Vol. 204, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 14, 1994 Pages 195-202

CLONING OF cDNAs AND THE MOLECULAR EVOLUTION OF A BOVINE MHC CLASS II *DRA* GENE+

Yoko Aida[1*], Chikara Kohda[1,2], Akira Morooka[1,3], Yutaka Nakai[2], Keiji Ogimoto[2], Takeshi Urao[4] and Masatoshi Asahina[1,5]

[1]Laboratory of Gene Technology and Safety and [4]Laboratory of Plant Molecular Biology, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan

[2]Faculty of Agriculture, Tohoku University, Sendai 981, Japan

[3]Nippon Veterinary and Animal Science University, 1-7-1 Kyonan-cho, Musashino, Tokyo 180, Japan

[5]Faculty of Agriculture, Iwate University, Morioka 020, Japan

Received August 30, 1994

SUMMARY: Two overlapping cDNA clones coding for bovine major histocompatibility complex (MHC) class II antigen DRA chain were isolated and characterized. The full-length cDNA clone, MR1, encoded a primary translated product of 253 amino acids, 24 of which were deduced to be a signal peptide and 229 which formed a mature polypeptide. The amino acid sequences deduced from this clone resembled those of class II A molecules from other species in both size and structure, but no potential consensus site of *N*-linked glycosylation comparable to those in the human, mouse, rat and swine proteins was found in the α2 domain, as well as ovine and equine DRA molecules. Comparison of amino acid sequences encoded by class II *A* genes among several species and a dendrogram constructed from these data places the *DRA* gene and the *DQA/DYA* genes on two distinct branches of a phylogenetic tree, with bovine *DRA* and ovine *DRA* being most similar on the *DRA* branch.

The major histocompatibility complex is an extensive genetic region of clustered genes that control the regulation of the immune response (1). In cattle, the class II genes, at the locus designated *BoLA-D*, have been found to control mixed lymphocyte reactivity (2). Studies of restriction fragment length polymorphism (3,4) and the characterization of cloned bovine class II genes (5) indicate that the organization of *BoLA* is similar to that of *HLA* in man. Evidence has been obtained for presence of a single *DRA* gene, multiple (at least three) *DRB* genes, one or two *DQA* and *DQB* genes (depending on the haplotype), single *DOB* and *DNA* genes, and nonorthologous genes *DYA*, *DYB* and *DIB* (3,4, 6-9). At present, however, very little information is available about the bovine α-chain sequence. The sequences of three partial genomic clones, representing

+Accession numbers for KQ1 and MQ1 are D37955 and D37956, respectively.

*To whom correspondence should be addressed. Fax: +81-298-36-9050.

0006-291X/94 $5.00

one *DRA*, one *DQA* and one *DYA* genes, and that of one partial cDNA clone representing *DRA* have been reported (10,11), whereas the sequence of exon 1, which encodes the signal peptide of these bovine α-chains, is not yet of record. Moreover, as compared to the human gene, little is known about the role of class II in the regulation of the immune response, or about associations between class II genes and disease susceptibility in cattle. Therefore, as a first step towards an understanding of these issues, we have isolated and characterized a full-length cDNA clone for bovine *DRA*. We have also studied phylogenic relationships based on nucleotide sequences.

MATERIALS AND METHODS

Preparation of cDNA library. Total RNA was isolated from bovine lymphoid cell line BLSC-KU-1 (12) by the guanidine isothiocyanate/CsCl method (13), and poly(A)$^+$ RNA was purified by chromatography on oligo(dT)-cellulose. Double-stranded cDNA was synthesized by a modified version of the Gubler-Hoffman method (14) and ligated with non-palindromic *Bst*X1 linkers (15). The cDNA was size-fractionated (>1.0 kb) by potassium acetate gradient centrifugation, ligated to the mammalian expression vector pCDM8 (16) and introduced into E. *coli* Mc1061/P3.

