Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 5616–5621, May 1999
Genetics

An imprinted, mammalian bicistronic transcript encodes two independent proteins

TODD A. GRAY*, SHINJI SAITOH*†, AND ROBERT D. NICHOLLS*‡

*Department of Genetics, Case Western Reserve University School of Medicine and Center for Human Genetics, University Hospitals of Cleveland, OH 44106; and †Department of Pediatrics, Hokkaido University School of Medicine, Kita 15 Nishi 7, Kita-ku, Sapporo 060, Japan

Communicated by Liane B. Russell, Oak Ridge National Laboratory, Oak Ridge, TN, March 23, 1999 (received for review February 4, 1999)

ABSTRACT **Polycistronic transcripts are common in prokaryotes but rare in eukaryotes. Phylogenetic analysis of the *SNRPN* (SmN) mRNA in five eutherian mammals reveals a second highly conserved coding sequence, termed *SNURF* (*SNRPN upstream reading frame*). The vast majority of nucleotide substitutions in *SNURF* occur in the wobble codon position, providing strong evolutionary evidence for selection for protein-coding function. Because *SNURF–SNRPN* maps to human chromosome 15q11–q13 and is paternally expressed, each cistron is a candidate for a role in the imprinted Prader-Willi syndrome (PWS) and PWS mouse models. *SNURF* encodes a highly basic 71-aa protein that is nuclear-localized (as is SmN). Because *SNURF* is the only protein-coding sequence within the imprinting regulatory region in 15q11–q13, it may have provided the original selection for imprinting in this domain. Whereas some human tissues express a minor *SNURF*-only transcript, mouse tissues express only the bicistronic *Snurf–Snrpn* transcript. We show that both *SNURF* and *SNRPN* are translated in normal, but not PWS, human, and mouse tissues and cell lines. These findings identify SNURF as a protein that is produced along with SmN from a bicistronic transcript; polycistronic mRNAs therefore are encoded in mammalian genomes where they may form functional operons.**

Imprinted genes carry a parental-specific gametic mark that results in differential expression of the maternally and paternally derived alleles during development of the mammalian organism (1). Among several genetic diseases that arise from abnormal imprinted gene expression, Prader-Willi syndrome (PWS) is a developmental and neurobehavioral disorder that results from a loss of function of paternally inherited genes in chromosome 15q11–q13 (2, 3). Several PWS candidate genes have been identified in this interval that are expressed from the paternally inherited allele (3). However, the identity and number of genes involved in the etiology of PWS is unknown. The best characterized of these is the *SNRPN* (*S*mall *N*uclear *R*ibonucleo*p*rotein *N*) gene. The human and mouse *SNRPN* genes have 10 exons, with exons 4–10 encoding the SmN spliceosomal protein (3–9).

This genetic locus appears to have a key role in the cis regulation of imprinting throughout chromosome 15q11–q13, because microdeletions that remove the 5′ end of the *SNRPN* gene are found in a subset of PWS patients in which the paternally inherited 15q11–q13 is otherwise intact, but inappropriately bears a maternal epigenotype (3, 6, 10). These microdeletions in PWS genetically define one component of a bipartite imprinting center (IC) that is required for proper germ-line imprint establishment over the entire 2-Mb imprinted domain (3, 5, 6). The homologous IC region in mice is functionally equivalent, as demonstrated by targeted mutagenesis of the 5′ end of the murine *Snrpn* locus (11). Although the mechanism of this process remains obscure, the immediate proximity of the IC with the *SNRPN* locus implies that the two are functionally connected.

To further elucidate the role of the *SNRPN* gene in imprinting and in PWS, we have examined the structure, function, and evolution of this locus. Mammalian phylogenetic analyses of *SNRPN* cDNAs have identified a second, evolutionarily constrained ORF, *SNURF* (a Human Gene Nomenclature Committee-approved symbol), that lies upstream of the *SNRPN* ORF. We have verified, by using immunodetection, that SNURF and SmN are both translated in various human and mouse tissues and cell lines. The unusual bicistronic gene structure of the *SNURF–SNRPN* locus adds yet another layer of complexity to this locus that is potentially central to PWS and the evolution of imprinting in chromosome 15q11–q13.

MATERIALS AND METHODS

Sequence Analysis. Cow and rabbit *SNURF* cDNAs were recovered in two phases by first amplifying the intercistronic region by reverse transcription–PCR of brain poly(A)$^+$ mRNA (CLONTECH) with primers corresponding to human and mouse conserved *SNURF* (RN625: 5′-GGCATTCTTAGCTGAGACACC-3′) and *SNRPN* (RN626: 5′-ACAATCACAGAGGATCAAATTCAT-3′) coding sequence. Products were gel-purified and extracted (Qiagen, Chatsworth, CA) and ligated into TA cloning vector pCR2.1 (Invitrogen). Insert sequences were determined (Macromolecular Resources, Fort Collins, CO). Species-specific intercistronic primers (rabbit, 5′-AAACAGTAGATGGAGCCTTGATATTC-3′; cow, 5′-CAAATAGTAGATGGAGCCTTGGTG-3′) were then used in 5′ rapid amplification of cDNA ends to recover the full-length *SNURF* coding sequence. A λ ZAP bovine brain cDNA library (Stratagene) pool was used as template in two successive rounds of amplification by first using the cow-specific intergenic primer and a T3 primer, and then the same cow-specific primer and an SK primer (Stratagene). Rabbit brain Poly(A)$^+$ mRNA (CLONTECH) was reverse-transcribed and Marathon linker-adapted (CLONTECH), then used for two rounds of amplification with the rabbit-specific intergenic primer and linker primers as detailed by CLONTECH. Fragments were isolated, cloned, and sequenced as above.

