The steroid receptor RNA activator is the first functional RNA encoding a protein

S. Chooniedass-Kothari[1], E. Emberley[1], M.K. Hamedani, S. Troup, X. Wang, A. Czosnek, F. Hube, M. Mutawe, P.H. Watson, E. Leygue*

Department of Biochemistry and Medical Genetics, Faculty of Medicine, University of Manitoba, P309 Pathology Building, 770 Bannatyne Avenue, Winnipeg, MB, Canada R3E 0W3

Received 11 March 2004; accepted 21 March 2004

Available online 20 April 2004

Edited by Robert Barouki

Abstract The steroid receptor RNA activator (SRA) has previously been characterized as belonging to the growing family of functional non-coding RNAs. However, we recently reported the Western blot detection of a putative endogenous SRA protein (SRAP) in breast cancer cells. Herein, we successfully suppressed the expression of this protein through specific RNA interference assay, unequivocally confirming its existence. Moreover, using database searches and Western blot analysis, we also showed that SRAP is highly conserved among chordata. Overall, our results suggest that SRA is the first example of a new class of functional RNAs also able to encode a protein.

Keywords: Steroid receptor RNA activator; Non-coding RNA; RNA interference; Steroid receptor RNA activator homologue; Steroid receptor RNA activator protein

1. Introduction

Even though RNAs have long been thought to be either messenger RNAs (mRNAs), transfer RNAs or ribosomal RNAs, it has become apparent in the last 20 years that many RNAs do not belong to any of these three subgroups [1]. The family of non-coding RNAs (ncRNAs), which groups all RNAs unable to encode a protein, is increasing exponentially [2]. Data for the new members of this growing family are actively gathered and corresponding information sorted on several websites such as http://biobases.ibch.poznan.pl/ncRNA/, http://rfam.wustl.edu/index.html or http://indiana.edu/~tmrna/. These RNAs which have no protein coding capacity have been shown to regulate several cellular processes as diverse as the subcellular distribution of RNAs and proteins, the modulation of protein function, or the transcriptional and translational regulation of gene expression [1,2]. The steroid receptor RNA activator (SRA) has recently been characterized as one such ncRNA that modulates steroid receptor transcriptional activity [3].

The originally described SRA sequences differed in their 5′ and 3′ ends, but were conserved in their central core region [3]. The core region was shown to be necessary and sufficient to increase the ligand-dependent transcriptional activation of target genes by steroid receptors. None of these original SRA sequences were successfully translated in vitro or in vivo [3], and SRA is still currently classified as belonging to the expanding family of functional ncRNAs [4]. Since 1999, data have accumulated regarding the possible mechanisms of action of SRA RNA. SRA RNA interacts with other proteins such as the co-repressor Sharp and the AF-1 specific activator p72/p68 protein to modulate steroid receptor activity [5,6]. Moreover, SRA RNA potentiates the estrogen-induced activation of both estrogen receptors α and β [7]. By introducing mutations in the SRA RNA sequence, Lanz et al. recently identified motifs participating in the RNA secondary structure that are involved in its ability to co-activate progesterone receptor [8,9].

We recently identified three new SRA RNA isoforms which corresponded to SRA except for an additional 37 nucleotides in the 5′ region [10]. This 5′ region contains two putative ATG codons, close together in the same open reading frame, that could encode putative 236/224 amino-acid SRA proteins (SRAPs). These isoforms, which contain the functional core region, encoded a stable protein both in vitro and in vivo. Using reverse polymerase chain reaction of RNA extracts, we were able to confirm the presence of these endogenous coding isoforms in breast cancer cell lines [10]. Furthermore, using an antibody raised against a peptide corresponding to amino acids 20–34 of the putative human SRAP, we were able to specifically detect a doublet at 30 kDa by Western blot analysis of total protein lysate from these same cell lines [10].

