cDNA cloning of U1, U2, U4 and U5 snRNA families expressed in pea nuclei

Brian A. Hanley and Mary A. Schuler*

Departments of Biochemistry and Plant Biology, University of Illinois, Urbana, Illinois 61801, USA

Received November 30, 1990; Revised and Accepted February 18, 1991 EMBL accession nos X15926 – X15936 (incl.)

ABSTRACT

Differences observed between plant and animal pre-mRNA splicing may be the result of primary or secondary structure differences in small nuclear RNAs (snRNAs). A cDNA library of pea snRNAs was constructed from anti-trimethylguanosine ($m_3^{2,2,7}G$) immunoprecipitated pea nuclear RNA. The cDNA library was screened using oligo-deoxyribonucleotide probes specific for the U1, U2, U4 and U5 snRNAs. cDNA clones representing U1, U2, U4 and U5 snRNAs expressed in seedling tissue have been isolated and sequenced. Comparison of the pea snRNA variants with other organisms suggest that functionally important primary sequences are conserved phylogenetically even though the overall sequences have diverged substantially. Structural variations in U1 snRNA occur in regions required for U1-specific protein binding. In light of this sequence analysis, it is clear that the dicot snRNA variants do not differ in sequences implicated in RNA:RNA interactions with pre-mRNA. Instead, sequence differences occur in regions implicated in the binding of small ribonucleoproteins (snRNPs) to snRNAs and may result in the formation of unique snRNP particles.

INTRODUCTION

Eukaryotic cells contain a class of nuclear RNAs known as snRNAs (1,2). The most abundant species, U1, U2, U3, U4, U5 and U6 snRNAs, are rich in uridine, have a $m_3^{2,2,7}G$ cap structure at the 5' end (except for U6 snRNA) and have lengths between 100 and 220 nucleotides (2). Each of these snRNAs is present at approximately 10^5 – 10^6 molecules/nucleus in human cells and are 10 to 100-fold less abundant in plant nuclei (2). They are functional in RNA processing when complexed with proteins in small nuclear ribonucleoprotein particles (snRNPs) (3). In animal and yeast pre-mRNA splicing, the U1 and U2 snRNPs are involved in recognition of the 5' splice site and the branch site, respectively (3–5). The U5 snRNP has been implicated in recognition of the 3' splice site (3).

The structure and expression of the snRNA components of snRNPs have been well documented in animal and yeast systems

(2,5,6). Comparisons of the major plant snRNAs with their animal counterparts indicate that plant snRNAs are less abundant and more variant (7–10). Direct sequencing of stably expressed snRNAs has revealed that some of the differences between plant and animal snRNAs can be accounted for by families of sequence variants which exist for U2 snRNA in wheat, U4 snRNA in broad bean and U5 snRNA in pea (9,11,12). Most of the other available plant snRNA sequences have been derived from gene analysis. These include two genes for soybean U1 snRNA (13), one gene for bean U1 snRNA (14), nine genes for tomato U1 snRNA (15), one gene for maize U2 snRNA (16), one gene for *Arabidopsis* U5 snRNA (17) and six genes for *Arabidopsis* U2 snRNA (18). In cases where multiple genes exist for a particular snRNA, each gene represents a unique sequence variant. Three of the *Arabidopsis* U2 snRNA genes, one tomato U1 snRNA gene and one *Arabidopsis* U5 snRNA gene have been expressed in transfected protoplasts, suggesting that at least some of the isolated genes are expressed *in vivo* (15,17–19). Others, such as a tomato U1 snRNA gene, represent pseudogenes not expressed *in vivo* (20).

This heterogeneous collection of snRNA sequences indicates that extensive similarities do exist between plant and animal snRNAs, especially in sequences or structural elements required for pre-mRNA splicing (21). In particular, the region of U1 snRNA that binds the 5' splice site is identical in all organisms (21). Sequences in stem-loops I and II, of U1 snRNA, which are known to interact with the U1-C, U1-70K and U1-A polypeptides (22–26), and the Sm-antigen binding motifs are also conserved (21). In U2 snRNA, the single stranded region interacting with the branch site (3) is conserved. The regions of U4 and U6 snRNA which interact to form the U4/U6 base paired complex are phylogenetically conserved (27). U5 snRNA represents the most highly diverged snRNA in that only the sequences in loop I and the Sm-binding site are conserved in plants and animals (2). Although these comparisons suggest that gross structures required for pre-mRNA splicing have been phylogenetically conserved in plants and animals, none of these studies have documented the fine structure diversity of the snRNAs expressed within one particular plant and compared these variations with the minor snRNA variants that exist in mammalian tissue.

* To whom correspondence should be addressed at Department of Plant Biology, University of Illinois, 289 Morrill Hall, 505 S. Goodwin Avenue, Urbana, IL 61801, USA

Figure 1. High-Resolution Analysis of Pea snRNAs. A. Nuclear RNAs were immunoprecipitated with anti-m₃G antibody, 3' end-labeled using ³²P-cytidine biphosphate and T4 RNA ligase and were fractionated on 15% acrylamide:bisacrylamide (38:2), 8.3 M urea gels (40 cm length), and autoradiographed as described in Methods. Lane 1, anti-m₃G immunoprecipitated pea nuclear RNA; lane 2, anti-m₃G immunoprecipitated HeLa cell nuclear RNA. The pea snRNA variants are identified by brackets at the left of the panel. HeLa cell snRNAs, 5S rRNA and two single stranded DNA strands (220 and 154 nt.) are designated on the right. B. Anti-m₃G^{2,2,7} immunoprecipitated pea nuclear RNA was fractionated on a 15% acrylamide:bisacrylamide, 8.3 M urea gel (40 cm in length), blotted onto Gene-Screen and sequentially hybridized with the ³²P-labeled oligonucleotide probes described in Methods. The probes used for each hybridization are designated at the top of each autoradiograph. snRNAs are identified at the left; the positions of the 220 and 154 nt. single-stranded DNA standards are shown at the right.

