

Alternative Processing as Evolutionary Mechanism for the Origin of Novel Nonprotein Coding RNAs

Dingding Mo¹, Carsten A. Raabe¹, Richard Reinhardt², Juergen Brosius^{1,*}, and Timofey S. Rozhdestvensky^{1,*}

¹Institute of Experimental Pathology, ZMBE, University of Muenster, Muenster, Germany

²Max Planck Genome Centre Cologne, Cologne, Germany

*Corresponding author: E-mail: rozhdest@uni-muenster.de, RNA.world@uni-muenster.de.

Accepted: October 3, 2013

Abstract

The evolution of new genes can ensue through either gene duplication and the neofunctionalization of one of the copies or the formation of a de novo gene from hitherto nonfunctional, neutrally evolving intergenic or intronic genomic sequences. Only very rarely are entire genes created de novo. Mostly, nonfunctional sequences are coopted as novel parts of existing genes, such as in the process of exonization whereby introns become exons through changes in splicing. Here, we report a case in which a novel nonprotein coding RNA evolved by intron-sequence recruitment into its structure. cDNAs derived from rat brain small RNAs, revealed a novel small nucleolar RNA (snoRNA) originating from one of the Snord115 copies in the rat Prader–Willi syndrome locus. We suggest that a single-point substitution in the Snord115 region led to the expression of a longer snoRNA variant, designated as L-Snord115. Cell culture and footprinting experiments confirmed that a single nucleotide substitution at Snord115 position 67 destabilized the kink-turn motif within the canonical snoRNA, while distal intronic sequences provided an alternate D-box region. The exapted sequence displays putative base pairing to 28S rRNA and mRNA targets.

Key words: evolution of novel nonprotein coding RNA variants, Prader-Willi syndrome (PWS), rat Snord115, processing mutant; snoRNA biogenesis, K-turn motif.

Introduction

Usually, novel genes are not generated de novo but evolve by duplication of existing genes and, if not inactivated and decaying as so-called pseudogenes, copies might change in a more or less gradual manner (Ohno 1970). A frequent mechanism of amplification is segmental duplication of one or several genes in a locus by unequal crossing over. A rare extreme is whole chromosome or even whole genome duplication. A different route of duplication is restricted to single genes and occurs via RNA intermediates by the mechanism of retroposition. This involves conversion of usually mature RNAs, for example, mRNA into cDNA accompanied by more or less random integration into the genome. Most often, this mechanism yields inactive retropseudogenes; for example, because of the lack of promoter elements necessary for expression. Should such regulatory elements fortuitously be coopted at the genomic locus of integration, a functional retrogene might evolve (Brosius 1991). True de novo formation of a gene out of hitherto neutrally evolving DNA is considered to be rare (Levine et al. 2006), but de novo evolution might be more frequent as expected (Neme and Tautz 2013). Interestingly, a mechanism termed

overprinting can generate novel protein products out of minimally altered preexisting genes, simply by shifting the open reading frame (Keese and Gibbs 1992). More common is the recruitment (exaptation) of novel modules to existing genes, such as exonization of intronic sequences (Lev-Maor et al. 2003), as predicted by Gilbert (1978). Generally, at the onset, such exons are alternatively spliced only, yielding low amounts of the novel mature mRNA in addition to the original mRNA. Furthermore, as the alternative exon is usually slightly deleterious, neutral or at best slightly advantageous, persistence over tens or even hundreds of millions of years is the exception rather than the rule (Krull et al. 2007). Functional nonprotein coding RNAs also arise by gene duplication including retroposition, as is the case for small nucleolar RNAs (snoRNAs) (Brosius 2003; Vitali et al. 2003; Weber 2006; Zemmann et al. 2006; Schmitz et al. 2008). Furthermore, neuronal BC1 RNA arose in the common ancestor of rodents by retroposition of a mature tRNA^{Ala}. The fortuitous location of a distal RNA polymerase III transcription terminator provided an additional 75 nt to contribute the 3' domain of BC1 RNA (DeChiara and Brosius 1987; Kim et al. 1994).

© The Author(s) 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

The snoRNAs constitute a large family of small nonprotein coding RNAs in eukaryotes and Archaea. The majority of snoRNAs, in complex with proteins as ribonucleoprotein particles (RNPs, snoRNPs), are involved in posttranscriptional processing and maturation of RNAs. Except for U3, U8, U14, U17, and U22 snoRNAs that have been proposed to function as RNA chaperones to regulate preribosomal RNA (pre-rRNA) folding and mediate correct nucleolytic processing (maturation) (Kiss 2004), most of the remainder direct site-specific posttranscriptional modifications on 18 S, 28 S, 5.8 S rRNAs, and some U spliceosomal small nuclear RNAs (snRNAs) (Dragon et al. 2006; Gagnon et al. 2007; Dieci et al. 2009). In addition, for a smaller subset of snoRNAs that exhibit base complementarity to pre-rRNAs but do not guide endonucleolytic cleavages or nucleotide modifications, a chaperone-like function was also suggested but not experimentally validated (Vitali et al. 2003). For most of the known snoRNAs, the transient interaction with complementary regions in RNA targets mediates function.

Based on conserved sequence and structural motifs, snoRNAs are divided into two subclasses, the C/D box snoRNAs and the H/ACA box snoRNAs, respectively (Kiss et al. 2006). The majority of C/D-box snoRNAs guide 2'-O-methylation of RNA ribose moieties, while H/ACA-box snoRNAs are involved in the isomerization of uridine to pseudouridine. The 2'-O-methylation guide snoRNAs harbor conserved C (RUGAUGA consensus) and D (CUGA) box motifs, located near to the 5'- and 3'-ends of the RNA, respectively (many snoRNAs also contain internal copies of these elements that are termed C' and D' boxes). An interaction between snoRNA-termini results in the formation of a stem structure, whereas C and D-boxes are involved in kink-turn (K-turn) motif assembly that is recognized by the 15.5 kDa protein in vertebrates (homolog of Snu13p in yeast and L7Ae in Archaea) (Watkins et al. 2000; Kuhn et al. 2002). Three further proteins, fibrillarin (a methyltransferase), Nop56p, and Nop58p, participate in the canonical core C/D box snoRNP assembly (Kiss et al. 2006).

In vertebrate genomes, most snoRNAs are encoded in introns of either protein coding or nonprotein coding host genes (Dieci et al. 2009). Together with exons, they are transcribed by RNA polymerase II as hnRNA. Biogenesis of C/D box snoRNAs is a complex process that involves posttranscriptional snoRNP assembly coupled with nucleolytic processing of host gene pre-RNA introns and intranuclear trafficking (Filipowicz and Pogacic 2002). In mammals, the majority of C/D box RNAs maps to intronic regions located ~70–80 nt upstream from the acceptor splice site. They are processed in a splicing-dependent manner, involving general splicing factors (Hirose and Steitz 2001; Hirose et al. 2003, 2006). The remaining C/D box snoRNAs, including those in the repetitive clusters on human chromosome 14 and 15, are located more distantly from the acceptor splice site and considered not to interact directly with general pre-RNA splicing factors during processing (Cavaille et al. 2000, 2002; Hirose et al. 2003). However, in both cases to prevent

snoRNA degradation during posttranscriptional processing, binding of the core-snoRNP proteins is essential (Richard and Kiss 2006). The 15.5 kDa protein recognizes the terminal K-turn motif formed between C and D-boxes of pre-snoRNA and provides the scaffold for the other core-protein components to bind (Watkins et al. 2000; Cahill et al. 2002; Watkins et al. 2002; Kiss et al. 2006). The assembled pre-snoRNPs undergo 5'- and 3'-RNA exonucleolytic trimming, resulting in mature snoRNA-protein complexes that are transported to the nucleolus (Tycowski et al. 1993; Kiss and Filipowicz 1995; Cavaille and Bachelierie 1996; Watkins et al. 1996).

