Nucleic Acids Research, 1999, Vol. 27, No. 23 **4577–4584**

Concerted regulation and molecular evolution of the duplicated *SNRPB′/B* and *SNRPN* loci

Todd A. Gray, Megan J. Smithwick[1], Melanie A. Schaldach, Diana L. Martone, Jennifer A. Marshall Graves[1], John R. McCarrey[2] and Robert D. Nicholls*

Department of Genetics, Case Western Reserve University School of Medicine, and Center for Human Genetics, University Hospitals of Cleveland, 10900 Euclid Avenue, Cleveland, OH 44106-4955, USA, [1]Department of Genetics, La Trobe University, Melbourne, Victoria, 3083 Australia and [2]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX 78228, USA

Received August 16, 1999; Revised October 11, 1999; Accepted October 22, 1999 DDBJ/EMBL/GenBank accession nos+

ABSTRACT

The human small nuclear ribonucleoprotein *SNRPB′/B* gene is alternatively spliced to produce the SmB or SmB′ spliceosomal core proteins. An ancestral duplication gave rise to the closely related *SNRPN* paralog whose protein product, SmN, replaces SmB′/B in brain. However, the precise evolutionary and functional relationship between these loci has not been clear. Genomic, cDNA and protein analyses presented here in chicken, two marsupials (South American opossum and tammar wallaby), and hedgehog, suggest that the vertebrate ancestral locus produced the SmB′ isoform. Interestingly, three eutherians exhibit radically distinct splice choice expression profiles, producing either exclusively SmB in mouse, both SmB and SmB′ in human, or exclusively SmB′ in hedgehog. The human *SNRPB′/B* locus is biallelically unmethylated, unlike the imprinted *SNRPN* locus which is unmethylated only on the expressed paternal allele. Western analysis demonstrates that a compensatory feedback loop dramatically upregulates SmB′/B levels in response to the loss of SmN in Prader–Willi syndrome brain tissue, potentially reducing the phenotypic severity of this syndrome. These findings imply that these two genes encoding small nuclear ribonucleoprotein components are subject to dosage compensation. Therefore, a more global regulatory network may govern the maintenance of stoichiometric levels of spliceosomal components and may constrain their evolution.

INTRODUCTION

The excision of introns from the pre-mRNA of the vast majority of eukaryotic transcripts requires the action of a series of ribonucleoprotein (RNP) complexes comprising the spliceosome (1). Each of the constituent RNP complexes is centered around a small RNA molecule for which that particular complex is named (e.g. U1, U2, U4, U5 or U6). This RNA serves as a scaffold for a set of 'core' proteins present in each complex, as well as additional complex-specific proteins. Many of the proteins present in these small nuclear RNP (snRNP) complexes share a common epitope, designated Sm, that is recognized by autoantibodies produced in patients with systemic lupus erythematosus (SLE) (2). In humans, eight core proteins have been described: SmB, SmB′, SmD1, SmD2, SmD3, SmE, SmF and SmG (1).

SmB and SmB′ represent isoforms encoded by alternatively spliced mRNAs from a single gene, called *SNRPB* (3–5). The SmB′/B proteins are expressed robustly in all tissues except for postnatal brain where they are replaced by SmN (6,7). SmN is encoded by the imprinted and paternally expressed *SNRPN* gene in chromosome 15q11–q13, and may play a role in the imprinted Prader–Willi syndrome (PWS) (8). A high degree of sequence similarity strongly suggests that the *SNRPN* and *SNRPB* loci are related by a gene duplication event (9,10). Recently, we have shown that the *SNRPN* locus produces a bicistronic transcript that encodes SmN as well as an independent protein termed SNURF (11). Furthermore, microdeletions of the promoter and first exon of the *SNURF–SNRPN* locus occur in one class of PWS mutation, called an 'imprinting mutation'. Imprinting mutations lead to a failure to establish the paternal imprint throughout the chromosome 15q11–q13 imprinted domain. These small deletions therefore define an imprinting center (IC) that putatively controls the establishment of imprinting in the male and female germline for all genes throughout a 2 Mb domain in 15q11–q13 (8,12). The complex gene structure of this locus (IC–*SNURF–SNRPN*) implies that the expression of the encoded gene products may require precise regulation. Similar detailed studies on the structure and expression of *SNRPB* have not been performed.

SmB was originally described in the rodent (13). Subsequently, the slightly larger SmB′ was identified in human tissues in addition to SmB (14,15). The characterization of the *SNRPB′/B* cDNAs suggested that they represent alternative splicing of the

*To whom correspondence should be addressed. Tel: +1 216 368 3331; Fax: +1 216 368 3432; Email: rxn19@po.cwru.edu

+AF134822–AF134834, AF176323–AF176326

final exon of the same locus (3), which was verified by intronic sequence analysis (4,5). Since SmB′ had only been defined in human, but SmB in human and mouse, it was hypothesized that SmB represents the ancestral form of this protein (16). In order to better define the evolutionary pathway giving rise to the *SNRPB′/B/N* gene family we have characterized the orthologous *SNRPB* loci from key phylogenetic branch points. We show that evolutionarily distant eutherian (hedgehog), metatherian (South American opossum and tammar wallaby) and avian (chicken) species all express *SNRPB′* indicating that in the vertebrate lineage, *SNRPB′*, not *SNRPB*, represents the ancestral archetypal gene. We also show that a strong compensatory feedback mechanism exists to modulate the levels of SmB with the loss of SmN protein in PWS brain tissue. This suggests that these closely related paralogs are functionally equivalent and that this feedback mechanism ensures that adequate levels of SmB′/B and SmN are maintained within the cell.

MATERIALS AND METHODS

Human *SNRPB′/B* genomic locus

A bacterial artificial chromosome (BAC) clone spanning the human *SNRPB′/B* gene was identified by reiterative PCR amplification of arrayed BAC pools (Research Genetics, Birmingham, AL). The primers used were RN743 (5′-GCCACAGAGTATGGAAGTAGCTCC-3′) and RN744 (5′-CCTGTAAGGAAACCAGACAATCCC-3′). *Hin*dIII or *Bam*HI fragments were subcloned into pZERO-2.1 (Invitrogen, Carlsbad, CA) for sequence and fine-structure restriction analyses. Sequence was determined by standard dye-terminator chemistries (Sequinet, Fort Collins, CO, or Cleveland Genomics, Cleveland, OH).

