CONSERVATION OF THE UPSTREAM OPEN READING FRAMES IN THE GENES FOR THE ERB-B2 RECEPTOR TYROSINE KINASE 2 IN VERTEBRATES

Luchezar Karagyozov

Abstract. In humans a short open reading frame (ORF) precedes the sequence of the erb-b2 receptor tyrosine kinase 2 (ERB-B2). It is established that the upstream ORF plays an important role in *ERB-B2* gene expression. In this report we show that an upstream ORF is present in all vertebrate genes coding for erb-b2 receptor tyrosine kinase 2. The position and the structure of *erb-b2* upstream ORF is compared between six different vertebrates. Amphibians and all land vertebrates share the same uORF characteristics. In bony fishes, however, the upstream ORF is noticeably different. This suggests that the regulation of *erb-b2* gene by uORF originated in its present form with the amphibians.

Keywords: receptor tyrosine kinase; upstream open reading frame; evolutionary conservation

Introduction

Structure of the messenger RNA in eukaryotes

The structure of the messenger RNA (mRNA) in eukaryotes is well known. At the 5′-end of the mRNA there is a specific "cap" – a modified guanosine nucleotide (m7GpppN-mRNA). The cap is followed by a leader – a 5′- untranslated region with varying length. The start codon (AUG) marks the beginning of the open reading frame (ORF), which ends with a stop codon (UAA, UAG or UGA). Downstream is the 3'-untranslated region. Most mRNAs end with a poly(A) tail.

Scanning model of translation

Translation, in addition to ribosomes, involves an impressive number of protein factors (Kozak, 1999). Briefly, in eukaryotes translation is initiated by the binding of the small 40S ribosomal subunit to the cap-structure at the 5′-end of the mRNA. The mRNA is then inspected towards its 3′-end. At a certain point the scanning machinery encounters an AUG codon in a favorable context (Kozak, 1987). Then the large 60S ribosomal subunit is recruited and synthesis of the peptide chain begins.

Translation is terminated at a stop codon and the ribosome dissociates into 40S and 60S subunits.

Upstream open reading frames

An important refinement of the scanning model was the finding by Kozak (1987) that sometimes the ORF of the major protein is preceded by a short ORF. The upstream ORF (uORF) is delineated by an upstream AUG codon (uAUG), followed by a stop codon.

Consequences of the presence of an uORF

There are two possible outcomes of the presence of an uORF. Firstly, it is possible that the uAUG is partially bypassed (leaky scanning). In other instances, however, all ribosomes may start translating the uORF. In this case after the stop codon the ribosomes should be able to reinitiate at the downstream AUG (Fig.1). However, reinitiation is inefficient, particularly if the uORF is long or the distance between the upstream stop codon and the main AUG is long (Sachs & Geballe, 2006).

Occurrence of uORFs

Both leaky scanning and re-initiation allow initiation at the AUG of the main protein. However, if an uORF is present, the translation of the main protein is less efficient as the number of translating ribosomes is decreased. Nevertheless, it was noted that in vertebrates a large fraction of important proteins (including protein kinases, growth factors, oncogenes, receptors and transcription factors) are synthesized by mRNA templates with uORF (Kozak, 1987). The reasons for this are not entirely clear. The assumption is that down-regulation is a difficult undertaking for the cell. Therefore, it appears necessary to use repressed, translation-deficient mRNAs. In spite considerable effort, the mechanisms to overcome the negative effects of the uORFs are not well elucidated (Barbosa et al., 2013; Somers et al., 2013).

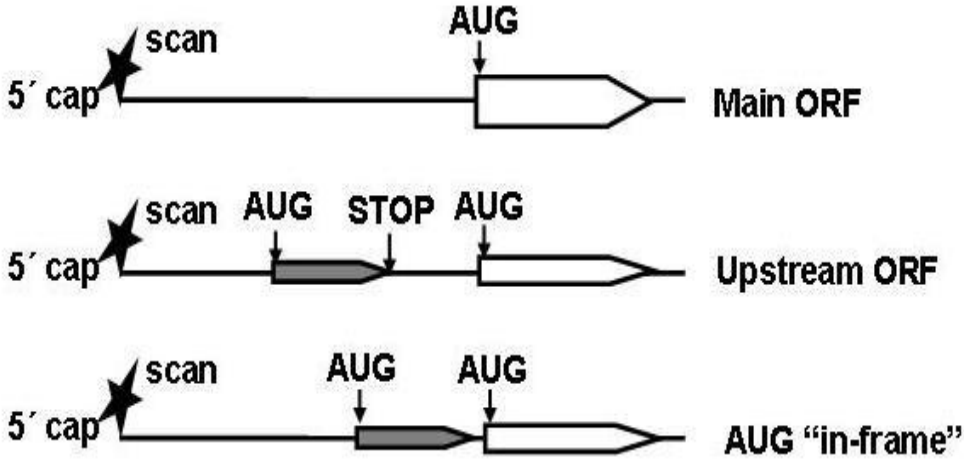

Fig. 1. The scanning model for the mRNA translation. The scanning of the mRNA starts at the 5´cap-structure. The protein synthesis starts when the scanning machinery encounters an AUG codon. *Top:* No AUG codon precedes the open reading frame (ORF) of the main protein. *Middle:* The principal ORF is preceded by an AUG followed by a stop codon. This arrangement makes the translation of the main protein deficient. *Bottom:* The principal ORF is preceded by an AUG, which is in-frame with the main protein. Sometimes this also represses translation