In humans, the Prader–Willi syndrome (PWS) is a neurogenetic disorder caused by deletion or inactivation of imprinted genes within the PWS locus on paternally inherited chromosome 15. Apart from several protein-coding genes, this locus harbors two large tandemly repeated clusters of C/D box snoRNAs: *SNORD116* and *SNORD115*, with 24 and 47 gene copies, respectively, generated from introns of the U-UBE3A-AS long nonprotein coding RNA (the typical arrangement is one SNORD gene copy per intron; a few introns harbor two copies of SNORD116 genes) (Cavaille et al. 2001; Wirth et al. 2001; Yin et al. 2012). Although deletion of the *SNORD116* gene cluster resembles key characteristics of the PWS-phenotype in patients and causes growth retardation in mice (Skryabin et al. 2007; Ding et al. 2008; Sahoo et al. 2008; de Smith et al. 2009), *SNORD115* deletion appears to lack a phenotype (Runte et al. 2005). *Snord115* and *Snord116* belong to a subclass termed “orphan” snoRNAs as they lack apparent base complementarities to common RNA targets, suggesting functions apart from rRNA and snRNA processing (Bachelierie et al. 2002).

On the basis of analysis of small nonprotein coding RNA enriched cDNA libraries from rat brain, we uncovered a novel snoRNA derived from the imprinted PWS locus. Our data suggest that this RNA arose by a single nucleotide substitution in an ancestral *Snord115* gene copy. Experimental analysis indicated that the nucleotide exchange lead to destabilization of the original *Snord115* K-turn motif. In addition, the downstream intron provided an alternative D-box motif and sequences that enable alternate K-turn formation. These structural alterations trigger alternative pathways of snoRNA maturation and lead to the utilization of *Snord115* 3'-flanking sequence resulting in a novel snoRNA variant. Our experimental data reveal additional mechanistic insight into nonprotein coding RNA origin and evolution.

Materials and Methods

Generation of Recombinant Plasmid Constructs

The L-*Snord115* expression constructs were generated via polymerase chain reaction (PCR) amplification of rat genomic templates. The DNA extraction was conducted according the standard proteinase K method (Maniatis et al. 1989). The

design of PCR primers to amplify the Snord115 gene flanked by G2 and G1 exons was based on available rat cDNAs (CB616315) (Landers et al. 2004). The resulting PCR products were cloned into pDrive vectors using the QIAGEN PCR cloning kit according to the manufacturer's recommendations. Inserts were verified by sequencing and subcloned into the pcDNA3.1+ vector (Invitrogen) via *Bam*HI and *Hind*III restriction sites. Mutations in the L-Snord115 or Snord115 coding regions were introduced by 5' overlapping PCR (Warrens et al. 1997). The resulting PCR fragments were cloned into pcDNA3.1+ vector and verified by sequencing. All recombinant plasmids were purified by cesium chloride (CsCl) gradient centrifugation (Maniatis et al. 1989).

Cell Culture and DNA Transfection

HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM) (Sigma) medium, supplemented with 10% fetal bovine serum (BioWest), 10 mM sodium pyruvate (LifeTech), 100 U/ml penicillin (LifeTech), and 100 U/ml streptomycin (LifeTech) at 37 °C in 5% (v/v) CO₂. Transient transfections were performed with the lipotransfection reagent (Lipofectamine-2000, Invitrogen) according to the manufacturer's recommendations, at 70–85% cell confluence in six-well plates using 2 µg of plasmid DNA per transfection. Lipofectamine-2000–DNA complexes were formed for 20 min at room temperature in OptiMEM (LifeTech) buffer. The complexes were transferred to HeLa cells and incubated for 6 h at 37 °C. Subsequently, the OptiMEM buffer was replaced by DMEM medium (see earlier), and total RNA was extracted 24–36 h posttransfection.

Total RNA Extraction and Northern Blot Hybridization

Total RNA from HeLa cells and rat tissues was extracted using the TRIzol reagent (Invitrogen) according to the manufacturer's recommendations. Approximately 6 µg of total RNA per sample was size fractionated on 8% (w/v) polyacrylamide (29:1 acrylamide/bis), 7 M urea gels and electrotransferred to positively charged nylon membranes (BrightStar-Plus, Ambion, Bio-Rad). Before hybridization, the RNA was UV-crosslinked to membranes (Stratalinker UV Crosslinker 2400, Stratagene). Membranes were prehybridized in 20 ml of 0.5 M sodium phosphate (pH 6.5 at 58 °C), 7% (w/v) sodium dodecyl sulfate (SDS) buffer at 56 °C for 40 min. Subsequently, northern blot hybridizations were performed with 50 pmol of 5'-³²P labeled oligonucleotides (fig. 1A) in prehybridization buffer at 56 °C overnight. Membranes were washed three times in 0.1 M sodium phosphate (pH 6.5), 1% (w/v) SDS buffer for 3 min each at 46 °C and BioMax MS films (Kodak) were exposed at –80 °C overnight.

In vitro Transcription of Different RNA Templates

L-Snord115 RNA and mutants were in vitro transcribed by T7 RNA polymerase (Fermentas). The corresponding runoff

templates for in vitro transcription were generated by PCR amplification. The PCR forward primer included sequences of the T7 RNA polymerase promoter (supplementary table S1, Supplementary Material online). In vitro transcribed fragments of 28 S rRNA were generated in a similar way using human genomic DNA isolated from placenta as a template for PCR reactions (supplementary table S1, Supplementary Material online). In vitro transcription was conducted in 100 µl total reaction volume, supplemented with 40 mM Tris–HCl pH 7.9, 6 mM MgCl₂, 10 mM dithiothreitol (DTT), 10 mM NaCl, and 2 mM spermidine. Template concentrations ranged between ~0.5 and 1 µg per reaction. Each reaction was performed with 2.5 mM NTPs, 40 U RNase inhibitor (Fermentas), and 2,000 U of T7 RNA polymerase (Fermentas). Transcription proceeded for 2 h at 37 °C. The synthesized RNAs were separated on 8% (w/v) polyacrylamide (29:1 acrylamide/bis) 7 M urea gels and eluted in 0.3 M NaOAc (pH 5.2) buffer at 4 °C overnight. Subsequently, RNAs were EtOH precipitated and dissolved in ddH₂O.

Lead (II)-Footprinting Analysis

L7Ae protein was purified as described previously (Rozhdetsvensky et al. 2003). In vitro transcribed RNAs were dephosphorylated by Antarctic Phosphatase treatment (New England BioLabs). The resulting RNAs were subjected to T4 polynucleotide kinase (T4 PNK) (New England BioLabs) treatment to incorporate [γ -³²P]-ATP (Perkin Elmer) label at 5'-ends. Lead acetate cleavage was performed with minor modifications (Youssef et al. 2007). In brief, 5'-³²P labeled RNAs were heat-denatured at 90 °C for 1 min and immediately chilled on ice for at least 2 min. RNA-L7Ae complex formation was performed in 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)–KOH (pH 7.0), 200 mM potassium acetate, 1.5 mM magnesium acetate, 2.5 µg/µl tRNA, and 10 U RNase inhibitor (Fermentas); specific concentrations of L7Ae protein are indicated in figure 3A. Footprinting analyses were performed with freshly prepared 15 mM lead acetate at room temperature. Cleavage was terminated after 10 min by ethylenediaminetetraacetic acid (EDTA) addition. All reactions were ethanol precipitated and separated on 8% (w/v) polyacrylamide (38:2 acrylamide/bis) 7 M urea gels. RNase T1 and alkaline hydrolysis RNA ladders were generated according to the manufacturer's instructions (Ambion). MS films (Kodak) were exposed to the gels overnight at –80 °C.