DNA methylation analysis was performed on 10 µg of genomic DNA prepared from lymphoblastoid cell lines from normal individuals. DNA was digested to completion with either *Hin*dIII or *Hin*dIII plus *Sac*II, electrophoresed and Southern blotted to Genescreen (DuPont, Boston, MA). The membrane was crosslinked, and hybridized at 60°C in ExpressHyb (Clontech, Palo Alto, CA) with a ^{32}P-labeled (Rediprime; Amersham, Arlington Heights, IL) PCR-amplified (RN956; 5′-AGGGTGATGGGAAATGAGGAC-3′, and RN884 5′-CGGTTCTGATGGCTCTGATAC-3′) fragment corresponding to the *SNRPB′/B* proximal promoter and first exon. The filter was washed in 0.1× SSC/0.1% SDS at 60°C and used to expose autoradiographic film.

cDNA isolation and analysis

Lambda ZAP cDNA libraries from chicken pigmented retinal epithelium (courtesy of Dr Sue Semple-Rowland, University of Florida; 17) or hedgehog liver (generously provided by Dr Richard Lawn, Stanford University; 18) were screened by hybridization with full-length ^{32}P-labeled human *SNRPN* cDNA until phagemid purity was achieved. The pBluescript derivatives were excised (Stratagene, LaJolla, CA) and used for sequence analysis. Aliquots of a testis lambda ZAP cDNA library from the South American opossum *Monodelphis domestica* (J.R.McCarrey, unpublished) were 5′ and 3′ RACE-amplified using conserved *SNRPN*/*SNRPB* primers (RN957, 5′-ACATGAACCTGATCCTGTGTGAC-3′ for 5′ RACE; RN958, 5′-CCTGTGGAGTCATCACCTGTTG-3′ for 3′ RACE) paired with vector primers (SK, 5′-CGCTCTAGAACTAGTGGATC-3′ for 5′ RACE; T7, 5′-TAATACGACTCACTATAGGG-3′ for 3′ RACE). The recovered products were gel purified (Qiagen, Valencia, CA) and cloned into pCR2.1 (Invitrogen, Carlsbad, CA) for sequence analysis. A tammar wallaby (*Macropus eugenii*) pouch young cDNA λZAPII library (Clontech Laboratories, Inc., Sydney, Australia) was screened by hybridization with human *SNRPN* cDNA. 5′ and 3′ RACE was performed as above using species-specific primer *snrp2* (5′-GTGGCCCCTCAACGGTCATT-3′) for 5′ RACE and primer *snrpb3′* (5′-TGCCGGTGTCCCAATGCCCCAG-3′) for 3′ RACE. Amplified RACE products were cloned into pGEM-T Easy (Promega, Madison, WI) or pCR2.1 as above.

Sequences for human and mouse *SNRPB′/B* were compiled from multiple dbEST entries, as well as independent sequencing of dbEST clones AA478087 (human *SNRPB′*) and AA259557 (mouse *Snrpb*). All sequence alignments were performed using the Clustal algorithm in the MacVector 6.0 analysis package.

Intron analysis

Genomic sequences homologous to the sixth intron of the human *SNRPB′/B* locus were recovered from hedgehog, opossum, wallaby and chicken by PCR amplification using primers derived from exons 6 and 7 as determined by cDNA sequence. For hedgehog, an aliquot of a genomic lambda library (provided by Dr Richard Lawn; 18) was amplified with primers RN928 (5′-CCGAGGAACTCCAATGGGC-3′) and RN978 (5′-CAGGCAGGGAGCTGAGGAGG-3′). Opossum genomic DNA was prepared from *M.domestica* liver and amplified with primers RN973 (5′-CCWGGCATGAGACCACCACC-3′; W = A,T) and RN971 (5′-GTGGCAAGCATCTGAGCAAAG-3′). Tammar wallaby liver genomic DNA was amplified with primers SnrpB3′end (F) (5′-GAGGGGCACCAATGGGCATT-3′) and SnrpB3′end(R) (5′-CCAGGTGGCAAACATAGAAAGAATATC-3′). Chicken genomic DNA was prepared from DT40 cells and amplified with RN973 and RN972 (5′-TATTAGTTCTTCTCACAGTTCTAAG-3′). All amplified products were cloned into pCR2.1 (Invitrogen, Carlsbad, CA) prior to sequence analysis. Human and mouse intron 6 sequences were obtained from GenBank entries HSSMB67 (accession no. X59746) and MMSMB67 (accession no. X59747), respectively. Human *SNRPN* intron 9 was recovered by amplification of human YAC 11H11(19) with primers RN644 (5′-TGGTATGAGACCACCCATGGG-3′) and RN562 (5′-GCAGAATGAGGGAACAAAAAGCTC-3′). Sequences were determined by dye-termination (Sequinet or Cleveland Genomics).

Western analysis

Human brain tissues were procured from the Miami Brain and Tissue Bank for Developmental Disorders (which is funded under NICHD contract #NO1-HD-8-3284). PWS tissue samples were derived from patients whose diagnosis had been confirmed by standard (8,12) *SNRPN* DNA methylation diagnostic analyses. Mouse brain was obtained from internally maintained colonies, as were *M.domestica* tissues. All tissues were cryopulverized and extracted with RIPA buffer supplemented with 1 mM dithiothreitol and a protease inhibitor cocktail (Calbiochem, LaJolla, CA) then clarified by centrifugation. Approximately 10 µg of clarified protein was electrophoresed in a 12.5% (37.5:1)