Mouse and rat cDNA sequences were derived from overlapping dbEST sequences, and mouse cDNA clones. Human cDNA sequence was obtained from GenBank accession no. U41303. Compiled *SNURF* cDNA and protein sequences used for phylogenetic analysis have been deposited in the GenBank

Abbreviations: ES, embryonic stem; GFP, green fluorescent protein; IC, imprinting center; PWS, Prader-Willi syndrome.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF101040–AF101044).

‡To whom reprint requests should be addressed at: Department of Genetics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4955. e-mail: rxn19@po.cwru.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

PNAS is available online at www.pnas.org.

database: cow, AF101040; rat, AF101041; mouse, AF101042; rabbit, AF101043; human, AF101044. Genomic sequence for human exon 3b was derived from GenBank accession no. U41384, whereas the equivalent mouse sequence was obtained from a BAC subclone of mouse 129/Sv DNA (7). All sequence alignments were performed with the CLUSTAL algorithm (MacVector Version 6.0).

Northern Blot and Reverse Transcription–PCR. Multitissue Northern blots from human and mouse (CLONTECH) were probed with *SNURF* exons 1–3, or separately with *SNRPN* exons 4–10, as per the manufacturer's protocol. Poly(A)$^+$ mRNA from human brain, skeletal, and heart muscle (CLONTECH), or prepared from normal or PWS lymphoblasts, were reverse transcribed and subjected to amplification with primers corresponding to exon 1 (RN683: 5′-TGACGCATCTGTCTGA-3′) and exon 3b (RN982: 5′-TGAATAATATTTTTATTACATTGT-3′). Products were gel-isolated, cloned, and verified by sequence analysis.

Expression Constructs and Antibody Generation. mAbs were raised against purified bacterially expressed SNURF protein. Restriction sites flanking the *SNURF* coding sequence were incorporated by using PCR (RN292, 5′-GACGCCATGGAGCGGGCAAGG-3′; RN293, 5′-CAAGATCTCCACCTCTTGGTGTTC-3′), providing a source for subsequent expression constructs. The entire 213-bp human *SNURF* coding sequence was transferred into a bacterial expression vector (pET32a, Novagen). SNURF expression was induced with isopropyl β-D-thiogalactoside (IPTG), and thioredoxin-$(His)_6$-SNURF fusion protein was purified under denaturing conditions by nitrilotriacetic acid–nickel chromatography by using standard procedures (Qiagen, Chatsworth, CA). Anti-SNURF mAbs were generated against this immunogen (Monoclonal Antibody Core Facility of Case Western Reserve University). Hybridoma supernatants were assayed by ELISA with purified glutathione *S*-transferase–SNURF (Amersham Pharmacia). The reactive epitopes were assigned to the N or C termini by Western analyses with glutathione *S*-transferase fusions with either the full-length 71-aa protein, or a truncated derivative containing only the N-terminal 36 amino acids.

The human SNURF coding sequence was modified to include an in-frame C-terminal hexahistidine motif and a translation initiation sequence (RN447: 5′-CTAGACTCGAGAGGAGGGTTTTTAC-3′ and RN448: 5′-CATGGTAAAAACCCTCCTCTCGAGT-3′) then transferred to the eukaryotic expression vector pREP9 (Invitrogen). 293c18 cells (American Type Culture Collection) were transiently transfected by using Lipofectamine (Life Sciences, Arlington Heights, IL) as per the manufacturer's instructions, and extracts were prepared 48 hours later.

For the intracellular localization of green fluorescent protein (GFP)-tagged SNURF, the human SNURF coding sequence, preceded by a translation initiation sequence, was transferred in-frame into pEGFP (CLONTECH) to produce full-length SNURF C-terminally fused to GFP. This SNURF–GFP or the underivatized construct were transiently transfected into HT1080 cells in chamber slides (Lab-tek) by using Lipofectamine, allowed to culture 48 hours, and inspected by fluorescence microscopy.

Protein Analyses. Samples in this study were obtained from the University of Miami Brain and Tissue Bank for Developmental Disorders (which is funded under National Institute of Child Health and Human Development Contract NO1-HD-8-3284). Human lymphoblast lines (Normal, AS139, PWS109, and GM09113) were cultured in RPMI medium 1640 under standard conditions. Embryonic stem (ES) cells were generously provided, free of feeder cells, by B. T. Lamb (Case Western Reserve University). The PWS mouse model has a fortuitous transgene insertion that generated a large deletion of all homologous loci from the PWS chromosomal region (unpublished data). Paternal transmission of this deletion leads to failure to thrive and lethality by postnatal day 7. Animal care was in accordance with institutional guidelines.

Mouse and human tissues were pulverized in liquid nitrogen. Pulverized tissues or tissue culture cells were homogenized at 100 mg/ml in RIPA buffer (0.15 mM NaCl/0.05 mM Tris·HCl, pH 7.2/1% Triton X-100/1% sodium deoxycholate/0.1% SDS) supplemented with 1 mM DTT and a protease inhibitor mixture (Calbiochem). Relative protein concentration and integrity of each extract was verified by SDS/PAGE followed by Coomassie staining.

Western analysis of Sm antigens used ≈10 μg of clarified RIPA extract electrophoresed on a 12.5% (37.5:1) tricine SDS/PAGE and electrotransferred in Bjerrum and Schafer-Nielson buffer (48 mM Tris/39 mM glycine/0.0375% SDS/20% methanol) to poly(vinylidene difluoride) membrane (Immobilon). For Western detection, a 1:250 dilution of anti-Sm antibody (Immunovision) in blocking buffer (PBS/0.1% Tween-20/0.5% casein) was used, followed by horseradish peroxidase-linked goat anti-human IgG (Sigma) and visualization by chemiluminescence (Amersham Pharmacia).