To identify the snRNA structural variants stably expressed in pea nuclei, we have constructed a cDNA library enriched in pea snRNAs. Initial screening of the cDNA library has led to the isolation of numerous U1, U2, U4 and U5 snRNA clones. Sequence analysis of the clones has allowed us to demonstrate that four of the five snRNAs required for splicing are represented by sequence variants expressed in pea nuclei. This analysis defines the fine structure variations in these sequences.

MATERIALS AND METHODS

Reagents

The radioisotopes 5'-[γ-³²P] ATP (7000 Ci/mmol), 5'-[α-³²P] cytidine 3',5'-bisphosphate (3000 Ci/mmol) (pCp), 5'-[α-³⁵S] dATP (1320 Ci/mmol) and Gene Screen were from New England Nuclear. T4 RNA Ligase and poly (A) polymerase were from P-L Biochemicals. All other enzymes were from Bethesda Research Laboratories.

RNA Isolation

Nuclei from 100 grams of pea (*Pisum sativum* var. Progress #9) epicotyl tissue were sedimented over a Percoll gradient as described by Gallagher and Ellis (28). The nuclei (approximately 2 × 10⁸ nuclei) were extracted with phenol:chloroform (1:1) and the nucleic acids were recovered from the aqueous phase by several ethanol precipitations at -70°C. The snRNAs were purified from 2 × 10⁸ nuclei by immunoprecipitation with 150 μg of anti-m₃G^{2,2,7} IgG as described by Adams and Herrera (29).

	10	20	30	40	50
U1.1	ATACTTACCTGGATGGGTCGATGGGTGATCATGAAGGCCCATGGC	TAG			
U1.2	ATACTTACCTGGATGGGTCaATGGGTGATCaAGAGGCCCATGGC	TAG			
U1.3	ATACTTACCTGGATGGGTCaATGGGTGATCATGAAGGCCCATGGC	TAG			
U1.36	ATACTTACCTGGATGGGTCaATGGTTCATCATGAAGGCaCCATGGCcTAG				
U1.15	ATACTTACCTGGATGGGTCGATGGGTGATCATGAAGGCCCATGGCcTAG				
U1.66	ATACTTACCTGGATGGGTCaATGGTTCATCATGAAG	aCCATGGCcTAG			
U1.137	ATACTTACCTGGATGGGTCaATGGGTGATCATcAgGGTCCATGGCcTAG				
	60	70	80	90	100
U1.1	GATTGTGACCTCCATTGCACTTAGGAGGGGTGCTTTCCTAAGGTCTACCC				
U1.2	GcaaGTGACCTCCATTGCACTTAGGAGGGGTGCTagCCTAAGGTCTACCC				
U1.3	GcaaGTGACCTCCATTGCACTTAGGAGGGGTGCTTgCCTAAGGTCTACCC				
U1.36	GagaGTGACCTCCATTGCACTTAGGAGGGGTGCTaTCCTAAGGTCTACCC				
U1.15	GATTGTGACCTCCATTGCACTTAGGAGGGGTGCTTTCCTAAGGTCTACCC				
U1.66	Gga GTGACCTCCATTGCACTTAGGAGGTGCTaTCCTAAGGTCTACCC				
U1.137	Gaa GTGACCTCCATTGCACTTAGGAGGGGTGCTagCCTAAGGTCTACCC				
	110	120	130	140	150
U1.1	AAGTGGTGGAGCCTACATCATAATTGTTGCGCTGAGGGGGCTGCGTTGG				
U1.2	AAGTGGTGGAGCCTACATCATAATTGTTGCG TGtGGGGGCGCTGtGTTGG				
U1.3	AAGTGGTGGaACCTACATCATAATTGTTGCG TGtGGGGGCGCTGCGTTGG				
U1.36	AtGTGGTGGaACCTACATCATAATTGTTGCG aGtGGGGGCGCTGCGTcGG				
U1.15	AAGTGGTGGAGCCTACATCATAATTGTTGCG TGAGGGGGCGCT CG				
U1.66	AAGTGGTGGaACCTACATCATAATcTGTGCG TGAGGGGGCGCTGCGTTGG				
U1.137	AAGTGGTGGAGCCTACATCATAATTGTTGCG TGtGGGGGCGCTGCGTTGG				
	160				
U1.1	GGCGGGCCCCCACC	162 nt			
U1.2	GGCGGGCCCCCtC	160 nt			
U1.3	GGCGGGCCCCtCt	160 nt			
U1.36	GGCGGGCCCCtCt	161 nt			
U1.15	GGCGGGCCCCCACC	157 nt			
U1.66	GGCGGGCCCCCACC	160 nt			
U1.137	GGCGGGCCCCCtC	161 nt			

Figure 2. U1 snRNA Sequence Variants Expressed in Pea Seedling Nuclei.