Evolutionary conservation of uAUG and uORF

Investigation of the evolutionary conservation of the upstream control elements (uAUGs and uORF) may clarify their importance and mechanisms of action (Iacono et al., 2005; Somers et al., 2013; Chew et al., 2016). In an earlier publication (Karagyozov & Böhmer, 2009) it was shown that in the protein tyrosine phosphatases the uAUGs and uORF are differentially conserved between human and mouse genomes.

The erb-b2 receptor tyrosine kinase 2

The receptor tyrosine kinases are a large group of cell surface receptors, which bind various regulatory polypeptides: growth factors, hormones, and cytokines. The gene *erb-b2* encodes a member of the family of epidermal growth factor (EGF) receptors. In humans under-expression or over-expression of the ERB-B2 protein is associated with embryonic defects and development of cancers (Citri, & Yarden, 2006). The structure and the mechanism of derepression of the uORF in *ERBB2* are investigated in human normal and cancer cells (Mehta et al., 2006).

Aim of present work

This short communication investigates the conservation of the upstream ORF of the erb-b2 receptor tyrosine kinase 2 in vertebrates.

Materials and methods

Data bases

The genes orthologous to the human *ERBB2* gene and their mRNAs were identified using the public resources of the National Center for Biotechnology, US National Library of Medicine.[1)] Data were collected for model organisms: human (*Homo sapiens*), mouse (*Mus musculus*), chicken (Aves, *Gallus gallus*), tropical clawed frog (Amphibia, *Xanopus tropicalis*), zebra fish (Actinopterygii, *Danio rerio*). Data for marsupials, opossum (Mammalia, Marsupialia, *Monodelphis domestica*) was also included.

Data analysis

The nucleotide sequences were compared and managed by Clone Manager Suite 8 (Scientific and Educational Software). The Kozak consensus sequence - A/GCCaugG - was used for estimation of the favourable start codon surrounding context; for mammals the most important determinants are the G at position +4 (immediately following the AUG codon), and the purine (preferably A) at -3 (three nucleotides upstream of the AUG codon) (Kozak, 1987).

Results and discussion

Characterization of the erb-b2 gene *transcripts*

The main features of the orthologous *erb-b2* gene transcripts from six vertebrate species are presented in Table 1.

Inspection of the cDNA sequences showed that the erb-b2 receptor tyrosine kinase 2 proteins are rather large – more than twelve hundred amino acids. In all genes the principal protein ORFs are preceded by a short uORF six codons long. The only exception is the uORF in zebrafish (Danio rerio); its uORF is only three codons. The distance between the end of the uORF and the next ORF is rather short (mostly five). This suggests that after the stop at the last codon of the uORF reinitiation still may occur at the next ORF (Sachs & Geballe, 2006).

Table 1. The erb-b2 receptor tyrosine kinase 2 genes in six vertebrate species. The cDNA GenBank accessions, the position of the uORF and the length of the erb-b2 protein are given. (aas – amino acids; bps – base pairs)

Species	GenBank accession (cDNA)	uORF (aas) (codons)	uORF (codons + stop) (bps)	Distance uSTOP - main ORF (bps)	main ORF (aas)
Homo (human)	NM_004448	6	21	5	1,255
Mus (mouse)	NM_001003817	6	21	5	1,256
Opossum (marsupial)	XM_001370775	6	21	5	1,257
Gallus (bird)	NM_001044661	6	21	5	1,235
Xenopus (frog)	XM_002940374	6	21	4	1,316
Danio (bony fish)	NM_200119	3	12	3	1,275

Conservation of the uORF in receptor tyrosine kinase 2 erb-b2 genes

The genes for receptor tyrosine kinase 2 (*erb-b2*) from six vertebrate species were compared. Model species from four main classes of Vertebrata were selected (*cf*. Materials and nethods). All investigated genes possessed an uORF, details of which are presented in Fig. 2.

From data inspection several conclusions are obvious.