Immunoprecipitations were performed by using standard methods with 0.5 ml of clarified RIPA extract and 10 μl protein A/G (Boehringer Mannheim) preabsorbed with anti-SNURF hybridoma supernatant. The immunoprecipitates were separated by tricine SDS/PAGE, 16.5% (37.5:1), and transferred and detected as for Sm except that the primary anti-SNURF antibody was used at 1:25, and detected with horseradish peroxidase-linked goat anti-mouse IgG (Pierce).

RESULTS

Identification of a Highly Conserved ORF Upstream of *SNRPN*. Dot-plot analysis of the 1.4-kb cDNAs for human and mouse *SNRPN* reveals two distinct regions of significant conservation, suggesting a possible bicistronic structure (Fig. 1*a* and *b*). The highly conserved 720-nt *SNRPN* ORF is seen as a diagonal from approximately nucleotides 450 to 1200 and encodes identical 240-aa SmN proteins in the human and mouse. A second region of homology lies upstream of the SmN ORF, extending from nucleotides 80–300 (Fig. 1*b*) encompassing exons 1–3 (8), and corresponds to the *SNURF* ORF of 71 aa. We extended the phylogenetic analysis upstream of the SmN initiation codon to include five eutherian mammals (Fig. 1*c*). Amino acid identity with human SNURF ranges from 93% to 100%, with most substitutions being quite conservative. Regions of reduced nucleotide identity, including insertions and deletions, occur in the intercistronic region between the two ORFs, as well as in the 5′ and 3′ untranslated regions. Most importantly, analysis of the total number of nucleotide substitutions in the *SNURF*-coding sequence shows that the vast majority occur in the third, or "wobble," position of the codon (Fig. 1*d*), preserving amino acid sequence. Taken together, these data provide strong genetic evidence that the *SNURF–SNRPN* transcript of eutherian species is bicistronic and can encode two independent, evolutionarily conserved proteins.

Northern Blot Analysis of *SNURF–SNRPN* Transcripts. To examine mRNA expression of the bicistronic locus, Northern analyses were performed. Extended autoradiogram exposures identify, in addition to the primary 1.6-kb *SNURF–SNRPN* transcript, an ≈0.5-kb mRNA species that hybridizes with an exon 1–3 probe (Fig. 2*a*) but not exons 4–10 (ref. 8, and data not shown). However, the 0.5-kb transcript is not seen in brain, placenta, lung (Fig. 2*a*), or human lymphoblasts (data not shown). Consistent with a 0.5-kb *SNURF*-only transcript in muscle and kidney (Fig. 2*a*), several dbEST entries (accession nos. AI017249, AA613634, and F18566) were identified from these tissues that correspond to an mRNA including exons 1–3 with an additional alternative 3′ exon (exon 3b) that contains a canonical AATAAA polyadenylation signal (Fig. 2*b*). This *SNURF*-only transcript was recovered by using reverse transcription–PCR from human heart and skeletal muscle and

