Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 85 (2005) 431–440

GENOMICS

www.elsevier.com/locate/ygeno

Intergenic transcripts in genes with phase *I* introns☆

Jana Královičová, Igor Vořechovský*

Division of Human Genetics, University of Southampton School of Medicine, Southampton University Hospital, MP808, Tremona Road, Southampton SO16 6YD, UK

Received 16 July 2004; accepted 3 December 2004
Available online 27 January 2005

Abstract

Nonsense-mediated mRNA decay (NMD) is a eukaryotic quality-control mechanism that detects and degrades aberrant transcripts prematurely terminating translation. NMD may be elicited by intergenic transcripts that contain premature termination codons (PTCs), but chimeric mRNAs of genes that have introns of identical phase would be predicted to lack PTCs and escape NMD. We examined intron phase *I*-containing *HLA* class II genes for the presence of intergenic mRNAs and found an extraordinary diversity of correctly spliced and polyadenylated intergenic transcripts. They lacked a significant homology at the chimeric joins and had no PTCs. Their expression levels were very low and positively correlated with the expression of natural transcripts. In contrast, pair-wise mixtures of separately transcribed plasmids carrying full-length *HLA-DQB1*, *-DQA1*, *-DRB1*, and *-DRA* cDNAs produced only hybrid molecules that lacked canonical exon boundaries, had homologous chimeric joins, and occasionally contained PTCs, implicating in vitro artifacts generated by template switching of *Taq* polymerase and reverse transcriptase. A differential exon structure of hybrid molecules observed in vitro and in cellular RNA preparations suggests that intergenic mRNAs with canonical exon boundaries arise in vivo during exon joining and/or transcription. Since the observed intergenic mRNAs may encode mixed class II heterodimers that were previously shown to present antigens it will be interesting to determine functional properties of such molecules in future studies.

Keywords: mRNA; Splicing; Exon; Lymphocytes; Intergenic; MHC; HLA; RNA surveillance

Initial sequencing and analysis of the human genome showed that the gene number could not be the sole determinant of high proteomic complexity and highlighted the importance of mechanisms that expand the RNA repertoire [1,2]. Although alternative pre-mRNA splicing, or a formation of multiple mRNAs from a single pre-mRNA molecule, is a major source of RNA and proteomic diversity in vertebrates (reviewed in [3]), there is also mounting evidence for the existence of atypical pre-mRNA processing events that result in transcripts in which exons are joined in an incorrect order or orientation. They include exon scrambling [4–7], exon repetitions [8,9], and antisense [10,11] or intergenic transcription/splicing [12–16]. In addition to these mechanisms, intergenic transcripts may be formed by a simple transcriptional termination bypass of tandemly arranged genes [17–21]. Although the biological and evolutionary significance of these processes is poorly understood, such transcripts have a potential to increase further the functional diversity of the genome [10,14,15].

Heterologous *trans*-splicing, or a generation of single mRNAs from independent primary transcripts, is widespread in intron-free trypanosomes and is manifested as the addition of a splice leader sequence to most, if not all, mRNAs. This mechanism has not been found in higher eukaryotes, but independent studies in vertebrates have supported the existence of a similar phenomenon designated "alternative" *trans*-splicing [8,14,15,22,23]. This mechanism does not involve an obligatory splice leader addition to achieve RNA maturation, but simply joins correctly spliced exons from separate pre-mRNAs. Alternative *trans*-splicing was also reported for hybrid viral and cellular mRNAs in

☆ Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AF522250 and AF522251.

* Corresponding author. Fax: +44 2380 794264.
E-mail address: igvo@soton.ac.uk (I. Vořechovský).

0888-7543/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2004.12.001

infected cells [24]. However, the evidence for *trans*-splicing in vertebrates has been questioned, because rare preexisting hybrid mRNAs are difficult to distinguish from chimeric molecules arising in vitro by template-switching activities of reverse transcriptase (RT) or *Taq* polymerase [15,25]. On the other hand, the rarity of "nonlinear" RNA processing events can be explained by the presence of premature termination codons (PTCs) in hybrid mRNAs and their degradation by RNA surveillance mechanisms such as nonsense-mediated mRNA decay (NMD). In addition, circular RNAs that may arise as intermediate molecules during atypical RNA processing events [4,5] are not polyadenylated and are likely to be underrepresented in cDNA libraries. Furthermore, cleavage followed by polyadenylation of such circular intermediates may result in potentially unstable linear RNAs due to their uncapped 5′ termini.