Isolation and characterization of BoLA-DRA cDNAs. To screen the cDNA library, Hybond^{TM}-N nylon filters (Amersham, UK) containing approximately 5 x 10^5 colonies of the cDNA library were screened with the 1.1 kbase pairs (bp) *Pst*I fragment of *HLA-DPA* cDNA clone, pDAα1-4 (17), by colony hybridization under low-stringency conditions. Plasmid DNAs were isolated and subcloned into pBluescript II SK (+) (Stratagene, Heidelberg, Germany), and the nucleotide sequences of both strands were determined by the dideoxy chain-termination method (18) using a *Bca*BEST sequencing kit (Takara Shuzo Co., Shiga, Japan). Computer analyses were performed with the program package from the Wisconsin Genetics Computer Group, which included FASTA and TFASTA (19).

Construction of dendrograms. Amino acid sequences were aligned, genetic distances between proteins were calculated by Kimura's two-parameter method (20) and dendrograms were constructed by the neighbor-joining method of Saitou and Nei (21).

RESULTS AND DISCUSSION

Approximately 5 x 10^5 colonies were screened with a human *DPA* cDNA probe by colony hybridization under low-stringency conditions, and one positive clone KR1 (1,051 bp) was isolated. The nucleotide sequence of this clone was identical to the sequences of exons 2, 3, and 4 of a genomic *BoLA-DRA* clone, W3, reported by van der Poel *et al.* (10). The cDNA did not, however, contain the sequences of the leader peptide and part of the α1 domain, suggesting that the 5' end of the gene in the KR1 clone had been deleted. To obtain a full-length cDNA clone that encoded *BoLA-DRA*, we rescreened the library (approximately 1 x 10^5 colonies) with a 0.6-kb DNA fragment of the 5' end of KR1 as a probe under high-stringency conditions and a total of three positive clones were isolated. Clone MR1, with the largest insert, was selected and

characterized. Analysis of the nucleotide sequence of MR1 demonstrated that MR1 and KR1 were identical over the 1,007 overlapping bases, and that the MR1 clone was an analogue of the bovine *DRA* gene.

The complete nucleotide sequence and deduced amino acid sequence of MR1 are shown in Figure 1. The cDNA sequence is 1, 240 nucleotides (nt) long with a single open reading frame

```
      5'UT                                    ↓sp
1     AAGACACTGACTCTCATCGAGACACCGAAGAAGAAAATGGCCATAACCAGGGTCCCAATA
                                              MetAlaIleThrArgValProIle
                                                              ↓α1
61    TTAGGACTTTTCATCACTGTCCTGATCGGCCTACAGGAATCGTGGGCTATTAAAGAGAAT
      LeuGlyLeuPheIleThrValLeuIleGlyLeuGlnGluSerTrpAlaIleLysGluAsn

121   CATGTGATCATCCAAGCTGAGTTCTATCTGAAACCTGAGGAATCAGCCGAGTTTATGTTT
      HisValIleIleGlnAlaGluPheTyrLeuLysProGluGluSerAlaGluPheMetPhe

181   GACTTTGATGGTGATGAGATTTTCCACGTGGATATGGGGAAGAAGGAGACGGTGTGGCGG
      AspPheAspGlyAspGluIlePheHisValAspMetGlyLysLysGluThrValTrpArg

241   CTTCCAGAATTTGGACATTTTGCCAGCTTTGAGGCTCAGGGTGCCCTGGCCAATATGGCT
      LeuProGluPheGlyHisPheAlaSerPheGluAlaGlnGlyAlaLeuAlaAsnMetAla

301   GTGATGAAAGCCAACCTGGACATCATGATAAAGCGCTCCAACAACACCCCAAACACCAAT
      ValMetLysAlaAsnLeuAspIleMetIleLysArgSer AsnAsnThr ProAsnThrAsn
      ↓α2
361   GTTCCTCCAGAAGTGACTCTGCTCCCAAACAAGCCTGTGGAACTGGGAGAGCCCAACACA
      ValProProGluValThrLeuLeuProAsnLysProValGluLeuGlyGluProAsnThr

421   CTCATCTGCTTCATTGACAAGTTCTCCCCACCCGTGATCAGTGTCACATGGCTTCGAAAT
      LeuIleCysPheIleAspLysPheSerProProValIleSerValThrTrpLeuArgAsn

481   GGCAAACCTGTCACTGATGGAGTGTCACAGACGGTCTTCCTGCCCAGGAATGACCACCTT
      GlyLysProValThrAspGlyValSerGlnThrValPheLeuProArgAsnAspHisLeu

541   TTCCGCAAGTTCCACTACCTCCCCTTCCTGCCCACAACAGAGGATGTCTATGACTGCAAG
      PheArgLysPheHisTyrLeuProPheLeuProThrThrGluAspValTyrAspCysLys
                                                      ↓CP
601   GTGGAGCACTTGGGTTTGAATGAGCCTCTTCTCAAGCACTGGGAGTATGAAGCTCCAGCC
      ValGluHisLeuGlyLeuAsnGluProLeuLeuLysHisTrpGluTyrGluAlaProAla
                                  ↓TM
661   CCCCTCCCAGAGACCACAGAGAATGCAGTGTGTGCCCTGGGCCTGATTGTGGCTCTGGTG
      ProLeuProGluThrThrGluAsnAlaValCysAlaLeuGlyLeuIleValAlaLeuVal
                                              ↓CY
721   GGCATCATTGCAGGGACCATCTTCATCATCAAGGGCGTGCGCAAAGCCAACACCGTTGAA
      GlyIleIleAlaGlyThrIlePheIleIleLysGlyValArgLysAlaAsnThrValGlu
                                  ↓3'UT
781   CGCCGAGGGCCTCTGTGAGGCGCCTGCAGGTAATGGACTTTGTTACAGAGAAGATCAATG
      ArgArgGlyProLeu***

841   AAGATATTTCTGCCTTAATAGCTTTACAAACCCGGCAATTCTCCAATTGTTCACCTCACT
901   GAAGACCACCATGCTTCAGCACTTCCCAGTCCTTTACTTACCCTAAGAGTAAGATGCCTT
961   CCACAATCTCCAGATTCTAATCTCTAGTTATTCCCTGTCTACTGCTCCTTCTTGTATTGG
1021  TTTTCCCTCCATTTTCCACTGTCTTCTTATTATCACCATGTAATGCCTTTGGAATGGACC
1081  CCATAGGATCCTTTTTTTTCCTGCTACAGAACCTTTGGAAAGTTCTATGGGTGCATCTCT
1141  TGTACACTTATTACTTGGAGTTTCTTCTAACTGTGATTGTCATTTTTCTGAATACAGTAA
1201  ACTTCAAGTGGGTCTCAGGTTGAAAAAAAAAAAAAAAAAAA 1240
```