To date, all other reports describe and discuss human SRA as a ncRNA molecule. To our knowledge, no functional RNA has been described to have a protein coding capacity. It, therefore, became important to confirm unequivocally the existence of such an SRAP. In the present study, we demonstrate that the human SRA gene not only encodes for a protein but that the sequence of this protein is conserved among vertebrates.

* Corresponding author. Fax: +204-789-3900.
E-mail address: eleygue@cc.umanitoba.ca (E. Leygue).

[1] These authors have equally contributed.

Abbreviations: SRA, steroid receptor RNA activator; RNAi, RNA interference; ncRNA, non-coding RNA; SRAP, steroid receptor RNA activator protein

0014-5793/$22.00
doi:10.1016/j.febslet.2004.03.104

2. Materials and methods

2.1. RNA interference (RNAi) vector construction and SRAP knockdown assay

pSuper.retro-SRA construct was generated by hybridization of oligonucleotides SRARNAif (5′-ccccaagtttccctcacgagtcttcaagagagactcg actgggaaacttgtttt-3′) and SRARNAir (5′-aaacaagtttcccagtcgagtctctctt gaagactcgactgggaaacttgggg-3′), which were then cloned between the *Bgl*II and *Hin*dIII sites of the pSuper.retro vector (Oligoengine, Seattle, WA). The SRARNAi oligonucleotide sequence was selected following analysis with the Oligoengine RNAi design tool (http://www.oligoengine.com/). HeLa cells were transiently transfected with either the pSuper.retro-SRA construct or pSuper.retro vector (empty vector) using the transfection agent Effectene (Qiagen, Missisauga, ON) according to the manufacturer's protocol. At indicated times, cells were washed in 1× PBS and lysed with 2× SDS buffer (62.5 mM Tris, 2% SDS, 10% glycerol, 0.1% bromophenol blue and protease inhibitor cocktail (Roche, Laval, QB). Protein sample concentration was determined by the Micro BCA Protein Assay Kit (Pierce, Rockford, IL). Protein samples were analyzed by Western blot.

2.2. Western blot analysis

Proteins were extracted from skeletal tissue [11] or cell lines [10] as described previously. We chose total skeletal muscle as Lanz et al. [3] had shown that SRA RNA is highly expressed in human skeletal muscle. Protein concentration for each sample was determined using Bio-Rad Protein Assay Kit (Bio-Rad, Hercules, CA) and equal amounts of total protein were analyzed by SDS–PAGE and immunoblotting. Supernatant samples containing equal amount of total protein were mixed 1:1 with sample buffer [1.25 mM Tris–HCl (pH 6.8), 20% glycerol, 4% SDS, 0.02% (w/v) bromophenol blue and 20 mM DTT], boiled for 5 min, electrophoresed using a 5% stacking gel and a 10% resolving polyacrylamide gel, and electrophoretically transferred to nitrocellulose membrane. Membranes were blocked with 5% non-fat dry milk in Tris-buffered saline (pH 7.6) and incubated with rabbit polyclonal antibodies raised against amino acids 20–34 of human SRA sequence [10] at a dilution of 1:1000 in 5% non-fat dry milk in Tris-buffered saline containing 0.1% Tween. Secondary horseradish peroxidase-linked goat antirabbit antibodies (1 μg/μl, Bio-Rad) were then used and signals were analyzed by SuperSignal West Pico Chemiluminescent Substrate (Pierce). Chemiluminescence signal was captured by video image analysis using the Quantity One system (Bio-Rad). To determine equal loading, the SDS–PAGE gels were either stained in Coomassie blue solution (50% methanol, 10% acetic acid and 0.1% w/v Coomassie powder) following transfer or blots were stripped and reprobed with anti-β-actin antibody (Sigma, Oakville, ON). In order to detect the specificity of the detection in the protein extracts from different species, we incubated duplicate blots with anti-SRA antibody premixed with the peptide used to raise it (1/10 v/v).