Electrophoresis of RNA

Immunoprecipitated snRNAs were 3' end-labeled with pCp and T4 RNA ligase (30) and fractionated on 15% acrylamide:bisacrylamide (38:2), 8.3 M urea gels (31 × 38 × 0.05 cm) containing 90 mM Tris-borate, 2 mM EDTA (1 × TBE) buffer at 45 W for 15–16 hours. Samples were diluted with an equal volume of 10 M urea, 0.1% xylene cyanol, 0.1% bromophenol blue and heated at 100°C for 1 minute prior to loading. Northern analysis was performed essentially as described by Egeland et al. (7). Northern blots were probed with oligodeoxyribonucleotides 5' end-labeled with [γ-³²P] ATP and T4 polynucleotide kinase (31). Blots were hybridized at room temperature for 16 hours in 6 × SSC, 5 × Denhardt's solution, 0.1% SDS and washed three times for 20 minutes at room temperature in 6 × SSC, 0.1% SDS. All oligodeoxyribonucleotide sequences were synthesized at the University of Illinois Biotechnology Center and were gel-purified or HPLC-purified prior to use. The oligodeoxyribonucleotide probes used in these studies are complementary to human and pea snRNAs as follows: U1a (5'-GCCAGGTAAGTAT), HeLa nt. 1–13; U2_{L15} (5'-C-AGATACTACACTTG), pea and HeLa nt. 28–42; U5_{L25} (5'-TAGTAAAAGGCGAAAGATAGTTTCGC), pea nt. 23–47; U4a (5'-TTTCAAYYAGCAATAA), HeLa nt. 52–67.

cDNA Cloning

For cloning, approximately 1 μg of immunoprecipitated snRNAs was poly (A) tailed using poly (A) polymerase (32). cDNA synthesis and cloning into oligo dT-tailed pCGN 1703 vector were performed as described by Alexander (33). pCGN 1703 is a derivative of pBS(-) (Stratagene) in which the lac and T3 RNA polymerase promoters have been deleted and replaced with a linker, which allows for convenient cloning of cDNAs.

Figure 3. Secondary Structures of the Pea U1 snRNA Variants. Bottom: The complete sequence for the U1.1 snRNA variant expressed in pea seedlings is drawn according to the model proposed by Mount and Steitz (36). Arrows followed by subscripted nucleotides (or deletions) are used to designate base variations and deletions in the U1.2, U1.3, U1.15, U1.36, U1.66 and U1.137 snRNA variants. Nucleotides phylogenetically conserved in other organisms (21) are circled. Nucleotides in the U1.1 variant which differ in phylogenetically conserved positions are designated in lower case. The Sm binding site is boxed. Top: Alternate duplexed structures for stem-loop II of the individual pea U1 snRNA variants are shown with their corresponding free energy coefficients designated below.

Colony Hybridizations

The cDNA plasmid library was amplified and screened for snRNA clones by colony hybridization using 5' end-labeled oligodeoxynucleotide probes (34) and hybridization conditions were performed as described for Northern analysis.

Sequencing

Clones were sequenced using modified T7 DNA polymerase (Sequenase, U.S. Biochemicals), 5'-[α -³⁵S] dATP and the T7 RNA polymerase promoter primer (P-L Biochemicals) as

described by the manufacturer. Free energy coefficients for stem structures were calculated using the secondary structure program PC fold 3.0 of Zuker and Stiegler (35).

RESULTS

RNA Isolation and Analysis

Pea snRNAs were purified from epicotyl tissue by immunoprecipitation with anti-m³2,2,7G IgG (29). Fractionation on a 10% acrylamide: bisacrylamide (38:2), 8.3 M urea gels

Figure 4. Secondary Structures for U2 snRNA Variants Expressed in Pea Nuclei. The sequence of the U2.1 snRNA variant expressed in pea seedlings is drawn according to the model described in Reddy and Busch (2). Arrows followed by subscripted nucleotides (or deletions) are used to designate base variations and deletions in the U2.2, U2.3 and U2.4 variants. U2.1, U2.2, U2.3 represent snRNAs expressed in seedling nuclei; U2.4 corresponds to the U2 snRNA gene sequence reported in Hanley and Schuler (42). Asterisks designate nucleotide insertions, deletions or variations in the partial pea U2 snRNA sequence reported by Krol et al. (9). The Sm antigen binding site is boxed. Alternate structures for hairpin III in U2.2 and U2.4 are shown above. Phylogenetically conserved nucleotides (21) are circled. Nucleotides in the U2.1 variant which differ in phylogenetically conserved positions are designated in lower case.

revealed a pattern of snRNAs very similar to those found in tobacco, tomato, cucumber, and broad bean nuclei (10) (not shown). Because plant U4 snRNAs migrate in the same molecular weight range as U1 snRNAs (7,12), the 3' end-labeled anti- $m_3^{2,2,7}G$ immunoprecipitated pea nuclear RNA were separated on high resolution 15% acrylamide: bisacrylamide (38:2), 8.3 M urea gels that fractionate on the basis of size and nucleotide composition (Fig. 1A). Although the U1 and U4 snRNAs still overlap to some extent, it is apparent that the molecular weights of the major pea snRNAs, especially U4 snRNA, vary significantly from those found in HeLa cell nuclei, as previously reported (7,10). Therefore, to identify individual snRNAs, total nuclear RNA was subjected to Northern analysis and sequential screening using oligodeoxyribonucleotide probes specific for sequences conserved in the U1, U2, U4, and U5 snRNAs. In order to detect all snRNA variants, including those with multiple mismatches, low stringency conditions were used for Northern analysis and colony screening (7). Hybridizations performed at higher temperatures, 5°C below the T_m of the probes, detect the same U1, U2, U4 and U5 snRNAs indicating that all closely-related variants have been detected. As shown in Figure 1B, this type of analysis reveals multiple length variants of U1, U4, and U5 snRNAs, but only one major and two minor species of U2 snRNA. Each of the snRNAs hybridizing with the oligodeoxyribonucleotide probes could be 3' end-labeled with pCp (Fig. 1A) indicating that the snRNA variants retain a 3'-hydroxyl group and do not represent RNA degradation products. Identical snRNA profiles were obtained with different nuclear RNA preparations and under alternate extraction

conditions (not shown) indicating that the snRNAs expressed in pea nuclei are as heterogeneous in size as those expressed in other plants (7,10).

cDNA Cloning

Immunoprecipitated snRNAs were poly (A) tailed with poly (A) polymerase (32) and cloned into oligo dT-tailed pCGN 1703 using the vector-primer technique described by Alexander (33). Sequencing indicated that approximately 70 adenosine monophosphate residues were polymerized onto each snRNA. Approximately 0.7×10^6 ampicillin-resistant transformants were obtained per microgram of immunoprecipitated snRNA. Restriction analysis indicated that approximately 40% of the randomly chosen transformants from the unamplified library had inserts greater than 200 base pairs, the expected size for cDNA clones encoding snRNAs. The library was amplified and screened by colony hybridization using 5' end-labeled oligodeoxyribonucleotide probes specific for the U1, U2, U4, and U5 snRNAs. The sequences reported here represent snRNA clones obtained from our initial screening of the library. Additional snRNA variants certainly exist.

