Ribosomal protein SA and its pseudogenes in ruminants: an extremely conserved gene family

A. Van den Broeke, M. Van Poucke, A. Van Zeveren, L.J. Peelman

Department of Nutrition, Genetics and Ethology, Faculty of Veterinary Medicine, Ghent University, Merelbeke, Belgium

ABSTRACT: The ribosomal protein SA (RPSA), also known as 37-kDa laminin receptor precursor/67-kDa laminin receptor (LRP/LR), has been identified as a multifunctional protein, playing an important role in multiple pathologies like cancer and prion diseases. Since RPSA is involved in the binding and internalization of the prion protein, mutations in the ovine *RPSA* gene, influencing the RPSA-PrP^{C}/PrP^{Sc} binding, can potentially play a part in the resistance to prion diseases. Our goal was to further characterize the complex *RPSA* gene family and to detect structural mutations which can play a role in this disease. In a prior study, 11 ovine pseudogenes were detected experimentally. As the whole genome shotgun ovine genome became accessible, an *in silico* genome-wide screening was performed and 37 new pseudogenes (36 processed and one semi-processed pseudogene) were detected, bringing the total to 48 ovine *RPSA* pseudogenes. Additionally, the complete bovine genome was screened *in silico* and 56 pseudogenes were identified. Once these sequences were known, it was possible to analyze the presence of mutations in the coding sequence and exon-flanking regions of the ovine functional full-length *RPSA* gene without the interference of pseudogenic sequences. Nineteen mutations were found: one in the 5' UTR, a silent one in the coding region, and seventeen in the exon-flanking regions, including an interesting mutation in the *SNORA62* gene, localized in intron 4 of *RPSA*, leading to potential ribosomal defects. Structural mutations of the *RPSA* gene can be ruled out to play a role in transmissible spongiform encephalopathies but regulatory mutations still can have an effect on these diseases.

Keywords: laminin receptor; mutation detection; prion; polymorphism; RPSA; sequence conservation; 37 kDA laminin receptor precursor/67-kDA laminin receptor

The ribosomal protein SA (RPSA), also designated as the 37-kDa laminin receptor precursor (LRP)/67- kDa laminin receptor (LR), is a protein that is involved in a broad range of functions. It is located on the cell surface as well as in the cytoplasm, the perinuclear compartment, and the nucleus (Nelson et al., 2008). In the cytoplasm, the protein is involved in the maturation of the 40S ribosomal subunit, and in the nucleus it is associated with histones (Ardini et al., 1998; Kinoshita et al., 1998). The function in the perinuclear compartment has yet to be elucidated but Sato et al. (1996) postulate that RPSA is a ligand protein between the nuclear envelope and chromatin DNA. As cell surface receptor it binds with high affinity to laminin, an extracellular matrix protein, hence playing an important role in tumor invasion and metastasis. Furthermore, it binds to other extracellular matrix molecules like elastin and carbohydrates. Besides cellular ligand, it acts as a receptor for different pathogens, e.g. viruses and respiratory tract pathogens (Nelson et al., 2008; Orihuela et al., 2009). RPSA also plays an important role in transmissible spongiform encephalopathies (TSEs). It not only acts as a receptor for cellular prion proteins (PrP^{C}) and infectious prions (PrP^{Sc}), but it is also involved in the propagation of prion diseases as well (Gauczynski et al., 2001, 2006; Leucht et al., 2003).

The RPSA-PrP^{C}/PrP^{Sc} interaction was proven by using the yeast two-hybrid technology and by *in*

Supported by the Agency for Innovation by Science and Technology in Flanders (IWT), Belgium.

vivo experiments using various cell lines or tissue lysates (Gauczynski et al., 2001, 2006). Additionally, the interacting binding domains of RPSA and PrP^C were determined. On RPSA there are two binding domains, a direct binding domain located between aa 161 and 179 (encoded by the last part of exon 4 and the first part of exon 5) and an indirect heparan sulfate proteoglycan (HSPG)-dependent binding domain that is presumed to be located between aa 180 and 285 (encoded by the last part of exon 5, exon 6, and the first part of exon 7) (Hundt et al., 2001). On PrP^C, the domain between aa 144 and 179 was identified as the direct binding domain to RPSA and the domain between aa 53 and 93 as the indirect HSPG-dependent binding domain to RPSA (Hundt et al., 2001).

Differences in the amino acids involved in the RPSA-PrP^C/PrP^{Sc} interaction could lead to variability in scrapie susceptibility. It is well established that the susceptibility to scrapie is influenced by polymorphisms of the *PRNP* gene encoding the PrP protein. In classical scrapie, codons 136 (A or V), 154 (R or H), and 171 (R, Q or H) are the major polymorphisms associated with variability in scrapie susceptibility. In atypical scrapie on the other hand, it is codon 154 (R or H), besides codon 141 (L or H), that has the greatest influence (Moum et al., 2005; Hunter, 2007).

It was already established that there is a high degree of sequence conservation between mammalian RPSA proteins, all of which are almost identical. The human RPSA protein has an amino acid identity of 98.99% with its porcine and murine ortholog and 98.3% with its bovine ortholog, as demonstrated by Knorr et al. (2007). RPSA exhibits multiple functions that are important for the cell viability, if not essential, and this is often reflected in extreme sequence conservation. RPSA proteins in vertebrates are sharing the ribosomal functions with their homologs in invertebrates and consequently have the lowest divergence in the first two-thirds of the protein where that function is localized. Later on in evolution, they acquired a laminin binding potential that is situated at the C-terminal of the protein and that is also essential for cell viability, as has recently been demonstrated by Scheiman et al. (2010). This explains why also the last part of the sequence is highly conserved in mammals.