Human histocompatibility leukocyte antigens (HLA) are characterized by an extraordinary degree of selection-driven polymorphism generated by a variety of genetic processes, including gene conversion, recombination, and mutation [26]. HLA molecules are encoded by closely linked genes in the major histocompatibility complex (MHC) at 6p21 and comprise telomeric class I and centromeric class II loci. The two classes of MHC molecules are expressed differentially on cells and have distinct subunit organizations, but similar three-dimensional structures. The *HLA* genes also have a similar exon organization, with the class II exon 1 coding for signal peptide (SP), exons 2 and 3 encoding extracellular domains, and the remaining variable exons coding for a single-pass transmembrane and cytoplasmic domains [26]. These genes contain only phase *I* introns (i.e., introns that interrupt codons between the first and the second nucleotides), and putative intergenic mRNAs (IR_{HLA}) would be predicted to lack PTCs, escape NMD, and be detectable in cellular RNA preparations at unbiased levels. In addition, the closely linked class II genes are in both tandem and opposite transcriptional orientations [26], which would permit detection of putative IRs that originate both by *trans*-acting (alternative *trans*-splicing) and *cis*-acting (inefficient transcription termination) mechanisms. Finally, a high number of coding SNPs/alleles [27] would improve unambiguous identification of exons that contribute to putative IRs and also facilitate distinction between IRs generated in vitro by enzymatic template switching and IRs with canonical exon boundaries.

In the present study, we have examined the *HLA* class II locus for the presence of IR_{HLAII} and discuss their origin, abundance, and putative functional significance.

Results

To detect putative chimeric mRNAs of closely linked *HLA-DQB1*, *-DQA1*, *-DRB1*, and *-DRA* (Fig. 1a), we designed a series of nested inverse RT-PCRs in combinations shown in Table 1 and Fig. 1b. Haplotype-specific products were amplified using cDNA samples homozygous for three haplotypes (designated H1, H2, or H3; see Materials and methods), gel-purified, and sequenced. The sequence analysis of 60 consecutive, IR_{HLAII}-containing products showed two categories of chimeric molecules: 75% amplicons had correctly spliced exons (Figs. 1d, 1e, 2a), whereas the remaining 25% lacked canonical exon–exon boundaries at the chimeric joins (Fig. 2b).

Correctly spliced IR_{HLAII} were found for most pair-wise combinations of *DQB1*, *DQA1*, *DRB1*, and *DRA* (Fig. 1d). They all lacked PTCs and were derived from genes in both transcriptional orientations, including head-to-head (*HLA-DRB1* and *-DQA1*) and tail-to-tail (*HLA-DRB1* and *-DRA*; Figs. 1b and 1d). Since several gene combinations were undetected in the initial analysis, which may have been due to PCR failures, we designed a further set of PCR oligonucleotides according to the expected chimeric mRNA (in bold in Table 1) and identified $IR_{DRA\text{-}DQB1}$, although genuine $IR_{DQB1\text{-}DQA1}$ were not found (Fig. 1d). These results suggested that a failure to detect all IR_{HLAII} pairs was likely to be due to primer design and/or a lower expression of DQ compared to DR, rather than the previously proposed directionality in the process of intergenic splicing [14].

In addition to IRs that contained *DRB1*, a gene present on all *HLA* haplotypes, we also found rare hybrid molecules containing haplotype-specific *DRB3* (IR7-13_{103}, IR7-13_{104}) or *DRB5* (IR7-13_{11}) transcripts in samples homozygous for *H1* and *H2*, respectively (Figs. 1b and 1d). In the *HLA-DP* locus, we identified canonical IRs between closely linked *DPA1* and *DPB1*, but not between the *DP* and the *DQ* genes located almost 0.5 Mb apart (Fig. 1e, Table 1).

In contrast, the latter category of hybrid molecules had only incomplete exons at the chimeric joins (Fig. 2b), suggesting that they originated by template switching of viral RT [28] or *Taq* polymerase [29]. We found that the combined size of incomplete joining exons was occasionally identical to the length of canonical exons (IR1-9_{41}, IR1-9_{87}, and IR1-9_{88} in Fig. 2b). To confirm formally that these molecules can be formed artifactually and to test if correctly spliced IR_{HLAII} can originate in vitro, we used the same primer pairs and PCR conditions to amplify pair-wise mixtures of linearized plasmids carrying full-length *DQB1*050101*, *DQA1*010101*, *DRB1*010101*, and *DRA*0101* cDNAs. Inverse PCR produced distinct fragments for most pair-wise plasmid mixtures. Sequencing of 30 products showed that they all lacked full-length exons at the chimeric joins and five had PTCs. As with noncanonical hybrids from cellular RNA (Fig. 2b), the chimeric joins were always in homologous regions shared by the partner genes, implicating the template switching activity of *Taq* polymerase (data not shown).

Next, we separately transcribed plasmids carrying canonical *HLA-DQB1*050101*, *DQA1*010101*, *DRB1*010101*, and *DRA*0101* in vitro and treated them twice with DNase