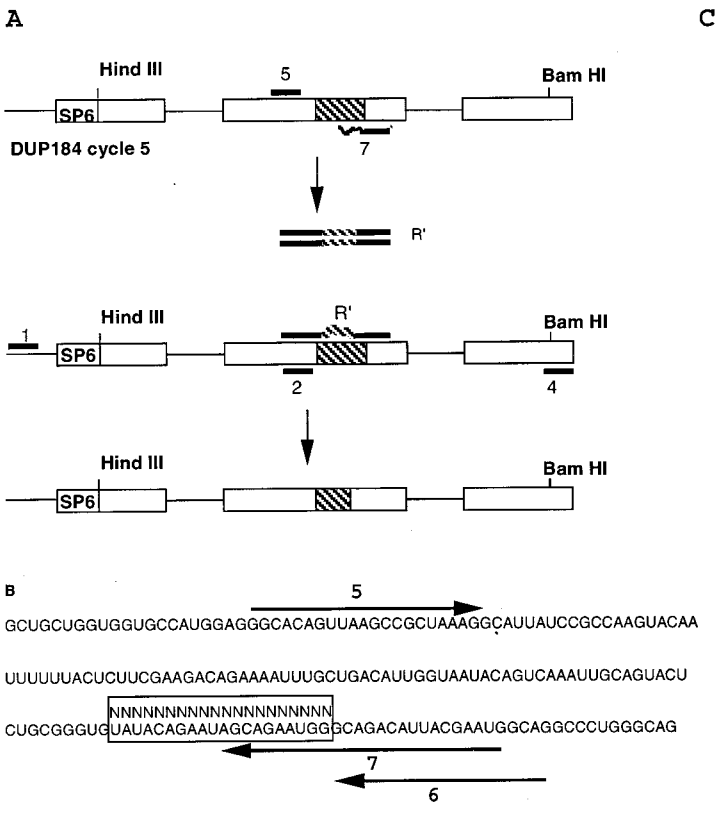
RESULTS

Refinement of Sequence Motifs from a Complex Pool through Selective Amplification—Through an iterated enrichment process, we have obtained largely diverse exonic splicing enhancers that promote exon inclusion during splicing of a human β-globin-based pre-mRNA (9). We have then attempted to isolate particular splicing enhancer motifs by selectively amplifying homologous sequences from the complex pool of enhancers. As the starting material for the experiments described in this study, we used the previously obtained pre-mRNA library containing a 20-nt random sequence in an internal exon (DUP184 cycle 5 in Ref. 9) to generate a less diverse pre-mRNA library. Different from the previous primer 6, which was completely complementary to the exon sequence flanking the random 20 nt, we extended the 3′-end of primer 6 by adding a 10-nt leader of the original DUP184 sequence, which showed no splicing enhancer activity. When annealed to the exon, this 10-nt leader of the modified primer (primer 7, Fig. 1*B*) would overhang from the adjacent flank into the random 20 nt and freely hybridize with any short complementary sequence present within or outside the random 20 nt of pre-mRNAs (Fig. 1*B*). If the DUP184 cycle 5 pool contains sequences complementary to the 3′-end of primer 7, they would be amplified through PCR and therefore become templates for the new pre-mRNA library.

To test this concept, the DNA template of DUP184 cycle 5 was subjected to PCR with primers 5 and 7, and the amplified fragment (R′ in Fig. 1*A*) was used to reconstruct the full-length DNA template for pre-mRNA (Fig. 1*A*). Sequencing of 30 individual clones isolated from the reconstructed pool (Pool LP0) showed that primers 5 and 7 have indeed amplified sequences from the DUP184 cycle 5 pool. In these clones, the random sequences were greatly varied in length with a roughly even distribution from 0 to 18 nt (Fig. 1*C*), although the "in-frame" random sequences should have been 10 nt in length. Only 5 clones contained random 10 nt (clones LP0.4, LP0.8, LP0.11, LP0.28, LP0.29) while three clones contained random sequences longer than 10 nt, and the majority had shorter or no random sequences. Surprisingly, no clones with truncations outside the random region were found through the preliminary sequencing. We conclude that the variable number of random nucleotides in different clones resulted from annealing of the 3′-leader of primer 7 at variable positions within the random region of the internal exon during PCR.