On the other hand, it is possible that there are polymorphisms that do not affect the laminin binding capacity or ribosomal functions but have an effect on the RPSA-PrP^C/PrP^{Sc} interaction. They could partially explain the species barrier which makes it more difficult to transmit a certain prion strain from one species to another which leads to longer incubation times. Furthermore, some species like rabbits, pigs, and dogs seem to be completely resistant to natural infection of prion diseases and this resistance is probably a consequence of the conformation of the prion protein of the host species and therefore the amino acid sequence of PrP (Lysek et al., 2005). Nevertheless, as the RPSA protein is necessary for prion propagation, differences in both the RPSA conformation and the PRP conformation can affect the binding between both proteins and have a putative role in the strength of the species barrier (Marcos-Carcavilla et al., 2008).

Mutation detection of the *RPSA* gene, however, has been hampered by the presence of multiple pseudogenes, with sequences highly similar to the functional full-length gene. Previously, we identified 11 ovine *RPSA* pseudogenes experimentally and named them *RPSAP1* to *RPSAP11* (Van den Broeke et al., 2010). Now that the sheep genome became accessible, we wanted to identify *in silico* as many additional ovine pseudogenes as possible with a sequence similarity with the functional full-length gene of at least 60% and study the conservation with their bovine orthologs. Once these sequences were known, it was possible to analyze the presence of mutations in the coding sequence (CDS) and exon flanking regions of the functional full-length ovine *RPSA* gene without the interference of pseudogenic sequences.

MATERIAL AND METHODS

In silico identification of ovine and bovine *RPSA* pseudogenes

The whole genome shotgun (WGS) assembly *Ovis_aries*_1.0 (GenBank: ACIV010000000) is the current draft assembly of the sheep genome, composed of short ovine sequences (± 100–1000 bp long), aligned using the bovine genome (reference assembly, based on Btau_4.0) as the initial guide, but still containing a lot of gaps (Archibald et al., 2010). This database was screened via BLAST (Altschul et al., 1990) with the ovine *RPSA* mRNA sequence (GenBank: EF649775), the ovine *RPSA* genomic sequence (GenBank: GQ202529), and the ovine and bovine *RPSA* pseudogene sequences previously described by Van den Broeke et al. (2010).

The bovine genome on the other hand is completely sequenced (The Bovine Genome Sequencing and Analysis Consortium et al., 2009). This genome (reference assembly, based on Btau_4.6.1) was screened via BLAST (Altschul et al., 1990) with the bovine *RPSA* mRNA sequence (GenBank: NM_174379), the bovine *RPSA* genomic sequence (GenBank: NC_007320.5), and the bovine and ovine *RPSA* pseudogene sequences described by Van den Broeke et al. (2010).

Potential ovine and bovine pseudogenic matches were classified as pseudogenes if they had a homology of at least 60% with their *RPSA* gene. In one case (*RPSAP28)*, the homology between the ovine pseudogene and *RPSA* was less than 60% but the pseudogenic sequence was classified as pseudogene since the homology of its bovine ortholog with bovine *RPSA* was 64%. A second requirement was a minimum length of 100 bp and a homology with at least 2 exons of the functional full-length *RPSA* gene.

Open reading frames were detected with the online program NCBI Open Reading Frame Finder (www.ncbi.nlm.nih.gov/gorf/gorf.html) and sequences repeated with the RepeatMasker program (www.repeatmasker.org).

It was not possible to identify an ortholog of a particular ovine or bovine pseudogene by performing a screening via BLAST (Altschul et al., 1990) on the bovine or ovine genome respectively, using the sequence of that pseudogene because this resulted in multiple pseudogenes with sometimes higher sequence similarity with each other than with their putative ortholog. As orthologs of genes are located in syntenic regions, the flanking sequences of the pseudogenes were used to identify the corresponding orthologs. Pseudogenes were classified as orthologs if their sequence and 200 bp upstream and downstream of their sequence shared a homology of at least 85%.

Mutation detection of the *RPSA* gene in sheep

Genomic DNA was isolated from 100 µl blood, via a proteinase K lysis as described in Van Poucke et al. (2005), of 33 unrelated Belgian sheep covering 7 breeds (5 Ardense Voskop, 4 Bleu du Maine, 5 Hampshire Down, 5 Rouge de l'Ouest, 4 Suffolk, 5 Texel, and 5 Vlaams Kuddeschaap), 9 *PRNP* genotypes (based on codon 136, 154, and 171), and both sexes (Table 1). Five primer pairs, amplifying the whole coding and the exon-flanking region of *RPSA*, were developed with the software Primer3 (Rozen and Skaletsky, 2000) taking into account potential secondary structures of the amplicon by analysis with Mfold (Zuker, 2003). All primer

Table 1. Characteristics of the samples used in the mutation analysis

Sample	Breed	Sex	Genotype PRNP	NSP class*
1	Bleu du Maine	F	ARR/ARR	NSP1
2	Vlaams Kuddeschaap	M	ARR/ARR	NSP1
3	Ardense Voskop	F	ARR/ARR	NSP1
4	Suffolk	M	ARR/ARR	NSP1
5	Rouge de l'Ouest	M	ARR/ARR	NSP1
6	Rouge de l'Ouest	M	ARR/ARR	NSP1
7	Hampshire Down	F	ARR/ARR	NSP1
8	Ardense Voskop	M	ARR/ARQ	NSP2
9	Texel	F	ARR/ARQ	NSP2
10	Suffolk	M	ARR/ARQ	NSP2
11	Suffolk	F	ARR/ARQ	NSP2
12	Rouge de l'Ouest	F	ARR/ARQ	NSP2
13	Rouge de l'Ouest	F	ARR/ARQ	NSP2
14	Hampshire Down	F	ARR/ARQ	NSP2
15	Hampshire Down	F	ARR/ARQ	NSP2
16	Texel	M	ARR/ARH	NSP2
17	Ardense Voskop	F	ARR/AHQ	NSP2
18	Vlaams Kuddeschaap	F	ARR/AHQ	NSP2
19	Bleu du Maine	F	ARQ/ARQ	NSP3
20	Vlaams Kuddeschaap	F	ARQ/ARQ	NSP3
21	Vlaams Kuddeschaap	M	ARQ/ARQ	NSP3
22	Texel	F	ARQ/ARQ	NSP3
23	Suffolk	F	ARQ/ARQ	NSP3
24	Hampshire Down	F	ARQ/ARQ	NSP3
25	Hampshire Down	M	ARQ/ARQ	NSP3
26	Ardense Voskop	M	ARQ/ARQ	NSP3
27	Texel	F	ARQ/ARH	NSP3
28	Ardense Voskop	F	ARQ/AHQ	NSP3
29	Bleu du Maine	F	VRQ/ARR	NSP4
30	Vlaams Kuddeschaap	M	VRQ/ARR	NSP4
31	Texel	M	VRQ/ARR	NSP4
32	Rouge de l'Ouest	M	VRQ/ARR	NSP4
33	Bleu du Maine	M	VRQ/ARQ	NSP5

