doi:10.1016/j.jmb.2004.06.070 J. Mol. Biol. (2004) 341, 883–886

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

COMMUNICATION

From "Junk" to Gene: *Curriculum vitae* of a Primate Receptor Isoform Gene

Silke S. Singer[1]*, Daniela N. Männel[1], Thomas Hehlgans[1] Jürgen Brosius[2] and Jürgen Schmitz[2]

[1]*Institute of Immunology University of Regensburg Franz Josef Strauss Allee 11 D-93053 Regensburg, Germany*

[2]*Institute of Experimental Pathology (ZMBE), University of Münster, Von-Esmarch-Str. 56, D-48149 Münster, Germany*

**Corresponding author*

Exonization of Alu retroposons awakens public opinion, particularly when causing genetic diseases. However, often neglected, alternative "Alu-exons" also carry the potential to greatly enhance genetic diversity by increasing the transcriptome of primates chiefly *via* alternative splicing. Here, we report a 5′ exon generated from one of the two alternative transcripts in human tumor necrosis factor receptor gene type 2 (*p75TNFR*) that contains an ancient Alu-SINE, which provides an alternative N-terminal protein-coding domain. We follow the primate evolution over the past 63 million years to reconstruct the key events that gave rise to a novel receptor isoform. The Alu integration and start codon formation occurred between 58 and 40 million years ago (MYA) in the common ancestor of anthropoid primates. Yet a functional gene product could not be generated until a novel splice site and an open reading frame were introduced between 40 and 25 MYA on the catarrhine lineage (Old World monkeys including apes).

Keywords: Alu exonization; gene evolution; primates; tumor necrosis factor receptor gene; differentially spliced transcripts

The pleiotropic activities of the cytokine tumor necrosis factor (TNF) are mediated by two distinct membrane receptors, p55 and p75TNFR.[1] While studying the transcriptional regulation of the *p75TNFR* gene, a novel intracellular p75 receptor isoform (icp75TNFR) was identified.[2] The two isoforms are produced *via* alternative transcripts generated from two start sites within the *p75TNFR* gene and are differentially spliced. In mRNA II, which encodes the icp75TNFR isoform, exon 1a replaces exon 1 of mRNA I that contributes the signal peptide and provides an alternative N terminus derived from the center of an Alu element instead (Figure 1).

Alu elements represent a primate-specific family of short interspersed elements (SINEs), which, as such, propagate *via* an RNA intermediate.[3] They are about 300 bp in length and have an estimated copy number of more than 10^6/human genome. Interspersed Alu sequences are present in numerous transcripts of protein-coding genes.[4] Because they often cause deleterious mutations or genetic diseases, Alu element insertions are generally regarded as disadvantageous or at most selectively neutral. However, host genomes might also benefit from Alu integrations, as they are implicated in gene regulation and transcriptome variability and versatility as exaptations.[5,6]

In humans at least 5% of all alternatively spliced exons within protein coding regions contain sequences from Alu elements.[7] Alu consensus sequences have been shown to harbor up to ten potential 5′ splice sites (donor) and 13 potential 3′ splice sites (acceptor).[7] The vast majority of splice sites (19 out of 23) is located on the antisense strand.[7] Furthermore, the adenosine-rich regions of Alu elements can provide, again in the opposite orientation, pyrimidine-rich tracts that are important components of 3′ splice sites.[8] Consequently, most Alu-containing exons (85%) involve the minus strand.[7] In most instances, where an Alu element is inserted within the protein-coding region, it results in a shortened protein *via* frameshift or in-frame stop codon.[7] In case of icp75TNFR, the Alu element that delivers the novel exon is located proximal to the major transcript. Hence, a functional 3′ splice

Abbreviations used: MYA, million years ago; TNF, tumor necrosis factor; SINE, short interspersed element; ORF, open reading frame.

E-mail address of the corresponding author: silke.singer@klinik.uni-regensburg.de