U1 snRNA Clones

Seven U1 snRNA clones detected by colony hybridization were sequenced and identified as full-length clones by comparison with other U1 snRNA sequences. The sequences of these, designated U1.1 to U1.137, are aligned with each other in Figure 2. The positions of nucleotide variations within the U1 snRNA secondary structure originally proposed by Mount and Steitz (36) are shown

Figure 5. Secondary Structure of the Pea U4 snRNA Variants. The complete nucleotide sequence of the U4.1 snRNA variant is shown base-paired to the broad bean U6 snRNA in the model proposed by Brow and Guthrie (27). Arrows followed by subscripted nucleotides (or deletions) are used to designate base variations and deletions. Phylogenetically conserved nucleotides (21) are circled. Nucleotides in the U4.1 variant which differ in phylogenetically conserved positions are designated in lower case. The Sm binding site is boxed.

in Figure 3. The phylogenetically conserved sequences that are present in a variety of organisms including yeast, *Drosophila*, *Xenopus*, and mammals, are indicated in Fig. 3 (21). The pea U1 snRNA clones represent sequence and length variants from 157 to 162 nucleotides which share 80–90% similarity with the common bean U1 snRNA gene, 70–80% similarity with the soybean U1 snRNA genes and 63% similarity with HeLa cell U1 snRNA (2,13,14). The overall degree of nucleotide conservation within the pea U1 snRNA family is 85–90%. Nearly all of the length variation results from deletions of the 3' terminal cytosine and internal insertions/deletions at nucleotides 38, 47, 53, and 131. The smallest variant, U1.15, has a five nucleotide deletion in stem/loop IV. Eighty-three nucleotides at the 3' terminus of U1.1 are identical to the partial pea U1 snRNA sequence obtained by Krol et al. (9). Although this might represent the same U1 snRNA variant, stem II, which contains nearly all variations found in the pea U1 snRNAs, is not represented in the partial U1 snRNA sequence.

Regions of the U1 snRNA that have been shown to be important either from phylogenetic and/or experimental data are conserved in the pea U1 snRNAs (Figs. 2 and 3). For example, nucleotides 1–11 in the seven U1 snRNA variants are identical to those present in all but two of the available U1 snRNA sequences (2,21,37,38). The Sm-antigen binding sites (boxed nucleotides, Fig. 3) are nearly identical among the pea U1 snRNA variants, and are similar to the Sm-antigen binding sites of other U1 snRNAs at six out of nine positions (2,21). Loops I and II correspond to the regions of the vertebrate U1 snRNA that interact with the U1-C, U1-70K and U1-A proteins, (22–26).

Six of the seven pea U1 snRNA variants contain fourteen nucleotides in loops I and II that are phylogenetically conserved in other organisms (Fig. 3) (21). U1.137 has a base variation at nucleotide 36 which is absolutely conserved in other organisms.

The pea U1 snRNAs vary primarily in looped, bulged or single stranded regions (Fig. 3). Many of the differences in the pea U1 snRNA variants occur in the same regions that variations exist in other organisms, namely stems II and III and loop I (2,21,39–41). In contrast to the vertebrate U1 snRNA variants, prominent variations exist in loop IV and in the single stranded regions at the 3' end of the pea U1 snRNAs. Changes, which affect the stability of duplexed structures occur but, in general, are compensatory. Alternate structures for stem/loop II in six of the clones and their associated free energy values are shown in Fig. 3.

U2 snRNA Clones

The full-length U2 snRNA clones, designated U2.1, U2.2 and U2.3, were 194 or 195 nucleotides in length and, in Figure 4, are folded according to the secondary structure described for U2 snRNA (2). The U2 snRNA genomic clone (42), designated U2.4, is 200 nt. in length. The pea U2 snRNA variants are 95% similar to the broad bean U2 snRNA (43), approximately 80% similar to *Arabidopsis* U2 snRNAs (18), and 70% similar to HeLa cell U2 snRNA (2). Ninety-five nucleotides at the 3' end of a partial pea U2 snRNA (9) differ from the U2 snRNA variants described here in the insertion of two compensatory nucleotides in stem III, a substitution in the stem IV bulge and a deletion in stem III (asterisks, Fig. 4). Thus, this sequence potentially

Figure 6. Secondary Structure of the Pea U5 snRNA Variants. The sequences of the U5.3 and U5.4 snRNA variants defined in this paper are drawn according to the model for U5 snRNA (2). The pea U5 snRNA variants described by Krol et al. (9) are designated as follows: U5.5 (U5.5), U5.6 (U5.6a), U5.7 (U5.7a), U5.8 (U5.8). Arrows followed by subscripted nucleotides (or deletions) are used to designate base variations and deletions. The positions of nucleotides phylogenetically conserved in other organisms (21) are circled. Nucleotides in the U5.3 variant which differ in phylogenetically conserved positions are designated in lower case. The Sm binding site is boxed.

