Duplicated Splice Site (in a Human Class I Histocompatibility Antigen, HLA-CW3 Gene)

Jerzy Jurka and Temple Smith
Dana-Farber Cancer Institute, Harvard University

To our knowledge, no one has heretofore shown a clear-cut example of a duplication of the splicing site. Such events could play an important role in the evolution of alternative splicing and/or the migration of splice sites. In addition, this particular case may be of interest to those studying the splicing mechanism in vivo.

Koller and Orr (1985) aligned homologous regions from HLA-CW3 (Sodoyer et al. 1984), HLA-A2 (Koller and Orr 1985), and HLA-A3 (Strachan et al. 1984). However, a 47-nucleotide tandem repeat between positions 2712-2758 and 2759-2805 in HLA-CW3 (fig. 1), was unreported both in this alignment and in the original sequence. The repeat includes 21 bp from the 3' end of exon 6 and 26 bp from the following intron. Of these 26 bases the last 20 are unique to HLA-CW3. The corresponding

```
exon 6
GlyGlyLysGlyGlySerCysSerGlnAlaAlaAlaS
C  ysLys STOP

intron
HLA-CW3 GTGGAAAGGAAGGAGCTGCTCTCAGGCTGCGTG  GTAAG TGA TGGCGGTGGGCGTGTTGGA  2758
HLA-A3 ATAGAAAGGAAGGAGTTACACTCAGGCTGCAA GTAAG T-- -----------------  2796
HLA-A2 ATAGAAAAGGAAGGAGCTACTCTCAGGCTGCAA GTAAG T-- -----------------  3082

GAP REGION I

HLA-CW3 --GAGCTCA-----------------------------GGAGCTCA  2813
HLA-A3 ATGAAAGGAGCTGATGATGCTACTTCGAGGATATTGTTTGGGAGCCCATGGGGGAGCCCA  2859
HLA-A2 ATGAAAGGAGCTGATGATGCTACTTCGAGGATATTGTTTGGGAGCCCATGGGGGAGCTCA  3145

HLA-CW3 CCCACCCCATAATTCTCCTGTG-CCACATCTCTGCGGCTGCTGACCAGCTCTTGTGTTTTTTTTTG  2875
HLA-A3 CCCACACCTACAATTCTCTCTCTAGGCCATCTTTCTGCGGATCTGACCAGGT----TCTGTTTTTTG  2921
HLA-A2 CCCACCCCACAATTCTCCTTCTAGGCCCATCTTTCTGCGGATCTGACCAGGT----TCTGTTTTTTG  3207

er (3'-end of exon 7)

HLA-CW3 TTCTACCCAG GC  2888
HLA-A3 TTCTACCCAG GC  2934
HLA-A2 TTCTACCCAG GC  3220
```

FIG. 1.—Alignment of HLA-CW3, HLA-A3, and HLA-A2 showing duplication in HLA-CW3. The amino acid sequence is for HLA-CW3. The space preceding GTAAG in the uppermost sequences separates exon 6 from the 5' end of the intron. The last two nucleotides mark the 3' end of exon 7. The upper amino acid sequence is that expected if the splice sites used are the same in HLA-CW3 as in HLA-A2 and HLA-A3. The continuation on the line below shows the result if the 5' end excision site in HLA-CW3 were to occur 47 nucleotides farther downstream. Gaps (—) were inserted by eye to improve the alignment.

Address for correspondence: Dr. Jerzy Jurka, MBCRR, Dana-Farber Cancer Institute, Harvard University, 44 Binney Street, Boston, Massachusetts 02115.

© 1987 by The University of Chicago. All rights reserved.
0737-4038/87/0402-4208$02.00
region in HLA-A2 and HLA-A3 has 57 bases that show no obvious similarity to HLA-CW3. The similarity begins at position 2806, immediately after the end of the repeat (gap region II). This indicates that, prior to the duplication, part of the intron similar to region II in HLA-A2 and HLA-A3 has been replaced by a 20-basc-long fragment present only in HLA-CW3. An alternative explanation, involving the expansion of this region in an ancestor common to HLA-A2 and HLA-A3, seems less likely.

The HLA-CW3 gene contains eight exons. The first five exons encode a hydrophobic signal peptide removed from the mature protein, three extracellular domains, and the transmembrane region. Exon 6, analyzed in this paper, is followed by exon 7, coding for 15 amino acids, and by exon 8, sharing terminal alanine with exon 7. The last three exons encode the cytoplasmic portion of the protein (Strachan et al. 1984). The additional splicing site created by the duplication of the exon 6/intron boundary opens the possibility for an alternative splicing. The new mRNA would include 47 extra nucleotides between the first and second splicing sites. The translation of this mRNA, if it occurred, would be prematurely terminated, owing to the presence of the terminating codon TGA following exon 6. The new protein would replace, with cysteine and lysine, serine encoded by exons 6 and 7 and the last 16 amino acids encoded by exons 7 and 8 (fig. 1). The HLA-CW3 provides the first sequence information about the HLA-C locus, whose protein products are very poorly expressed on the cell surface (Sodoyer et al. 1984). It remains to be established whether the additional splicing site after exon 6 is related to this low level of expression.

One possibility is that the truncated product of the HLA-CW3 gene is secreted. We note in this context that a class I(H-2)-related gene in mouse encodes a transplantation-like antigen that has the entire cytoplasmic portion of the protein truncated. This antigen is produced only in liver cells and is secreted (Kress et al. 1983).

Acknowledgment

We thank Allan Maxam for helpful discussions and pointing out the truncated antigen in mouse.

LITERATURE CITED