Characterization of Splicing Enhancer Motifs—The pool of LP0 pre-mRNAs was subjected to four cycles of selection with previously described methods (9). Individual pre-mRNAs from the pool LP0 and final pool LP4 were tested for splicing activity. Nearly all LP0 pre-mRNAs showed less than 40% inclusion of the internal exon in splicing probably because of the disruption of previously selected splicing enhancers during reconstruction of the pre-mRNA pool (data not shown). In contrast, the majority of the LP4 pre-mRNAs showed much greater exon inclusion (Fig. 2*A*), indicating that the reconstruction had isolated and further enriched a subpopulation of splicing enhancers from the pool of extremely diverse pre-mRNAs.

Analysis on the sequences of 45 LP4 clones showed that the selection had narrowed down the length distribution of the random sequences. Clones with fewer than 6 randomized nt were eliminated from the pool LP4 during selection, and thirty-five of the clones (78%) contained 10 ± 2 random nt (Table I). We had divided the sequences into three groups based on homology. In Group I, 5 clones (Fig. 2*B*, clones LP4.17, LP4.19, LP4.27, LP4.37, LP4.44) contained an identical 12 nt sequence, ACGACGAGGAGC. Unexpectedly, when combined with the 5′- and the 3′-flanking nucleotides, this redundant sequence formed a symmetrical motif gACGACGAGGAGCagcag (the



C

Pool LP.0 Clones

DUP184 cycle 5

UCUGCGGGUG NNNNNNNNN NNNNNNNNNGCAGAC

Clones	Flanking	Random	Flanking
0.3	G	---------	AGCAGAAUGGGCAGAC
0.10	G	---------	A
0.20	G	---------	A
0.9	G	A--------	A
0.26	G	A--------	A
0.7	G	AGG------	A
0.30	G	AAGC-----	A
0.17	G	AGCUA----	A
0.15	G	UCGACG---	A
0.22	G	AUGGAC---	A
0.18	G	AAUUACA--	A
0.19	G	AAUCGCC--	A
0.23	G	ACGACAA--	A
0.1	G	CUGACCGU-	A
0.2	G	UUCAUGGA-	A
0.6	G	AUGGCGAA-	A
0.25	G	AUCGCUCG-	A
0.27	G	AACUGUGA-	A
0.5	G	GACCGACAU-	A
0.13	G	ACGAACGGA-	A
0.16	G	GGAGGACGG-	A
0.24	G	AACCUACGC-	A
0.4	G	AGAACGCUAC	A
0.8	G	AAACGUAACG	A
0.28	G	GAUGGAAUG	A
0.29	G	GUGGUCGCUA	A
0.11	G	ACGUCGUCGG	A
0.12	G	ACGAGAACGAAG	A
0.21	G	ACGUCGACAAGA	A
0.14	G	AGAGACGACGUCCGGAAC	A

FIG. 1. *A,* scheme for generating length polymorphism of randomized nucleotides. *Open boxes*, exon sequences; *hatched box*, randomized sequence; *thin lines*, introns; *SP6 box*, SP6 promoter; *heavy lines*, PCR primers. The PCR template represents the sequence library obtained by 5 cycles of selection described previously (cycle 5, Ref. 9). Primer 7 corresponds to the sequence indicated in Fig. 1*B*. The 10 nucleotides from its 3′-end are able to hybridize with complementary stretches present within the random sequence of the individual clones. PCR with primers 5 and 7 generated a product of ~150 nt (R′), which was used with primer 4 in subsequent PCR to generate the 3′-half of the DNA template. This fragment combined with the product of PCR with primers 1 and 2 was further amplified with primers 1 and 4 leading to the full-length DNA template used subsequently for *in vitro* transcription of pre-mRNA. *B,* sequence of the internal exon of DUP184 pre-mRNA. The randomized sequence is indicated by *N* and is *boxed*. *Arrows* indicate primers 5 and 7, which were used to alter the length of the randomized sequence, as shown in *A* and throughout the selection cycles to reconstruct the DNA templates. Primer 6, which had been used previously (9) to obtain a 20-nucleotide, fixed length random sequence is shown for comparison. *C,* sequence of the randomized region from 30 individual LP0 clones. The central portion of the figure shows random sequences of variable length. The constant flanking sequences are shown for clone LP0.3. *Dashes* indicate deletions within the expected 10 random nucleotides.

flanking nucleotides are in lowercase). Sequences of this group indicated the consensus [gACGAC····Cagcag] with a variable purine-rich core in most clones, although a few contained no purine-rich core (*e.g.* LP4.7, Fig. 2*A*). Fourteen clones from pool LP4 contain this motif, or 31% of the pool, although only 1 clone (LP0.14) from pool LP0 matches this consensus, or 3.3%. Therefore, this motif has been enriched approximately 10-fold from pool LP0 to LP4. In Group II, sequences point to the consensus [gAC··purine-rich··cag] with a variable purine-rich core in the middle (Fig. 2*A*). There are 21 clones in LP4 in this group, or 46.7% of the pool, while 13 clones in pool LP0, or 43.3%. Therefore, there is nearly no enrichment for this motif from pool LP0 to LP4. Group III contains both purine-rich and non-purine-rich sequences without a consensus. This study has focused on investigating whether the Group I motif represents strong splicing enhancers and whether the symmetrical structure *per se* plays a role in exon inclusion.