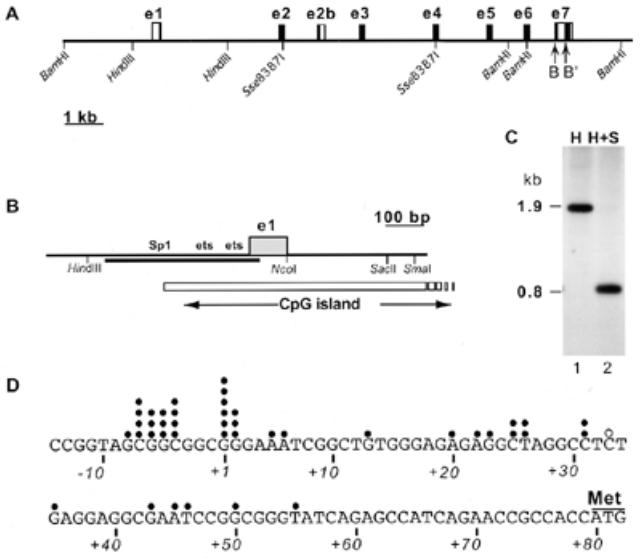

Figure 1. Genomic structure of the human *SNRPB′/B* locus. (**A**) Exon/intron organization and partial restriction map (not all *Hin*dIII restriction sites are shown). The locus consists of seven exons (e1–e7), spanning ~10 kb (open box, untranslated; filled box, translated). An additional exon, exon 2b, is occasionally used but creates a truncated protein product as indicated by its partial shading in the schematic. Splicing into the 5′-most acceptor of exon 7 (B, arrow) produces the *SNRPB* transcript and the SmB protein, whereas splicing into the more 3′ acceptor (B′, arrow) produces the *SNRPB′* mRNA and SmB′ protein. (**B**) Features of the promoter and first intron. Restriction sites for the methylation-sensitive *Sac*II and *Sma*I enzymes are embedded in a CpG island (open box) that extends well into the first intron, and an *Nco*I site includes the initiating methionine, the final codon of the first exon (e1). Potential *cis*-elements conforming to consensus binding sites for transcription factors Sp1 and ets-1 are shown in their relative positions. Black bar indicates the hybridization probe used in methylation analysis. (**C**) Methylation analysis of the *SNRPB′/B* CpG island. Genomic DNA prepared from a normal lymphoblastoid cell line was digested to completion with *Hin*dIII (H) or *Hin*dIII plus *Sac*II (H+S) enzymes, Southern transferred and hybridized with a probe shown in (B). A band of 1.9 kb corresponds to a *Hin*dIII fragment, while a 0.8 kb fragment indicates cleavage with the methylation-sensitive *Sac*II enzyme. (**D**) Distribution of putative transcription initiation sites. Filled circles represent 42 dbEST clone entries. Open circle is human *SNRPB* GenBank entry HSSNRNPB (accession no. X17567). The initiating methionine is also indicated. Human *SNRPB′/B* genomic sequences have been deposited in GenBank: AF134822, AF134823, AF134824 and AF134825. After these data were obtained, the Sanger Centre Chromosome 20 Mapping Group has deposited the complete nucleotide sequence of a BAC encompassing the *SNRPB′/B* genomic locus (accession no. AL049650).

tricine–SDS–polyacrylamide gel and electrotransferred in Bjerrum-Schafer/Nielson buffer (48 mM Tris, 39 mM glycine, 0.0375% SDS, 20% methanol) to ImmobilonP (Millipore, Bedford, MA). Sm proteins were detected using anti-Sm antibodies (Immunovision, Springdale, AR) diluted 1:250 in blocking buffer (phosphate buffered saline, 0.1% Tween-20, 0.5% casein) followed by horseradish peroxidase-linked goat anti-human secondary antibodies (Sigma, St Louis, MO) and chemiluminescent visualization (Amersham, Arlington Heights, IL).

RESULTS

The BAC 345N4 containing the human *SNRPB′/B* locus was identified by sequential PCR amplification of arrayed library pools. Sequence analysis of subclones derived from the BAC demonstrated a 7 exon structure (Fig. 1A and Table 1; this confirms and corrects data in ref. 10). Exons 1–7 of the *SNRPB′* locus are homologous to exons 4–10 of the *SNRPN* gene, with conservation of intron positions. Restriction mapping and Southern analysis shows that the *SNRPB′/B* locus is compact, spanning ~10 kb, with introns of 3.5, 1.9, 1.9, 0.6, 0.3 and 0.9 kb (Fig. 1A). The *SNRPB′/B* gene has been mapped by GeneMap '98 to human chromosome 20p13 (stSG10927) (20).

Table 1. Exon/intron boundaries of the human *SNRPB′/B* gene

5′ Sequence	Exon	3′ Sequence
tccggtagcggcggcGGGAA	1	CCATG**gt**aaggagggcacag
catttctccttg**ag**ACGGT	2	ATCAA**gt**gaggagtggactg
atgttctccccac**ag**GTGTG	2b	TCAGG**gt**aagtgctcagctc
taaacctctctttt**ag**GCCAA	3	AAGAT**gt**aagtcagagcagg
ttgtttatttttc**ag**ACTGG	4	AACAG**gt**gaggagttgggga
cttatgtcctctt**ag**GTGAT	5	TCCAG**gt**gaggagcccataa
gcttttctcttg**ag**GCATG	6	GCGAG**gt**aagtgcctgcagt
ttactcttactt**ag**GCCTT	7(B)	
tggtttcccttac**ag**GGCCC	7(B′)	AATTGcacctatgtatcttt
yyyyyyyyyyyny**ag**GNNNN	cons	NNMAG**gt**ragt

Sequences flanking each exon are shown in lower case text, with exon sequences displayed in upper case text. Sequences 5′ represent splice acceptors, except for exon 1 which corresponds to promoter sequence. Sequences 3′ are splice donors, except exon 7, which shows genomic sequence following the site of polyadenylation. An alternative exon 2b is occasionally included between exons 2 and 3. Exon 7 has two active splice acceptors in human, producing the B and B′ subtypes. Consensus (cons) splice acceptor and donor sites are shown below, with conserved intronic boundary 'ag' and 'gt' dinucleotides in bold text in all intron/exon boundaries.