*NSP classification system used in the United Kingdom for genetic resistance to scrapie with NSP1 class, the genetically most resistant class, and NSP5, the genetically most susceptible class
F = female, M = male

Table 2. Characteristics of the primerpairs used in the mutation analysis

Primer	Primer sequence	Annealing temperature (°C)	Amplicon length (bp)	Location	Position (GenBank: GQ202529)
Amplicon 1 F	CAGAGGTTTGTTCAGTGCTTTCC	60	804	5'UTR	
Amplicon 1 R	AAATGGGGTGTGCGTGTGT			intron 1	649–631
Amplicon 2 F	TTAGAAGGAATGCTGAAGG	58	817	intron 1	1041–1059
Amplicon 2 R	ACGACACATAACCTACCAGTT			intron 2	1957–1837
Amplicon 3 F	GAGATAGAAGCACGGAAGGATT	62	712	intron 2	2269–2290
Amplicon 3 R	GCGGCTCAAGAAAATACACA			intron 3	2980–2961
Amplicon 4 F	GCTTGCTTGGTGACATTGC	62	748	intron 3	11504–11522
Amplicon 4 R	AACCTCTGCCCCGTTCTTAT			intron 4	12252–12233
Amplicon 5 F	ACCAAGGGACCTAGACGATGA	64	867	intron 4	12477–12497
Amplicon 5 R	CCCATAGACGGAAATAAATGAACAC			3'UTR	

pairs were designed to amplify the *RPSA* gene and none of the *RPSA* pseudogenes (for amplicon characteristics see Table 2).

PCR was performed with 0.5 U FastStart Taq DNA Polymerase (Roche Diagnostics Belgium, Vilvoorde, Belgium), 2.0mM $MgCl_2$, 200µM (each) dNTPs (Bioline Reagents Ltd., London, UK), 500nM of each primer, and 200 ng DNA. PCR conditions were 5 min at 95°C, 40 cycles of 30 s at 95°C, 30 s at the annealing temperature, 1 min at 72°C, and a final 10 min elongation at 72°C. All amplicons were sequenced as previously described in Van den Broeke et al. (2010). Sequencing data were analyzed with the Clustal W program (Larkin et al., 2007).

Promoter elements and putative transcription factor binding sites were identified with programs as neural network promoter prediction (Reese, 2001), Cister (Frith et al., 2001), Signal scan (Prestridge, 1991), and TFSEARCH (Heinemeyer et al., 1998). miRBase was used to find possible targets of microRNAs in the *RPSA* gene (Kozomara and Griffiths-Jones, 2011).

RESULTS AND DISCUSSION

In silico identification of ovine and bovine *RPSA* pseudogenes

The availability of the *Ovis aries* 1.0 genome assembly (Archibald et al., 2010) enabled us to identify *in silico* 37 *RPSA* pseudogenes additional to the ones that were already discovered experimentally (Van den Broeke et al., 2010). This brings the total to 48 ovine *RPSA* pseudogenes (Table 3, Figure 1), a number that is in the same range of number of pseudogenes discovered by Balasubramanian et al. (2009) in their *in silico* genome-wide screening study in fully sequenced genomes, namely 63 processed *RPSA* pseudogenes in human, 45 in mice, 52 in chimpanzee, and 45 in rat. Not all the experimentally discovered pseudogenes of our previous study were found in the whole genome shotgun sheep genome (Van den Broeke et al., 2010). The recent species specific ovine pseudogenes *RPSAP1,2,5,7-10* were not present in the WGS sheep genome build. Because the International Sheep Genomics Consortium (ISGC) aligned ovine sequences to the bovine genome to create the whole genome shotgun sheep genome, we suppose that the pseudogenic sequences, which did not show any similarity with the bovine genome, could not be aligned to the bovine genome and thus were not included in the WGS build (Archibald et al., 2010). The ovine pseudogenes share from less than 60–98% nucleic acid sequence identity with the ovine functional full-length *RPSA* gene. The pseudogenes with a lower similarity were not detected due to our study design.