Figure 1. The nucleotide sequence of the *BoLA-DRA* cDNA clone MR1 and the deduced amino acid sequence. The numbers indicate the numbering of the nucleotide sequence. *Arrows* designate the putative α-chain domains: 5'UT, 5'-untranslated region; SP, signal sequence; α1, first domain; α2, second domain; CP, connecting peptide; TM, transmembrane domain; CY, cytoplasmic domain; and 3'UT, 3'-untranslated region are indicated. The potential sites of *N*-linked glycosylation and two conserved cysteines are indicate in *bold face* and the nucleotides forming a stop codon are indicated by *asterisks*.

that begins with an ATG initiation codon (nt 37 to 39) and ends with a TGA stop codon (nt 796 to 798), predicting a coding sequence of 253 amino acids. The *N*-terminal region of the putative protein begins with the codon ATT for isoleucine at nucleotide position 109, suggesting that there is a signal sequence of 24 predominantly non-polar amino acids, so that the mature protein contains only 229 amino acids. The mature BoLA-DRA polypeptide contains an amino-terminal α1 extracellular domain (positions 1 to 84), an α2 extracellular domain (positions 85 to 178) and a stretch of residues that contains a connecting peptide of 13 amino acids (positions 179 to 191), a highly hydrophobic membrane-anchoring domain of 23 amino acids (positions 192 to 214), and a cytoplasmic tail of 16 amino acids (positions 215 to 229). There is a 3'-untranslated region of 424 nucleotides before a poly(A) tail. No consensus polyadenylation signal, ATTAAA, was found in the 3'-untranslated region. It is of interest here that no polyadenylation signal was also found in the ovine *DRA* gene described by Fabb *et al.* (22). Thus, MR1 encoded the complete precursor molecules encoded by *BoLA-DRA*, with a calculated Mr of 28,269 for the nonglycosylated polypeptide that includes the signal sequence.