Because the purine-rich core GAGGAG of LP4.19 conforms to the exon-splicing enhancer GARGAR present in other pre-mRNAs (*e.g.* Ref. 32), we attempted to determine the contribution of the purine-rich core to splicing enhancer activity. As a comparison, we used LP4.7, which lacks the purine-rich core but has the same motif and a single uridine present at the center (Fig. 2*B*). As shown in Fig. 3, the LP4.19 pre-mRNA included the internal exon very efficiently during splicing (*lane 4*), and the LP4.7 pre-mRNA showed only slightly lower exon inclusion (*lane 5*). This indicates that the motif [GACGAC·····CAGCAG] largely retained enhancer activity in the absence of the purine-rich core.

To examine the effect of the symmetrical structure of the motif on splicing enhancer activity, LP4.19 and LP4.7, and several modified LP4.7 sequences either retaining or lacking a symmetrical structure (Fig. 3, LP4.7RB, LP4.7RA, and LP4.7RC), as well as the sequence corresponding to the 5′ portion of the motif (GACGAC, LP4.7DR) were introduced via site-directed mutagenesis to the equivalent positions in the internal exon of the construct DUP184REV, in which a major portion of the internal exon was in reverse orientation relative to that of DUP184. As shown in Fig. 3, while the original DUP184REV pre-mRNA included the middle exon even less efficiently than DUP184 (Fig. 3, *lanes 3* and *2*, respectively), the LP4.19 and LP4.7 sequences placed in DUP184REV enhanced exon inclusion as efficiently as in DUP184 (*lanes 4–7*), indicating that they retained splicing enhancer activity in a

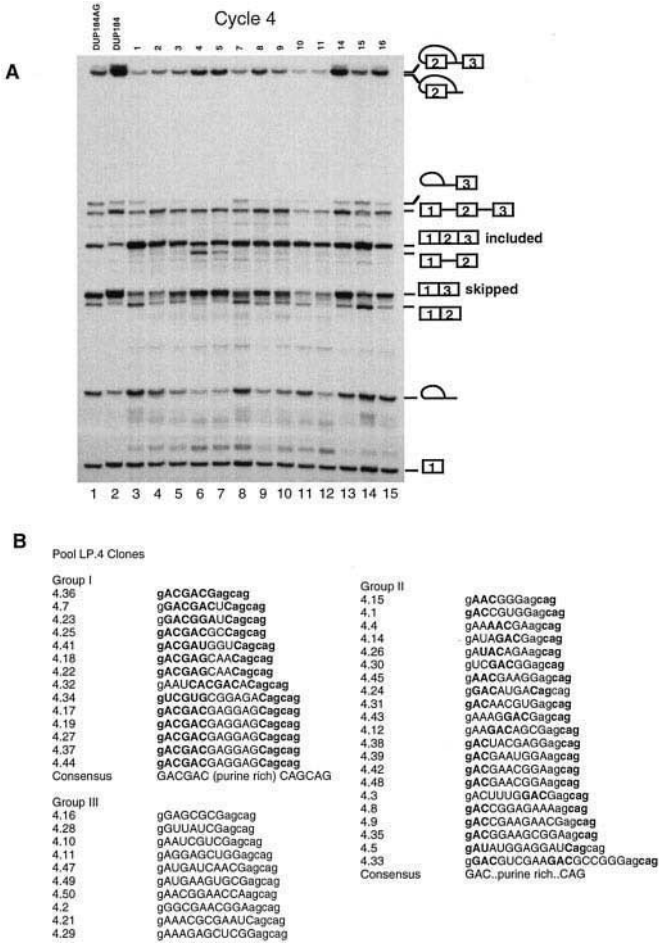


FIG. 2. *A,* splicing of the pre-mRNAs isolated from pool LP4. Splicing of control pre-mRNAs which include (DUP184AG) or exclude (DUP184) the internal exon (Ref. 2) is shown in *lanes 1* and *2*, respectively. *Lanes 3–15*, splicing of pre-mRNAs transcribed from individual LP4 clones indicated at the *top* of the figure. The structure of splicing products and intermediates is shown on the right. *B,* sequence of the randomized region of 45 individual LP4 clones. The number of nucleotides in the randomized region ranges from 6 to 18. Clones 4.17, 4.19, 4.27, 4.37, and 4.44 contain an identical sequence. The missing clone numbers (*e.g.* 4.13) are due to non-readable sequencing gels. The consensus motifs are indicated in *bold face* and flanking sequences are in *lowercase*.

