J Mol Evol (1999) 48:248–251

JOURNAL OF MOLECULAR EVOLUTION

The Long Terminal Repeat of an Endogenous Retrovirus Induces Alternative Splicing and Encodes an Additional Carboxy-Terminal Sequence in the Human Leptin Receptor

Vladimir V. Kapitonov, Jerzy Jurka

Genetic Information Research Institute, 1170 Morse Avenue, Sunnyvale, CA 94089, USA

Received: 17 June 1998 / Accepted: 12 August 1998

Abstract. The evolution of mammalian protein structure and regulation, specifically transcriptional and posttranscriptional regulation, may include among its tools the use of abundant retroviral long terminal repeats (LTRs). In particular, LTRs may be turned into switches for alternative splicing. This type of regulatory pathway is illustrated by the alternative splicing in the human leptin receptor (OBR). The human leptin receptor is involved in the control of important biological processes including energy expenditure, production of sex hormones, and activation of hemopoietic cells. OBRa and OBRb are the two major, alternatively spliced forms of the leptin receptor, called the ''short form'' and the ''long form,'' respectively. We report that the OBRa short form is the result of a double splicing event which occurs within the LTR of the endogenous retrovirus HERV-K. Working as a switch of alternative splicing, this LTR also encodes the terminal 67 amino acid residues in OBRa. We suggest the possibility of transcriptional and posttranscriptional regulation of OBR expression by steroids that bind the LTR.

Key words: Leptin receptor — Retrovirus — Long terminal repeats — Alternative splicing

Correspondence to: V.V. Kapitonov; *e-mail:* vladimir@charon.girinst.org

The present report provides an example of the impact of retroviral elements on the evolution of mammalian genomes.

Leptin, OB, is a hormone secreted by adipose tissue and encoded by the obese (*ob*) gene (Zhang et al. 1994). *ob/ob* mice, homozygous for mutation in the obese gene, display obesity because of leptin secretion failure. The intensive research over the last 3 years demonstrates that leptin affects regulation of body weight, energy expenditure, puberty, and production of sexual hormones, hemopoiesis, and macrophage functions (Pelleymounter et al. 1995; Halaas et al. 1995; Campfield et al. 1995; Yu et al. 1997; Chehab et al. 1997; Gainsford et al. 1996). Leptin's effects are thought to be mediated by its interaction with the leptin receptor, OBR, related to the class I cytokine receptor family (reviewed by Tartaglia 1997).

The human leptin receptor gene consists of 20 exons and spans over ~70 kb on chromosome 1q31. A major fraction of OBR intronic sequences is still unknown. There are two major forms of leptin receptor expressed in human cells: (short) OBRa and (long) OBRb (Tartaglia 1997; Tartaglia et al. 1995). The OBRb exon 20 encodes a 303-amino acid-long intracellular domain (see Fig. 1), presumed to be essential for intracellular signal transduction by activation of JAK tyrosine kinase and STATs [signal transducers and activators of transcription (Chen et al. 1996; Lee et al. 1996)]. In OBRa, the translation of exon 20 is replaced by translation of an unrelated sequence generating new 67 terminal amino acids (aa). The remaining exons are identical in both proteins.