2.3. Sequence analysis and database searches

Search of the NCBI protein database (www.ncbi.nlm.nih.gov) led to the identification of the SRA sequences for three species: *Homo sapiens* (GenBank Accession Nos. AF293024, AF293025 and AF293026), *Mus musculus* (GenBank Accession No. NP_079567) and *Rattus norvegicus* (GenBank Accession Nos. NP_000035 and AAG02116). Upon examining the mouse and rat expressed sequence tag (EST) database, a longer 5′ *M. musculus* sequence (GenBank Accession No. CB274276) and a *R. norvegicus* sequence (GenBank Accession No. CB771552) were identified. These sequences were used to deduce the theoretical sequences for the rat and mouse SRAPs. Search of the Unigene database and ProtESTs (http://www.ncbi.nlm.nih.gov/UniGene/protest.cgi?SORT=4&ORG=Hs&XID=114234) led to the identification of SRA EST sequences for the following species: *Xenopus laevis* (GenBank Accession Nos. BG364002.1, BG551872.1 and AW642449.1), *Silurana tropicalis* (GenBank Accession No. AL969036.1), *Sus scrofa* (GenBank Accession Nos. CF366666 and CF368085), and *Oryzias latipes* (GenBank Accession No. AU170197.1). We blasted the human SRA3 amino-acid sequence against the translated EST nucleotide sequences using tblastn, limiting the search to vertebrates and excluding mouse and human sequences. This search identified ESTs for the following additional species: *Macaca mulata* (GenBank Accession No. CD766957), *Equus caballus* (GenBank Accession Nos. B1961443 and B1961063), *Bos taurus* (GenBank Accession Nos. CB422540, AW654516, CB450664 and CB457765P), *Danio rerio* (GenBank Accession Nos. BQ258955, CB352395), *Occhonrhyncus Mykiss* (BX860673) and *Gallus gallus* (GenBank Accession Nos. CR338992, CR338823, CR338795). Species-specific sequences were assembled using an EST assembler tool (http://bio.ifom-firc.it/ASSEMBLY/assemble.html) giving contigs that were translated using the DNA to amino-acid translational tool (http://ca.expasy.org/cgi-bin/dna_aa.html). All the putative SRAP sequences were aligned using the Multalin alignment tool (http://prodes.toulouse.inra.fr/multalin/) and two conserved regions were identified. These two conserved sequences were used to identify the SRA analogue in Fugu fish by doing a Blast search against the Fugu fish genome. Blast search using the human SRA1 sequence against the ascidian *Ciona intestinalis* genome in the TIGR database (http://tigrblast.tigr.org/tgi/) resulted in the identification of an SRA analogue (GenBank Accession No. BW276199) in *C. intestinalis*. Blast searched at the NCBI site http://www-genome.wi.mit.edu/cgi-bin/annotation/ciona/blast_page.cgi using the *C. intestinalis* SRA sequence resulted in a *Ciona savignyi* SRA analogue. Nuclear localization signals were search on each individual sequence using Psort II (http://psort.nibb.ac.jp/form2.html).

2.4. PCR cloning and in vitro translation of rat SRA cDNA

PCR primers (5′-agtgagctaccaccccggaa-3′ and 5′-tatagaagtcatgtggaggt-3′) designed by analyzing the rat theoretical SRA sequence were used to amplify cDNA from rat skeletal muscle. The resulting product was sequenced (GenBank Accession No. AY542868) and cloned in pcDNA3.1 expression vector. RNA isolation from rat skeletal tissue and reverse transcription-polymerase chain reaction were conducted as described previously [9]. pcDNA3.1 (Invitrogen, Carlsbad, CA) expression plasmid containing either human SRA cDNA or rat SRA cDNA was used for in vitro translation/transcription reaction. [^{35}S]methionine labelled SRAPs were generated using wheat germ lysate coupled transcription/translation reactions by the TnT System (Promega, Madison, WI) according to the manufacturer's instructions. The V5-tagged human SRA cDNA construct, previously [10] cloned in pcDNA3.1/V5-His© (Invitrogen, Carlsbad, CA), was used as a positive control. Lysates were then subjected to SDS–PAGE separation, after which gels were dried and [^{35}S]methionine labelled protein bands visualized by exposing overnight to a Molecular Imager™-FX Imaging screen (Bio-Rad) and subsequently scanned using a Molecular Imager™-FX (Bio-Rad).