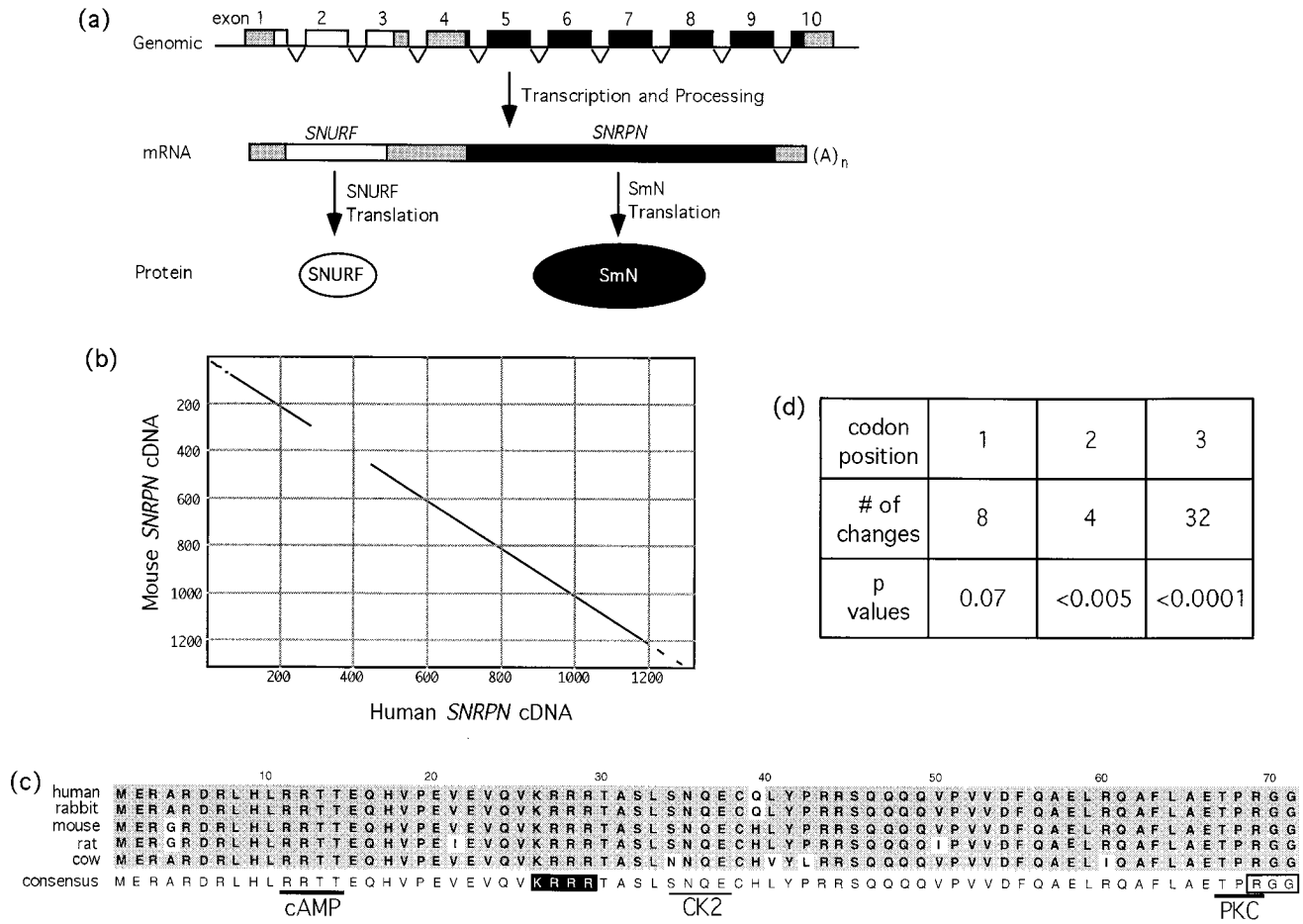

FIG. 1. *SNURF–SNRPN* gene structure and sequence analysis. (*a*) Schematic representation of the bicistronic *SNURF–SNRPN* locus depicting the progression from a contiguous nonoverlapping series of exons to a single transcript with two ORFs, to independent protein products. Open and filled structures correspond to *SNURF* or *SNRPN*, respectively, whereas the shaded regions indicate untranslated sequences. (*b*) Dot-plot analysis of full-length mouse and human *SNRPN* cDNAs; conserved sequences (window = 30, 65% minimum) appear as a diagonal line. (*c*) Amino acid alignment of putative SNURF proteins from five eutherian mammals. The derived consensus is shown below the compilation; residues consistent with the consensus are shaded in gray, changes are unshaded. Potential features appended to the consensus are: a nuclear localization signal (white letters on a black background), a C-terminal RGG motif (boxed), and phosphorylation sites that are absolutely conserved (heavy underline; cAMP, cAMP-dependent kinase; PKC, protein kinase C) or partially conserved (light underline; CK2, casein kinase II). (*d*) Codon position of the nucleotide changes that occur in the *SNURF* coding sequence of the five species examined. *P* values were derived from a χ^2 analysis at each position.

verified by sequencing to confirm its presence in these tissues (data not shown). Exon 3b maps 712 bp downstream of exon 3 and is preceded by a potential lariat branch point and consensus splice acceptor. This result suggests a model in which a putative ancestral *SNURF* gene may have been an autonomous locus that terminated with exon 3b.

Genomic analysis of the homologous region in mice reveals a nucleotide change in the invariant splice acceptor motif and degeneration of the oligopyrimidine tract (ref. 7; Fig. 2*b*). Consistent with this observation, no ≈0.5-kb *Snurf*-only transcript was detected by Northern analysis in any murine tissue (Fig. 2*c*). Furthermore, a primer corresponding to the exon 3b-homologous region in the mouse failed to amplify products in reverse-transcribed mRNA from murine heart muscle (data not shown). Therefore, in mouse tissues, *Snurf* mRNA is only present as part of a larger bicistronic transcript with *Snrpn.*

SNURF and SmN Are Translated in Normal Human and Mouse Tissues and Cell Lines. To demonstrate the presence of SNURF protein *in vivo*, mAbs were generated and used in immunoprecipitations followed by Western analyses. A single band corresponding to SNURF protein is present in extracts derived from normal human and mouse tissues (Fig. 3 *a* and *b*). In human lymphoblasts (lanes 1, 3, and 11), mouse ES cells (lane 4), and mouse brain (lane 5), SNURF migrates at ≈9 kDa, which agrees well with the calculated molecular mass of 8.4 kDa. In contrast, no SNURF protein is detected in lymphoblasts from PWS patients (lanes 2 and 12) or a PWS-deletion mouse model (lane 6), validating the specificity of the antibodies for a protein derived from this locus. Human heart (Fig. 3*b*, lane 7) and skeletal muscle (lane 9) have an ≈11-kDa SNURF species that comigrates with hexahistidine-tagged SNURF expressed in human kidney 293 cells (lane 13), but that is absent in PWS tissue (lanes 8 and 10). Therefore, the native ≈11-kDa form most likely arises from a posttranslational modification, such as phosphorylation (see Fig. 1*c*). These observations clearly verify that SNURF is normally translated *in vivo* as an imprinted and paternally expressed gene that is not expressed in PWS or a mouse model of PWS.

Intriguingly, in spite of high mRNA levels (8), SNURF protein is not detected in normal adult human brain ($n = 2$ females; age 44, 45 years; data not shown), although it is present in mouse brain up to at least 145 days of age (Fig. 3*a*, lane 5; data not shown). Also, SNURF protein was detected in adult skeletal muscle and heart from one, but not a second, normal individual (Fig. 3*b*, lanes 7 and 9); the SmN protein is present in both of these normal individuals, as expected from