The promoter and first intron bear the hallmarks of a typical CpG island, and include sites for the methyl-sensitive restriction enzymes *Sac*II and *Sma*I (Fig. 1B). A fragment corresponding to sequences from –291 to +73 in the first exon was used to probe a Southern blot of genomic DNA from normal human lymphoblasts. Autoradiographic patterns observed in DNA cut with *Hin*dIII alone (1.9 kb) or *Hin*dIII and *Sac*II enzymes (0.8 kb) show that both alleles of the *SNRPB′/B* locus are unmethylated at this site (Fig. 1C). The 5′ flanking sequence, corresponding to the putative proximal promoter, contains multiple *cis*-elements identified as consensus binding sites by Transfac (21) database analysis. Noteworthy examples are an Sp1 site (–125[GGCCCCGCCTCTG]–113), and a pair of inverted ets-1 sites at (–62[ACCGGAAGTA]–53) and (–19[GTAAAGGCCA]–10) (Fig. 1D); the functional significance of these and other potential binding sites awaits future empirical studies. One characteristic of promoters lacking a TATA-box is that the site of transcription initiation is quite heterogeneous. A survey of 43 dbEST clones that include the 5′ UTR of the *SNRPB′/B* shows a wide range of putative transcript initiation sites with a mode of 7 at one guanine (Fig. 1D), though no identifiable major start site is evident. This guanine is designated as +1 throughout this report.

To examine the evolutionary origin of the *SNRPB′/B* gene family, we isolated and sequenced orthologs from extant species representing pivotal fissures in vertebrate evolution. Open reading frames (ORFs) derived from the cDNA sequences from hedgehog, opossum, wallaby and chicken were then compared with existing sequences for mouse SmB, human SmB, SmB′ and SmN (Fig. 2). The putative protein products from hedgehog, opossum, wallaby and chicken are all

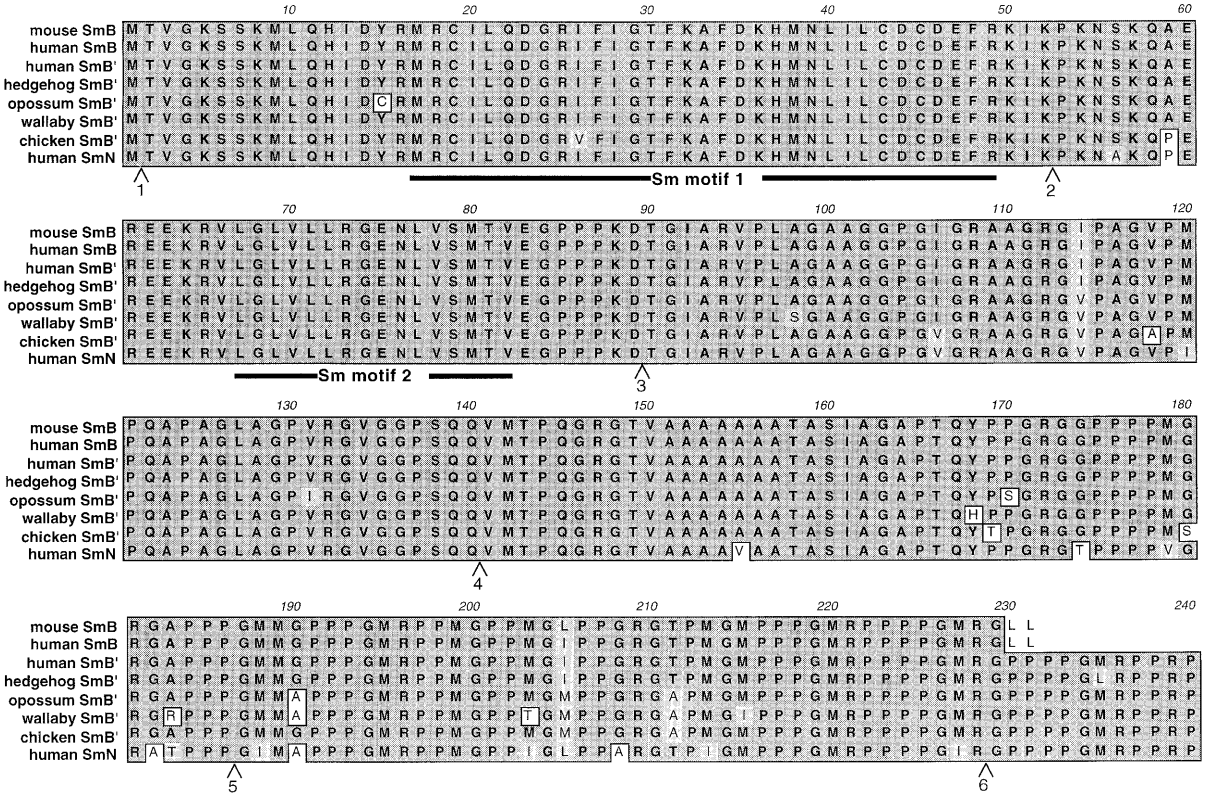

Figure 2. Alignment of vertebrate SmB′/B polypeptide sequences. Putative amino acid sequences, derived from their respective cDNA sequences are aligned along with human SmN. Residues conserved with the consensus are shown on a dark gray field, conservative changes on a light gray background and non-conservative changes are unshaded. Sm motifs 1 and 2 (22) are indicated by a black bar. Numbered chevrons appearing below the alignment correspond to intron placement in the human *SNRPB′/B* genomic locus. Hedgehog, opossum, wallaby and chicken *SNRPB′* cDNA sequences have been deposited in GenBank: AF134826, AF134827, AF176323 and AF134830, respectively.

240 amino acids in length, with C-termini reminiscent of human SmB′ and SmN (Fig. 2). The degree of amino acid identity with human SmB′ is >99, 97, 96 and 96% in hedgehog, opossum, wallaby and chicken, respectively. The respective values in comparison to human SmN are 93, 94, 93 and 92% identity. These data clearly identify the four newly cloned sequences as vertebrate *SNRPB′* orthologs.

These sequences served to derive a consensus vertebrate SmB′ sequence that was used to search databases to identify invertebrate homologs. Sequences were compiled from GenBank, dbEST and genomic sequencing entries for homologs from *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana* and *Schizosaccharomyces pombe*. These putative products were aligned with the projected ancestral vertebrate SmB′ sequence (supplementary data Fig. S1). The *D.melanogaster*, *C.elegans*, *A.thaliana* and *S.pombe* loci are 59, 53, 46 and 42% identical with human SmB′, respectively, with notable conservation of the two Sm motifs (22) and proline clusters in the central and C-terminal regions.