First of all, the uORF in humans and the mouse are identical in nucleotide and in amino acid sequence. This is not unusual as both species belong to the group of placental mammals. The opossum is an American marsupial; although a mammal, it is phylogenetically rather distant from placentals. The opossum nucleotide sequence, compared to humans, contains four base pair mismatches, each in a different codon. Two of these mismatches, however, do not change the codon meaning. Thus, in the opossum uORF only two amino acids are different, compared to humans and mouse.

Secondly, the chicken and frog sequences are diverse; however, they retain the same basic features of the mammalian uORF. They share the same number of codons (six), they use the same stop codon (TGA). Notably, the distance between the main ORF and the upstream stop codon is such (five and four) that both ORF are not in-frame. This diminishes severely the ability of the scanning ribosomes to reinitiate. Thus, the *erb-b2* genes appear to use other mechanisms to soften the repression exerted by the presence of an uORF (Mehta et al., 2006).

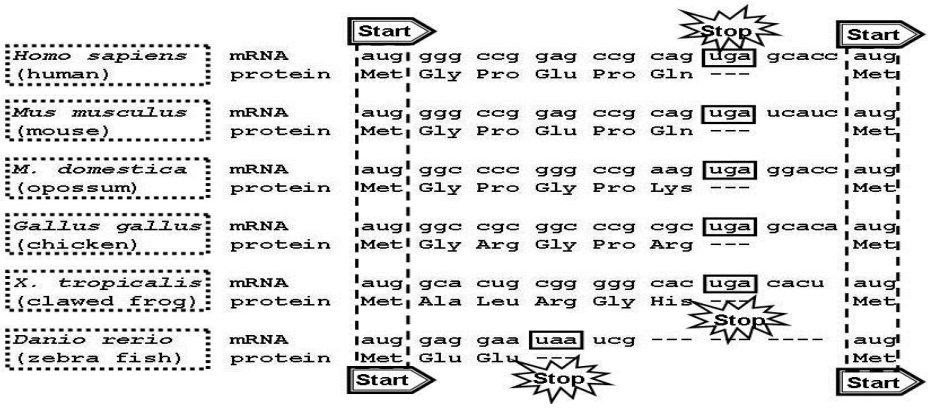

Fig. 2. The upstream open reading frames in receptor tyrosine kinase 2 (*erb-b2*) genes from six vertebrate species. Indicated are the AUG start codons of the uORFs and of the main ORFs. The stop codons of the uORFs are also shown

On the contrary, the upstream ORF in bony fishes is markedly different. It is shorter; its position is in-frame with the ORF of the main protein; it uses a different stop codon. This suggests a different type of regulation in bony fishes, compared to other vertebrates (Wethmar et al., 2016).

The overall conclusion is that the erb-b2 receptor tyrosine kinase 2 regulation as it exists in placentals appeared after the split of the amphibians. The fossil record and the molecular clock indicate that this occurred some 370 million years ago, in the Devonian period.[2)]

NOTES

1. https://www.ncbi.nlm.nih.gov/
2. http://sci.waikato.ac.nz/evolution/AnimalEvolution shtml#evolutionofamphibian

REFERENCES

Barbosa, C., Peixeiro, I. & Romão, L. (2013). Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet.*, *9*(8), e1003529.

Chew, G.-L., Pauli, A. & Schier, A F. (2016). Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. Nature Commn., *7,* art.no. 11663.

Citri, A. & Yarden, Y. (2006). EGF–ERBB signalling: towards the systems level. *Nature Rev. Mol. Cell Biol. 7*, 505 – 516.

Iacono, M., Mignone, F. & Pesole, G. (2005). uAUG and uORFs in human and rodent 5′ untranslated mRNAs. *Gene, 349*, 97 – 105.

Karagyozov, L. & Böhmer, F.D. (2009). Conservation of the upstream AUGs and ORFs in the human and mouse 5′untranslated region of the mRNAs for protein tyrosine phosphatases. *Biotech. & Biotechnol. Equipment, 23*(Sup1), 458 – 462.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res., 15*, 8125 – 8148.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene*, *234*, 187 – 208.

Mehta, A., Trotta, C.R. & Peltz, S.W. (2006). Derepression of the Her-2 uORF is mediated by a novel post-transcriptional control mechanism in cancer cells. *Genes & Dev.*, *20*, 939 – 953.

Sachs, M.S. & Geballe, A.P. (2006). Downstream control of upstream open reading frames. Genes *& Dev.*, 20, 915 – 921.

Somers, J., Pöyry, T. & Willis, A.E. (2013). A perspective on mammalian upstream open reading frame function. *Int. J. Biochem. & Cell Biol., 45*, 1690 – 1700.

Wethmar, K., Schulz, J., Muro, E.M., Talyan, S., Andrade-Navarro, M.A. & Leutz, A. (2016). Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. *Oncogene, 35*, 1736 – 1742.

✉ **Dr. Luchezar Karagyozov**
6-A, Razlaka St.
1111 Sofia, Bulgaria
E-mail: lucho3k@gmail.com