represents a fifth U2 snRNA variant expressed in pea nuclei. The first 71 nucleotides of the pea U2 snRNA variants are identical to each other and to broad bean, *Arabidopsis*, wheat, maize, and HeLa cell U2 snRNAs and other phylogenetically conserved U2 snRNA sequences (2,11,16,18,43). The Sm-antigen binding site is identical in the pea U2.1, U2.2 and U2.3 snRNA variants, but differs slightly from the Sm binding site in U2.4 and other plant U2 snRNAs. The branch point binding sequence (GUA-GUA) is absolutely conserved (Fig. 4). The base variations between pea U2 snRNAs occur in the last 120 nucleotides of the molecules primarily in duplexed and bulged structures of stems III and IV. Within stem III, the extensive variations in U2.2 and U2.4, contribute to the formation of the alternate hairpin structures shown in Fig. 4. The unusual 4 nt. insertion in loop III of U2.4 can base pair with sequences to form an extended hairpin III containing an alternate loop sequence or, if the tetraloop structure contributes increased stability to stem III, additional nucleotides are positioned in bulged sequences.

U4 snRNA Clones

Three full-length U4 snRNA clones, U4.1, U4.2 and U4.3, are 152, 152, and 150 nucleotides in length, respectively (Fig. 5). The U4 snRNA variants are 88–95% similar to the broad bean U4 snRNAs (12) and 60–65% similar to HeLa cell U4 snRNAs (2). The region of interaction between U4 and U6 snRNAs (21) is phylogenetically conserved in the three U4 snRNA variants.

The 5' halves of the pea U4 snRNA variants contain the most highly conserved region of each molecule in parallel with the highest degree of sequence conservation found in animal and yeast U4 snRNAs (2,21). Stems I and II, which form in the U4/U6 hybrid, and the 5' stem/loop are essentially invariant in the pea snRNA variants. The Sm-antigen binding site is conserved in all of the pea variants, although U4.2 has a T to C replacement in the pyrimidine stretch (Fig. 5).

Variations between pea U4 snRNAs are localized mainly in the central and 3' stem/loops in the U4 molecule (Fig. 5). Most of these variations are compensatory, and maintain base-pair interactions. In only one case is a C:G base pair in the central stem replaced by a G:A, thus abolishing a base pair interaction. This change enlarges the single stranded loop region without introducing a bulge into the stem structure.

U5 snRNA Clones

The two full length U5 cDNA clones which have been isolated (U5.3 and U5.4) are 122 nucleotides in length. In Figure 6, we have compared the sequences of these two snRNA clones with four of the U5 snRNA sequences (U5.5–U5.8) previously reported by Krol et al. (9), which range in size from 119 to 122 nt. Our U5 cDNA clones are only 75 to 96% similar to the pea U5.5–U5.8 sequences and 55% similar to rat U5 snRNA (2). Nucleotides 33 to 52, in loop C of stem I, are absolutely conserved among the pea U5 snRNAs, and in fact, all known U5 snRNAs (2,9,21). The proportions of the hairpin stems and bulges in the secondary structure are well conserved in the pea U5 snRNAs and an *Arabidopsis* U5 snRNA (17) and less well conserved when compared with the mammalian and yeast structures (21).

Most of the differences in the pea U5 snRNA variants occur in the center of hairpins I and II, in loops A and D and near the Sm binding site (Fig. 6). In most cases, nucleotide variations in the stem structures involve compensatory changes or nucleotide substitutions in G-U base pairs to strengthen duplexed structures. Variations within loops A and D and in single stranded regions are essentially random.

DISCUSSION

The participation of snRNAs and snRNPs in pre-mRNA splicing has been clearly established (3). Although plant and animal snRNAs have similar conserved regions in their primary and secondary structures, the functionality of the snRNAs and/or snRNPs are not interchangeable in that the recognition of introns and the splicing of pre-mRNA appear to be modulated by slightly different factors in plants and animals (44,45). These differences result in the inefficient *in vitro* excision of some plant introns in HeLa cell extracts (46–48). Nucleotide differences within the snRNAs may directly or indirectly account for the splicing deficiencies in heterologous systems. Direct effects potentially result from variations in the conserved snRNA sequences specifically interacting with pre-mRNA sequences (49–53). Indirect effects might result from the formation of alternate snRNP structure due to variations in the secondary and/or tertiary structure of snRNAs. Effects, such as these, result in the inability of a bean U1 snRNA to bind *Xenopus* U1-specific proteins *in vitro* (24) and in the differential assembly of mouse U1 snRNAs (26).

A direct comparison of HeLa cell and pea snRNAs reveals that pea snRNAs exhibit more length and/or sequence variations

than their animal counterparts (Fig. 1). Using oligodeoxyribonucleotide probes directed against conserved sequences in each snRNA, we have demonstrated that length variants exist for U1, U2, U4, and U5 snRNAs. The sequences chosen for the U1 and U2 snRNA oligodeoxyribonucleotide probes hybridize with conserved sequences in the snRNAs critical for splicing (50). The oligodeoxyribonucleotide complementary to nucleotides 53–68 of the pea U4 snRNA overlaps nucleotides 64–75 of HeLa cell U4 snRNA, another region indispensable for splicing (54). The oligonucleotide chosen for the U5 snRNA probe is complementary to U5 snRNA sequences absolutely conserved in all organisms (21).