An alternative splice form that incorporates an additional novel exon, exon 2b, was identified in opossum, wallaby and human (Fig. 1A, Table 1, and supplementary data Fig. S2). Interestingly, by PCR amplification of the opossum and wallaby liver cDNA libraries, this splice form appears to be nearly as abundant as the normal splice form, although only two of 52 human dbEST clones have been identified that contain exon 2b (accession nos AA127120 and AA380572), indicating a less prominent role in humans. This splice form possesses multiple stop codons in all reading frames, and could only yield a truncated polypeptide of ~8 kDa. The truncated protein would contain only the first of two Sm motifs (supplementary data Fig. S2), both of which are required for efficiently interacting with SmD (22), suggesting that this polypeptide would lack spliceosomal function. Nevertheless, the degree of nucleotide conservation between human, opossum, and wallaby (supplementary data Fig. S2), and its exceptionally high representation in marsupials (data not shown), indicates a possible regulatory or non-coding role for this novel exon.

Since the sole determinant of the production of the B or B′ derivatives in mouse and human relies on the presence and activity of functional splice acceptors (5), phylogenetic analysis of these critical sequences may define the evolutionary event giving rise to these two derivatives. PCR primers were designed in each species corresponding to sequences projected to be in exons 6 and 7 based on the human genomic structure. Genomic DNA sources along with these primers were then used to amplify the intervening intron from each of the species (Materials and Methods). Sequence data obtained from the cloned products were analyzed for splice donor and splice acceptor potential for the *SNRPB* and *SNRPB′* cDNA variants (Fig. 3). All discernible exon 6 splice donor junctions (Fig. 3A)

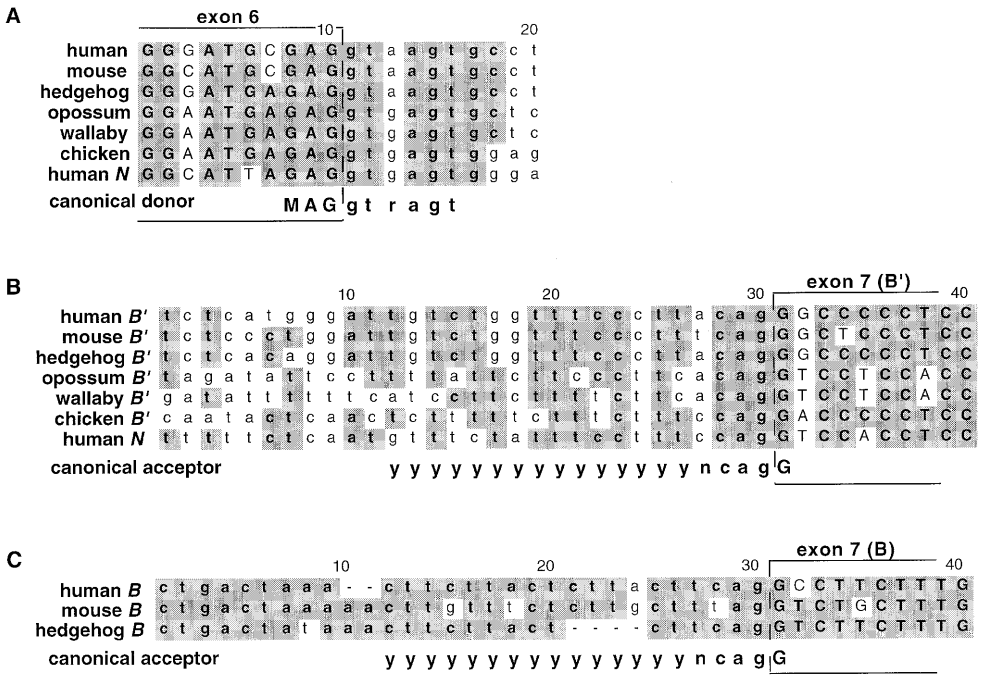

Figure 3. Alignment of intron 6 splice donor and acceptor boundaries. (**A**) Splice donor alignment. Sequences derived from *SNRPB′/B* exon 6 (or *SNRPN* exon 9) are shown in upper case, while adjacent intron sequences are lower case. Conserved nucleotides are shaded, and a canonical donor site is shown below (M = C or A, R = A or G, Y = C or T). (**B**) Potential splice acceptor sites corresponding to the *SNRPB′* splice form. Human *SNRPN* intron 9/exon 10 is also shown for comparative purposes. Intron, exon and consensus styles are applied as in (A). (**C**) Potential splice acceptor sites corresponding to the *SNRPB* splice form. Intron, exon and consensus styles are applied as in (A). Gaps inserted to improve the alignment are indicated (–). Hedgehog, opossum, wallaby and chicken *SNRPB′* intron 6 and human *SNRPN* intron 9 sequences have been deposited in GenBank: AF134829, AF134833, AF134833, AF134831, AF176325, AF176326 and AF134832, respectively.

and exon 7 splice acceptor junctions of *SNRPB′* (Fig. 3B) conform to consensus sequences as shown. Possible splice acceptors (a polypyrimidine tract followed by an AG dinucleotide) corresponding to the human and mouse *SNRPB*-form (Fig. 3C) were not detectable by visual inspection or several alignment algorithms in opossum, wallaby or chicken (data not shown). Similarly, the small 88 nt intron 9 of human *SNRPN* contains no sequence equivalent to the *SNRPB* splice acceptor site that is 146 nt upstream of the *SNRPB′* acceptor in the human.

In contrast, the *SNRPB*-like junction appears to be conserved in hedgehog, although there is a 4-bp deletion close to the expected position of a *SNRPB* splice acceptor site (Fig. 3C). PCR amplification of an aliquot of a hedgehog kidney cDNA library with exon 6 and exon 7 primers, however, produced a single product consistent with a *SNRPB′*-form splice product, with no evidence of a *SNRPB*-form splicing event (data not shown).