different sequence context. Deletion of the right-half portion of the motif drastically reduced its enhancer activity (*lane 8*), indicating that the 3′ portion of the motif (CAGCAG) is also required for the splicing enhancer activity even though the 3′ portion by itself has no splicing enhancer activity. In addition, it appears that the particular selected sequence rather than the symmetrical structure *per se* contributed to the splicing enhancer activity, as nucleotide substitutions either retaining or disrupting a symmetrical structure also led to dramatically reduced exon inclusion (*lanes 9–11*).

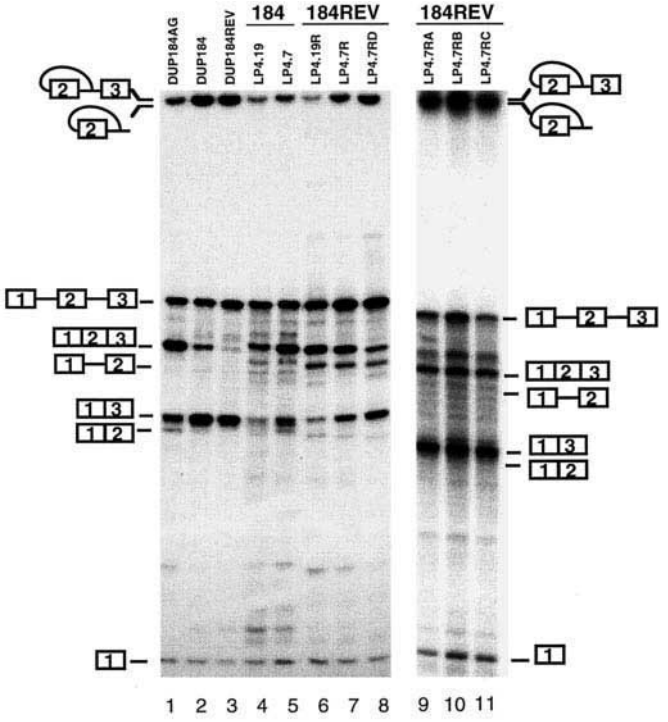
Interaction of the Selected Splicing Enhancer Motif with SR Proteins—To test whether the selected splicing enhancer motifs interact with protein factors in the HeLa nuclear extract, we used a double filter binding method (33). Internally ^{32}P-labeled RNA probes of the selected motifs (LP4.19 and LP4.7) and several mutated sequences lacking the motifs (LP4.19A, LP4.7A, LP4.7B, LP4.7C, and LP4.7D), as well as the equivalent DUP184 sequence were incubated with HeLa nuclear extract. The mixtures were passed through the stacking nitrocellulose and nylon filters. As shown in Fig. 4*A*, around 60% of the input LP4.19 and LP4.7 RNA probes were present in RNA-protein complexes and bound to the nitrocellulose membrane, while only 9–17% of the input DUP184, LP4.19A, LP4.7A, LP4.7B, LP4.7C, and LP4.7D RNA probes were retained on the nitrocellulose membrane. To determine whether the selected splicing enhancers interact with SR proteins, a mixture of fully internally ^{32}P-labeled RNA probes (DUP184, LP4.19, LP4.19A, LP4.7, 4.7A, LP4.7B, LP4.7C, and LP4.7D) and HeLa nuclear extract were irradiated with UV after incubation. The cross-linking products were immunoprecipitated with the monoclonal antibody 1H4 that recognizes all members of the SR protein family. The immunoprecipitates were treated with RNase before being resolved on an SDS-polyacrylamide gel for autoradiography. As shown in Fig. 4*B*, both LP4.7 and LP4.19 probes were cross-linked to SRp30. In contrast, mutations that changed the LP4.19 and LP4.7 enhancer motifs (LP4.19A, LP4.7A, B, C, and D) abolished their interaction with SRp30. Therefore, the specific interaction of the selected LP4.7 and LP4.19 motifs and SRp30 strongly correlates with their splicing enhancer activity. Consistent with its lack of splicing enhancer activity, the DUP184 sequence was not recognized by SR proteins. In separate experiments, depletion of the SRp30 from HeLa nuclear extract by affinity columns tagged with either the LP4.7 or LP4.19 RNA motifs has led to the inhibition of splicing of the LP4.7 and LP4.19 pre-mRNAs, while depletion with the DUP184 RNA sequence showed nearly no effect on splicing (data not shown). It is possible that the SRp30 is an essential splicing factor or that by binding to exonic enhancers the SRp30 interacts with other general splicing factors in the early spliceosome, so that depletion of the SRp30 also removed these general splicing factors and abolished the splicing activity of the nuclear extract.