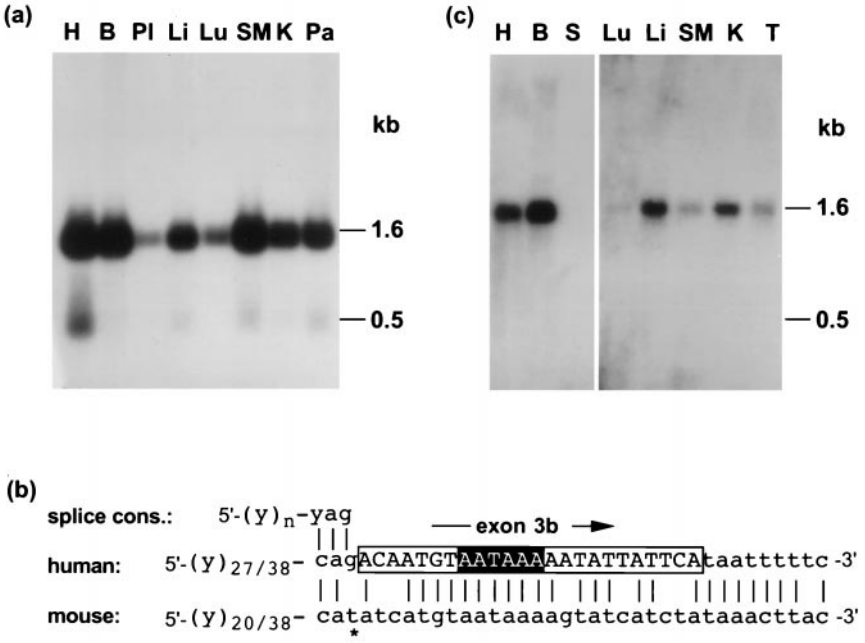

FIG. 2. SNURF-specific mRNA analyses. (*a*) Extended exposure of a human multitissue Northern blot probed with *SNURF* exons 1–3. A major 1.6-kb product is present in all tissues, and a minor 0.5-kb species is seen in a subset of tissues. Only the 1.6-kb product is seen with probes from *SNRPN* exons 4–10 (8). Tissue types are: H, heart; B, brain; Pl, placenta; Lu, lung; Li, liver; SM, skeletal muscle; K, kidney; Pa, pancreas. (*b*) *SNURF* exon 3b. The exon is shown as boxed, capital letters, with flanking genomic sequence shown as lower case. Consensus sequences for polyadenylation (white letters in a black box) and splice acceptor (splice consensus, conformity indicated by a vertical line; y, pyrimidine) also are shown. An alignment of the homologous mouse genomic sequence is shown below; ∗ marks a point mutation of the splice acceptor consensus in mice. (*c*) Extended exposure of a mouse multitissue Northern blot probed with mouse *Snurf* exons 1–3. The 1.6-kb product is present in most tissues, but a 0.5-kb band is not observed in any tissue. Identical results are obtained for probes from *Snrpn* exons 4–10 (data not shown). Tissue types are as for *a*. S, spleen; T, testis.

its up-regulation in postnatal brain (12). Multiple ($n = 7$) normal lymphoblast and fibroblast cell lines translate *SNURF* without exception (data not shown), implying that external factors may influence *SNURF* translation or stability. Identifying the exact parameters governing *SNURF* translation (e.g., age, sex, weight), or detection (e.g., protein half-life, postmortem procedures) must await systematic and comprehensive studies encompassing a large number of individuals.

To demonstrate that both protein products of the bicistronic transcript are translated in the same cell types, human lymphoblast and murine ES cell extracts that contain SNURF protein (see above) were examined for SmN. Three separable Sm species in the 25- to 30-kDa molecular mass range are detected by Western analyses with anti-Sm antibodies (Fig. 4). Human lymphoblasts produce relatively high levels of SmB and SmB′, partially obscuring the SmN protein, which is expressed at low levels in this cell type; nevertheless, discernible levels of SmN are present in normal and Angelman syndrome but not PWS lymphoblasts (Fig. 4*a*, lanes 1–3). Mouse samples, which lack the obscuring SmB′, clearly show the presence of SmN in ES cell extracts (Fig. 4*b*, lane 4) and normal brain (Fig. 4*b*, lane 6) but not PWS mouse model brain (Fig. 4*b*, lane 5). Taken together, these data provide significant evidence that both SNURF and SmN are translated from the same polycistronic mRNA.

SNURF–GFP Intracellular Localization. The predicted SNURF polypeptide sequence (Fig. 1*c*) bears no overt homology to other proteins. SNURF possesses a C-terminal Arg-Gly-Gly (RGG), motif which is a distinctive feature of the bioactive C terminus of ubiquitin and the ubiquitin-like family of proteins (13). However, RGG motifs are also found in a subset of RNA-binding proteins and the very basic nature of SNURF with 17% arginine residues (pI = 10.7) suggests it could interact with RNA (14). Additional studies are needed to determine whether SNURF functions as a ubiquitin-like or RNA-binding protein or has other biochemical functions. Nevertheless, to gain further insight into the potential function for SNURF, the subcellular localization was determined. A nuclear localization (KRRR) motif is conserved in SNURF (Fig. 1*c*), implying that the protein may be partitioned in the nucleus. C-terminally tagged SNURF–GFP is indeed targeted to the nucleus when ectopically expressed, whereas unfused GFP is distributed throughout the cell (Fig. 3*c*).

DISCUSSION

PWS is thought to be a contiguous gene syndrome, although the number and identity of contributing genes is currently undefined (3). A number of genetic aberrations previously pointed to a key role for the *SNRPN* locus in PWS and imprinting in chromosome 15q11–q13 (3, 5, 6, 10, 15, 16). Our finding that two independent proteins (SNURF and SmN) are encoded by a single mRNA renames this as the bicistronic *SNURF–SNRPN* locus and suggests that the *SNURF* cistron may specifically participate in these key roles.