To ensure that the DNA and cDNA analyses accurately predict the protein products, anti-Sm western blots were performed with a broad tissue spectrum from *Monodelphis*. While human and mouse brain produce detectable SmB and SmN (the small amount of human SmB′ is obscured by the large amount of SmN in the brain; see below and ref. 11), adult opossum brain and liver, as well as total pup head or body, do not produce the SmB-form (Fig. 4A). All opossum tissues produce robust amounts of a 29 kDa protein consistent with SmB′ (Fig. 4A). Although this analysis unequivocally demonstrates an absence of opossum SmB, it is possible that the observed 29 kDa band represents a composite of SmB′ and a putative opossum SmN protein. It should be noted that these proteins typically exhibit retarded migration rates in SDS–PAGE, relative to their calculated molecular masses (23.7, 24.6 and 24.6 kDa, for human SmB, SmB′ and SmN, respectively), perhaps due to their extremely basic isoelectric points (pI~11) (3).

One interesting observation that may shed light on the highly conserved nature of the *SNRPB′/N* gene family is that the encoded proteins appear to be coordinately regulated. Normal human brain extracts subjected to anti-Sm western analysis reveal three discernible proteins with M_r's consistent with proteins of 26 kDa for SmB, 28.5 kDa for SmB′ and 29 kDa for SmN (Fig. 4B, lanes 1 and 4). PWS patients, however, do not have a transcriptionally active *SNRPN* locus (8), and therefore do not produce the SmN protein (Fig. 4B, lanes 2, 3, 5 and 6). The SmB and SmB′ proteins in the PWS extracts are markedly increased (Fig. 4B, lanes 2 and 3); while accurate quantitation is precluded due to possible unequal affinity of the anti-Sm antibodies for SmN and SmB′/B, the bolstered SmB′/B levels appear to be similar to total SmB′/B/N in normal tissue. The size of the upregulated SmB′ and SmB isoforms correspond to identically migrating proteins in normal brain (Fig. 4) and in lymphoblasts where *SNRPB′/B* are highly expressed relative to *SNRPN* (data not shown). Extended exposures indicate that other Sm proteins (e.g., SmC, SmD) are not upregulated, testifying to the specificity of the phenomenon (Fig. 4B, lanes 4–6).

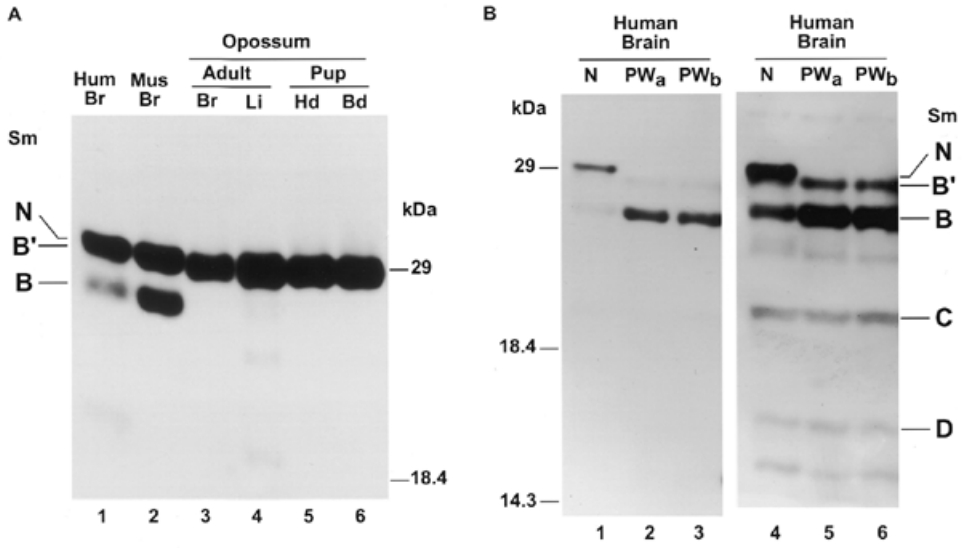

Figure 4. Anti-Sm western analyses. (**A**) Human (Hum), mouse (Mus), and opossum tissues showing SmN, SmB′ and SmB proteins. Tissue extracts were derived from brain (Br), liver (Li), head (Hd) or body (Bd). Protein molecular weight standard migration distances are indicated in the right margin. (**B**) Normal (N) and Prader–Willi syndrome (PW_a and PW_b) brain Sm protein profiles. Lanes 1–3 and 4–6 correspond to short and long chemiluminescent exposures, respectively. Sm family members SmN, SmB′, SmB, SmC and SmD are indicated in the right margin, while molecular weight standard migration distances are shown in the left margin.

DISCUSSION

With this work we have shown that the ancestral gene of the human *SNRPB′/B* locus is of the *SNRPB′*-form. We propose that duplication of an ancestral *SNRPB′* locus likely gave rise to the *SNRPN* gene prior to the emergence of the *SNRPB* splice form. We base the latter conclusion on our observation that the *SNRPN* gene has a final intron splice structure similar to the putative ancestral *SNRPB′* gene, although we cannot rule out that *SNRPN* was originally formed after origin of the *SNRPB*-like splice choice and subsequently deleted this intronic region prior to the eutherian radiation 60–80 million years ago (23). As expected for chromosomally dispersed duplicated genes of this age, the two genes have diverged in their mRNA expression regulatory pathways, with *SNRPB′/B* being ubiquitously expressed and likely non-imprinted. In contrast, the derivative *SNRPN* locus is most highly expressed in postnatal brain and transcribed only from the paternally inherited allele. Nevertheless, PWS brain does not express SmN but strikingly upregulates SmB′/B, revealing a compensatory feedback mechanism. Therefore, we suggest that these two genes are evolving in concert under selective pressure to preserve this regulatory pathway that maintains stoichiometric levels of spliceosomal components.