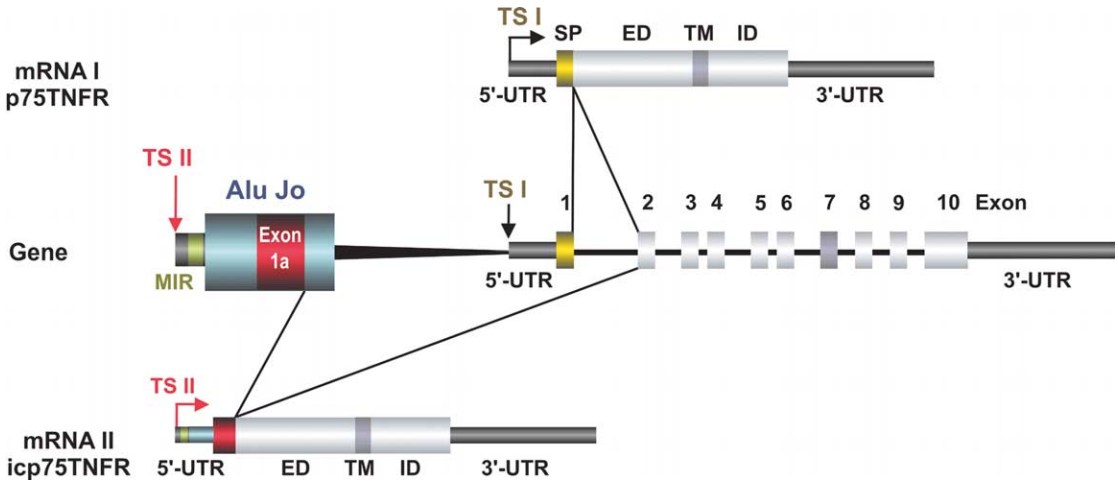

Figure 1. Structure of the *p75TNFR* gene (center) and the two p75TNFR mRNA splice variants. mRNA I and mRNA II originate at the transcriptional start sites TSI and TSII, respectively. Introns are depicted as thin lines, 5′ and 3′ untranslated regions (5′ UTR and 3′ UTR) of the terminal exons as grey bars (intermediate thickness) and protein-coding regions of exons as thick bars. The extracellular domain (ED), transmembrane domain (TM), and intracellular domain (ID) are indicated. The part of exon 1 in mRNA I that contributes the signal peptide (SP) is highlighted in yellow. The proximal portion of the gene is enlarged. A mammalian interspersed repeat (MIR, green) precedes the Alu Jo element (blue), part of which contributes the ATG start codon of mRNA II (at the left border of the protein coding domain, shown in red, of exon 1a). The right border of the exon 1a protein-coding domain (red) is determined by a 5′ splice site. The differential splicing of the first exons of the alternative transcripts to exon 2 is indicated by the respective lines.

site is not required. The Alu element, however, provides a functional ATG start codon and a 5′ splice site linking this Alu-derived 5′ exon *via* exon 2 to the remainder of the open reading frame (ORF).

Consistent with the observation that older Alu-subfamilies are over-represented in exons compared to younger ones,[7] the Alu element that gave rise to exon 1a is a member of the Alu Jo subfamily, one of the oldest dimeric Alu subfamilies.[9] Since older elements accumulate mutational changes over a longer period of time they have a higher chance of acquiring the changes required for exonization.

We compared the human Alu sequence with a consensus Alu Jo sequence and identified three alterations essential for creating a functional translated 5′ exon (Figure 2(a)). An A→G transition at position 101 of the Alu Jo consensus sequence yielded the start codon ATG. A C→T transition at position 188 generated the GT at the 5′ end of intron 1, which form part of a consensus splice donor site.[10] A 7 bp deletion from positions 119 to 125 of the Alu Jo consensus enabled an ORF that includes the subsequent exons.

The Alu Jo element is thought to have retroposed into its genomic locus during early primate evolution at least 40 million years ago (MYA).[9] Based upon the unique and irreversible character of Alu transposition,[11,12] copies of the same element present in different taxa provide evidence of an integration event that took place in the germ line of a common ancestor. From the viewpoint of phylogeny, orthologous sequences of the living primates represent different evolutionary stages along the way to the human exon 1a. Thus, by comparative sequence analysis we reconstructed, the evolution of exon 1a of *p75TNFR*. Using various sets of primers we obtained PCR amplicons representing all major primate clades. By sequencing, we confirmed the presence of the Alu Jo element in Catarrhini and Platyrrhini and its absence in *Tarsius* and *Lemur* (GenBank accession numbers AY548516–AY548527). Therefore, we assigned the integration event to the lineage leading to anthropoid primates after the Tarsioidea diverged.

Thus, an evolutionary scenario was delineated as follows (Figure 2(b)): after the divergence of tarsiers, an Alu Jo transposed upstream of the *p75TNFR* gene in the common ancestor of all anthropoids between 58 and 40 MYA. Following the Alu insertion but before the platyrrhines' divergence, an A→G transition occurred that

Biosystems) according to the manufacturer's instructions. Sequences obtained from non-human primate species were aligned to the human cDNA ortholog using Clustal W. Alu elements were detected and classified by the RepeatMasker Server (Smit & Green, RepeatMasker at http://www.repeatmasker.org/). We verified reading frames and checked for possible frame shifts and stop codons with the aid of the biological sequence alignment editor (BioEdit; available from http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The results revealed that Alu insertion and start codon mutation date back to a common ancestor of anthropoid primates. However, the splice site mutation and the ORF occurred after separation of Old World and New World monkeys on the lineage leading to the Old World monkeys. The alternative intracellular protein product of the *p75TNFR gene* can therefore potentially be expressed only in catarrhines but not in platyrrhines and prosimians.