Mapping of 2'-O-Methylation on 28 S rRNA

Mapping of possible 2'-O-methylation sites on 28 S rRNA was done by reverse transcription (RT) assays at low deoxyribonucleotide triphosphate (dNTP) concentration (Maden 2001). Briefly, ~0.5 µg total RNA, isolated from transiently transfected HeLa cells (see Results and Discussion), was mixed with 0.5 pmol of 5'-³²P-labeled oligonucleotide primer for RT. The mixtures were denatured at 85 °C for 2 min and

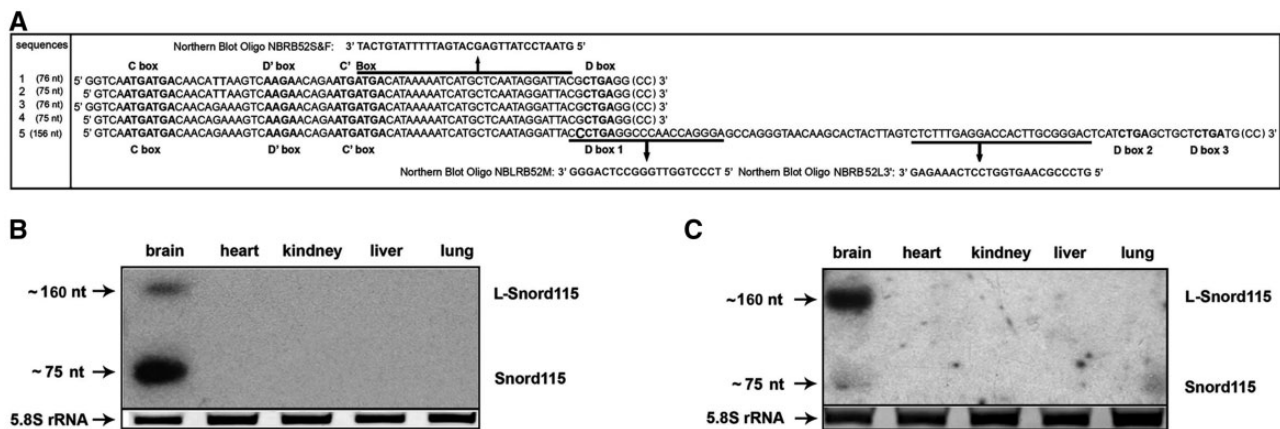


Fig. 1.—Identification and expression analysis of L-Snord115 RNA. (A). Sequence alignment between known rat Snord115 variants (1–4) and the novel L-Snord115 RNA (5). snoRNAs sizes in nucleotides are indicated on the left in parentheses. Putative C, D, C', and D' boxes are in bold letters and, in addition, designated above and below the alignment. snoRNA regions complementary to the northern blot probes are underlined. Hybridization probes and their sequences are also indicated. (B and C) northern blot analysis with NBRB52S&F (B) and NBRB52M (C) probes to examine L-Snord115 expression in different rat tissues indicated above the blot lanes. L-Snord115, Snord115 RNAs and their estimated sizes (in nt) are indicated on the margins. As a loading control, a negative image of an ethidium bromide stained 5.8 S rRNA signals is shown at the bottom.

allowed to anneal at room temperature. RT was performed in 20 μ l reaction volume, containing 50 mM Tris-HCl (pH 8.5), 30 mM KCl, 8 mM MgCl₂, 20 U RNase inhibitor (Fermentas), and 2.5 U of transcript reverse transcriptase (Roche). Different concentrations of dNTP (10 μ M, 100 μ M, and 1 mM) (Roche) were added to individual reactions. RT was performed at 55 $^{\circ}$ C for 40 min (supplementary fig. S1, Supplementary Material online) and terminated by the addition of 2 \times RNA loading dye (Ambion). Aliquots were separated on 8% (w/v) polyacrylamide (38:2 acrylamide/bis) 7 M urea gels. To monitor potential stops of RT caused by RNA secondary structures, RT reactions conducted with the identical primer and 0.02 pmol of in vitro transcribed 28 S rRNA fragments served as control (fig. 3A). MS films (Kodak) were exposed to resulting gels overnight at -80° C.

Results and Discussion

Identification of Novel snoRNA from the Rat PWS-Locus

On the basis of specialized rat brain cDNA libraries enriched for small nonprotein coding RNAs (Raabe CA, Brosius J and Rozhdestvensky TS, unpublished data), we identified numerous isoforms of Snord115 snoRNA (supplementary fig. S2, Supplementary Material online). The library design favored full-length cDNAs because the synthesis relied on RNA 5'- and 3'-end modifications by adapter ligation and C-tailing, respectively (Raabe et al. 2010). We identified a novel nonprotein coding RNA candidate, almost twice as long as the previously known Snord115 RNAs (fig. 1A), designated as long-Snord115 (L-Snord115). The RNA, represented by 26 cDNAs, is 156 nt long and maps to the rat Snord115 gene cluster. L-Snord115 displays sequence identity to an annotated

rat canonical Snord115 isoform (Ensembl Transcript: ENSRN00000052941) throughout its entire 5'-domain, except for a single G to C nucleotide substitution located adjacent to the D-box of the known snoRNA (fig. 1A). Close inspection of the 3'-region of L-Snord115 RNA revealed two additional putative D-box elements (CUGA-sequences) located 14 and 4 nt upstream from the RNA 3'-end, respectively (fig. 1A). Therefore, sequence and structural analysis suggests that the identified nonprotein coding RNA candidate is a potentially novel C/D-box snoRNA. For further validation and to establish the expression profile of L-Snord115 RNA across different tissues, northern blot hybridization on total RNA isolated from rat brain, heart, kidney, liver, and lungs was carried out (fig. 1B and C). Three specific oligonucleotide probes complementary to the 5', central, and 3'-regions of L-Snord115 RNA were designed (fig. 1A). We could detect brain-specific expression of the novel snoRNA candidate (fig. 1B and C and data not shown) paralleling the expression profile of canonical Snord115 snoRNAs. The oligonucleotide probe (NBRB52S&F) corresponding to the 5'-domain of L-Snord115 detected the canonical Snord115 isoforms as well as L-Snord115 RNA (fig. 1B). Probes complementary to the central and 3'-portion of the novel snoRNA isoform (NBRB52M and NBRB52L3', respectively) identified brain-specific signals of ~160 nt, indicative of L-Snord115 RNA (fig. 1C and data not shown). Genomic analysis revealed that the 5'-region of L-Snord115 RNA only differed by a single nucleotide substitution from the rat-annotated Snord115. We therefore investigated the potential impact of single nucleotide transversion on the biogenesis of a novel snoRNA variant.

Notably, Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, last accessed October 28, 2013) and Blat (<http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>, last accessed October

28, 2013) rat genome searches did not reveal a perfect match for the L-Snord115 query sequence as well as its shorter form. The best Blat hit displayed only 99.4% sequence identity to the available annotated sequence of the rat Snord115 locus, with a single C → G transversion adjacent to the D-box.

Because of the repetitive structure of the snoRNA clusters within PWS loci in mammals, the genomic assembly of this region often remains incomplete (Nahkuri et al. 2008). We therefore PCR amplified the rat PWS genomic region, presumed to contain the novel variant, with primers located in the flanking G1 and G2 exons. The resulting PCR products were inserted into the pDrive vector and subsequently sequenced to verify the L-Snord115 containing insert.