TABLE I
Length distribution of the selected sequences

Pool	LP0	LP4
Length	Total no. of clones	
nt	*30*	*45*
0	3	–
1	2	–
2	–	–
3	1	–
4	1	–
5	1	–
6	2	2
7	3	7
8	5	8
9	4	9
10	5	6
11	–	5
12	2	7
13	–	–
14	–	–
15	–	–
16	–	–
17	–	–
18	1	1

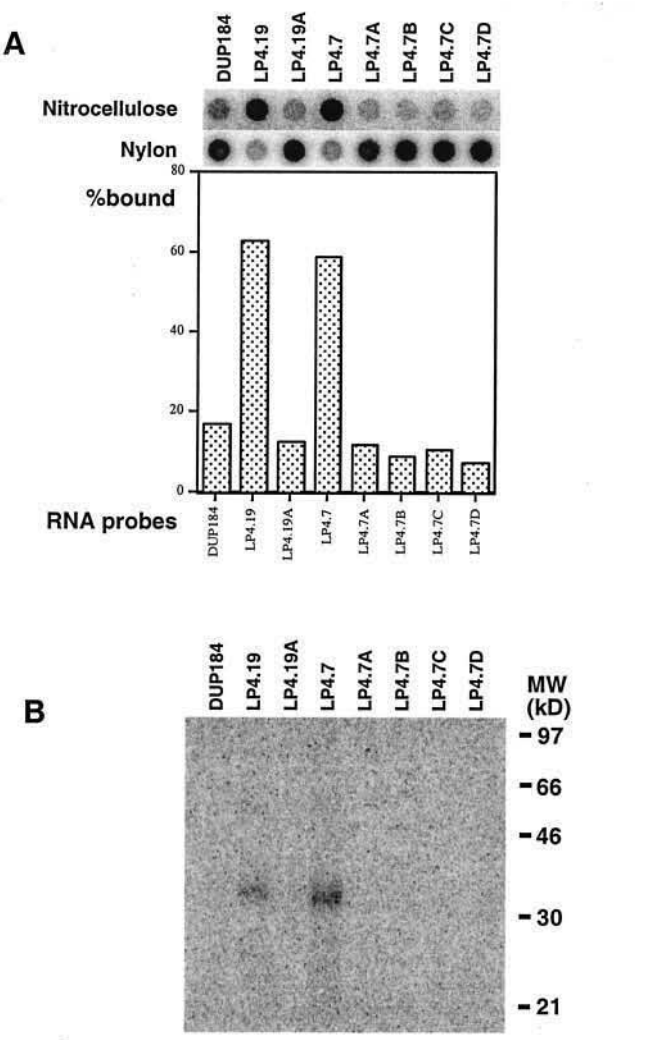
DISCUSSION

Identification of Particular Splicing Enhancer Motifs from a Complex Pool—Selection of particular RNA molecules from a random sequence library provides a powerful tool for identifying binding motifs or for selecting new catalysts (34–37). Because the exact number of nucleotides involved in binding or catalysis is mostly unknown prior to selection, the number of residues to be randomized is in most cases arbitrarily chosen. The "consensus" sequences identified by such method are the homologous regions present among the multiple selected winners but may not necessarily represent the complete motifs for binding or function. In most cases, sequences outside the consensus must also be included for proper activity. Our study shows that complete sequence motifs can be directly identified



DUP184	ggug	uauacagaau	agcagaaugg
LP4.19		ACGACGAGGAGC	
LP4.7		GACGACUC	
DUP184REV	auaa	ugccuuuagcggcuua	acugugcccu
LP4.19R		GACGACGAGGAGCAGCAG	
LP4.7R		GACGACUCAGCAGa	
LP4.7RD		GACGACUa	
LP4.7RA		CAGCAGUCAGCAGa	
LP4.7RB		GACCACUCACCAGa	
LP4.7RC		CACCAGUCACCAGa	