3. Results and discussion

We were previously able to detect a putative endogenous SRAP in breast cancer cell lines with an antibody targeted against amino acids 20–34 of the hypothetical human SRAP sequence [10]. As we are still today the only ones to have reported the existence of such a putative protein, it became essential to definitively confirm the identity of the protein recognized by our antibody. In order to do so, we used the recently developed RNAi technology in an attempt to knockdown its expression. Indeed, we reasoned that the specific degradation of SRA RNA should result in a decrease in the expression of the protein recognized by our antibody. Upon 24 and 48 h, transfection of HeLa cells with a RNAi specifically targeted against SRA RNA resulted in a significant decrease in the doublet detected by Western blot and believed to correspond to the SRAP (Fig. 1A). In contrast and as expected, RNAi treatments had no significant effect on β-actin levels (Fig. 1B) as well as a non-specific protein detected by our anti-SRA antibody (Fig. 1A, 45 kDa). This result links the expression of the SRA gene to the detection of the suspected SRAP by Western blot. This unequivocally confirms for the first time the existence of an endogenous SRAP in human cells. It should be stressed that no noticeable phenotypic changes were observed in cells treated with SRA-specific RNAi over the 48-h post-transfection. This absence of apparent effect

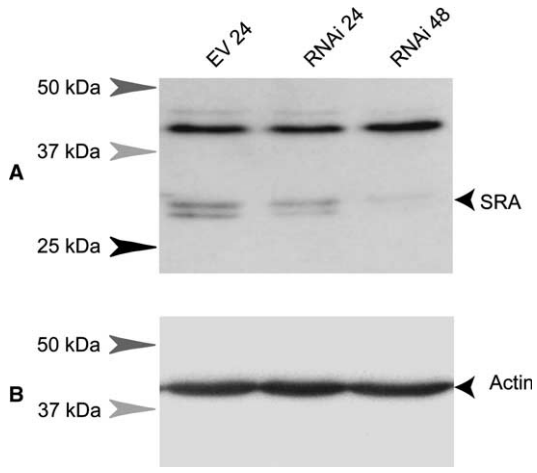

Fig. 1. Decrease of SRAP expression by RNAi specifically targeted against SRA RNA. HeLa cells were transiently transfected with either SRA RNAi (RNAi) or control (EV) vectors, and SRA or actin protein expressions assessed by Western blot 24 and 48 h after transfection as described in Section 2. Positions of the molecular size markers are indicated on the left.

likely results from the short length of time and the transient nature of these experiments have been conducted. Further experiments are needed to investigate longer term effects of the knockdown of SRA gene on phenotypic changes such as growth rate.

We have then investigated the possible existence of this SRAP in other species. As of today, Blink, the NCBI sofware which groups putative species homologues (http://www.ncbi.nlm.nih.gov/sutils/blink.cgi?pid = 9930614), only gathers human, rat and mouse SRA sequences. Rat SRA sequences (GenBank Accession Nos. NP_000035 and AAG02116) correspond to the recently published 146 amino-acid rat SRAP sequence [12]. This putative rat SRAP sequence was successfully translated in vitro and also expressed in vivo when fused with the C-terminal extremity of green fluorescence protein. SRAP is much smaller (16 kDa) than the putative human SRAP we identified and its endogenous existence has not yet been demonstrated. Close analysis of the SRAP mRNA sequence revealed that it is analogous to the human SRA sequences starting at exon 2. We thus suspected the SRAP mRNA sequence described so far to be incomplete. Indeed, analysis of the rat chromosome 18 sequence revealed the presence of a putative additional exon 1 present in a single rat EST sequence (GenBank Accession No. CB771552). To confirm the existence of this longer SRA sequence, we PCR amplified rat muscle cDNA with specific primers as described in Section 2. We cloned a new rat cDNA sequence (GenBank Accession No. AY542868), which unlike the shorter SRAP sequence contains two possible methionines possibly initiating the translation of a 222/230 amino-acid proteins. As shown in Fig. 2, this sequence is translatable in vitro, generating a visible doublet at 31/32 kDa. As expected and as a result of an additional V5-tag (4.8 kDa), in vitro translated SRA-V5 protein (35 kDa) has a higher apparent molecular size than the non-tagged human SRA doublet (30/31 kDa) and rat SRA doublet (31/32 kDa). As previously described [10], the observed molecular masses (35, 30/31 and 31/32 kDa) are slightly higher than those predicted (30.5, 25.7 and 25.3 kDa for the V5-tagged human, non-tagged human SRAP and rat SRAP,