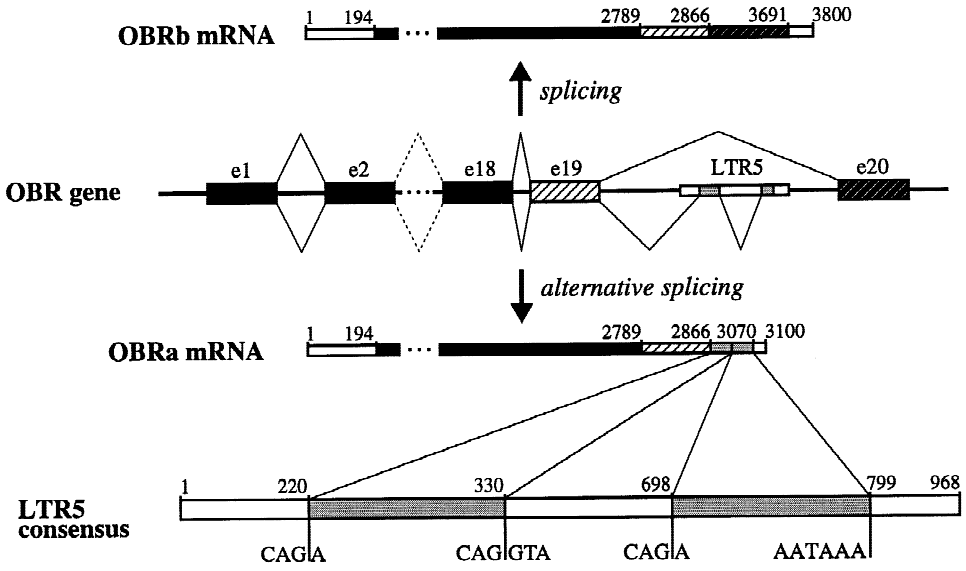

Fig. 1. Putative model of alternative splicing of leptin receptor gene (OBR). The ~71-kb OBR gene consists of 20 exons (e1, e2, . . ., e18, e19, and e20). Exons 1–18 are indicated as *black boxes;* exons 19 and 20 are shown as *light* and *dark hatched boxes,* respectively. Nucleotide sequences of the exons were deposited in GenBank as sequences U59246–U59263. The 5′ and 3′ untranslated regions of the OBR mRNA are shown as *white terminal boxes.* Two alternative splicing events leading to the expression of the OBR long and short forms are indicated schematically *above* and *below* the OBR gene, respectively. The *angular lines* indicate the joining of the exons. There is no difference in splicing of the first 19 exons in the OBRa and the OBRb transcripts (GenBank sequences U66496 and U43168, respectively). Two putative terminal exons in the OBRa mRNA (*shaded boxes*) come from the shaded portions of the LTR5 sequence inserted between exon e19 and exon e20. Nucleotide positions are indicated on *top* of the schematic LTR5 consensus sequence and the OBRa and the OBRb mRNAs. Sequences of acceptor and donor splice sites, as well as of the poly(A) signals, are indicated *below* the LTR5 consensus sequence.

The short form is not likely to be involved in activation of STATs (Ghilardi et al. 1996) but may be important for transfer, clearance, and regulation of concentration of free leptin in the bloodstream.

We report that the nucleotide sequence encoding the terminal 67 aa in OBRa is 88% identical to the consensus sequence of LTR5 (Figs. 1 and 2), which is a long terminal repeat from primate-specific endogenous retrovirus HERV-K10 (Ono 1986). As shown here, the LTR5 sequence is responsible for the generation of the short form by induction of alternative splicing. We found the alternative splicing to be mediated by the OBR donor splice site at the 3′ end of exon 19 and the three splice signals encoded by the LTR5 sequence (CAGA acceptor splice signals at positions 217–220 and 695–698 and CAGGTA donor splice signal at position 328–333; see Fig. 2). As a result, two short internal LTR5 fragments become terminal exons in OBRa coding for 67 terminal amino acids (Figs. 1 and 2), which replace the intracellular domain encoded by exon 20 of the OBR gene. Thus, working as a switch of alternative splicing, this LTR also encodes the terminal 67 aminio acid residues in OBRa.

The portions of the LTR5 consensus sequence homologous to the OBRa terminal exons contain two stop codons interrupting the 67-aa-long ORF. Only two point mutations are needed to make this region translatable. One is deletion of T (position 288) and the other is a T → G substitution at position 786. Our search of the NCBI database using BLAST server did not reveal any proteins similar to the 67 terminal amino acid sequence. It is therefore likely that this sequence was generated from the usually untranslatable LTR5 insert. In addition to the splice signals and the ORF, the LTR5 encodes a putative polyadenylation signal and a translation stop codon in OBRa (Fig. 2). Regardless of the lack of evidence of short form function, it is unlikely that fixation of the LTR5 was a neutral event since the alternative splicing changes the concentration of the long form, which is evidently an important regulatory element involved in many functions.