Figure 2 shows a comparison of the predicted amino acid sequences from the MR1 clone and from genes for MHC class II α-chains from ovine, human, swine, equine, mouse, rat and cow. In the signal peptide region, MR1 encodes two amino acids fewer than those encoded by the human, mouse and rat *DRA*, but two more amino acids than are encoded by the swine *DRA* gene. Likewise, the lengths of all the other functional domains (α1, α2, CP, TM, and CY) encoded by MR1 cDNA are quite similar to those encoded by the human, equine and swine genes, while being one amino acid shorter than those encoded by the rat and mouse genes. By contrast, all domains encoded by MR1 seem to be of the same size as those encoded by the ovine *DRA*, to which the greatest similarity was noted. As shown in Table 1 and Figure 2, it is evident that the signal peptide contains most of the variable region of the molecule in mammalian equivalents. In contrast, although virtually every amino acid site shows variations between species, there are some blocks of highly conserved residues composed of positions 26-30, 102-107, 120-125, 161-170, 173-178, and 195-199. In particular, in the α2 domain there are two conserved cysteine residues at positions 107 and 163 which form a disulfide loop of 55 amino acids resulting in an immunoglobulin-like structure of alternating β-pleated sheets and bends. All of ten residues that are highly conserved in the immunoglobulin chain (positions 142, 145, 151, 161, 163, 165, 167, 170, 173 and 178) (23,24) are present in the protein encoded by MR1, as indicated by dotted boxes. In contrast to the DRA molecules of human, swine, rat and mouse, which have the two potential sites of *N*-linked glycosylation with the conserved positions (positions 78 and 118) in both the α1 and α2 domains, the bovine DRA molecule has only a single potential site of *N*-linked glycosylation in the α1 domain, as do also the ovine and equine DRA molecules.

A comparison of the predicted amino acids in each domain encoded by MR1 and genes for MHC class II α-chains of bovine, ovine, human, swine, equine, mouse and rat origin is shown in Table 1. The MR1-encoded protein exhibits a clone which has 100% identity in terms of the previously published sequences of exons 2, 3, and 4 of a bovine *DRA* gene, W3 (10), but only 47.8% and 54.2% identity to those encoded by bovine *DYA* and *DQA* genes, respectively. In addition, the extent of the identity between the putative bovine protein and putative ovine, human,