To investigate the unusual processing pattern of L-Snord115, we subcloned the corresponding construct into the pcDNA 3.1 eukaryotic expression vector (fig. 2A, sequence 1) and performed transient transfection experiments in HeLa cells. Total RNA was examined by northern blot hybridization using NBRB52S&F probe from the 5'-region of the RNA. In HeLa cells, human SNORD115 was not detected, as in this cell-type the endogenous genes are silent (fig. 2B, vector lane). In transfected HeLa cells two signals corresponding to Snord115 and L-Snord115 RNAs were observed (fig. 2B, lane 1). Presumably, alternative posttranscriptional processing generates both snoRNAs from a single gene repeat unit.

Analysis of Posttranscriptional Processing of Snord115

Canonical C/D box snoRNAs including all Snord115 isoforms contain consensus C and D boxes at their 5'- and 3'-ends, respectively (Cavaille et al. 2000; Bachellerie et al. 2002; Nahkuri et al. 2008). These elements are part of the K-turn structure motif that is a hallmark of eukaryal and archaeal snoRNAs of this type (Kiss et al. 2006; Gagnon et al. 2007; Dieci et al. 2009). The double-stranded K-turn motif resembles a variation of the "helix, internal-loop, helix" type of RNA secondary structures. Typically, it contains a 5'-canonical stem forming the base of the motif that is followed by a 3 nt long asymmetric bulge and ending in A•G and G•A sheared base pairs that constitute the noncanonical stem formation (Klein et al. 2001). The phosphodiester backbone of the bulge nucleotides forms a sharp turn (kink) of ~120° in the helix toward the minor groove. The K-turn conformation is stabilized by base stacking interactions between adenosines of the A•G base pairs and the canonical stem (Klein et al. 2001). In addition, bases located in the bulge also participate in stacking interactions with nucleotides of the canonical and noncanonical stems and further contribute to stabilize the motif (Lin et al. 2011). To form K turns, the RNA requires interaction of specific proteins or/and metal ions (Matsumura et al. 2003; Goody et al. 2004; Turner and Lilley 2008). K-turn motifs in mammalian C/D box snoRNAs are associated with the 15.5 kDa protein, which is also required for the assembly of core-snoRNP (Watkins et al. 2000). Deletions or mutations

within the terminal C- or D-boxes disturb K-turn motif formation and protein binding and therefore will result in snoRNA exonucleolytic degradation during processing (Darzacq and Kiss 2000; Filipowicz and Pogacic 2002).

To gain further insight into mechanistic details of L-Snord115 processing, we generated and analyzed snoRNA mutants. We postulated that processing of L-Snord115 is dependent on the formation of a K-turn motif involving the 5'-C-box and the potential D-box2 or D-box3 of L-Snord115 RNA (figs. 1A and 2A). We designed experiments to identify the functional 3'-terminal D-box of L-Snord115. Two constructs were generated harboring mutations in the D-box2 or D-box3, respectively (fig. 2A–C, sequences 6 and 7). Mutation of D-box2 did not interfere with L-Snord115 post-transcriptional processing and stability (fig. 2B, lane 6). In contrast, a GpA to CpU (positions 151 and 152) substitution within D-box3 completely abolished the expression of the long L-Snord115 form but not of the shorter canonical form (fig. 2B, lane 7; fig. 2C). Expectedly, mutation of both boxes abolished L-Snord115 expression (fig. 2B, lane 8). To further investigate the importance of D-box3, we stabilized the 5'-canonical stem by replacing cytosine at position 155 with adenosine (fig. 2C). The corresponding base exchange would generate an additional U-A Watson–Crick base pair instead of a U-C mismatch at the base of the canonical stem in the K-turn motif (fig. 2A, construct 2; fig. 2C). Accordingly, northern blot analysis shows an increase of L-Snord115 accumulation. This observation further supports the involvement of D-box3 in formation of the K-turn motif within the longer snoRNA structure. Stabilization of this motif resulted in a shift of the snoRNA processing equilibrium toward L-Snord115 (fig. 2B, lane 2). To further analyze the post-transcriptional processing equilibrium between canonical Snord115 and L-Snord115 RNA, we mutated D-box1 to abolish the competition with D-box3 for K-turn motif formation (fig. 2A, construct 5; fig. 2D). In transient transfections conducted with the D-box1 mutant, we detected only L-Snord115 RNA in northern blots (fig. 2A, B lane 5; fig. 2D).

All cDNAs representing L-Snord115 RNA detected in our cDNA library screens contain cytosine adjacent to D-box1 at position C67. Most of the known mammalian Snord115 isoforms harbor guanine (G67) at this position (Nahkuri et al. 2008). Therefore, we investigated the potential influence of G67 to C67 substitution on L-Snord115 RNA maturation. A construct containing the corresponding rat genomic sequence with G67 instead of C67 adjacent to the D-box1 (fig. 2A, construct 3) only yielded Snord115 RNA (fig. 2B, lane 3). Secondary structure analysis of the potential K-turn motif formed between C-box and D-box1 of Snord115 suggested that G67 might stabilize the noncanonical stem of the K-turn by base pairing with the C12 nucleotide (fig. 2D). The increase in stability is likely to explain the exclusive generation of Snord115 RNA in transient transfections. To examine whether structural stabilization of the alternative K-turn motif formed