One newly found pseudogene (*RPSAP19*) is a rare "semi-processed" pseudogene still possessing a remnant intron 4 bearing the *SNORA62* gene. The term "semi-processed pseudogene" was first proposed by Zhang et al. (2008) and is used for pseudogenes generated by retrotransposition of partially spliced premature mRNA. The other two "semi-processed" pseudogenes of the ovine *RPSA* gene family, *RPSAP8* and *RPSAP9* (GenBank: GQ202537 and GQ202538), also bear intron 4 including the sequence of the *SNORA62* gene. snoRNAs are often localized within introns

Figure 1. Schematic overview of ovine (ov) and bovine (bov) pseudogenes

Exons present in both ovine and bovine pseudogenes (in black), exons only found in one of the orthologs (dotted bar), regions with missing sequences from the NCBI database (stripped bar), repeated sequences that are inserted in the pseudogenes and the *SNORA62* gene (in grey)

Table 3. Characteristics of the ovine and bovine *RPSA* pseudogenes

Name	Ovine pseudogene			Bovine pseudogene				
	accession No.	SI* ovine *RPSA* (%)	SI* bovine ortholog (%)	accession No.	range	SI* bovine *RPSA* (%)	location	feature
RPSAP2	ACIV010279956/57	74	91	NC_007301.5	16174977–5951	75	in *DAP3*	
RPSAP3	ACIV010712079/80	80	95	NC_007305.5	51656775–7870	81	intergenic	
RPSAP4	ACIV011799502/03	83	91	NC_007320.5	2904735–5521	78	intergenic	*LOC100336434*
RPSAP6	ACIV011534191/92	80	92	NC_007315.5	65680450–1539	81	intergenic	
RPSAP11	ACIV011903422	76	91	NC_007325.5	22503990–5049	77	intergenic	
RPSAP12	ACIV010009098	71	93	NC_007299.5	107388186–9062	72	intergenic	
RPSAP13	ACIV010059570/71	75	94	NC_007299.5	21075516–6747	74	intergenic	*LOC781313: RPSA*-like
RPSAP14	ACIV010071859	81	94	NC_007299.5	35555129–6675	76	intergenic	
RPSAP15	ACIV010172107	67	91	NC_007300.5	14516593–7247	69	intergenic	
RPSAP16	ACIV010240662	87	97	NC_007300.5	9868953–9366	77	intergenic	
RPSAP17	ACIV010343175/76/77	85	95	NC_007301.5	87426328–7261	77	intergenic	
RPSAP18	ACIV010344762/63	77	95	NC_007301.5	89553151–4114	74	intergenic	
RPSAP19	ACIV010319172	79	92	NC_007301.5	60009518–937	86	intergenic	*LOC100295707*
RPSAP20	ACIV010349106	74	88	NC_007301.5	93522075–339	85	intergenic	*LOC100295233*
RPSAP21	ACIV010291498/99	70	94	NC_007301.5	29984816–5704	75	intergenic	
RPSAP22	ACIV010384986	82	95	NC_007302.5	19529087–496	83	intergenic	*LOC100336308: RPSA*-like
RPSAP23	ACIV010435878/79/80	74	95	NC_007302.5	74842869–3978	77	in *ZNF804B*	*RPSAP1*
RPSAP24	ACIV010491485	70	92	NC_007303.5	20906129–7028	70	intergenic	
RPSAP25	ACIV010527707	90	90	NC_007303.5	60546599–7030	92	intergenic	
RPSAP26	ACIV010495714	73	96	NC_007303.5	24897357–908	73	intergenic	
RPSAP27	ACIV010583245/46/47	72	91	NC_007304.5	12231035–946	71	intergenic	
RPSAP28	ACIV010645497	< 60	94	NC_007304.5	8752937–349	64	intergenic	
RPSAP29	ACIV010706330/31/32	70	94	NC_007305.5	46521540–2253	69	intergenic	
RPSAP30	ACIV010688073/74	75	97	NC_007305.5	26665996–6856	69	intergenic	
RPSAP31	ACIV010870568/69	80	92	NC_007307.5	25005295–6244	78	intergenic	*LOC100139221: RPSA*-like
RPSAP32	ACIV010965166/67	75	95	NC_007308.5	31709930–10694	74	intergenic	
RPSAP33	ACIV011173083	73	93	NC_007310.5	59551557–2032	71	intergenic	

RPSAP34	ACIV011182489/855/856	73	92	NC_007310.5	7464603–5384	72	intergenic	
RPSAP35	ACIV011141860/61	73	93	NC_007310.5	25331128–989	70	intergenic	
RPSAP36	ACIV011267035	74	92	NC_007311.5	76527703–8027	70	in *EYA2*	
RPSAP37	ACIV011422416/17	75	94	NC_007314.5	15416879–7876	71	intergenic	
RPSAP38	ACIV011480743	79	95	NC_007314.5	8270309–823	82	intergenic	
RPSAP39	ACIV011465309	73	90	NC_007314.5	6475545–850	73	intergenic	
RPSAP40	ACIV011803302/03	70	93	NC_007320.5	25190706–754	73	intergenic	
RPSAP41	ACIV011873056	73	93	NC_007324.5	39260430–710	78	intergenic	
RPSAP42	ACIV0120867/68/69	82	92	NC_007329.5	23314339–3486	81	intergenic	*LOC100139311: RPSA*-like
RPSAP43	ACIV012088199/200	72	94	NC_007329.5	25035168–6187	73	intergenic	
RPSAP44	ACIV012097249/50/51	80	95	NC_007329.5	33593767–2851	71	intergenic	*LOC786412: RPSA*-like
RPSAP45	ACIV012126348	72	93	NC_007330.5	23521114–2634	72	intergenic	
RPSAP46	ACIV012192492	75	92	NC_007331.4	59466130–5191	75	intergenic	*LOC782805: RPSA*-like
RPSAP47	ACIV012199192	75	95	NC_007331.4	67228718–9530	74	intergenic	*LOC786360: RPSA*-like
RPSAP48	ACIV012190019/20/21	82	94	NC_007331.4	55328763–9702	79	in *CCDC22*	
RPSAP49				NC_007299.5	108688755–9571	70	intergenic	
RPSAP50				NC_007301.5	125970082–69189	71	intergenic	*LOC100141114: RPSA*-like
RPSAP51				NC_007302.5	19122655–781	83	intergenic	*LOC100297616: RPS24*-like
RPSAP52				NC_007302.5	123658944–9840	71	intergenic	*LOC100296290: RPSA*-like
RPSAP53				NC_007308.5	64883466–670	74	intergenic	
RPSAP54				NC_007314.5	16378105–596	70	intergenic	
RPSAP55				NC_007317.5	64776546–7462	72	intergenic	*LOC781120: RPSA*-like
RPSAP56				NC_007325.5	62698525–9833	73	intergenic	*LOC786678*
RPSAP57				NC_007326.5	44073985–4896	74	intergenic	*LOC783961: RPSA*-like
RPSAP58				NC_007327.5	5989909–197	74	intergenic	*LOC786170: RPSA*-like
RPSAP59				NC_007329.5	378146–9070	78	intergenic	*LOC789763: RPSA*-like
RPSAP60				NC_007329.5	45961803–2621	73	intergenic	*LOC784951: RPSA*-like
RPSAP61				NC_007330.5	23846784–6989	84	intergenic	
RPSAP62				NC_007330.5	20780510–1413	66	intergenic	