FIG. 3. **Splicing enhancer activity of LP4.19 and LP4.7 motifs and their mutants in a heterogeneous pre-mRNA.** The symmetrical LP4.19 and LP4.7 sequences as well as their mutated sequences (listed below) were inserted into the DUP184REV construct via site-directed mutagenesis, and pre-mRNAs transcribed from these constructs were spliced. *Lanes 1–3*, splicing of control DUP184AG, DUP184, and DUP184REV pre-mRNAs, respectively; *lanes 4* and *5*, splicing of LP4.19 and LP4.7 pre-mRNAs, respectively; *lanes 6* and *7*, splicing of LP4.19R and LP4.7R pre-mRNAs; *lane 8*, splicing of LP4.7RD pre-mRNA, which has a truncated LP4.7 motif. Other altered LP4.7 sequences, which either retain (4.7RB) or eliminate (4.7RA and 4.7RC) the symmetrical LP4.7 motif were also inserted into DUP184REV. *Lanes 9–11*, splicing of 4.7RA, 4.7RB, and 4.7RC pre-mRNAs, respectively.



RNA Probes	
DUP184	GUAUACAGAAUAGCAGAAUG
LP4.19	GACGACGAGGAGCAGCAG
LP4.19A	CAGCAGGAGGAGCAGCAG
LP4.7	GACGACUCAGCAG
LP4.7A	CAGCAGUCAGCAG
LP4.7B	GACCACUCACCAG
LP4.7C	CACCAGUCACCAG
LP4.7D	GACGACUAACUGUG

FIG. 4. **The splicing enhancer motif is specifically recognized by SRp30.** *A,* double filter binding assay with HeLa nuclear extract and fully internal ^{32}P-labeled RNA probes DUP184, LP4.19, LP4.19A, LP4.7, LP4.7A, LP4.7B, LP4.7C, and LP4.7D (sequences are listed below). RNA/protein complexes are retained on the nitrocellulose membrane (*top*) whereas the unbound RNA probes pass through the nitrocellulose membrane and bind to the nylon membrane (*below*). The nitrocellulose retention was calculated as the percentage of counts of the RNA probes on nitrocellulose membrane out of the sum on both membranes. *B,* UV cross-linking of the ^{32}P-labeled RNA probes with SR proteins. Fully internal ^{32}P-labeled RNA probes (listed on *top* of the gel) and HeLa nuclear extract were irradiated by UV on ice. The cross-linking products were immunoprecipitated with the monoclonal antibody 1H4 specific for SR proteins and treated with RNase. The immunoprecipitates were dissolved in SDS loading buffer and resolved on a 10% polyacrylamide/SDS gel. The gel was dried for autoradiography. The protein molecular size markers are on the *right*.

from a complex pool enriched for multiple targets by isolating a subpopulation of sequences of variegated length. This method uses a primer extended from the fixed flanking region to isolate and amplify a subpopulation that shares a common short sequence through annealing with the extended 3′-leader of the primer in PCR. Consequently, a library with variable number of nucleotides but a common short sequence in the random region is obtained. We believe that by using PCR primers with various leader sequences different subpopulations of the whole pool, which represent various regions of the sequence space, may be isolated and individually enriched to a further extent. As a result, multiple sequence identities from selections for very complex targets may be rapidly determined. Although in our study short sequences recognized by the 3′-leader of the selective primer are present in certain selected enhancers, they do not have to be part of the enhancer motifs.

In search for unknown splicing enhancers, RNA sequences have been selected against purified SR proteins through binding affinity (7, 27, 28). In these studies, however, the splicing enhancer activity has not been involved in selection mechanism. Not surprisingly, all selected high-affinity binding sequences are not splicing enhancers (*e.g.* Ref. 28). We and others

have used function-based SELEX to simultaneously select various exonic splicing enhancers regardless of the mechanism of action (9, 25, 26, 38). Although the *in vitro* selections have generated a much broader spectrum of splicing enhancers than those identified *in vivo*, it is difficult to identify particular motifs based on limited sequencing data due to the presence of extremely diverse enhancers (9, 26). In this study we have developed a unique method to isolate a motif through selective amplification of a subpopulation from a complex pool of splicing enhancers, which is shown to specifically interact with SRp30. We believe that with our method, various motifs recognized by different splicing factors can be isolated from a very complex pool of enhancers when the identity of particular motifs cannot be revealed through preliminary selection.