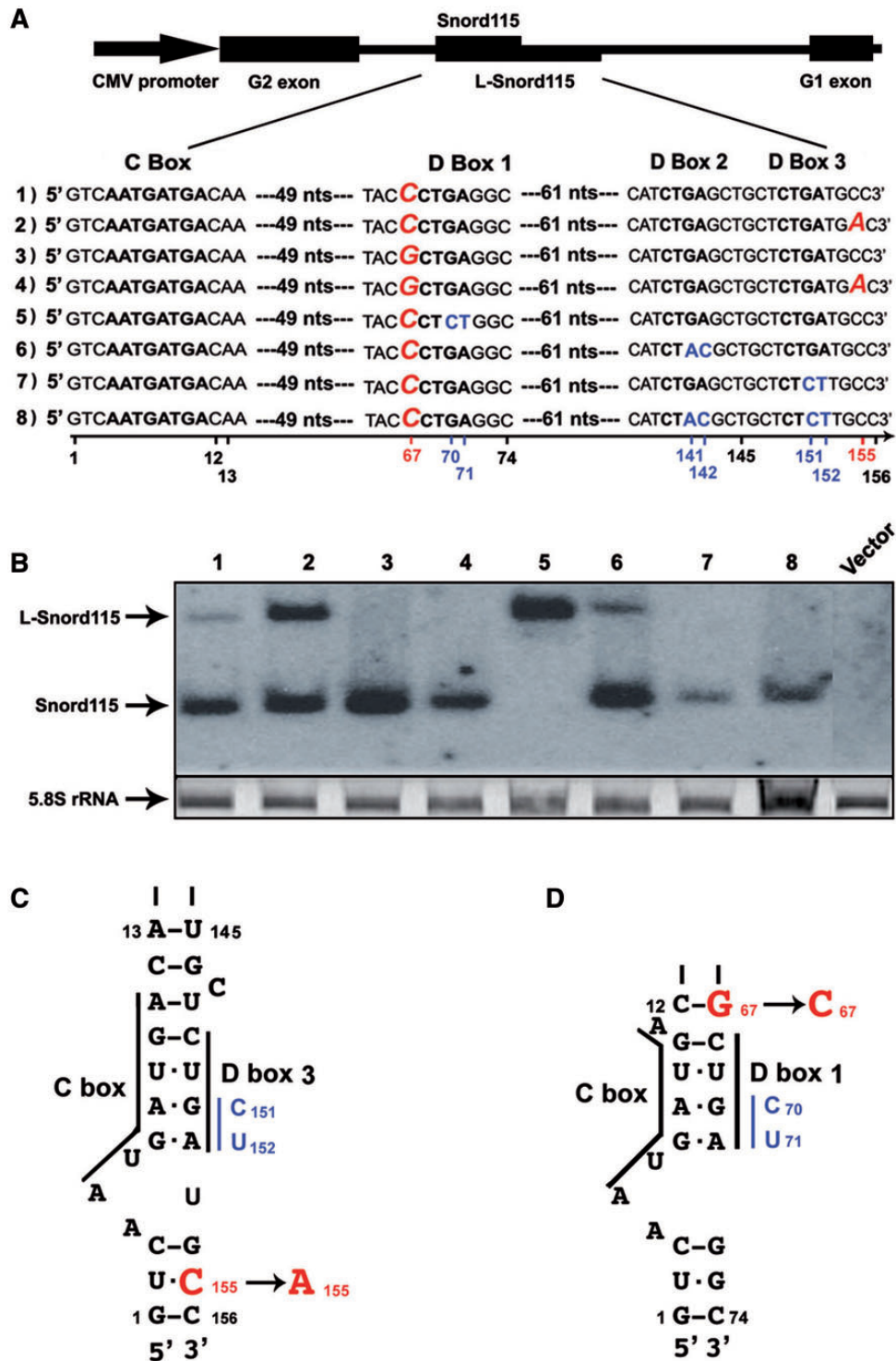


FIG. 2.—Analyses of L-Snord115 posttranscriptional processing. (A) Schematic representation of expression constructs containing the genomic L-Snord115/Snord115 repeat unit harboring the snoRNA and relevant sequences (1, 3) or snoRNA mutants (2, 4–8) used in expression studies. Nucleotides representing putative snoRNA boxes are in bold letters. The G67 to C67 substitution and C155 to A155 mutation leading to K-turn stabilization are highlighted in red. Mutations in putative D-boxes are in blue lettering. (B) Northern blot analysis of total RNA isolated from transfected HeLa cells. Transfected pcDNA 3.1 control vector or constructs 1–8 (as represented in A) are indicated above the respective lanes. Arrows indicate signals of snoRNAs and 5.8 S rRNA (as loading control, a negative image of an ethidium bromide stained gel is shown at the bottom). (C and D). Putative secondary structures for terminal K-turn motifs in L-Snord115 (C) and Snord115 RNA (D). Nucleotide substitutions (mutations) are indicated as in (A).

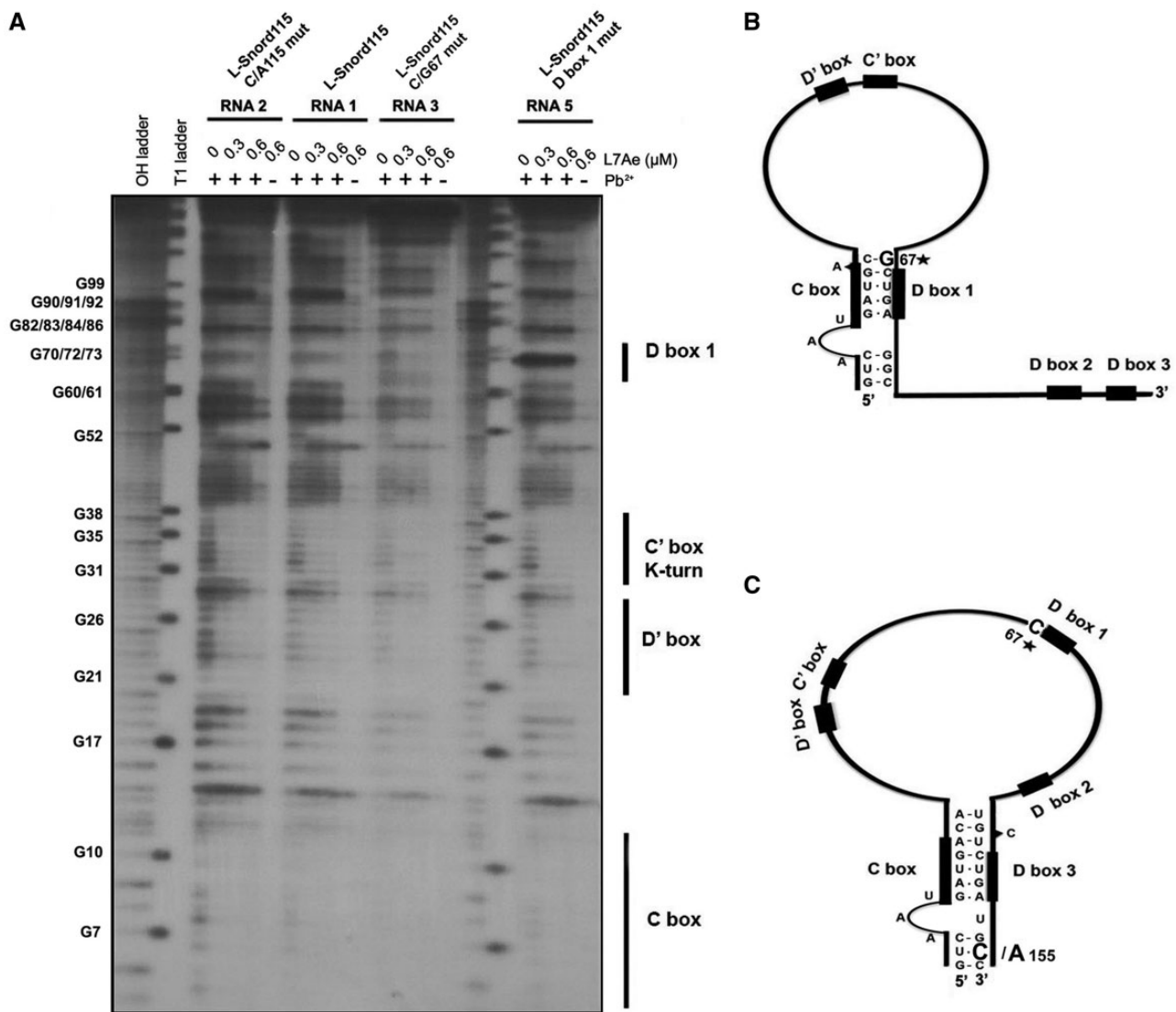


FIG. 3.—Structural analysis of snoRNA/L7Ae RNP complexes with lead acetate. (A) In vitro transcribed 5'-³²P labeled RNAs (indicated on top) were incubated with increasing concentrations of L7Ae protein (0, 0.3, and 0.6 μM) and treated with 15 μM of lead acetate (+ lanes). As control, the corresponding RNA incubated with 0.6 μM of L7Ae protein without lead cleavage was loaded (–lanes). To determine RNA cleavage sizes alkaline and RNase T1 digestions of L-Snord115 RNA were included (indicated as OH or T1 ladders, respectively). Positions of 3'-G residues of RNase T1 cleavage products are indicated on the left. On the right, the regions of putative snoRNA boxes are designated. (B and C) Structural models suggested for posttranscriptional processing of snoRNA resulting in Snord115 (B) or L-Snord115 (C) RNAs, respectively.

between the C-box and D-box3 results in generation of L-Snord115 RNA, we introduced a C155 to A155 base change into the G67 containing construct (fig. 2A, sequence 4). In HeLa cell transfections, the short Snord115 RNA was detected only, indicating that G67 is sufficient to shift the processing equilibrium to the canonical Snord115 variant (fig. 2B, lane 4). In conclusion, our data indicate that the single G to C nucleotide substitution in one of the snoRNA gene copies permits fortuitous recruitment of an external D-box in an appropriate sequence context, located in the 3'-flanking intron. This recruitment results in alternative

K-turn motif formation and permits L-Snord115 RNA generation.