experimentally verified pseudogenes are given in grey boxes

*sequence identity

of non-protein-coding genes (Dieci et al., 2009). They mostly do not have an independent promoter but are synthesized cotranscriptionally with their host genes. They are processed from pre-mRNA by exonucleolytic digestion of the debranched lariat (Kiss et al., 2004). "Semi-processed" pseudogenes are generated presumably from partially spliced premature mRNA. Normally, the spliceosome removes all the intronic fragments from the primary RNA transcripts. When an intron is still present in the pre-mRNA and it is reversely transcribed in cDNA, semi-processed pseudogenes arise. It is remarkable that all "semi-processed" pseudogenes of the *RPSA* gene family carry a snoRNA gene. Probably, the splicing of the introns with snoRNA gene is hampered in one way or another.

All the other pseudogenes are processed and are exhibiting one or more typical features of processed pseudogenes. Some pseudogenes are disrupted by inserted sequences, mostly SINE (e.g. *RPSAP38* and *RPSAP43*), while other pseudogenes have large sequence deletions (e.g. *RPSAP4* and *RPSAP35*) (Figure 1). Most pseudogenes carry frameshift mutations or have premature stop codons in their sequences. The ORF of *RPSAP7* (295 amino acids with 99% sequence identity to the functional RPSA protein), discovered in our previous study, was the longest ORF of all 48 pseudogenes.

A BLAST analysis of the updated bovine genome (reference assembly, based on Btau_4.6.1) identified 56 potential bovine *RPSA* family members. In our previous study 60 bovine *RPSA* family members were identified *in silico* (reference assembly, based on Btau_4.0) but some of them were withdrawn from the current reference assembly. For instance, some unplaced genomic scaffold turned out to be part of already annotated regions and therefore some pseudogenes with two LOC names were actually the same pseudogene (e.g. LOC789047 withdrawn from current assembly is the same pseudogene as LOC783961). Some parts of certain bovine pseudogenes are designated in GenBank as "ribosomal protein SA-like" by an automated computational analysis tool (see Table 3). However, the annotation is often not complete or incorrect. For instance, 3000 bp of the flanking intergenic region of the pseudogene *RPSAP47* are annotated as a component of the pseudogene in LOC786360. LOC100297616 (*RPSAP51)* is annotated as "ribosomal protein S24-like" but is actually a pseudogene of *RPSA*. In LOC781120 (*RPSAP55*), some parts of the sequence are labeled as exons and other parts as introns but this is a processed pseudogene. In this study, the position of the pseudogenes and their features (the present exons, insertions, and deletions) are determined through careful sequence analysis and differ largely from the automatically generated data.

The bovine pseudogenes are displaying 64–92% sequence identity with the bovine *RPSA* gene. Thirty-three bovine pseudogenes can be catalogued as "processed pseudogenes" and 17 as "pseudogenic fragments" as they have lengths less than 70% of the parent protein (Zhang et al., 2004). The ovine pseudogenes cannot be catalogued because their sequence is often incomplete due to the absence of a complete version of the sheep genome (Figure 1, gaps in the sequence in green).

Most processed pseudogenes described in literature are more truncated at the 5' site than at the 3' site. Processed pseudogenes arise by incorporation of transcribed cDNA. As the reverse transcription process starts at the poly-A tail (3' site), the 5' site will be incomplete when there is premature termination of the process (Zhang et al., 2002). However, in our group of processed pseudogenes, the truncation is the highest at the 3' site and more experimental data is needed to clarify this inconsistency.

Forty-two ovine pseudogenes have a bovine ortholog. The ovine and bovine orthologs share 88–97% nucleic acid identity with each other. Six of the experimentally discovered ovine pseudogenes (*RPSAP1*, *RPSAP5*, *RPSAP7*, *RPSAP8*, *RPSAP9*, and *RPSAP10*) do not have a bovine ortholog. The surrounding sequences were found without the interruption of a bovine pseudogene. For fourteen bovine pseudogenes, no ovine ortholog could be found. As the current sheep genome sequence is a pool of short ovine sequences with a lot of gaps, this is not evidence that the ovine ortholog does not exist. As the whole sheep genome becomes accessible, ovine orthologs of those bovine pseudogenes can be excluded completely. No conclusions on chromosomal rearrangements between the ovine and bovine genome can be made because the location on the ovine chromosomes is not experimentally verified.