First, rare, paternally inherited, balanced translocations that directly disrupt the *SNURF* portion of *SNURF–SNRPN* have been found in classical PWS patients (15, 16). Because in at least one of these cases, all other imprinted transcripts are expressed (16), and because ablation of the *Snrpn* ORF causes no obvious phenotype in the mouse (11), these translocations are consistent with the loss of *SNURF* expression as a potential critical etiologic factor in PWS. However, two other balanced translocations break distal of *SNURF–SNRPN* in atypical PWS patients (17, 18), indicating that the molecular basis of these balanced translocations may be complex and will require further studies to be understood. Second, the centrally located IC that regulates imprinting of all genes throughout the entire 2-Mb imprinted domain is minimally defined as a 4.3-kb region spanning the promoter and first exon of the *SNURF–SNRPN*

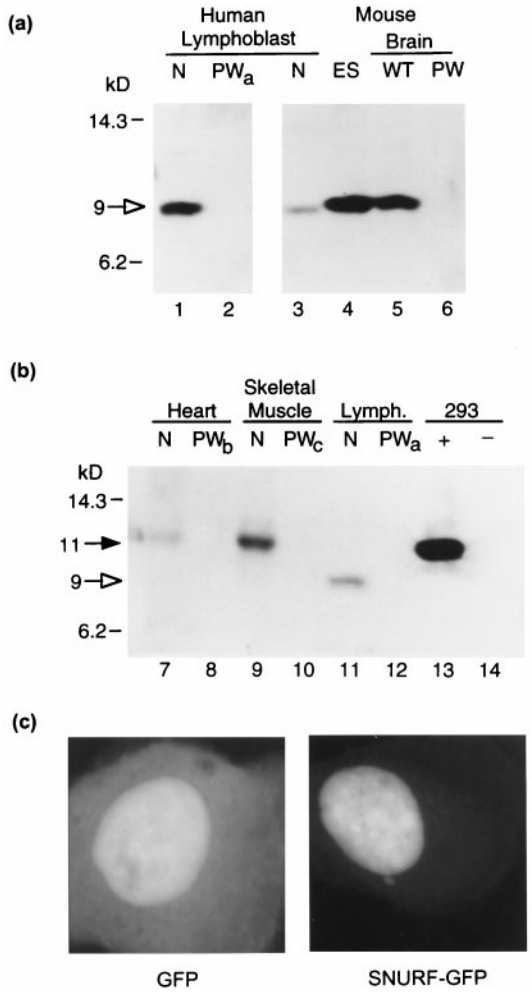

FIG. 3. SNURF protein analyses. (*a*) Representative immunoprecipitation-Western analyses of human and mouse tissues with an ≈9-kDa SNURF form. N, normal lymphoblast; PW_a, PWS109 lymphoblast line; ES, mouse embryonic stem cell line; WT, normal mouse brain; PW, Prader-Willi mouse model. (*b*) Immunoprecipitation–Western analysis of human muscle tissues with an ≈11-kDa SNURF isoform. N, normal heart and skeletal muscle from the same individual (no. 2,144); PW_b, heart muscle from PWS individual (no. 1,199), PW_c, skeletal muscle from PWS individual (no. 1,889). Lymphoblast designations are as in *a*. 293, human kidney 293 cells that have been transiently transfected with SNURF-$(His)_6$ (+), or mock empty vector (−). (*c*) Intracellular localization of ectopically expressed underivatized GFP or GFP fused to the C terminus of SNURF.

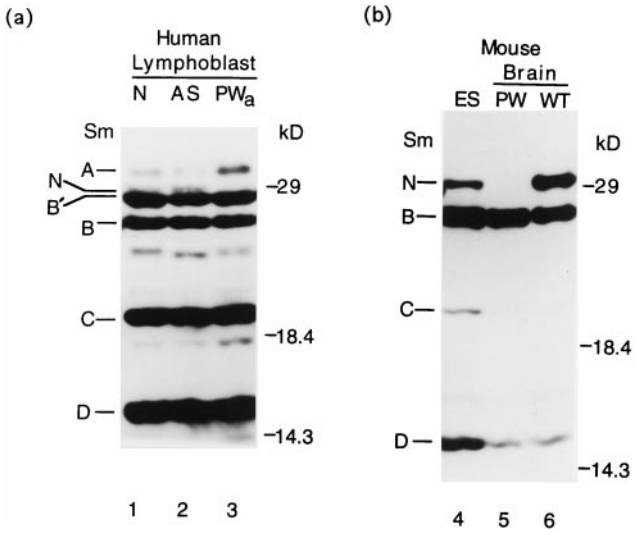

FIG. 4. Western analysis of SmN. (*a*) SmN in human lymphoblasts. N, normal; AS, AS139; PWS, PWS109. (*b*) SmN in mouse tissues. ES, embryonic stem cells; PW, murine PWS model brain extract; WT, normal mouse brain extract. SmA, SmN, SmB′, SmB, SmC, and SmD complex are indicated in the left margins of *a* and *b*.

locus (3, 10). Our analyses in the present work have determined that *SNURF* is the only protein-coding gene in this minimal region, implying an intrinsic relationship between the imprinting regulatory mechanism and the *SNURF* gene.

Further study of the *SNURF* gene is likely to provide insights into the impetus for the evolution of genomic imprinting within the PWS region of chromosome 15q11–q13. The profoundly intimate mammalian mother–offspring relationship is characterized by the direct dependence of the progeny, from conception through weaning, on maternal resources. Paternally expressed genes are postulated to exploit this relationship, endowing the offspring with a competitive advantage for those limited resources, whereas maternally expressed genes abrogate these effects to avoid compromising maternal long-term reproductive fitness (19). PWS neonates (2) and PWS mouse model pups (refs. 11 and 20; unpublished data) have a severe failure-to-thrive phenotype consistent with a growth-promoting role for an absent paternally expressed gene. Because *SNURF* is the only protein-coding gene in the minimal IC region associated with control of imprinting of all paternally expressed genes in 15q11–q13 (see above), we suggest that the imprinting mechanism in this chromosomal region evolved in response to a stringent regulatory requirement of SNURF function. This further implies that loss of SNURF may lead to the observed neonatal failure-to-thrive phenotype. The alternative hypothesis, that the IC evolved at the 5′ end of *SNURF* in response to selection for a distantly located gene in 15q11–q13, does not appear parsimonious with present molecular evolutionary understanding.