LTR5 of HERV-K10 retrovirus is known to contain a glucocorticoid response element followed by a transcription enhancer and its transcription is activated by steroids such as progesterone and estradiol (Ono 1987). The transcription of LTR from another HERV-K-related retrovirus, murine mammary tumor virus (MMTV), is also activated by steroids (Le Ricousse et al. 1996). Furthermore, it has been shown that glucocorticoid and progesterone may affect the posttranscriptional processing of various genes (Hayward-Lester et al. 1996; Ehretsmann et al. 1995) including alternative splicing of the human insulin receptor gene (Norgren et al. 1994) in a dose-dependent way. Finally, it has been shown that splicing may depend on transcription (Cramer et al. 1997; McCracken et al. 1997). In this context it is very striking that leptin expression and function depend on steroids (Shimizu et al. 1997; Zachow and Magoffin 1997; Chehab et al. 1996). Therefore, it is possible that the generation of OBRa by alternative splicing is regulated by

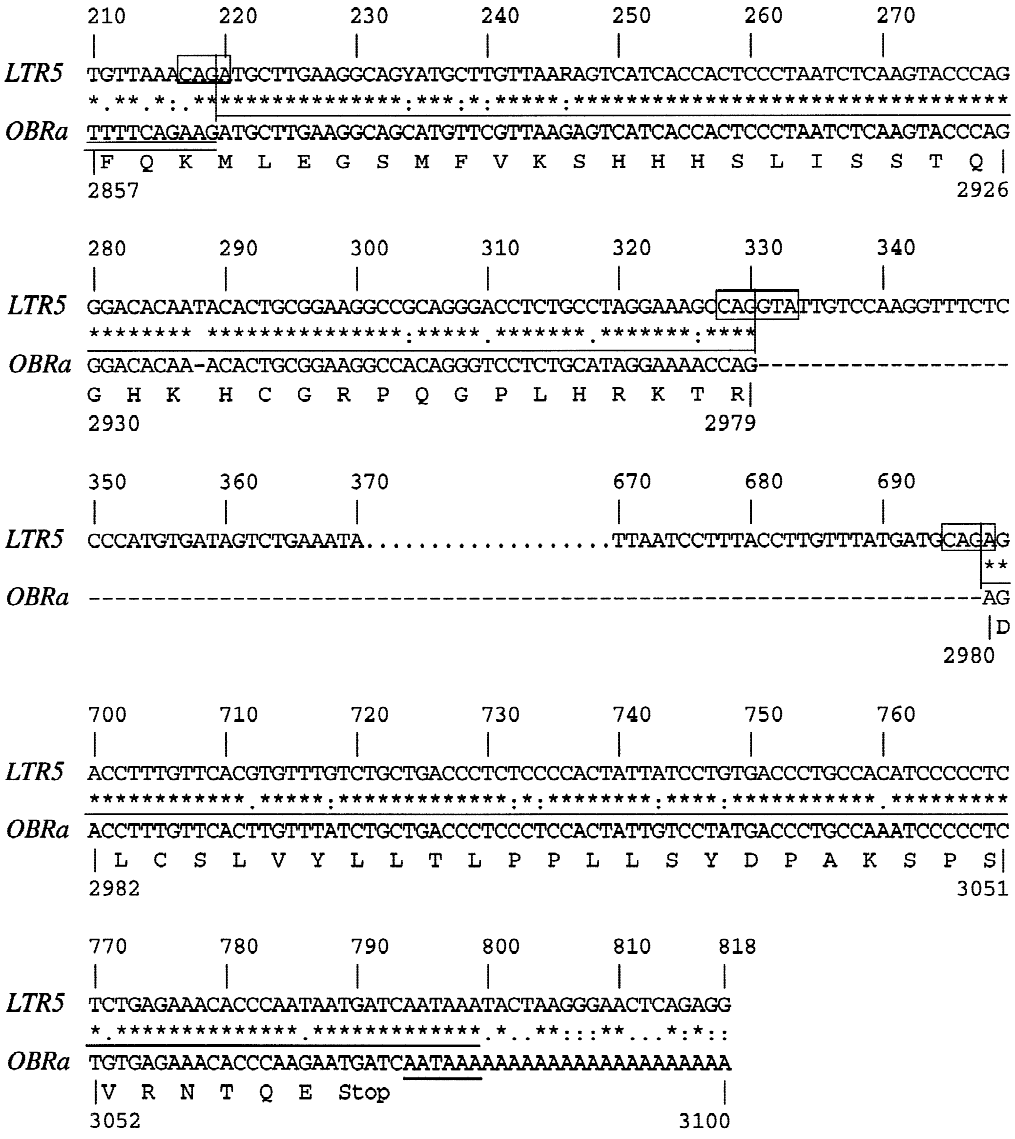

Fig. 2. Alignment of the LTR5 consensus and OBRa nucleotide sequences. Identical bases are indicated by *asterisks;* transitions and transversions are represented by *colons* and *dots*, respectively. *Hyphens* indicate alignment gaps. The LTR5 sequence is a fragment of a 974-bp-long consensus sequence deposited in RepBase Update (http://www.girinst.org); nucleotide positions in the aligned LTR5 fragment are shown on the *top*. The 3100-bp OBRa nucleotide sequence comes from GenBank (Accession No. U66496). The nucleotide positions in the aligned fragment are indicated on the *bottom*. The 3′ end of exon 19 is *double underlined,* and it is differentfrom the LTR5 sequence. Regions of OBRa mRNA that come from the LTR5 sequence are *overlined.* The internal part of the LTR5 sequence (positions 371–669) is *dotted* out. The poly(A) signal in the LTR5 is *underlined.* Acceptor and donor splice signals are *boxed. Vertical lines* within the boxes show alternative exon/intron boundaries. The 67-residue-long amino acid sequence encoded by the ORBa alternative exons is presented *below* the nucleotide sequences.

steroids which may interact with the LTR5 element fixed in the OBR gene. It has been noticed in experiments (reviewed by Woods et al. 1998) that glucocorticoid hormones are endogenous antagonists of leptin in the control of energy homeostasis. This is consistent with our hypothesis that glucocorticoids may induce alternative splicing in the leptin receptor, thereby switching expression from the long to the short form. Since the short form of the leptin receptor does not have the intracellular domain, the binding leptin cannot transduce regulatory signals.