Probing of Putative RNA Structural Conformations

For further analysis of structural elements underlying Snord115 posttranscriptional processing, we performed lead acetate cleavage experiments. Lead ions usually catalyze phosphodiester bond cleavage within unstructured single stranded (bulges, loops, etc) or flexible RNA regions (Huntzinger et al. 2008). Protein(s) interacting with RNA might protect the

phosphodiester backbone from cleavage (Huntzinger et al. 2008). As mentioned earlier, K-turn formation requires specific protein(s) to interact with RNA. Archaeal ribosomal protein L7Ae is a functional homolog of mammalian C/D box snoRNA 15.5 kDa protein (Kuhn et al. 2002). It has been reported to specifically recognize and stabilize different K-turn structural motifs (Rozhddestvensky et al. 2003). Therefore, we used recombinant L7Ae protein as structural component to form and protect K-turns within L-Snord115 and the corresponding RNA mutants. In addition to L-Snord115 (fig. 2A, construct 1), other *in vitro* transcribed RNAs were examined: 1) The L-Snord115 C67 to G67 substitution transcript only yielding canonical Snord115 during cell culture transfection (fig. 2A, construct 3; fig. 2B, lane 3); 2) L-Snord115 C155 to A155 mutant that resulted in increased generation of L-Snord115 RNA (fig. 2A, construct 2; fig. 2B, lane 2); and 3) L-Snord115 D-box1 mutant that exclusively generated L-Snord115 snoRNA (fig. 2A, construct 5; fig. 2B, lane 5).

When comparing the lead cleavage results of the investigated L7Ae RNP complexes, the most obvious differences were observed within the D-box1 region of snoRNAs (fig. 3A). We were unable to resolve the D-box3 RNA region because of its proximity to the 3'-end of the RNA. In L-Snord115 containing the C67 to G67 substitution, the D-box1 region appeared completely protected from cleavage. This indicated that these nucleotides are involved in K-turn motif formation and bound to L7Ae protein in the majority of RNA molecules (fig. 3A–C). Hence, the presence of G67 nucleotide in L-Snord115 sequence favors an RNA structure where the D-box1 forms a K-turn motif with the 5'-terminal C-box region (fig. 3A and B). In experiments with the L7Ae/L-Snord115 (C67) RNP-complex, we detected slight RNA backbone cleavage in the D-box1 sequence (fig. 3A). In agreement with our transfection experiments, at least two major RNA structural conformations resulted from competition between D-box1 and D-box3 for the 5' C-box sequence in both cases, leading to a K-turn motif (fig. 3A–C). Examining the L-Snord115 C155 to A155 mutant, we observed a slight increase in cleavage within the D-box1 region. This supported our previous observation that the C155 to A155 mutation stabilized the K-turn motif formation between C-box and D-box3 (fig. 2B and C) and is consistent with the accumulation of L-Snord115 as observed in northern blots (fig. 2B).

Finally, lead acetate footprinting with L-Snord115 harboring the GpA to CpU substitution in D-box1 showed strong cleavage in the mutated region (fig. 3A and C) correlating with our transfection studies, where only L-Snord115 RNA was generated from the construct when the D-box1 motif was deactivated (fig. 2B). Notably, D', C', and C-boxes were completely protected from lead cleavage in all tested RNPs, indicating that in the investigated RNAs those nucleotides were involved in K-turn formation and therefore bound to L7Ae. In summary, the G67 to C67 nucleotide substitution destabilized the canonical Snord115 terminal K-turn motif during

RNA processing and led to an additional RNA structure that allowed for L-Snord115 maturation during posttranscriptional processing.

Does the L-Snord115 RNA Variant Have a Function?

The majority of C/D-box snoRNAs exhibit complementary to rRNAs or snRNAs guiding posttranscriptional modification of their targets. We therefore performed computational analysis using a modified DNAMAN (version 6.015) software (Zemann et al. 2006) to screen for putative antisense elements located within the 3'-region of L-Snord115 that could potentially target rRNA or snRNA molecules. We identified an 8 nt sequence element adjacent to D-box3 of L-Snord115 that exhibits base complementarity to an evolutionary conserved region of 28 S rRNA. The analysis suggested that L-Snord115 might guide 2'-O-methylation of rat 28 S rRNA at G4737 corresponding to G4980 in human rRNA (fig. 4A). In mammalian rRNA, this nucleotide modification has not been reported (Lestrade and Weber 2006). Therefore, we experimentally analyzed the potential involvement of L-Snord115 in methylation of endogenous human (during HeLa transfection experiments) and rat brain 28 S rRNAs, respectively. The biochemical analysis to verify the potential modification by RT did not reveal stops at low dNTP concentrations, indicating that rat L-Snord115 does not guide G4980 2'-O-methylation in rat brain and HeLa cells at detectable levels (supplementary fig. S1A and S1B, Supplementary Material online).

The complementarity of L-Snord115 to 28 S rRNA theoretically extends up to 11 nt involving parts of the D-box element (fig. 4A). This might destabilize the K-turn motif upon binding and, in turn, leading to dissociation of 15.5 kDa protein. Interaction with 15.5 kDa protein is required to recruit the core snoRNP proteins, including the 2'-O-methyltransferase, fibrillarin (Lafontaine and Tollervey 2000; Watkins, et al. 2000; Dragon et al. 2006). This might be one of several potential explanations as to why this relatively young snoRNA variant does not, despite the theoretical complementarity, target modification of G4980 in 28 S rRNA. Alternatively, the novel variant might be involved in chaperone-like functions to stimulate RNA folding, as suggested for other snoRNAs (Vitali et al. 2003).

There are snoRNAs whose target or functions are unknown, including those encoded by the Snord115 and Snord116 gene clusters in the PWS locus. However, in Snord115 RNA, an 18 nt long complementary to the alternatively spliced exon Vb of the 5HT2c serotonin receptor pre-mRNA has been predicted (fig. 4B) (Cavaille et al. 2000). The targeted region is also subject to enzymatic posttranscriptional A-to-I editing by two proteins termed "adenosine deaminase acting on RNA" (ADAR1 and ADAR2) (fig. 4B) (Burns et al. 1997; Vitali et al. 2005). The alternative splice site is located 13 nt upstream from the predicted complementarity to