Two models can be proposed to explain the evolution of a bicistronic gene structure at this complex locus. The identification of the *SNURF*-only exon 3b in humans, together with its remnants in mice, suggests that an ancestral autonomous *SNURF* gene, consisting of four exons, may once have been present at this locus. In this model, the *SNRPN* exons were generated by a duplication of the highly related *SNRPB′* locus (unpublished data) and fused just downstream of the terminal *SNURF* exon (exon 3b). The terminating exon 3b degenerated thereby allowing transcription into the *SNRPN* exons and the putative *SNRPN* promoter may have been destroyed, or not included, in the initial gene duplication event, or progressively silenced by inactivating mutations. Alternatively, a new ORF-encoding SNURF may have been created at the time of the *SNRPB′*-to-*SNRPN* duplication, similar to a recently proposed mechanism in *Drosophila* (21). An examination of whether a *SNURF*-only or *SNRPN*-only locus exists in extant mammals may allow these two models to be distinguished.

Translation of two proteins from a common bicistronic mRNA may be developmentally or spatially coordinately regulated. Because *SNURF–SNRPN* transcription begins in the preimplantation embryo in both human and mouse (22–24), but *SNRPN* translation is greatly up-regulated in postnatal brain (12), there may be an ontogenic switch from *SNURF* translation at earlier stages. The absence of SNURF protein in brain from two aged individuals and the hypothesized neonatal role of SNURF (see above) are consistent with this model, but further detailed studies are required. Nevertheless, we have shown that both SNURF and SmN are produced in several human and mouse tissues and cell lines; thus, there may also be translational mosaicism at the cellular or ribosomal level, or both proteins may be simultaneously cotranslated from a single mRNA. The ribosome-scanning model of translation

initiation posits that the first AUG in a respectable Kozak consensus context will be the most efficient translation initiator (25). The human *SNURF* translation initiation site (**g**a**cGc**g*AUG***G**) conforms more closely to the Kozak consensus (gccRcc*AUG*G) than does the downstream *SNRPN* ORF (**gc**a**Atc***AUG*a), suggesting that *SNURF* should be translated at least as well as *SNRPN*. SmN could be translated by any of several possible mechanisms (26): (*i*) leaky ribosome scanning, in which only a portion of the scanning ribosomes begin translation at the *SNURF* ORF, while others bypass this and instead begin translation at the *SNRPN* ORF; (*ii*) ribosome reinitiation in which the ribosomes do not dissociate after *SNURF* translation termination, but continue tracking and initiate translation again at the *SNRPN* ORF; or (*iii*) by an internal ribosome entry site mechanism in which a cis element in the intercistronic region recruits ribosomal subunits that begin 3′ scanning and translation of *SNRPN*. Although the 5′ end of the intercistronic region is poorly conserved, the region preceding the *SNRPN* AUG is highly conserved in all five species examined, suggesting a functional role for this sequence at the mRNA level, perhaps as an internal ribosome entry site.

The bicistronic structure of the mammalian *SNURF–SNRPN* transcript is significant because it is atypical of eukaryotic genomes (27). Polycistronic transcripts are well described in prokaryotes, where they commonly encode proteins involved in the same functional pathway, thereby constituting an operon (28). In invertebrates, examples of polycistronic loci include numerous *Caenorhabditis elegans* loci that are reduced to monocistrons by trans-splicing before translation (27), as well as the bicistronic *stoned* locus in *Drosophila* (29). Selective pressure for diminutive genomes acting on mammalian viruses has led to the evolution of a plethora of economical transcriptional and translational mechanisms, including strategies referred to as polycistronic but involving complex alternative splicing, pre-protein proteolytic processing, or multiple sites for translation initiation (30, 31). In mammals, active L1 elements of retroviral origin produce two proteins from a single transcript (32).

Bona fide cellular polycistronic loci have been proposed in vertebrates (27, 33, 34), although these are unsubstantiated with the requisite molecular evidence for translation of two independent proteins from a single transcript. The human and mouse growth/differentiation factor (*GDF1*) gene lies 3′ of another highly conserved ORF (*UOG*) of unknown function on an apparently bicistronic transcript (33). However, to date there is no supportive *in vivo* evidence for translation of a separate *UOG* ORF (33), despite a conserved monocistronic homologue in other species (27). Recently, a transcript containing putative nonoverlapping *MOCS1A* and *MOCS1B* ORFs has been described (34); however, this structure was inferred only from human sequence and the identification of a single mutation in each putative ORF associated with a rare molybdenum metabolic disorder. Given the extremely short in-frame separation of the two ORFs, other mechanisms, such as selenocysteine suppression (35), may act to bypass the putative opal (TGA) stop codon of the 5′ ORF to produce a single, continuous polypeptide. Our demonstration *in vivo* of two independent proteins (SNURF and SmN) produced from a single mRNA represents definitive evidence for a cellular vertebrate polycistronic gene. The mechanism of translational regulation of such genes and the functional implications in terms of mammalian operons are questions of fundamental importance to an understanding of the structure, function, and evolution of the animal genome.

We thank Dr. D. J. Driscoll for patient lymphoblast lines, Dr. B. T. Lamb for mouse ES cells, M. A. Schaldach and N. A. Rebert for technical assistance, and Drs. H. F. Willard, A. G. Matera, and B. T. Lamb for helpful comments. This work was funded by National Institutes of Health Grants HD31491 and HD36079 (R.D.N.), by donations from anonymous Prader–Willi Syndrome Association members (R.D.N.), and by National Institutes of Health N.R.S.A. Postdoctoral Fellowship GM18306 (T.A.G.).