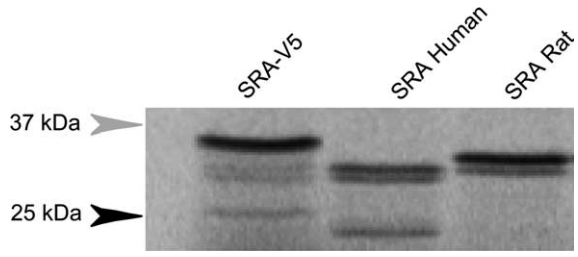

Fig. 2. In vitro translation of human and rat SRA cDNAs. In vitro transcription/translation reactions were performed using V5-tagged human SRA cDNA as control (SRA-V5), human SRA cDNA (SRA Human) and rat SRA cDNA (SRA Rat) as described in Section 2. Positions of the molecular size markers are indicated on the left.

respectively). We hypothesize that the doublet seen for human and rat SRAPs results from the alternative use of one initiating methionine instead of another (1 versus 12). Weaker lower molecular size bands (around 25 kDa) are likely resulting from translation at internal downstream methionines present in both the human and rat SRA sequences (see Fig. 3).

Through database analysis, we identified a *M. musculus* sequence (GenBank Accession No. CB274276), slightly longer than the one present in the protein database. When translated, this sequence contains an additional stretch of 12 N-terminal amino acids, 10 out of which are identical to their corresponding human counterparts (Fig. 3).

In an attempt to identify SRA analogues in other species, we searched several databases as described in Section 2. Obtained cDNA sequences were translated and aligned (Fig. 3). The alignment of putative SRA sequences from different species shows proteins of similar lengths, highly conserved in discrete domains and in two main regions (amino acids 15–39 and 180–208 of the human sequence). We have previously demonstrated that the stably transfected SRAP localizes to the nucleus in MCF-7 cells [10]. Analysis of the SRA sequences (Psort II), for nuclear localization signal domains, revealed a conserved pat-7 (P-x-[RK]-[RK]-[RK]) nuclear localization motif in *Gallus* (chicken), *Oryzias* and *Occonrhyncus* (two fish) and all the mammalian putative SRAP sequences (amino acids 155–160). Interestingly, the two *Ciona* species which do not contain pat-7 motif at this position, contain however another nuclear localization pat-4 motif (P-[RK]-[RK]-[RK]) at amino acids 39–42 with respect to the human SRA sequence.

Overall, putative SRAPs were found in all vertebrates in which SRA-related EST sequences could be detected. In addition, putative SRAP analogues were identified in two *Ciona* species, *Ciona savignyi* and *C. intestinalis*. *Ciona* belong to the urochordata subphylum, which together with Cephalochordata and Craniata (contains Vertebrata) subphyla, defines the *Chordata phylum*. We were unable to find any sequences closely related to SRA in any other phyla such as Arthropoda (*Drosophila melanogaster*), Nematoda (*Caenorhabditis elegans*) or Protobacteria (*E. coli*). Conservation of the SRAP sequence from an invertebrate Chordata (*Ciona*) to a higher vertebrate (human) suggests an important role possibly played by this molecule. Furthermore, the conservation of a nuclear localization signal in most of the SRAPs suggests a nuclear localization for this putative conserved function. In addition, the apparent absence of SRA homologues in all non-chordata phyla suggests that this protein might have been involved in the emergence of early Chordata.