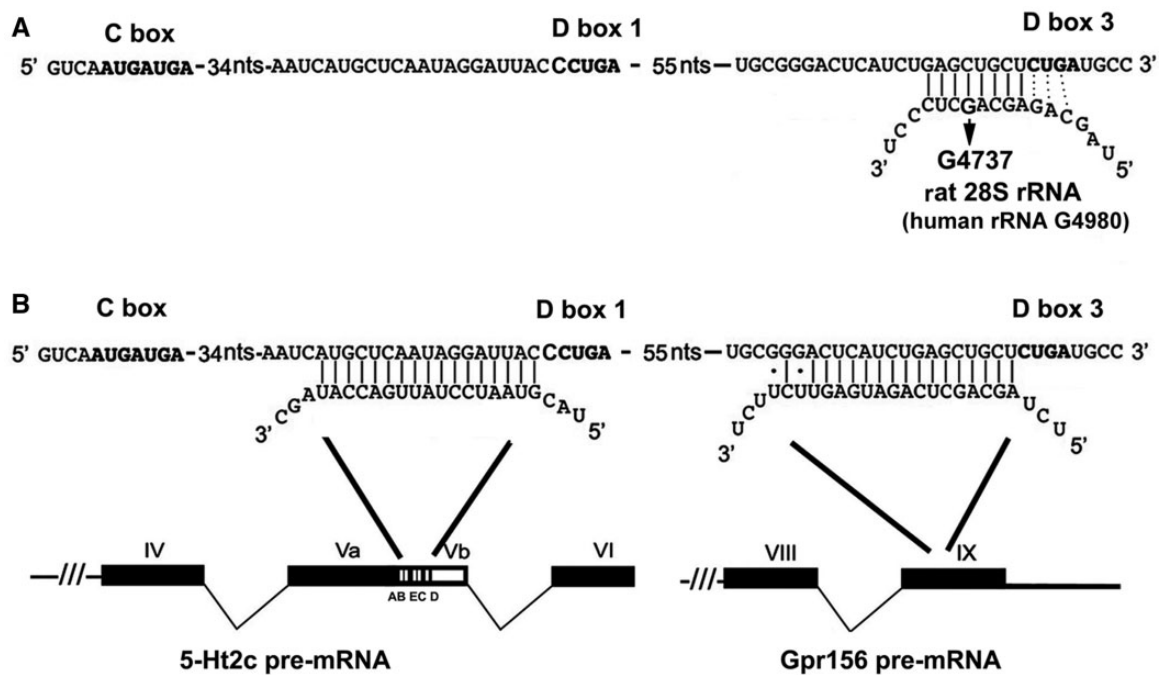


Fig. 4.—L-Snord115 snoRNA and its putative RNA targets. (A) Potential base pairing between L-Snord115 RNA and the 3'-region of 28 S rRNA. The predicted 2'-O-methylated nucleotide (G4980) is shown in bold. Additional base pairings involving 3 nt of the D-box are indicated by dotted lines. (B) Putative base pairing of L-Snord115 RNA with exon Vb of 5HT2c pre-mRNA and exon IX of Gpr156 mRNA, respectively. Parts of the alternatively spliced exon Vb and A to I editing sites in 5HT2c pre-mRNA are indicated (A, B, E, C, D).

SNORD115 RNA and leads to a truncated serotonin receptor (Cavaille et al. 2000). The E, C, and D editing sites on pre-mRNA overlap with the targeted region. Posttranscriptional A to I editing has been reported to decrease the efficiency of G-protein coupling and therefore generates 5HT2c receptor variants with reduced activity (Berg et al. 2001; Vitali et al. 2005). Hence, perfect base complementarities displayed by the antisense element of Snord115 RNA to the alternatively spliced and posttranscriptionally edited exon of 5HT2c pre-mRNA suggest a tempting model for regulation of serotonin receptor biogenesis by snoRNAs (Cavaille et al. 2000). Although, *in vitro* analysis suggested potential involvement of SNORD115 RNA to regulate alternative splicing and editing of 5HT2c pre-mRNA (Vitali et al. 2005; Kishore and Stamm 2006), *in vivo* confirmation remains elusive, thus far (Doe et al. 2009).

Similar to Snord115, the L-Snord115 RNA variant contains the antisense element to the 5HT2c pre-mRNA located in the 5'-portion of snoRNA (fig. 4B). Computational searches to identify potential mRNA targets for the guide element located in the 3'-part of L-Snord115 identified a 19 nt complementarity between a region directly adjacent to the D-box3 and protein-coding exon 9 of the metabotropic glutamate receptor Gpr156 mRNA (fig. 4B). Interestingly, both 5HT2c and Gpr156 proteins are members of the G protein-coupled receptor

family (Stam et al. 1994; Calver et al. 2003). However, until there is sound *in vivo* evidence for a functional interaction of Snord115 or L-Snord115 RNAs with mRNA targets, the complementarities should be considered fortuitous.

Conclusion

We identified a novel brain specific C/D-box snoRNA variant in the rat PWS locus. The potential to generate L-Snord115 RNA from one of the Snord115 copies hinges on sequences in two separate regions. An intronic sequence provided an alternative D-box motif, while the canonical snoRNA coding region acquired a crucial G67 to C67 transversion adjacent to the Snord115 canonical D-box. The latter change led to a slight destabilization of the K-turn motif formed between 5'-C-box and 3'-D-box regions of Snord115 RNA. The presence of an additional D-box region in the 3' flanking sequence provided nucleotides for an alternative K-turn formation. This structure is assembled between the Snord115 5'-C-box and the distal intronic D-box and is necessary to express the novel L-Snord115 RNA variant. However, the changes did not completely abolish canonical Snord115 production. Instead, they resulted in a posttranscriptional processing equilibrium yielding Snord115 as well as the novel L-Snord115 RNA variant.

All snoRNAs encoded within the PWS-locus lack significant base complementarities to the classical rRNAs or snRNAs targets (Cavaille et al. 2000). However, L-Snord115 exhibits a complementarity of 8–11 nt to the 3'-domain of 28 S rRNA. Experimental approaches failed to identify the corresponding 2'-O-methylation at detectable levels. Despite proposals that members of Snord115 snoRNA family are involved in regulation of A → I editing or alternative splicing, solid in vivo evidence is still lacking. In any event, the novel snoRNA variant is restricted to rat but absent in mouse, and hence, at most ~25 million years old. By analogy, when we studied Alu element exonizations out of introns dating back between 20 and 60 Ma, many of them were lost again on their way through the various Old World, New World monkey, and Ape lineages (Krull et al. 2005). Once more, this analogy is not surprising as most such events initially are slightly deleterious or neutral and rarely more or less advantageous, and persistence of novel parts of existing genes is rather the exception than the rule. Significantly, older events such as exonizations of mammalian-wide repetitive elements, exhibited evidence for purifying selection (Krull et al. 2007).

However, despite the low odds, one should not underestimate the significance of exaptations of genetic novelties. For example, exaptation of a recombinase of a DNA transposon, perhaps was for many million years near neutral. Nevertheless, it was a key event for the evolution of the immune system in jawed vertebrates (Kapitonov and Jurka 2005).

Here, we revealed a mechanism by which new isoforms of nonprotein coding RNAs evolve. Based on our current understanding of snoRNA evolution, new members arise by cis- or trans-duplication of ancestral snoRNA genes (Vitali, et al. 2003; Weber 2006; Zemmann, et al. 2006; Schmitz et al. 2008). Cis-duplications are considered to be generated by recombination and lead to integration of new snoRNA copies into neighboring introns of the same host gene. Trans-duplications are mediated by retroposition and result in random integration of snoRNAs retrotransposons. The mechanism of L-Snord115 generation is different from the above. It demonstrates that the corresponding pre-snoRNA structure during alternative posttranscriptional processing is subject to length variation resulting in extension or reduction of snoRNA sequences. Based on our data, it is tempting to suggest that many of the known snoRNAs larger or smaller than the canonical structures arose by similar mechanisms. In summary, our data demonstrate new aspects in nonprotein coding RNA evolution and biogenesis.

Supplementary Material

Supplementary figures S1–S2 and table S1 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

This work was supported by the Nationales Genomforschungsnetz (NGFNIII 01GS0808-4 and 01GS0805).