A subpopulation of exonic enhancers isolated from our previously enriched pool of splicing enhancers contained a variable number of random nucleotides ranging from 0 to 18. After a few additional rounds of selection, the length of the random sequences in 45 randomly chosen clones ranged from 6 to 12 nucleotides (Table I). The pre-mRNAs with fewer or no random sequences (*e.g.* LP0.3, LP0.7, and LP0.9) were eliminated through further selection. The progress of selection is associated with increasing splicing enhancer activity of the pool and enrichment of the most efficient splicing enhancers, *e.g.* LP4.19 and LP4.7 (Figs. 2 and 3). It is unlikely that enhanced exon inclusion is due to the removal of a negative element from exon as observed in other studies (39–41), because replacement of the original DUP184 sequence with either random sequences prior to selection or various defined sequences largely maintained predominant exon skipping (9).[3] Therefore we conclude that specific splicing enhancers have been enriched through the cycled selection employed in our studies.

Splicing Enhancer Activity of the Selected Motifs and Interaction with SR Proteins—The alternative splicing of pre-mRNA involves the selection of a particular pair of 5′- and 3′-splice sites from other possible combinations of potential splice sites present in the same pre-mRNA. This process appears to be largely facilitated by SR proteins that link the appropriate splice sites in the early spliceosome (*e.g.* Ref. 18). From a very complex pool of splicing enhancers, we have isolated a splicing enhancer motif (GACGAC···CAGCAG) that stimulates exon inclusion and specifically interacts with SRp30. Two studies have independently identified similar but shorter consensus (RGACKACGAY, R = A/G, K = G/U, Y = C/U) in exonic enhancers that are recognized by the SR protein 9G8 and stimulate splicing of an upstream intron (26, 28). However, in either case the role of the sequences adjacent to the consensus in splicing enhancer activity or in binding with 9G8 has not been characterized. Therefore, the consensus identified in these studies may not represent the complete splicing enhancers, as the adjacent sequences obtained from selection were also inserted in the pre-mRNAs when the enhancer activity was assayed. In contrast, our approach allowed direct identification of the complete enhancer motifs through systematic comparison of a selectively amplified subpopulation of enhancers variegated in length. Although the enhancer motifs identified in our study are mostly symmetrical in contents, it is apparent that the particular 5′ and 3′ sequences rather than the symmetrical structure present in these motifs are essential for splicing enhancer activity, because nucleotide substitutions that either disrupted the original symmetrical structure or created a new symmetrical sequence dramatically reduced exon inclusion and abolished their interaction with SR proteins (Figs. 3 and 4). In our preliminary experiments, the RNA probe GACGACGA representing the 5′ portion of our selected motif was rapidly degraded in HeLa nuclear extract, indicating that this sequence itself may not be sufficient to specifically interact with any proteins (data not shown). We have also found that a 3′ adjacent sequence AGCAGCAG is required for both splicing enhancer activity and binding with SRp30, because replacement of this sequence by the sequence at further downstream of the exon abolished exon inclusion and binding with SRp30 (LP4.7RD, Fig. 3; LP4.7D, Fig. 4). Therefore, the 3′ portion AGCAG, although by itself shows a lack of enhancer activity in the original DUP184 pre-mRNA, makes a significant contribution to the enhancer activity and therefore part of the enhancer motif. We are not certain whether the SRp30 cross-linked with the RNA probes in our study is 9G8. However, it is apparent that although we and others have obtained a similar splicing enhancer consensus, degenerate but particular 3′ adjacent sequences are also required for enhancer activity and interaction with SR proteins. The SRp30 observed in our study may participate in the early spliceosome assembly through direct interaction with the enhancer motifs of the pre-mRNAs and with other general splicing factors, as depletion of the SRp30 from the nuclear extract through affinity chromatography using the enhancer motifs has dramatically reduced splicing activity (data not shown). We are now characterizing particular proteins purified by affinity chromatography tagged with the selected enhancers, by protein mass spectrometry or microsequencing. This will allow us not only to identity the SRp30 but also to find other associated general splicing factors in the early spliceosome.

Although the purine-rich core GAGGAG is present in many selected enhancers in our study (Fig. 2*B*), it is dispensable for their splicing enhancer activity, as the absence of the purine-rich core only slightly reduced exon inclusion (Fig. 3) and did not abolish the cross-linking of the enhancer motif with SRp30 (Fig. 4). This is in contrast with observations by others in which purine-rich GARGAR enhancers specifically interact with SR proteins (*e.g.* Refs. 1, 7, 32). We conclude that the selected enhancer motif (GACGAC···CAGCAG) largely contributes to the splicing enhancer activity observed in our study.

Acknowledgment—We thank John P. Jones III for critically reading the manuscript.

[3] H. Tian and R. Kole, unpublished data.