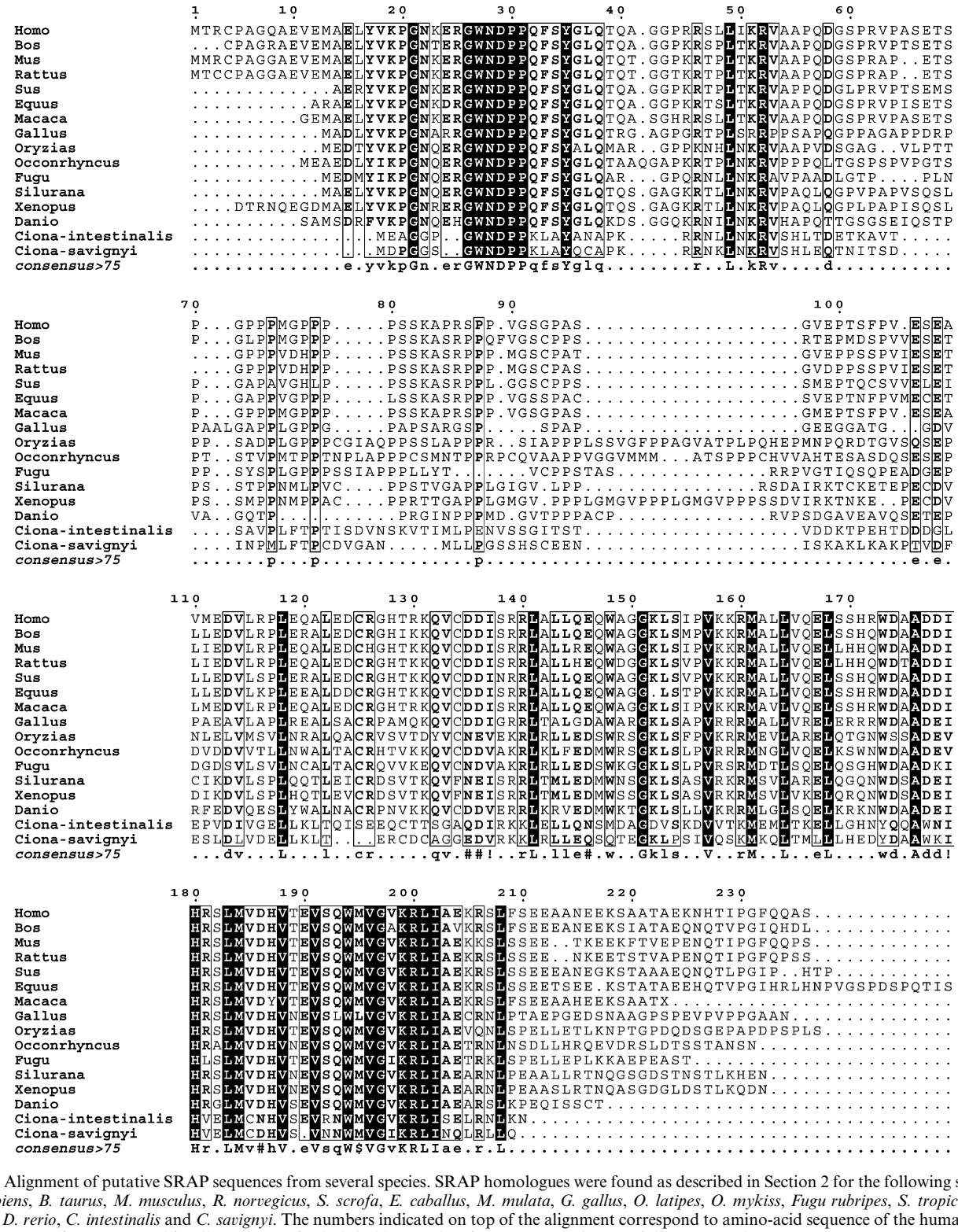

Fig. 3. Alignment of putative SRAP sequences from several species. SRAP homologues were found as described in Section 2 for the following species: *H. sapiens*, *B. taurus*, *M. musculus*, *R. norvegicus*, *S. scrofa*, *E. caballus*, *M. mulata*, *G. gallus*, *O. latipes*, *O. mykiss*, *Fugu rubripes*, *S. tropicalis*, *X. laevis*, *D. rerio*, *C. intestinalis* and *C. savignyi*. The numbers indicated on top of the alignment correspond to amino-acid sequence of the human SRA isoform 1. Regions conserved in all species are in dark bold. Boxed regions correspond to partially conserved sequences.

As most of the putative sequences obtained were well conserved in the region between amino acids 15 and 39, we hypothesized that our antibody, raised against the peptide 20–34 of the human SRAP sequence, could recognize SRAPs from other species. Western blot analysis of SRAP expression was therefore performed on protein extracted from skeletal muscles (tissue known to contain high levels of SRA RNA [3]) of several non-human vertebrates.

As shown in Fig. 4A, we were able to specifically detect a band of similar size (around 32 kDa) in *Bos* (cow), *Sylvilagus* (rabbit), *Sus* (pig), *Gallus* (chicken), *Meleagris* (turkey) and *Ovis* (sheep). Equal amounts of total protein extracts were