It had previously been theorized that *SNRPB* was the ancestral form of the *SNRPB′/B* locus and that *SNRPB′* may be a human-specific variant (16). By a combination of cDNA, genomic sequence and protein analyses from several distant vertebrate species, we have shown here that *SNRPB′* is the archetypal *SNRPB* gene (Fig. 5). From a molecular standpoint, the derivation of *SNRPB* from *SNRPB′* required the introduction of an alternative splice acceptor in the final intron, corresponding to intron 6 of the human gene. Curiously, even though all three

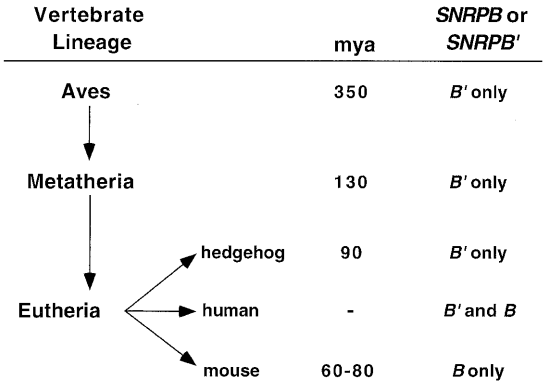

Figure 5. Evolution of the vertebrate *SNRPB′/B* locus. Aves, as represented by the chicken, metatheria, as represented by the South American opossum and tammar wallaby, and eutherian divisions are shown, along with their respective *SNRPB′/B* locus classifications as determined by cDNA, protein and genomic sequence analyses. The approximate number of millions of years (mya) since the last common ancestor with humans (18,23) is indicated.

eutherian species studied possess the basic elements required for functional splice acceptors at both *SNRPB*-like and *SNRPB′*-like junctions, all conceivable splice permutations are observed; human produces both *SNRPB* and *SNRPB′* (i.e. both junctions are active), while mouse produces only *Snrpb* (i.e. from the more 5′ junction) and hedgehog only *SNRPB′* (i.e. from the more 3′ junction). It has been proposed that the discrepant human and mouse *SNRPB′/B* splicing profiles result from subtle contextual differences creating permissive or restrictive stem–loop structures at the splice acceptors in these species (5).

Alternatively, deletions of 2 and 4 nt in the human and hedgehog, respectively, near the *SNRPB*-like splice acceptor may reduce its efficacy in these species.

The conservation of the orthologous vertebrate *SNRPB′* genes is notably high suggesting that the encoded protein is functionally constrained. This is not surprising since *SNRPB′* is involved in a fundamental cellular process and makes contacts with other constituents of the spliceosome (22,24). The paralogous SmN is also very well conserved and is itself 100% identical in human and mouse. The high degree of sequence similarity between SmB′/B and SmN suggests that they may be functionally redundant, a hypothesis which is supported by protein data showing a compensatory feed-back loop to control SmB′/B/N levels (see below). Nevertheless, the question of functional redundancy will need to be addressed using murine knockout models of the *Snrpb* gene in combination with *Snrpn* ablation.

As described in the Introduction, the establishment and regulation of imprinting in chromosome 15q11–q13 intimately involves the *SNURF–SNRPN* locus (11). Consequently, it is of interest to discern the evolutionary progression of this complex locus to determine the origin of the two ORFs and of imprinting. The highly conserved sequence and gene structure indicate that *SNRPN* arose from *SNRPB′* by a gene duplication event, but it is unknown whether this event included regulatory elements as well. Sequence comparison of the *SNRPB′/B* first exon and promoter with the promoter of the *SNURF–SNRPN* locus reveals no detectable similarity using several alignment algorithms (data not shown). However, since *SNRPB′/B* exon 1 corresponds to exon 4 of the *SNRPN* locus, a more judicious comparison would involve the third intron and fourth exon of the *SNRPN* locus with the *SNRPB′/B* promoter region. No homology is detected 5′ of the initiating methionine in these two regions. Furthermore, whereas *SNRPN* is regulated by genomic imprinting (25), we have shown that both alleles of the *SNRPB′/B* promoter/CpG island are unmethylated, and therefore that this gene is unlikely to be imprinted. Taken together, these data suggest that either the 5′ flanking region was lost in the duplication event or that the *SNRPN* sequences subsequently diverged beyond recognition and with *SNPRN* becoming differently regulated in the process.

Despite having different transcriptional (and possibly translational; 11) mechanisms, the functions of *SNRPB′/B* and *SNRPN* appear to be undergoing evolution in concert, and hence are functionally restricted from divergence. First, the *SNRPB′* to *SNRPN* duplication occurred over 90 million years ago, but the SmN coding sequence has remained nearly identical to the ancestral SmB′ sequence. Second, there is an apparent regulatory interaction of the two genes with reciprocal transcriptional and translational activities in most tissues, which is particularly pronounced in brain. Indeed, SmN is most highly expressed in postnatal brain where it replaces SmB′/B in spliceosomal complexes (6,10,26). PWS patients do not transcribe the *SNRPN* locus resulting in an absence of *SNRPN* mRNA in PWS tissues (8) thereby precluding the presence of SmN or its isoforms (this report; 11). We have shown here that while brain extracts from PWS individuals lack SmN, there is a significant upregulation of SmB′/B protein levels in what appears to be a compensatory feedback loop. A similar phenomenon of much lower maganitude occurs in *Snrpn*-ablated mice (27) and a PWS deletion mouse model (11). The lesser degree of SmB upregulation in mouse may reflect the relatively comparable SmN and SmB levels in wild type mouse brain, requiring a much less pronounced compensation in the mutant mice than in PWS patients.

It is not known whether the upregulation of SmB′/B protein in postnatal brain occurs at the transcriptional, post-transcriptional, translational or post-translational level. Such a feedback mechanism must be capable of first assessing the aggregate level of SmN/B′/B and then adjusting the steady-state protein levels accordingly. The common recognition of SmN and SmB′/B in this process may mandate that these proteins retain sequence or structural identity. Moreover, the subsequent adjustment in protein level may require element(s) that are shared between SmN and SmB′/B, or their genetic loci. These may be *cis*-elements in the DNA if the mechanism of this regulation is transcriptional, mRNA elements if it is post-transcriptional or translational, or protein motifs if it is post-translational. Therefore, we propose that the evolution of these two loci is constrained by the regulatory feedback mechanism linking them. While the molecular mechanism of this feedback loop has not yet been determined, these data form the foundation of future studies that will focus on the mechanism of regulation. Whether a reciprocal dosage compensation mechanism exists awaits empirical demonstration by modulation of SmB expression level (e.g., ablation or overexpression).