Mutation detection of the *RPSA* gene in sheep

In the present work, a mutation analysis was conducted to detect structural mutations of the

ovine *RPSA* gene. When carrying out a mutation study of genes from gene families, one has to pay attention that only the desired gene is amplified and none of the related (pseudo)genes. In this mutation study, the strategy was to use exon-spanning primers situated in the introns as all of the discovered ovine pseudogenes were processed or semi-processed pseudogenes bearing a single intron. Because the genome of the sheep has not been fully sequenced yet, there is a chance that some ovine pseudogenes have not yet been discovered. However, we screened the fully sequenced bovine genome and discovered only processed or semi-processed pseudogenes bearing a single intron (all ± 90% sequence identity with their ovine ortholog) and therefore we can assume that it is very likely that the ovine *RPSA* gene family also consists of those types of pseudogenes. This strategy was successful as none of the obtained data gave any evidence of co-amplifying pseudogenes. The mutation analysis of the whole coding and the exon-flanking region of *RPSA* was carried out on 33 unrelated sheep covering 7 different breeds, varying in *PRNP* genotype at codons 136, 154, and 171. Nineteen mutations were found: one in the 5' UTR, seventeen in different introns, and one in the coding region (Table 4). In four out of the nineteen mutations (10, 12, 15, and 19), a high percentage of the individuals (> 84.8%) were homozygous for one genotype and no homozygotes of the other genotype were detected. This could imply that one homozygous genotype is lethal but this conclusion can only be made if more animals are tested. Mutation 1 is situated in the first exon that is a part of the 5' UTR. Transcription factor elements, including a TATA box, were detected using online prediction programs but mutation 1 was not situated in any of the detected transcription factor elements. The mutations 2–17 and 19 are mutations in introns. None of the mutations are disrupting a splice site. On the other hand, the SNP 10 and the indel 15 are part of the small nucleolar RNAs *SNORA6* and *SNORA62* respectively. These small nucleolar RNAs, located in introns 2 and 4 of *RPSA*, are H/ACA box snoRNAs that guide the isomerization of uridine into pseudouridine. They are characterized by two imperfect hairpins that contain two short antisense sequences that can base pair upstream and downstream of the targeted, unpaired uridine (Ganot et al., 1997). *SNORA62* forms a base pair with the uridine at position 3830 and 3832 of 28S rRNA. The indel 15 affects the binding site of *SNORA62* with the uridine at position 3830. Ni et al. (1997) demonstrated that a weakened binding site blocked the pseudouridylation of the uridine in question (Ni et al., 1997). Therefore, we can presume that the mutation causes a loss of pseudouridine in the 28S rRNA and this could result in misfolding of the rRNA with a reduced rate of processing and potential defects in the assembly of the ribosome as a consequence. As there were no homozygous animals detected carrying two indels, it is possible that the effect of the indel is lethal. SNP 10 on the other hand does not alter the H- or ACA-box or the 28S rRNA U3616 PU guide of *SNORA6*. The folding of the RNA stays unchanged, too.

Mutation 18, a T → C substitution at position 69 of exon 6 and a part of the indirect PrP-binding domain, is a silent mutation T232T. It is known for a while that synonymous mutations can affect protein expression levels (Sharp et al., 1986). Additionally, the protein folding can be influenced by a silent SNP, possibly changing the function of the protein (Kimchi-Sarfaty et al., 2007; Komar, 2007). It has yet to be examined if mutation 18 causes any of these events.

None of the mutations were part of a possible target of microRNA. Most of the mutations are equally represented in different breeds.

Mutation analyses of the *RPSA* gene were only carried out in Spanish and Chinese sheep breeds and recently one human study was published. Eight of our 19 mutations were also present in the Spanish breeds examined by Marcos-Carcavilla et al. (2008). They neither found polymorphisms that cause an amino acid change. The amino acid sequence of a local Chinese breed deposited by Qiao et al. (2009) however differs in four amino acids with the sequences mentioned above. These possible polymorphic amino acids cannot be confirmed by any EST of the NCBI databank. In the human mutation study 4 SNPs were observed, including a synonymous one in the coding region (exon 5) (Yun et al., 2011). An association study with sporadic Creutzfeldt-Jakob disease found no significant associations.

We can conclude that the CDS of the *RPSA* gene is extremely well conserved in sheep, even between sheep of very different breeds. We could not find polymorphisms in the coding region of the *RPSA* gene that can play a direct role in the RPSA-PrP^{C}/PrP^{Sc} interaction. Because the variability in scrapie susceptibility cannot be subscribed

to structural mutations in the *RPSA* gene, other strategies have to be examined for the treatment of scrapie. For example, several research groups are investigating the downregulation of RPSA in different species like human and mice (Leucht et al., 2003, 2004). Complete knock-out of RPSA is not possible because it results in apoptosis due to the multiple essential functions of the protein (Ardini et al., 1998; Scheiman et al., 2010).

The *RPSA* gene is very conserved not only in the ovine species, but also between different species. The three main ruminants, namely sheep, goat, and cattle, share 100% identity in their amino acid sequence (GenBank: ADE09296 (sheep), ADI56590 (goat), and DAA17125 (cattle)). Because the RPSA protein is 100% identical in the three species, the observed species barrier for transmission of certain prion strains cannot be ascribed to polymorphisms in the RPSA protein.

One can wonder if the two features of the *RPSA* gene observed in this study, namely a substantial amount of processed pseudogenes and extreme conservation, are linked to each other. Processed pseudogenes arise by retrotransposition of the mRNA of the ancestral gene into the genome. Because the fixation in the genome requires gene expression in the germ line, it is evident that the majority of gene families with several processed pseudogenes are often housekeeping genes which are highly expressed in the germ line (Zhang et al., 2004). Secondly, housekeeping genes are under stronger selective constraints than tissue-specific genes and, therefore, evolve more slowly, hence are more conserved (Zhang and Li, 2004). Moreover, housekeeping genes are significantly more likely to have orthologs in other species relative to other genes (She et al., 2009). It is therefore not surprising that characteristics like high conservation and multiple pseudogenes are associated with each other. Consequently, when performing molecular analysis of a conserved gene, one has to keep in mind that those genes are often housekeeping genes and that there is a high possibility that pseudogenes can interfere with for example sequencing, mapping, polymorphism detection, genotyping, association analysis, and mRNA expression studies.