In Northern analysis (Fig. 1), approximately ten to twelve equally abundant U1 snRNA variants and five to six U4 snRNA variants were detected. Pea U2 snRNA consists of one major and two minor variants. The U5 snRNA variants, which are harder to resolve, include at least six U5 snRNA variants. The number of variants reported for each snRNA family represents a minimal estimate since variants migrating with the same apparent molecular weights cannot be separated in this type of analysis. In addition, highly divergent snRNA variants may have escaped our low stringency Northern and cDNA screening analyses. The large number of individual snRNA variants, which has been demonstrated previously in several plants (7,10,12), is in sharp contrast to the limited diversity of snRNAs in human and yeast cells. Yeast cells contain the most defined population in that each snRNA is represented by a single species (21). In vertebrate cells, snRNAs generally exist as unique species (2). Some vertebrate cells do contain variants of the U1, U4 and U5 snRNAs, but these variants tend to be expressed in a select set of tissues and typically represent a small percentage of the total snRNA population (2,40,55). Our cloning and sequence analysis of the pea U1-U5 snRNAs indicates the degree of sequence divergence which exists in these snRNA populations. Additional variants certainly exist for each snRNA.

The overall secondary structures of the pea U1 snRNA variants are highly conserved when compared with each other and with phylogenetically conserved U1 snRNA sequences (Fig. 3; 21). The first eleven nucleotides of U1 snRNAs, including the pea variants sequenced here, appear to be absolutely conserved in all organisms studied to date with the exception of an *Arabidopsis thaliana* U1 snRNA gene that contains a single nucleotide insertion at nucleotide 2 and a *S. pombe* U1 snRNA that lacks the first two nucleotides following the 5' trimethylguanosine cap (37,38). This degree of conservation is consistent with *in vitro* splicing experiments indicating that the 5' terminal U1 snRNA sequences participate in recognition of 5' splice sites (3). However, in light of the heterogeneous 5' splice sites in plant introns (44), the absolute conservation of the 5' termini in the pea variants was unexpected. Sequence analysis on additional pea U1 snRNA variants will determine if any highly divergent variants with alternate base pairing interactions between the 5' termini of U1 snRNA and 5' splice site in introns have escaped detection with the U1 oligonucleotide.

Some of the more unusual sequence variations in the expressed U1 snRNA variants occur in the loop of stem I which is highly conserved in a variety of organisms (Fig. 3; 21). All plant U1 snRNAs contain uridine or adenine at nucleotide 33 (13–15), but the analogous position in *Xenopus*, human and mouse U1 snRNAs contains a cytosine or uridine residue (39–41). The nucleotides flanking this nucleotide in loop I are absolutely conserved in vertebrate U1 snRNAs (21), but vary in one out

of seven (nt. 34) pea U1 snRNAs that have been described in this paper, and eight out of nineteen (nt. 32 and 34) of the plant U1 snRNA variants that have been sequenced (13–15,38). Sequences in the loop of stem I are essential for binding the U1-C polypeptide and additional structures within stem I mediate the binding of the U1-70K and U1-A polypeptides (22,24,25). Because sequence variations at positions 32 and 34 in some plant U1 snRNAs have been shown to effectively block binding of the U1-70K, A and C polypeptides in *in vitro* snRNP assembly systems (24), the sequence variants that we have documented here, potentially form alternate snRNP structures. In reality, only one of the two pea U1 snRNAs (U1.2) that vary at these critical nucleotides has a sequence which predictably binds U1-specific proteins *in vitro*.

The majority of the sequence variations in the pea U1 snRNAs are localized in stem II, in positions similar to the U1 snRNA variants identified in *Xenopus*, mice and other plants (15,21,39,41). In vertebrates, stem/loop II is clearly involved in binding the U1-A protein, and sequences within the loop have been directly implicated in sequence-specific binding of U1-A (23,25). Variations in stem II of the pea U1 snRNA variants result in alternate stem II structures with substantially different free energies (Fig. 3). Thus, the differences in stem II structures potentially stabilize or destabilize *in vivo* interactions with a plant protein equivalent to the vertebrate U1-A protein, as observed for the mouse U1a (embryonic) and U1b (adult) snRNA variants (26,41). The stabilities of the pea U1 snRNA variants in stem II are similar to the stabilities that we have calculated for the differentially expressed mouse U1a and U1b snRNA variants (U1a1 = -16.7; U1b1 = -24.1; 41). This level of secondary structural variation in the mouse U1 snRNA variants results in the differential binding of snRNP and/or alternate snRNP proteins (26).

The first seventy nucleotides of the U2 snRNAs are phylogenetically conserved in animals, yeast and the pea U2 snRNA variants (21). Stem I appears to be conserved in U2 snRNAs because the 5' half of the stem is indispensable for formation of the pre-splicing complex (50). The nucleotides within the mammalian U2 snRNA that are implicated in branch site binding, GUAGUA, (50,56) are conserved even though the intron lariat structure is only inferred in plants and the highly variant branch site sequence described for plant introns (CYRAY; 57) shows limited complementarity to the conserved GUAGUA sequence present in plant U2 snRNAs.

Nucleotides in stem II and adjacent to the branch site binding region, which may be involved in forming alternate secondary structures in animal and yeast U2 snRNAs are identical in the pea variants sequenced here (Fig. 4). Six of the eight nucleotides in the Sm-antigen binding site are invariant. Most variations in the pea U2 snRNA variants occur in stems III and IV and create a set of alternate stem III structures (Fig. 4) containing single or double nucleotide bulges. The four additional nucleotides present in the U2 snRNA genomic sequence (42) include complementary bases which can form an extended stem III with an alternate loop sequence, or, if the tetraloop sequence is conserved for stability these sequences create an enlarged bulge in stem III. Base changes in the U2 variants include both transition and compensatory changes which maintain the helical structures in stems III and IV.

The terminal loop of stem IV is absolutely conserved in our pea U2 snRNA variants and nine of eleven positions are maintained in other plant U2 snRNAs (16,18,43). The U2-A'

and U2-B" polypeptides appear to interact with this loop, even though five of eleven positions vary in mammalian systems, and more weakly with Sm proteins situated at the Sm binding site (21,58).