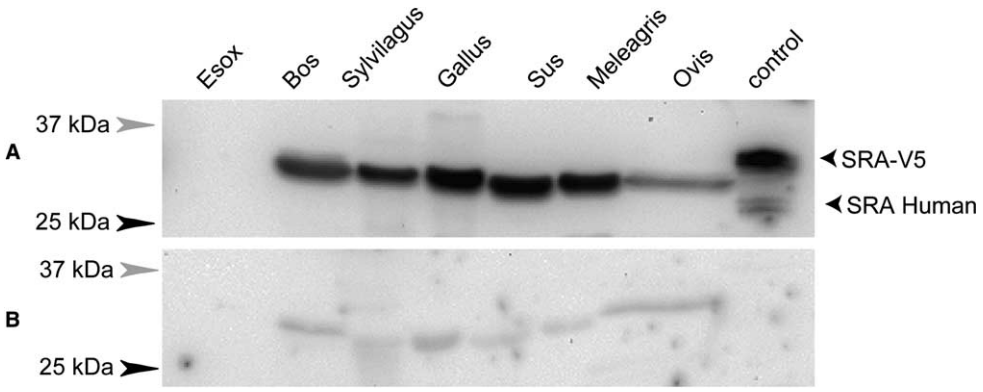

Fig. 4. SRAP expression in skeletal muscle of various species. As described in Section 2, Western blots of total protein extracts from the muscle of different species were performed in the absence (A) or presence (B) of neutralizing peptide. Positive control consists of protein extracts from MCF-7 cell line stably expressing V5-tagged human SRA. Positions of the molecular size markers are indicated on the left.

loaded as confirmed by Coomassie blue staining of the gels after transfer (data not shown). Protein extracts from MCF-7 cells stably expressing V5-tagged SRA were used as positive control (right lane). An upper band corresponding to the transfected V5-SRA is seen at 35 kDa. In addition, a doublet, corresponding to the endogenously expressed SRA is seen at 30/31 kDa. No signals were observed when the antibody was pre-neutralized by the corresponding peptide (Fig. 4B). Interestingly, only one band is detected in muscle extracts, suggesting that only one AUG codon is used to initiate SRAP translation in these tissues. This may result either from the existence of only one such codon in the *Sus*, *Bos*, *Gallus* putative mRNA sequences (see Fig. 3), or from the preferential tissue-related use of one codon over the other.