REFERENCES

1. Dirksen, W. P., Hampson, R. K., Sun, Q., and Rottman, F. M. (1994) *J. Biol. Chem.* **269,** 6431–6436
2. Dominski, Z., and Kole, R. (1994) *J. Biol. Chem.* **269,** 23590–23596
3. Lavigueur, A., Branche, H. L., Kornblihtt, A. R., and Chabot, B. (1993) *Genes Dev.* **7,** 2405–2417
4. Lynch, K. W., and Maniatis, T. (1995) *Genes Dev.* **9,** 284–293
5. Selvakumar, M., and Helfman, D. M. (1999) *RNA* **5,** 378–394
6. Sun, Q., Mayeda, A., Hampson, R. K., Krainer, A. R., and Rottman, F. M. (1993) *Genes Dev.* **7,** 2598–2608
7. Tacke, R., and Manley, J. L. (1995) *EMBO J.* **14,** 3540–3551
8. Tanaka, K., Watakabe, A., and Shimura, Y. (1994) *Mol. Cell. Biol.* **14,** 1347–1354
9. Tian, H., and Kole, R. (1995) *Mol. Cell. Biol.* **15,** 6291–6298
10. Watakabe A., Tanaka, K., and Shimura, Y. (1993) *Genes Dev.* **7,** 407–418
11. Amrein, H., Hedley, M. L., and Maniatis, T. (1994) *Cell* **76,** 735–746
12. Staknis, D., and Reed, R. (1994) *Mol. Cell. Biol.* **14,** 7670–7682
13. Tian, M., and Maniatis, T. (1992) *Science* **256,** 237–240
14. Tian, M., and Maniatis, T. (1993) *Cell* **74,** 105–114
15. Fu, X. D. (1995) *RNA* **1,** 663–680
16. Tacke, R., and Manley, J. L. (1999) *Curr. Opin. Cell Biol.* **11,** 358–362
17. Wang, Z., Hoffman, H. M., and Grabowski, P. J. (1995) *RNA* **1,** 21–35
18. Wu, J. Y., and Maniatis T. (1993) *Cell* **75,** 1061–1070
19. Zuo, P., and Maniatis, T. (1996) *Genes Dev.* **10,** 1356–1368
20. Côté, J., Simard, M. J., and Chabot, B. (1999) *Nucleic Acids Res.* **27,** 2529–2537
21. Zahler, A. M., and Roth, M. B. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2642–2646
22. Fu, X. D., and Maniatis, T. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 1725–1729
23. Hoffman, B. E., and Grabowski, P. J. (1992) *Genes Dev.* **6,** 2554–2568
24. Kuo, H. C., Nasim, F. H., and Grabowski, P. J. (1991) *Science* **251,** 1045–1050
25. Coulter, L. R., Landree, M. A., and Cooper, T. A. (1997) *Mol. Cell. Biol.* **17,** 2143–2150
26. Schaal, T. D., and Maniatis, T. (1999) *Mol. Cell. Biol.* **19,** 1705–1719

27. Tacke, R., Tohyama, M., Ogawa, S., and Manley, J. L. (1998) *Cell* **93,** 139–148
28. Cavaloc, Y., Bourgeois, C. F., Kister, L., and Stevenin, J. (1999) *RNA* **5,** 468–483
29. Konarska, M. M., Padgett, R. A., and Sharp, P. A. (1984) *Cell* **38,** 731–736
30. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489
31. Krainer, A. R. T., Maniatis, T., Ruskin, B., and Green, M. R. (1984) *Cell* **36,** 993–1005
32. Xu, R., Teng, J., and Cooper, T. (1993) *Mol. Cell. Biol.* **13,** 3660–3674
33. Wong, I., and Lohman, T. M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 5428–5432
34. Ekland, E. H., Szostak, J. W., and Bartel, D. P. (1995) *Science* **269,** 364–370
35. Ellington, A. D., and Szostak, J. W. (1990) *Nature* **346,** 818–822
36. Tsang, J., and Joyce, G. F. (1994) *Biochemistry* **33,** 5966–5973
37. Tuerk, C., and Gold, L. (1990) *Science* **249,** 505–510
38. Liu, H. X., Zhang, M., and Krainer, A. R. (1998) *Genes Dev.* **12,** 1998–2012
39. Amendt, B. A., Si, Z. H., and Stoltzfus, C. M. (1995) *Mol. Cell. Biol.* **15,** 4606–4615
40. McNally, L. M., and McNally, M. T. (1996) *J. Virol.* **70,** 1163–1172
41. Nemeroff, M. E., Utans, U., Kramer, A., and Krug, R. M. (1992) *Mol. Cell. Biol.* **12,** 962–970