CONCLUSION

Until now 48 ovine *RPSA* pseudogenes have been discovered. All of them are processed except for 3 semi-processed ovine pseudogenes bearing a snoRNA in their remnant intron 4. Fifty-six bovine *RPSA* pseudogenes were detected (55 processed and 1 semi-processed) out of which 42 are orthologs of ovine pseudogenes. In a mutation analysis of the whole coding and exon-flanking non-coding region of the functional full-length ovine *RPSA* gene, 19 mutations were discovered of out which 1 is positioned in the 5' UTR, 17 in the different introns, and 1 silent mutation in the coding region. An interesting mutation was revealed in the *SNORA62* gene, leading to potential ribosomal defects. No structural mutations that can play a direct role in the RPSA-PrP^{C}/PrP^{Sc} interaction were found but regulatory mutations in the ovine *RPSA* gene still can have an effect on prion diseases. Furthermore, it was established that sheep, goat, and cattle have 100% identical RPSA proteins. Consequently, differences in the RPSA proteins are not responsible for the observed species barrier in prion diseases.

Acknowledgement

The authors wish to thank Linda Impe for excellent technical assistance.

REFERENCES

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990): Basic local alignment search tool. Journal of Molecular Biology, 215, 403–410.

Archibald A.L., Cockett N.E., Dalrymple B.P., Faraut T., Kijas J.W., Maddox J.F., McEwan J.C., Hutton Oddy V., Raadsma H.W., Wade C., Wang J., Wang W., Xun X. (2010): The sheep genome reference sequence: a work in progress. Animal Genetics, 41, 449–453.

Ardini E., Pesole G., Tagliabue E., Magnifico A., Castronovo V., Sobel M.E., Colnaghi M.I., Ménard S. (1998): The 67-kDa laminin receptor originated from a ribosomal protein that acquired a dual function during evolution. Molecular Biology and Evolution, 15, 1017–1025.

Balasubramanian S., Zheng D., Liu Y.J., Fang G., Frankish A., Carriero N., Robilotto R., Cayting P., Gerstein M. (2009): Comparative analysis of processed ribosomal protein pseudogenes in four mammalian genomes. Genome Biology [serial online]. 10, R2. Available from http://genomebiology.com/content/10/1/R2 (accessed Feb 10, 2010).

Dieci G., Preti M., Montanini B. (2009): Eukaryotic snoRNAs: a paradigm for gene expression flexibility. Genomics, 94, 83–88.

Frith M.C., Hansen U., Weng Z. (2001): Detection of cis-element clusters in higher eukaryotic DNA. Bioinformatics, 17, 878–889.

Ganot P., Bortolin M.L., Kiss T. (1997): Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell, 89, 799–809.

Gauczynski S., Peyrin J.M., Haik S., Leucht C., Hundt C., Rieger R., Krasemann S., Deslys J.P., Dormont D., Lasmézas C.I., Weiss S. (2001): The 37-kDa/67-kDa laminin receptor acts as the cell-surface receptor for the cellular prion protein. EMBO Journal, 20, 5863–5875.

Gauczynski S., Nikles D., El-Gogo S., Papy-Garcia D., Rey C., Alban S., Barritault D., Lasmezas C.I., Weiss S. (2006): The 37-kDa/67-kDa laminin receptor acts as a receptor for infectious prions and is inhibited by polysulfated glycanes. Journal of Infectious Diseases, 194, 702–709.

Heinemeyer T., Wingender E., Reuter I., Hermjakob H., Kel A.E., Kel O.V., Ignatieva E.V., Ananko E.A., Podkolodnaya O.A., Kolpakov F.A., Podkolodny N.L., Kolchanov N.A. (1998): Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Research, 26, 362–367.

Hundt C., Peyrin J.M., Haik S., Gauczynski S., Leucht C., Rieger R., Riley M.L., Deslys J.P., Dormont D., Lasmézas C.I., Weiss S. (2001): Identification of interaction domains of the prion protein with its 37-kDa/67-kDa laminin receptor. EMBO Journal, 20, 5876–5886.

Hunter N. (2007): Scrapie: uncertainties, biology and molecular approaches. Biochimica et Biophysica Acta, 1772, 619–628.

Kimchi-Sarfaty C., Oh J.M., Kim I.W., Sauna Z.E., Calcagno A.M., Ambudkar S.V., Gottesman M.M. (2007): A "silent" polymorphism in the *MDR1* gene changes substrate specificity. Science, 315, 525–528.

Kinoshita K., Kaneda Y., Sato M., Saeki Y., Wataya-Kaneda M., Hoffmann A. (1998): LBP-p40 binds DNA tightly through associations with histones H2A, H2B, and H4. Biochemical and Biophysical Research Communications, 253, 277–282.

Kiss A.M., Jady B.E., Bertrand E., Kiss T. (2004): Human box H/ACA pseudouridylation guide RNA machinery. Molecular and Cellular Biology, 24, 5797–5807.

Knorr C., Beuermann C., Beck J., Brenig B. (2007): Characterization of the porcine multicopy ribosomal protein SA/37-kDa laminin receptor gene family. Gene, 395, 135–143.

Komar A.A. (2007): Genetics. SNPs, silent but not invisible. Science, 315, 466–467.

Kozomara A., Griffiths-Jones S. (2011): miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Research, 39, D152–D157.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. (2007): Clustal W and Clustal X version 2.0. Bioinformatics, 23, 2947–2948.

Leucht C., Simoneau S., Rey C., Vana K., Rieger R., Lasmezas C.I., Weiss S. (2003): The 37 kDa/67 kDa laminin receptor is required for PrP(Sc) propagation in scrapie-infected neuronal cells. EMBO Reports, 4, 290–295.

Leucht C., Vana K., Renner-Muller I., Dormont D., Lasmezas C.I., Wolf E., Weiss S. (2004): Knock-down of the 37-kDa/67-kDa laminin receptor in mouse brain by transgenic expression of specific antisense LRP RNA. Transgenic Research, 13, 81–85.