1. Bartolomei, M. S. & Tilghman, S. M. (1997) *Annu. Rev. Genet.* **31,** 493–525.
2. Cassidy, S. B. (1997) *J. Med. Genet.* **34,** 917–923.
3. Nicholls, R. D., Saitoh, S. & Horsthemke, B. (1998) *Trends Genet.* **14,** 194–200.
4. Schmauss, C., Brines, M. L. & Lerner, M. R. (1992) *J. Biol. Chem.* **267,** 8521–8529.
5. Sutcliffe, J. S., Nakao, M., Christian, S., Örstavik, K. H., Tommerup, N., Ledbetter, D. H. & Beaudet, A. L. (1994) *Nat. Genet.* **8,** 52–58.
6. Buiting, K., Saitoh, S., Gross, S., Dittrich, B., Schwartz, S., Nicholls, R. D. & Horsthemke, B. (1995) *Nat. Genet.* **9,** 395–400.
7. Gabriel, J. M., Gray, T. A., Stubbs, L., Saitoh, S., Ohta, T. & Nicholls, R. D. (1998) *Mamm. Genome* **9,** 788–793.
8. Glenn, C. C., Saitoh, S., Jong, M. T., Filbrandt, M. M., Surti, U., Driscoll, D. J. & Nicholls, R. D. (1996) *Am. J. Hum. Genet.* **58,** 335–346.
9. Özçelik, T., Leff, S., Robinson, W., Donlon, T., Lalande, M., Sanjines, E., Schinzel, A. & Francke, U. (1992) *Nat. Genet.* **2,** 265–269.
10. Ohta, T., Gray, T. A., Buiting, K., Gabriel, J. M., Rogan, P. K., Saitoh, S., Deng, G., Ishikawa, I., Weksberg, R., Driscoll, D. J., *et al.* (1999) *Am. J. Hum. Genet.* **64,** 397–413.
11. Yang, T., Adamson, T. E., Resnick, J. L., Leff, S., Wevrick, R., Francke, U., Jenkins, N. A., Copeland, N. G. & Brannan, C. I. (1998) *Nat. Genet.* **19,** 25–31.
12. Grimaldi, K., Horn, D. A., Hudson, L. D., Terenghi, G., Barton, P., Polak, J. M. & Latchman, D. S. (1993) *Dev. Biol.* **156,** 319–323.
13. Johnson, P. R. & Hochstrasser, M. (1997) *Trends Cell Biol.* **7,** 408–413.
14. Siomi, H. & Dreyfuss, G. (1997) *Curr. Opin. Genet. Dev.* **7,** 345–353.
15. Kuslich, C. D., Kobori, J. A., Mohapatra, G., Gregorio-King, C. & Donlon, T. A. (1999) *Am. J. Hum. Genet.* **64,** 70–76.
16. Sun, Y., Nicholls, R. D., Butler, M. G., Saitoh, S., Hainline, B. E. & Palmer, C. G. (1996) *Hum. Mol. Genet.* **5,** 517–524.
17. Conroy, J. M., Grebe, T. A., Becker, L. A., Tsuchiya, K., Nicholls, R. D., Buiting, K., Horsthemke, B., Cassidy, S. B. & Schwartz, S. (1997) *Am. J. Hum. Genet.* **61,** 388–394.
18. Schulze, A., Hansen, C., Skakkebaek, N. E., Brondum-Nielsen, K., Ledbeter, D. H. & Tommerup, N. (1996) *Nat. Genet.* **12,** 452–454.
19. Moore, T. & Haig, D. (1991) *Trends Genet.* **7,** 45–49.
20. Cattanach, B. M., Barr, J. A., Evans, E. P., Burtenshaw, M., Beechey, C. V., Leff, S. E., Brannan, C. I., Copeland, N. G., Jenkins, N. A. & Jones, J. (1992) *Nat. Genet.* **2,** 270–274.
21. Nurminsky, D. I., Nurminskaya, M. V., De Aguiar, D. & Hartl, D. L. (1998) *Nature (London)* **396,** 572–575.
22. Shemer, R., Birger, Y., Riggs, A. D. & Razin, A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 10267–10272.
23. Szabo, P. E. & Mann, J. R. (1995) *Genes Dev.* **9,** 3097–3108.
24. Huntriss, J., Daniels, R., Bolton, V. & Monk, M. (1998) *Am. J. Hum. Genet.* **63,** 1009–1014.
25. Kozak, M. (1996) *Mamm. Genome* **7,** 563–574.
26. Hershey, J. W. B., Mathews, M. B. & Sonenburg, N. (1996) *Translational Control* (Cold Spring Harbor Lab. Press, Plainview, NY).
27. Blumenthal, T. (1998) *BioEssays* **20,** 480–487.
28. Jacob, F. & Monod, J. (1961) *J. Mol. Biol.* **3,** 318–356.
29. Andrews, J., Smith, M., Merakovsky, J., Coulson, M., Hannan, F. & Kelly, L. E. (1996) *Genetics* **143,** 1699–1711.
30. Gupta, K. C., Ono, E. & Xu, X. (1996) *Biochemistry* **35,** 1223–1231.
31. Lamb, R. A. & Horvath, C. M. (1991) *Trends Genet.* **7,** 261–266.
32. Kazazian, H. H., Jr., & Moran, J. V. (1998) *Nat. Genet.* **19,** 19–24.
33. Lee, S. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 4250–4254.
34. Reiss, J., Cohen, N., Dorche, C., Mandel, H., Mendel, R. R., Stallmeyer, B., Zabot, M.-T. & Dierks, T. (1998) *Nat. Genet.* **20,** 51–53.
35. Low, S. C. & Berry, M. J. (1996) *Trends Biochem. Sci.* **21,** 203–208.