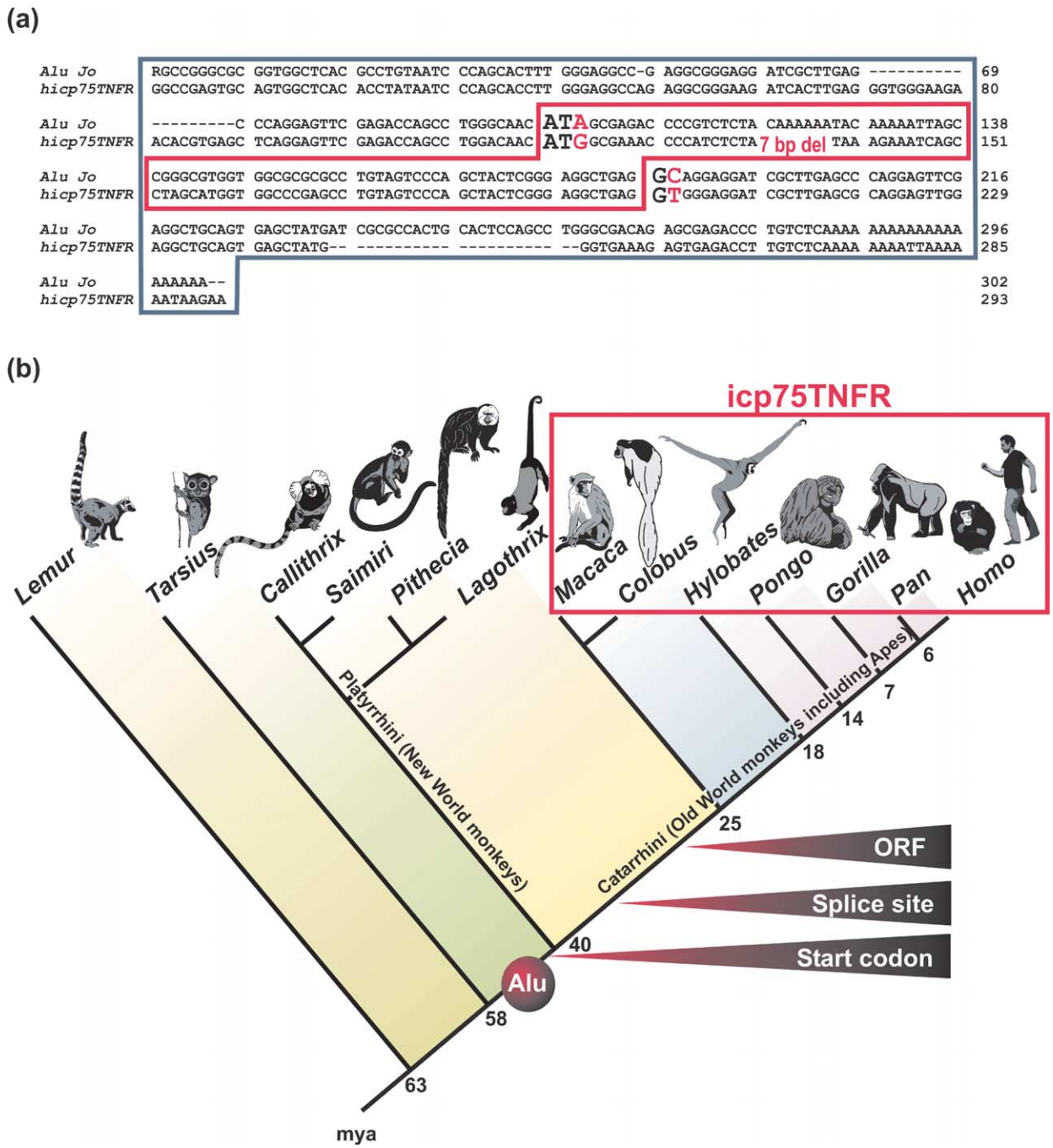

Figure 2. Evolution of exon 1a of the *p75TNFR* gene. (a) Genomic evolution: alignment of part of the human *p75TNFR* 5′ gene with an Alu Jo consensus sequence demonstrates that exon 1a (red border) is derived from the center of an Alu element (blue border). Three alterations yielding a start codon (ATG), a 5′ splice signal (GT), and an ORF (*via* a 7 bp del) produced an exon that contributes an alternative N terminus to the remainder of the protein. (b) Phylogeny: comparative sequence analysis of different primate genera permits mapping of the pivotal molecular events in addition to generation of the alternative transcript onto the primate tree (topology and ages drawn from Goodman *et al.*[13]). Genomic DNA was extracted from tissue samples of representatives of all major primate clades. In each species, various sets of primers derived from human and mouse cDNA sequences were tested for successful polymerase chain reaction (PCR) amplification of the human exon 1a ortholog. A first PCR was performed with a forward primer specific to a conserved human–mouse region and a reverse primer located in exon 1 to allow successful amplification even in evolutionarily distant taxa. A second PCR was done with nested primers selected from the human sequence only, that amplified fragments no longer than 600 bp in order to facilitate routine sequencing procedures. For standard gradient PCR reactions (Taq polymerase kit, Qiagen; Biometra T Gradient Cycler) we used the following conditions: four minutes at 94 °C pre-denaturation, 36 cycles of 30 seconds at 94 °C denaturation, 40 seconds at primer specific annealing temperature gradient and 60 seconds per 1 kb at 72 °C elongation (primers and annealing temperatures available from the authors upon request). The PCR products were separated on a 1% (w/v) agarose gel (1% SeaKem, Biozym), visualized by ethidium bromide staining and purified using the High Pure PCR Product Purification Kit (Roche). Positive PCR products were directly sequenced with the AmpliTaq FS BigDye Terminator Kit (version 2.0; PE

provided the potential start codon ATG. After the divergence of New World monkeys and Old World monkeys, but before the divergence of cercopithecoids and apes, about 40 to 25 MYA, a C→T transition yielded a potential splice donor site in Old World monkeys. In the same lineage, seven nucleotides of the Alu sequence were deleted, which enabled an ORF. New World monkeys show a range of variations at the respective loci (e.g. varying lengths of oligo(A), CA-dinucleotides; see GenBank entries) indicating that there has been no selection pressure on this sequence. We are not able to resolve the exact order of the C→T transition and the 7 bp deletion because there are no living links between the catarrhine and platyrrhine lineages. Likewise, we cannot determine when the alternative transcriptional start site (TS II) arose.