Lysek D.A., Schorn C., Nivon L.G., Esteve-Moya V., Christen B., Calzolai L., von Schroetter C., Fiorito F., Herrmann T., Guntert P., Wuthrich K. (2005): Prion protein NMR structures of cats, dogs, pigs, and sheep. Proceedings of the National Academy of Sciences of the United States of America, 102, 640–645.

Marcos-Carcavilla A., Calvo J.H., Gonzalez C., Serrano C., Moazami-Goudarzi K., Laurent P., Bertaud M., Hayes H., Beattie A.E., Lyahyai J., Martín-Burriel I., Torres J.M., Serrano M. (2008): Structural and functional analysis of the ovine laminin receptor gene (*RPSA*): Possible involvement of the LRP/LR protein in scrapie response. Mammalian Genome, 19, 92–105.

Moum T., Olsaker I., Hopp P., Moldal T., Valheim M., Benestad S.L. (2005): Polymorphisms at codons 141 and 154 in the ovine prion protein gene are associated with scrapie Nor98 cases. Journal of General Virology, 86, 231–235.

Nelson J., McFerran N.V., Pivato G., Chambers E., Doherty C., Steele D., Timson D.J. (2008): The 67 kDa laminin receptor: structure, function and role in disease. Bioscience Reports, 28, 33–48.

Ni J., Tien A.L., Fournier M.J. (1997): Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell, 89, 565–573.

Orihuela C.J., Mahdavi J., Thornton J., Mann B., Wooldridge K.G., Abouseada N., Oldfield N.J., Self T., Ala'Aldeen D.A., Tuomanen E.I. (2009): Laminin receptor initiates bacterial contact with the blood brain barrier in experimental meningitis models. Journal of Clinical Investigation, 119, 1638–1646.

Prestridge D.S. (1991): SIGNAL SCAN: a computer program that scans DNA sequences for eukaryotic transcriptional elements. Computer Applications in the Biosciences, 7, 203–206.

Qiao J., Su X., Wang Y., Yang J., Kouadir M., Zhou X., Yin X., Zhao D. (2009): Cloning and characterization of full-length coding sequence (CDS) of the ovine 37/67-kDa laminin receptor (RPSA). Molecular Biology Reports, 36, 2131–2137.

Reese M.G. (2001): Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Computers and Chemistry, 26, 51–56.

Rozen S., Skaletsky H. (2000): Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology, 132, 365–386.

Sato M., Kinoshita K., Kaneda Y., Saeki Y., Iwamatsu A., Tanaka K. (1996): Analysis of nuclear localization of laminin binding protein precursor p40 (LBP/p40). Biochemical and Biophysical Research Communications, 229, 896–901.

Scheiman J., Jamieson K.V., Ziello J., Tseng J.C., Meruelo D. (2010): Extraribosomal functions associated with the C terminus of the 37/67 kDa laminin receptor are required for maintaining cell viability. Cell Death and Disease [serial online]. 1, e42. Available from http://www.nature.com/cddis/journal/v1/n5/full/cddis201019a.html (accessed Oct 22, 2011).

Sharp P.M., Tuohy T.M., Mosurski K.R. (1986): Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Research, 14, 5125–5143.

She X., Rohl C.A., Castle J.C., Kulkarni A.V., Johnson J.M., Chen R. (2009): Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. BMC Genomics [serial online]. 10, 269. Available from http://www.biomedcentral.com/1471-2164/10/269 (accessed Sept 10, 2010).

The Bovine Genome Sequencing and Analysis Consortium, Elsik C.G., Tellam R.L., Worley K.C. (2009): The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science, 324, 522–528.

Van den Broeke A., Van Poucke M., Marcos-Carcavilla A., Hugot K., Hayes H., Bertaud M., Van Zeveren A., Peelman L.J. (2010): Characterization of the ovine ribosomal protein SA gene and its pseudogenes. BMC Genomics [serial online]. 11, 179. Available from http://www.biomedcentral.com/1471-2164/11/179 (accessed March 18, 2010).

Van Poucke M., Vandesompele J., Mattheeuws M., Van Zeveren A., Peelman L.J. (2005): A dual fluorescent multiprobe assay for prion protein genotyping in sheep. BMC Infectious Diseases [serial online]. 5, 13. Available from http://www.biomedcentral.com/1471-2334/5/13 (accessed Feb 21, 2009).

Yun J., Jin H.T., Lee Y.J., Choi E.K., Carp R.I., Jeong B.H., Kim Y.S. (2011): The first report of RPSA polymorphisms, also called 37/67 kDa LRP/LR gene, in sporadic Creutzfeldt-Jakob disease (CJD). BMC Medical Genetics [serial online]. 12, 108. Available from http://www.biomedcentral.com/1471-2350/12/108 (accessed Sept 20, 2011).

Zhang L., Li W.H. (2004): Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Molecular Biology and Evolution, 21, 236–239.

Zhang Z., Harrison P., Gerstein M. (2002): Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. Genome Research, 12, 1466–1482.

Zhang Z., Carriero N., Gerstein M. (2004): Comparative analysis of processed pseudogenes in the mouse and human genomes. Trends in Genetics, 20, 62–67.

Zhang Z.D., Cayting P., Weinstock G., Gerstein M. (2008): Analysis of nuclear receptor pseudogenes in vertebrates: how the silent tell their stories. Molecular Biology and Evolution, 25, 131–143.

Zuker M. (2003): Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research, 31, 3406–3415.

Received: 2012–02–17
Accepted after corrections: 2012–10–03

Corresponding Author

DVM Alice Van den Broeke, Department of Nutrition, Genetics and Ethology, Faculty of Veterinary Medicine, Ghent University, Heidestraat 19, B-9820 Merelbeke, Belgium
Tel. +329 264 78 01, e-mail: alice.vandenbroeke@UGent.be