Interestingly, leptin itself contains the MER11 repeat (Kaplan et al., 1991; Smit 1995), recently identified as HERV-K-related LTR (Kapitonov and Jurka 1997) in its 5′ UTR region (GenBank Accession No. U43589; positions 412–1382). It has been shown (Bi et al. 1997) that a 60-bp region of the MER11 sequence acts as an enhancer of leptin expression in placenta. Moreover, this MER11 contains a motif TGTTATcttcataaGCTAAG (see GenBank Accession No. U43589; positions 570–589), in which capitalized sequences are similar to the steroid responsive element TGTTAT and the enhancer core GTGCTAAG, immediately following each other in LTR5 (Ono 1987). Taken together, these observations

may imply that transcription of leptin and splicing of its receptor are regulated by steroids. Thus, our hypothesis regarding a role for LTR5 in the control by steroids of alternative splicing, combined with previous data (Hayward-Lester et al. 1996; Ehretsmann et al. 1995; Norgren et al. 1994; Cramer et al. 1997; McCracken et al. 1997; Shimizu et al. 1997; Zachow and Magoffin 1997; Chehab et al. 1996), suggests that the proportions of OBRa and OBRb forms of leptin receptor may be regulated by steroids, possibly through interaction with the LTR5 sequence. By using similar functional elements, LTR5 and MER11 may provide an evolutionary basis for the coordinate expression of leptin and its receptor. Viral LTRs, at least those carried by HERV-K endogenous retroviruses, may be among the tools used in evolution for coupling the expression of regulatory factors and their receptors.

Acknowledgments. This work was supported by Grant 1 P41 LM06252 from the National Institutes of Health. We thank Dr. Emile Zuckerkandl for fruitful discussions and critical reading of the manuscript. We are grateful to Jolanta Walichiewicz and John Ledesma for technical assistance.

References

Bi S, Gavrilova O, Gong DW, Mason MM, Reitman M (1997) Identification of a placental enhancer for the human leptin gene. J Biol Chem 272:30583–30588

Campfield LA, Smith FJ, Guisez Y, Devos R, Burn P (1995) Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. Science 269:546–549

Chehab FF, Lim ME, Lu R (1996) Correction of the sterility defect in homozygous obese female mice by treatment with the human recombinant leptin. Nat Genet 12:318–320

Chehab FF, Mounzih K, Lu R, Lim ME (1997) Early onset of reproductive function in normal female mice treated with leptin. Science 275:88–90.