From a functional point of view these findings indicate that the alternative exon 1a of human *p75TNFR* might also be active in apes and cercopithecoids. However, it cannot be active in New World monkeys. Although they do have an ATG that could serve as a start codon for translation, they cannot produce a functional protein because they are lacking the functional splice donor site of the catarrhines and the ORF is interrupted by stop codons or frameshifts. Thus, only the coincidental occurrence of several key mutations permitted the Alu sequence's exonization.

Acknowledgements

We thank Christian Roos for providing samples, Andy Scheffel for help with the Figures and Marsha Bundman for editorial assistance. This work was supported by the German Human Genome Project through the BMBF (01KW9966) (to J.B.) and the Deutsche Forschungsgemeinschaft (SFB585 TPB1) (to D.N.M.).

References

1. Hehlgans, T. & Männel, D. N. (2002). The TNF–TNF receptor system. *J. Biol. Chem.* **383**, 1581–1585.
2. Seitz, C., Müller, P., Krieg, R. C., Männel, D. N. & Hehlgans, T. (2001). A novel p75TNF receptor isoform mediating NFkappa B activation. *J. Biol. Chem.* **276**, 19390–19395.
3. Schmid, C. W. (1998). Does SINE evolution preclude Alu function? *Nucl. Acids Res.* **26**, 4541–4550.
4. Nekrutenko, A. & Li, W. H. (2001). Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621.
5. Brosius, J. & Gould, S. J. (1992). On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc. Natl Acad. Sci. USA*, **89**, 10706–10710.
6. Makalowski, W., Mitchell, G. A. & Labuda, D. (1994). Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* **10**, 188–193.
7. Sorek, R., Ast, G. & Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067.
8. Brow, D. A. (2002). Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**, 333–360.
9. Jurka, J. & Milosavljevic, A. (1991). Reconstruction and analysis of human Alu genes. *J. Mol. Evol.* **32**, 105–121.
10. Zhang, M. Q. (1998). Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* **7**, 919–932.
11. Schmitz, J., Ohme, M. & Zischler, H. (2001). SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. *Genetics*, **157**, 777–784.
12. Singer, S. S., Schmitz, J., Schwiegk, C. & Zischler, H. (2003). Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). *Mol. Phylogenet. Evol.* **26**, 490–501.
13. Goodman, M., Porter, C. A., Czelusniak, J., Page, S. L., Schneider, H. Shoshani, J. *et al.* (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* **9**, 585–598.

Edited by J. Karn

(Received 14 May 2004; received in revised form 28 June 2004; accepted 28 June 2004)