Because the U4 snRNAs in plant nuclei are nearly identical in length to U1 snRNAs (Fig. 1), plant U4 snRNAs have been difficult to characterize. The pea U4 snRNA variants that we have described here (Fig. 5) are nearly identical in their primary and secondary structures to the broad bean U4A and U4B snRNAs (12). As in most organisms (21), the 5' halves of the pea U4 snRNA variants are significantly more conserved than the 3' halves which contain nineteen variations in the last 62/64 nucleotides. Regions defined by base pairing to U6 snRNA, and participation in *in vitro* splicing reactions, are phylogenetically conserved in the pea U4 snRNA variants and other organisms. In particular, the 5' loop is conserved at most positions in plant and animal U4 snRNAs (2,21) possibly because it provides a site for the binding of snRNP polypeptides as suggested by RNAase H cleavage experiments (54).

Nearly all of the U4 snRNA variations occur in the central and 3' stems which appear to be masked with snRNP proteins in the U4/U6 snRNP complex (R. Luhrmann, communication). The central stem variations observed in the pea U4 snRNA variants (Fig. 5), occur in the region containing all of the broad bean U4 snRNA differences (12) and the sequence/length variations in other organisms (21). Nearly all of the variations in the central stem of the pea U4 snRNAs represent compensatory changes or transitions between alternate base pairs, which maintain the structure of this stem. In contrast, the 3' terminal stem in the pea U4 snRNA variants has few variations; those that occur do not substantially affect the length or base pairing interactions of the stem.

Most of the variations that occur in the pea U5 snRNAs reported here and by Krol et al. (9) are localized in the 5' terminal stem loop A and in loop D on the 3' terminal stem. Only two of these regions of extreme variability, loop A and loop D, correspond to regions of high variability in human U5 snRNA variants (Sontheimer and Steitz, communication). The 5' terminal stem sequences are particularly variant in plant, but not mammalian U5 snRNAs. Although the 3' terminal stem sequences vary in the pea U5 snRNAs, the length of the stem is well conserved. The single stranded regions, including the Sm binding site, are also well conserved in the plant variants. By far the most highly conserved region of U5 snRNA occurs in loop C which is invariant in nine out of eleven positions in plants and animals (Fig. 6; 21). Chemical modification experiments with human U5 snRNA and snRNPs reveal that loops A and C are accessible to base modifying reagents in the U5 snRNP particle (59). The high degree of conservation in the sequences of loop C suggests that they represent a protein recognition sequence. Correspondingly, the lack of conservation in loop A sequences suggests that they are not involved in protein recognition.

In conclusion, we have cloned and sequenced numerous variants for the U1, U2, U4, and U5 snRNAs expressed in pea nuclei, thus, defining the minimum complexity of dicot snRNA populations. In addition, this represents the first report documenting the diversity of U1 and U2 snRNA sequence variants expressed in plant nuclei. Comparison of the pea snRNA variants with snRNAs from other organisms indicates that, although numerous variants of U1, U2, U4, and U5 snRNAs exist, the variants maintain functionally conserved elements involved in RNA:RNA interactions with the 5' splice site and

branch point sequences. The sequences diverge in specific regions of their secondary structure which have been implicated in the binding of snRNP polypeptides. Further analysis has demonstrated that some of the U1, U2, U4 and U5 snRNA variants described in this paper are developmentally regulated in pea nuclei (Hanley and Schuler, submitted) and potentially represent snRNAs required for the splicing of developmentally regulated transcripts which, in some cases, possess extremely heterogenous splice sites.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Science Foundation DCB-84-02792 and GM39025 from the National Institutes of Health. We are especially grateful to Dr. Andrew McCullough and Dr. Van Anderson for their extensive help in completing this manuscript. We would like to thank Dr. Reinhard Luhrmann for generously supplying the anti-m₃^{2,2,7}G antiserum, Dr. JoAnn Wise for providing some of the oligodeoxyribonucleotides used in this project and Dr. Daniel Alexander for providing the pCGN 1703 vector.