Chen H, Charlat O, Tartaglia LA, Woolf EA, Weng X, Ellis SJ, Lakey ND, Culpepper J, Moore KJ, Breitbart RE, Duyk GM, Tepper RI, Morgenstern JP (1996) Evidence that the diabetes gene encodes the leptin receptor: Identification of a mutation in the leptin receptor gene in db/db mice. Cell 84:491–495

Cramer P, Pesce CG, Baralle FE, Kornblihtt AR (1997) Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci USA 94:11456–11460

Ehretsmann CP, Chandler LA, Bourgeois S (1995) A nuclear post-transcriptional mechanism mediates the induction of fibronectin by glucocorticoids. Mol Cell Endocrinol 110:185–194

Gainsford T, Willson TA, Metcalf D, Handman E, McFarlane C, Ng A, Nicola NA, Alexander WS, Hilton DJ (1996) Leptin can induce proliferation, differentiation, and functional activation of hemopoietic cells. Proc Natl Acad Sci USA 93;14564–14568

Ghilardi N, Ziegler S, Wiestner A, Stoffel R, Heim MH, Skoda RC (1996) Defective STAT signaling by the leptin receptor in diabetic mice. Proc Natl Acad Sci USA 93:6231–6235

Halaas JL, Gajiwala KS, Maffei M, Cohen SL, Chait BT, Rabinowitz D, Lallone RL, Burley SK, Friedman JM (1995) Weight-reducing effects of the plasma protein encoded by obese gene. Science 269: 543–546

Hayward-Lester A, Hewetson A, Beale EG, Oefner PJ, Doris PA, Chilton BS (1996) Cloning, characterization, and steroid-dependent posttranscriptional processing of RISH-1 alpha and beta, two uteroglobin promoter-binding proteins. Mol Endocrinol 10:1335–1349

Kapitonov VV, Jurka J (1997). RepBase Update 1.7, MER11, http://www.girinst.org

Kaplan DJ, Jurka J, Solus JF, Duncan CH (1991) Medium reiteration frequency repetitive sequences in the human genome. Nucleic Acids Res 19:4731–4738

Le Ricousse S, Gouilleaux F, Fortin D, Joulin V, Richard-Foy H (1996) Glucocorticoid and progestin receptors are differently involved in the cooperation with a structural element of the mouse mammary tumor virus promoter. Proc Natl Acad Sci USA 93:5072–5077

Lee GH, Proenca R, Montez JM, Carroll KM, Darvishzadeh JG, Lee JI, Friedman JM (1996) Abnormal splicing of the leptin receptor in diabetic mice. Nature 379:632–635

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley D (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature 385:357–361

Norgren S, Li LS, Luthman H (1994) Regulation of human insulin receptor RNA splicing in HepG2 cells: effects of glucocorticoid and low glucose concentration. Biochem Biophys Res Commun 199:277–284

Ono M (1986) Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. J Virol 58:937–944

Ono M (1987) Stimulation of expression of human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J Virol 61:2059–2062

Pelleymounter MA, Cullen MJ, Baker MB, Hecht R, Winters D, Boone T, Collins F (1995) Effects of the obese gene product on body weight regulation in ob/ob mice. Science 269:540–543

Shimizu H, Shimomura Y, Nakanishi Y, Futawari T, Ohtani K, Sato N, Mori M (1997) Estrogen increases in vivo leptin production in rats and human subjects. J Endocrinol 154:285–292

Smit AFA (1995) Structure and evolution of mammalian interspersed repeats, PhD dissertation. University of Southern California, Los Angeles

Tartaglia LA (1997) The leptin receptor. J Biol Chem 272:6093–6096

Tartaglia LA, Dembski M, Weng X, Deng N, Culpepper J, Devos R, Richards GJ, Campfield LA, Clark FT, Deeds J, et al. (1995) Identification and expression cloning of a leptin receptor, OB-R Cell 83:1263–1271

Woods SC, Seeley RJ, Porte D, Schwartz MW (1998) Signals that regulate food intake and energy homeostasis. Science 280:1378–1383

Yu WH, Kimura M, Walczewska A, Karanth S, McCann SM (1997) Role of leptin in hypothalamic-pituitary function. Proc Natl Acad Sci USA 94:1023–1028

Zachow RJ, Magoffin DA (1997) Direct intraovarian effects of leptin: Impairment of the synergistic action of insulin-like growth factor-I on follicle-stimulating hormone-dependent estradiol-17 beta production by rat ovarian granulosa cells. Endocrinology 138:847–850

Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM (1994) Positional cloning of the mouse obese gene and its human homologue. Nature 372:425–432