No SRAP was detected in the Esox (pickerel) protein extract. A careful examination of the four hypothetical fish sequences we gathered (*Fugu*, *Oryzias*, *Occhonrhyncus* and *Danio*) shows that they are similar to the human sequence in the region recognized by the antibody used (20–34, referenced to human SRA) except for the presence of a glutamine instead of a lysine at amino acid 23 (Fig. 3). The Esox SRA sequence, which remains unknown to date, should, if similar to the other fish sequences in this particular region, also contain this particular amino-acid substitution. Interestingly, our antibody was able to recognize SRAPs from the *Bos* (cow) and *Gallus* (chicken) tissue extracts despite the presence of a threonine (*Bos*) and an alanine (*Gallus*) instead of a lysine at this same residue 23. Lysine, alanine and threonine as opposed to glutamine are all hydrophobic amino acids. We therefore suspect that a switch in the hydrophobicity of the amino acid at residue 23 may be responsible for the impaired recognition of Esox SRAP by our antibody.

In this study, we have demonstrated that the human SRA gene encodes for a protein conserved among vertebrates. The high conservation of SRAP sequence underlines the possible important role played by this protein. Previous studies have shown that SRA RNA is fully functional independently of its protein coding capacity. SRA therefore appears to be the first example of a new class of functional RNAs also able to encode a protein.

The existence of an SRAP raises several important questions waiting to be addressed. What is the role of SRAP? What are the implications of SRAP on SRA RNA function? How are the expressions of SRA RNA and SRAP regulated? What other molecules function at dual protein/RNA levels? Since all of the functional studies on human SRA described to date ignore the existence of a protein, addressing the above questions is critical to fully understand SRA function. More importantly, development in our understanding of SRA RNA and protein function is in turn critical for a change in the current perspective of functional "non-coding" RNA molecules.

Acknowledgements: This work is supported by a USAMRMC Career Development Award Grant No. DAMD17-01-1-0308 and a CBCRA Grant No. 12326.

References

[1] Eddy, S.R. (2001) Nat. Rev. Genet. 2, 919–929.
[2] Storz, G. (2002) Science 296, 1260–1263.
[3] Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. (1999) Cell 97, 17–27.
[4] Lanz, R.B., Chua, S.S., Barron, N., Soder, B.M., DeMayo, F. and O'Malley, B.W. (2003) Mol. Cell Biol. 23, 7163–7176.
[5] Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M. and Evans, R.M. (2001) Genes. Dev. 15, 1140–1151.
[6] Watanabe, M., Yanagisawa, J., Kitagawa, H., Takeyama, K., Ogawa, S., Arao, Y., Suzawa, M., Kobayashi, Y., Yano, T., Yoshikawa, H., Masuhiro, Y. and Kato, S. (2001) EMBO J. 20, 1341–1352.
[7] Deblois, G. and Giguere, V. (2003) J. Steroid Biochem. Mol. Biol. 85, 123–131.
[8] Lanz, R.B., Razani, B., Goldberg, A.D. and O'Malley, B.W. (2002) Proc. Natl. Acad. Sci. U.S.A 99, 16081–16086.
[9] Leygue, E., Dotzlaw, H., Watson, P.H. and Murphy, L.C. (1999) Cancer Res. 59, 4190–4193.
[10] Emberley, E., Huang, G.J., Hamedani, M.K., Czosnek, A., Ali, D., Grolla, A., Lu, B., Watson, P.H., Murphy, L.C. and Leygue, E. (2003) Biochem. Biophys. Res. Commun. 301, 509–515.
[11] Troup, S., Njue, C., Kliewer, E.V., Parisien, M., Roskelley, C., Chakravarti, S., Roughley, P.J., Murphy, L.C. and Watson, P.H. (2003) Clin. Cancer Res. 9, 207–214.
[12] Kawashima, H., Takano, H., Sugita, S., Takahara, Y., Sugimura, K. and Nakatani, T. (2003) Biochem. J. 369, 163–171.