REFERENCES

1. Busch, H., Reddy, R., Rothblum, L. and Choi, Y.C. (1982) *Ann. Rev. Biochem.*, **51**, 617-654.
2. Reddy, R. and Busch, H. (1988) In Birnstiel, M.L. (ed.) *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp 1-37.
3. Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Kramer, A., Frendewey, D., and Keller, W. (1988) In Birnstiel, M.L. (ed.) *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 115-154.
4. Maniatis, T. and Reed, R. (1987) *Nature*, **325**, 673-678.
5. Guthrie, C. (1988) In Birnstiel, M.L. (ed.) *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 196-211.
6. Dahlberg, J.E. and Lund, E. (1988) In Birnstiel, M.L. (ed.) *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer-Verlag, Berlin, pp. 38-70.
7. Egeland, D.B., Sturtevant, A.P. and Schuler, M.A. (1989) *The Plant Cell*, **1**, 633-643.
8. Tollervey, D. (1987) *J. Mol. Biol.*, **196**, 355-361.
9. Krol, A., Ebel, J.-P., Rinke, T. and Luhrmann, R. (1983) *Nucl. Acid Res.*, **11**, 8583-8594.
10. Kiss, T., Toth, M. and Solymosy, F. (1985) *Eur. J. Biochem.*, **152**, 259-266.
11. Skuzeski, J.M. and Jendrisak, J.J. (1985) *Plant Mol. Biol.*, **4**, 181-193.
12. Kiss, T., Jakab, G., Antal, M., Palfi, Z., Hegyi, H., Kis, M. and Solymosy, F. (1988) *Nucl. Acids Res.*, **16**, 5407-5426.
13. van Santen, V., Swain, W. and Spritz, R.A. (1988) *Nucl. Acids Res.*, **16**, 4176.
14. van Santen, V. and Spritz, R. (1987b) *Proc. Natl. Acad. Sci. USA*, **84**, 9094-9098.
15. Abel, S., Kiss, T. and Solymosy, F. (1989) *Nucl. Acids Res.*, **17**, 6319-6337.
16. Brown, J.W.S. and Waugh, R. (1989) *Nucl. Acids Res.*, **17**, 8991-9001.
17. Vankan, P., Edoh, D. and Filipowicz, W. (1988) *Nucl. Acid Res.*, **16**, 10425-10440.
18. Vankan, P. and Filipowicz, W. (1988) *EMBO J.*, **7**, 791-799.
19. Vankan, P. and Filipowicz, W. (1989) *EMBO J.*, **8**, 3875-3882.
20. Kiss, T., Abel, S. and Solymosy, F. (1989) *Plant Mol. Biol.*, **12**, 709-711.
21. Guthrie, C. and Patterson, B. (1988) *Annual Rev. Genetics*, **22**, 387-419.
22. Query, C.C., Bentley, R.C. and Keene, J.D. (1989) *Mol. Cell. Biol.*, **9**, 4872-4881.
23. Scherly, D., Boelens, W., van Venrooij, J., Dathan, N.A., Hamm, J. and Mattaj, I.W. (1989) *EMBO J.*, **8**, 4163-4170.
24. Hamm, J., van Santen, V.L., Spritz, R.A. and Mattaj, I.W. (1988) *Mol. Cell Biol.*, **8**, 4787-4791.
25. Hamm, J., Dathan, N.A., Scherly, D. and Mattaj, I.W. (1990) *EMBO J.*, **9**, 1237-1244.
26. Bach, M., Krol, A. and Luhrmann, R. (1990) *Nucl. Acids Res.*, **18**, 449-457.
27. Brow, D.A. and Guthrie, C. (1988) *Nature*, **334**, 213-281.

28. Gallagher, T.F. and Ellis, R.J. (1982) *EMBO J.*, **1**, 1493–1498.
29. Adams, D.S. and Herrera, R.J. (1987) *Comp. Biochem. Physiol.*, **88B**, 415–420.
30. England, T.E., Bruce, A.G. and Uhlenbeck, O.C. (1980) *Methods in Enzymology*, **65**, 65–74.
31. Wallace, R.B. and Miyada, C.G. (1987) *Meth. Enzymol.*, **152**, 432–442.
32. Taniguchi, T., Palmiera, M., and Weissmann, C. (1978) *Nature*, **274**, 223–228.
33. Alexander, D.C. (1987) *Meth. Enzymol.*, **154**, 41–64.
34. Vogeli, G. and Kaytes, P.S. (1987) *Meth. Enzymol.*, **152**, 407–415.
35. Zuker, M. and Stiegler, P. (1981) *Nucl. Acids Res.*, **9**, 133–148.
36. Mount, S.M. and Steitz, J.A. (1981) *Nucl. Acids Res.*, **9**, 6351–6368.
37. Porter, G., Brennwald, P. and Wise, J.A. (1990) *Mol. Cell. Biol.*, **10**, 2874–2881.
38. Sturtevant, A.P. and Zielinski, R.E. (1990) Manuscript submitted.
39. Forbes, D.J., Kirschner, M.W., Caput, D., Dahlberg, J.E. and Lund, E. (1984) *Cell*, **38**, 681–689.
40. Lund, E. (1988) *Nucl. Acids Res.*, **16**, 5813–5825.
41. Lund, E., Kahan, B. and Dahlberg, J.E. (1985) *Science*, **229**, 1271–1274.
42. Hanley, B.A. and Schuler, M.A. (1989) *Nucl. Acids Res.*, **17**, 10106.
43. Kiss, T., Antal, M. and Solymosy, F. (1987) *Nucl. Acids Res.*, **15**, 1332.
44. Hanley, B.A. and Schuler, M.A. (1988) *Nucl. Acids Res.*, **16**, 7159–7176.
45. Goodall, G.J. and Filipowicz, W. (1989) *Cell*, **58**, 473–483.
46. Brown, J.W.S., Feix, G. and Frendeway, D. (1986) *EMBO J.*, **5**, 2749–2758.
47. Wiebauer, K., Herrero, J.-J., and Filipowicz, W. (1988) *Mol. Cell Biol.*, **8**, 2042–2051.
48. vanSanten, V. and Spritz, R. (1987a) *Gene*, **56**, 253–265.
49. Mount, S.M., Petterson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983) *Cell*, **33**, 509–518.
50. Black, D.L., Chabot, B. and Steitz, J.A. (1985) *Cell*, **42**, 737–750.
51. Krainer, A.R. and Maniatis, T. (1985) *Cell*, **42**, 725–736.
52. Zhuang, Y. and Weiner, A.M. (1986) *Cell*, **46**, 827–835.
53. Bindereif, A. and Green, M.R. (1987) *EMBO J.*, **6**, 2415–2424.
54. Black, D.L. and Steitz, J.A. (1986) *Cell*, **46**, 697–704.
55. Korf, G.M., Botros, I.W. and Stumph, W.E. (1988) *Mol. Cell Biol.*, **8**, 5566–5569.
56. Parker, R., Siliciano, P.G. and Guthrie, C. (1987) *Cell*, **49**, 229–239.
57. Brown, J.W.S. (1986) *Nucl. Acids Res.*, **14**, 9549–9559.
58. Pan, Z.-Q. and Prives, C. (1989) *Genes and Devel.*, **3**, 1887–1898.
59. Black, D.L. and Pinto, A.L. (1989) *Mol. Cell Biol.*, **9**, 3350–3359.