Literature Cited

- Bachelier JP, Cavaille J, Huttenhofer A. 2002. The expanding snoRNA world. *Biochimie* 84:775–790.
- Berg KA, et al. 2001. RNA-editing of the 5-HT(2C) receptor alters agonist-receptor-effector coupling specificity. *Br J Pharmacol.* 134:386–392.
- Brosius J. 1991. Retroposons—seeds of evolution. *Science* 251:753.
- Brosius J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118:99–116.
- Burns CM, et al. 1997. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 387:303–308.
- Cahill NM, et al. 2002. Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J.* 21:3816–3828.
- Calver AR, et al. 2003. Molecular cloning and characterisation of a novel GABAB-related G-protein coupled receptor. *Brain Res Mol Brain Res.* 110:305–317.
- Cavaille J, Bachelier JP. 1996. Processing of fibrillar-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure. *Biochimie* 78:443–456.
- Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachelier JP. 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet.* 11:1527–1538.
- Cavaille J, Vitali P, Basyuk E, Huttenhofer A, Bachelier JP. 2001. A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J Biol Chem.* 276:26374–26383.
- Cavaille J, et al. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A.* 97:14311–14316.
- Darzacq X, Kiss T. 2000. Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5',3'-terminal stem structure. *Mol Cell Biol.* 20:4522–4531.
- DeChiara TM, Brosius J. 1987. Neural BC1 RNA: cDNA clones reveal non-repetitive sequence content. *Proc Natl Acad Sci U S A.* 84:2624–2628.
- de Smith AJ, et al. 2009. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet.* 18:3257–3265.
- Dieci G, Preti M, Montanini B. 2009. Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 94:83–88.
- Ding F, et al. 2008. SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3:e1709.
- Doe CM, et al. 2009. Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. *Hum Mol Genet.* 18:2140–2148.
- Dragon F, Lemay V, Trahan C. 2006. snoRNAs: Biogenesis, Structure and Function. In: eLS. Chichester: John Wiley & Sons Ltd.
- Filipowicz W, Pogacic V. 2002. Biogenesis of small nucleolar ribonucleoproteins. *Curr Opin Cell Biol.* 14:319–327.
- Gagnon K, Zhang X, Maxwell E. 2008. The box C/D RNPs: evolutionary ancient nucleotide modification complexes. In: Smith H, editor. RNA and DNA editing: Molecular mechanisms and their integration into biological systems. Hoboken, NJ: John Wiley & Sons. p. 313–339.
- Gilbert W. 1978. Why genes in pieces? *Nature* 271:501.
- Goody TA, Melcher SE, Norman DG, Lilley DM. 2004. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* 10:254–264.
- Hirose T, Shu MD, Steitz JA. 2003. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell.* 12:113–123.

- Hirose T, Steitz JA. 2001. Position within the host intron is critical for efficient processing of box C/D snoRNAs in mammalian cells. *Proc Natl Acad Sci U S A*. 98:12914–12919.
- Hirose T, et al. 2006. A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol Cell*. 23:673–684.
- Huntzinger E, et al. 2005. Probing RNA structures with enzymes and chemicals in vitro and in vivo. In: Hartmann RK, et al. editors. *Handbook of RNA biochemistry*. Weinheim, Germany: Wiley-VCH. p. 151–171.
- Kapitonov VV, Jurka J. 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol*. 3:e181.
- Keese PK, Gibbs A. 1992. Origins of genes: “big bang” or continuous creation? *Proc Natl Acad Sci U S A*. 89:9489–9493.
- Kim J, Martignetti JA, Shen MR, Brosius J, Deininger P. 1994. Rodent BC1 RNA gene as a master gene for ID element amplification. *Proc Natl Acad Sci U S A*. 91:3607–3611.
- Kishore S, Stamm S. 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311:230–232.
- Kiss T. 2004. Biogenesis of small nuclear RNPs. *J Cell Sci*. 117:5949–5951.
- Kiss T, Fayet E, Jady BE, Richard P, Weber M. 2006. Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb Symp Quant Biol*. 71:407–417.
- Kiss T, Filipowicz W. 1995. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev*. 9:1411–1424.
- Klein DJ, Schmeing TM, Moore PB, Steitz TA. 2001. The kink-turn: a new RNA secondary structure motif. *EMBO J*. 20:4214–4221.
- Krull M, Brosius J, Schmitz J. 2005. Alu-SINE exonization: en route to protein-coding function. *Mol Biol Evol*. 22:1702–1711.
- Krull M, Petrusma M, Makalowski W, Brosius J, Schmitz J. 2007. Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res*. 17:1139–1145.
- Kuhn JF, Tran EJ, Maxwell ES. 2002. Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res*. 30:931–941.
- Lafontaine DL, Tollervey D. 2000. Synthesis and assembly of the box C+D small nucleolar RNPs. *Mol Cell Biol*. 20:2650–2659.
- Landers M, et al. 2004. Regulation of the large (approximately 1000kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. *Nucleic Acids Res*. 32:3480–3492.
- Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res*. 34:D158–D162.
- Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* 300:1288–1291.
- Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci U S A*. 103:9935–9939.
- Lin J, et al. 2011. Structural basis for site-specific ribose methylation by box C/D RNA protein complexes. *Nature* 469:559–563.
- Maden BE. 2001. Mapping 2'-O-methyl groups in ribosomal RNA. *Methods* 25:374–382.
- Maniatis T, Fritsch EF, Sambrook J. 1989. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press.
- Matsumura S, Ikawa Y, Inoue T. 2003. Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res*. 31:5544–5551.
- Nahkuri S, Taft RJ, Korbie DJ, Mattick JS. 2008. Molecular evolution of the HBII-52 snoRNA cluster. *J Mol Biol*. 381:810–815.
- Neme R, Tautz D. 2013. Phylogenetic patterns of emergence of new genes support a model of frequent de novo evolution. *BMC Genomics* 14:117.
- Ohno S. 1970. *Evolution by gene duplication*. New York: Springer-Verlag.
- Raabe CA, et al. 2010. A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum*. *Nucleic Acids Res*. 38:608–617.
- Richard P, Kiss T. 2006. Integrating snoRNP assembly with mRNA biogenesis. *EMBO Rep*. 7:590–592.
- Rozhdestvensky TS, et al. 2003. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res*. 31:869–877.
- Runte M, Varon R, Horn D, Horsthemke B, Buiting K. 2005. Exclusion of the C/D box snoRNA gene cluster HBII-52 from a major role in Prader-Willi syndrome. *Hum Genet*. 116:228–230.
- Sahoo T, et al. 2008. Prader-Willi phenotype caused by paternal deficiency for the HBII-85C/D box small nucleolar RNA cluster. *Nat Genet*. 40:719–721.
- Schmitz J, et al. 2008. Retroposed SNOfall—a mammalian-wide comparison of platypus snoRNAs. *Genome Res*. 18:1005–1010.
- Skryabin BV, et al. 2007. Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet*. 3:e235.
- Stam NJ, et al. 1994. Genomic organisation and functional expression of the gene encoding the human serotonin 5-HT_{2C} receptor. *Eur J Pharmacol*. 269:339–348.
- Turner B, Lilley DM. 2008. The importance of G.A hydrogen bonding in the metal ion- and protein-induced folding of a kink turn RNA. *J Mol Biol*. 381:431–442.
- Tycowski KT, Shu MD, Steitz JA. 1993. A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. *Genes Dev*. 7:1176–1190.
- Vitali P, et al. 2003. Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res*. 31:6543–6551.
- Vitali P, et al. 2005. ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J Cell Biol*. 169:745–753.
- Warrens AN, Jones MD, Lechler RI. 1997. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene* 186:29–35.
- Watkins NJ, Dickmanns A, Luhrmann R. 2002. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol*. 22: 8342–8352.
- Watkins NJ, Leverette RD, Xia L, Andrews MT, Maxwell ES. 1996. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. *RNA* 2:118–133.
- Watkins NJ, et al. 2000. A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* 103:457–466.
- Weber MJ. 2006. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet*. 2:e205.
- Wirth J, et al. 2001. A translocation breakpoint cluster disrupts the newly defined 3' end of the SNURF-SNRPN transcription unit on chromosome 15. *Hum Mol Genet*. 10:201–210.
- Yin QF, et al. 2012. Long noncoding RNAs with snoRNA ends. *Mol Cell*. 48:219–230.
- Youssef OA, Terns RM, Terns MP. 2007. Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP. *Nucleic Acids Res*. 35:6196–6206.
- Zemann A, op de Bekke A, Kiefmann M, Brosius J, Schmitz J. 2006. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res*. 34:2676–2685.

Associate editor: Dan Graur