It is noteworthy that the RNA components of the spliceosomal complexes also exhibit a regulatory compensation mechanism wherein ectopic overexpression of a U1 or U2 snRNA is balanced by a reduction in expression of the endogenous RNA (28,29). Since a compensatory effect is displayed by at least two of the RNAs and two of the core proteins, there may be similar evolutionarily conserved mechanisms regulating expression of the remaining spliceosomal components as well. Recent crystallographic work defining the interactions of Sm proteins with each other suggests that the seven core Sm proteins form a highly ordered heptameric ring-like structure that likely surrounds the Sm RNA motif of each U1, U2, U4 and U5 snRNA (24,30). Therefore, we propose that this putative core Sm protein ring may be sensitive to any stoichiometric imbalance thereby requiring stringently regulated steady-state levels of each component.

SUPPLEMENTARY MATERIAL

See Supplementary Material available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Richard Lawn for graciously supplying the hedgehog cDNA and genomic libraries, and Dr Sue Semple-Rowland for generously providing the chicken cDNA library. We are grateful to Dr Greg Matera for his many helpful comments. This work was funded by National Institutes of Health grants HD31491 and HD36079 (R.D.N.), by March of Dimes Birth Defects Foundation grant #6-FY99-390, by donations from anonymous PWSA members (R.D.N.), and by National Institutes of Health N.R.S.A. Postdoctoral Fellowship GM18306 (T.A.G.).

REFERENCES

1. Luhrmann,R., Kastner,B. and Bach,M. (1990) *Biochim. Biophys. Acta*, **1087**, 265–292.
2. Craft,J. (1992) *Rheum. Dis. Clin. North Am.*, **18**, 311–335.
3. van Dam,A., Winkel,I., Zijlstra-Baalbergen,J., Smeenk,R. and Cuypers,H.T. (1989) *EMBO J.*, **8**, 3853–3860.
4. Chu,J.L. and Elkon,K.B. (1991) *Gene*, **97**, 311–312.
5. Griffith,A.J., Schmauss,C. and Craft,J. (1992) *Gene*, **114**, 195–201.
6. McAllister,G., Roby-Shemkovitz,A., Amara,S.G. and Lerner,M.R. (1989) *EMBO J.*, **8**, 1177–1181.
7. Sharpe,N.G., Williams,D.G. and Latchman,D.S. (1990) *Mol. Cell. Biol.*, **10**, 6817–6820.
8. Nicholls,R.D., Saitoh,S. and Horsthemke,B. (1998) *Trends Genet.*, **14**, 194–200.
9. Schmauss,C. and Lerner,M.R. (1990) *J. Biol. Chem.*, **265**, 10733–10739.
10. Schmauss,C., Brines,M.L. and Lerner,M.R. (1992) *J. Biol. Chem.*, **267**, 8521–8529.
11. Gray,T.A., Saitoh,S. and Nicholls,R.D. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 5616–5621.
12. Ohta,T., Gray,T.A., Rogan,P.K., Buiting,K., Gabriel,J.M., Saitoh,S., Muralidhar,B., Bilienska,B., Krajewska-Walasek,M., Driscoll,D.J. *et al.* (1999) *Am. J. Hum. Genet.*, **64**, 397–413.
13. Lerner,M.R. and Steitz,J.A. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 5495–5499.
14. Hinterberger,M., Pettersson,I. and Steitz,J.A. (1983) *J. Biol. Chem.*, **258**, 2604–2613.
15. Matter,L., Schopfer,K., Wilhelm,J.A., Nyffenegger,T., Parisot,R.F. and De Robertis,E.M. (1982) *Arthritis Rheum.*, **25**, 1278–1283.
16. Brunet,C., Quan,T. and Craft,J. (1993) *Gene*, **124**, 269–273.
17. Semple-Rowland,S.L., Lee,N.R., Van Hooser,J.P., Palczewski,K. and Baehr,W. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 1271–1276.
18. Lawn,R.M., Boonmark,N.W., Schwartz,K., Lindahl,G.E., Wade,D.P., Byrne,C.D., Fong,K.J., Meer,K. and Patthy,L. (1995) *J. Biol. Chem.*, **270**, 24004–24009.
19. Christian,S.L., Bhatt,N.K., Martin,S.A., Sutcliffe,J.S., Kubota,T., Huang,B., Mutirangura,A., Chinault,A.C., Beaudet,A.L. and Ledbetter,D.H. (1998) *Genome Res.*, **8**, 146–157.
20. Deloukas,P., Schuler,G.D., Gyapay,G., Beasley,E.M., Soderlund,C., Rodriguez-Tome,P., Hui,L., Matise,T.C., McKusick,K.B., Beckmann,J.S. *et al.* (1998) *Science*, **282**, 744–746.
21. Wingender,E., Dietze,P., Karas,H. and Knuppel,R. (1996) *Nucleic Acids Res.*, **24**, 238–241.
22. Hermann,H., Fabrizio,P., Raker,V.A., Foulaki,K., Hornig,H., Brahms,H. and Luhrmann,R. (1995) *EMBO J.*, **14**, 2076–2088.
23. Andersson,L., Archibald,A., Ashburner,M., Audun,S., Barendse,W., Bitgood,J., Bottema,C., Broad,T., Brown,S., Burt,D. *et al.* (1996) *Mamm. Genome*, **7**, 717–734.
24. Kambach,C., Walke,S., Young,R., Avis,J.M., de la Fortelle,E., Raker,V.A., Luhrmann,R., Li,J. and Nagai,K. (1999) *Cell*, **96**, 375–387.
25. Glenn,C.C., Saitoh,S., Jong,M.T.C., Filbrandt,M.M., Surti,U., Driscoll,D.J. and Nicholls,R.D. (1996) *Am. J. Hum. Genet.*, **58**, 335–346.
26. McAllister,G., Amara,S.G. and Lerner,M.R. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5296–5300.
27. Yang,T., Adamson,T.E., Resnick,J.L., Leff,S., Wevrick,R., Francke,U., Jenkins,N.A., Copeland,N.G. and Brannan,C.I. (1998) *Nature Genet.*, **19**, 25–31.
28. Mangin,M., Ares,M.,Jr and Weiner,A.M. (1985) *Science*, **229**, 272–275.
29. Frey,M.R., Bailey,A.D., Weiner,A.M. and Matera,A.G. (1999) *Curr. Biol.*, **9**, 126–135.
30. Lamond,A.I. (1999) *Nature*, **397**, 655–656.