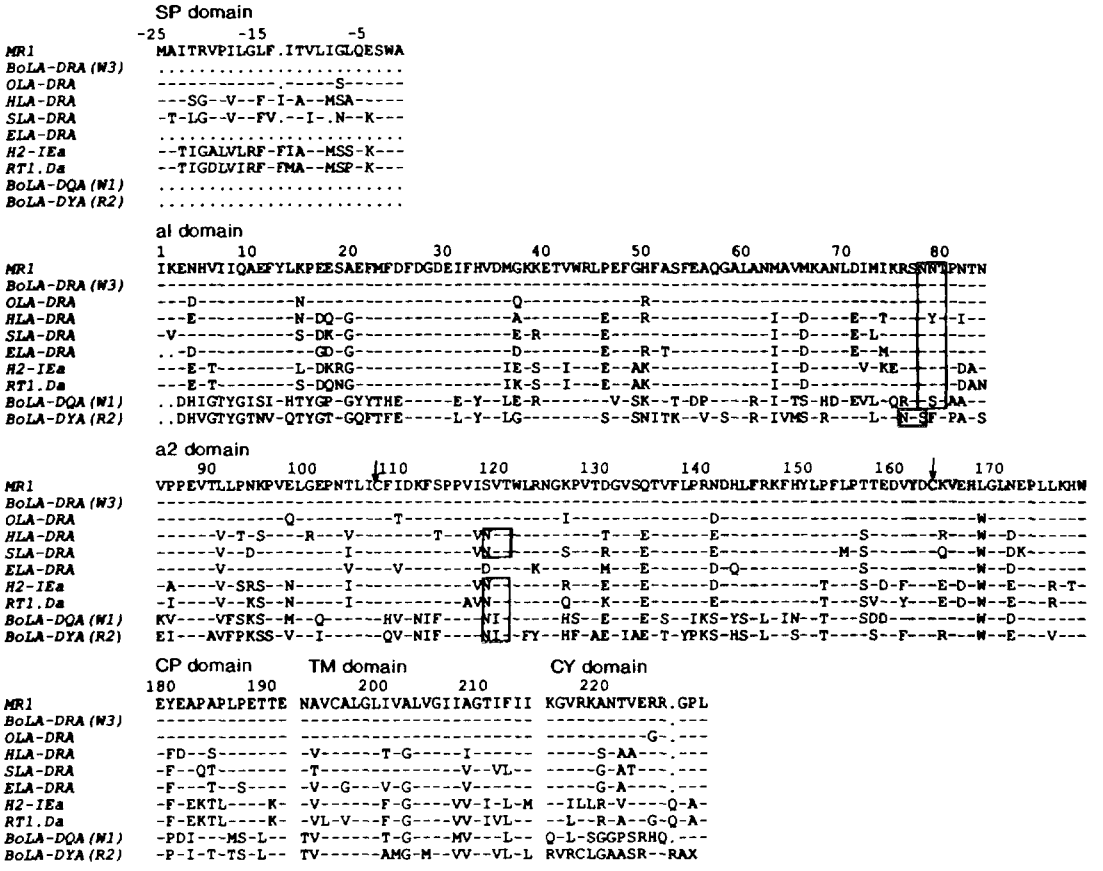

Figure 2. Comparison of the amino acid sequence predicted from the *BoLA-DRA* cDNA clone MR1 with those predicted from genes for the MHC class II α-chains of other species including ovine (*OLA*), human (*HLA*), swine (*SLA*), equine (*ELA*), mouse (*H2*), rat (*RT*) and bovine (*BoLA*). Substituted amino acids, gaps (.) and amino acid identities (-) are indicated. Sequences used for the alignments were taken from reports by Fabb *et al.* (22) for *OLA-DRA*, Koppelman and Cresswell (25) for *HLA-DRA*, Hirsch *et al.* (26) for *SLA-DRA*, Holowachuk *et al.* (27) for *RT1.Da*, Hyldig-Nielsen *et al.* (28) for *H2-IEa*, Albright et al. (29) for *ELA-DRA*, and van der *Poel et al.* (10) for *BoLA-DRA* (W3), *BoLA-DQA* (W1) and *BoLA-DYA* (R2). *Arrows* indicate the two cysteine residues that are presumed to form an interchain S-S bond. Potential sites of *N*-linked glycosylation are boxed. Highly conserved residues in the immunoglobulin-like region are indicated by shaded boxes.

swine, equine, mouse and rat proteins is rather high, ranging from 70% to 96%. Thus, the bovine DRA molecule exhibits greater homology to the DRA molecules of other mammalian species than to the bovine DQA and DYA molecules. The genetic relationships among MHC class II α-chains were further investigated by construction of a phylogenetic tree using the neighbor-joining method (Fig. 3). Phylogenetic analysis revealed the same tendencies as the abovementioned results. The shape of the tree clearly shows that the *DRA* gene and the *DQA/DYA* genes form two separate major branches, and the MR1 clone is on the *DRA* branch, rather than on the *DQA /DYA* branch. These results suggest that *DRA* genes of bovine and other mammalian species might have arisen independently after the divergence of these species from a common ancestor. The phylogenetic tree also indicates that the sequences of various mammalian *DRA* genes form three main branches

Table 1. Similarities in terms of predicted amino acid sequences# among MR1 cDNA and genes for class II α-chains from several species*, expressed as percentages

	SP	α1	α2	CP/TM/CY	Total
BoLA-DRA (W3)	-	100	100	100	100
OLA-DRA	95.8	95.2	94.7	98.0	95.7
HLA-DRA	66.7	83.3	84.0	80.4	81.4
SLA-DRA	58.3	85.7	84.0	80.4	79.8
ELA-DRA	-	85.4	87.2	80.4	85.0
H2-IEa	37.5	75.0	75.5	60.8	68.8
RT1.Da	37.5	77.4	76.6	58.8	69.6
BoLA-DQA (W1)	-	43.9	62.8	54.9	54.2
BoLA-DYA (R2)	-	43.9	54.3	44.0	47.8

#The regions indicated in the Table are defined in the legend to Figure 1.

*Amino acid sequences deduced from genes for class II α-chains were taken from the references listed in the legend to Figure 2.

(Fig. 3). One branch contains MR1, W3 and the ovine *DRA* gene, suggesting that the *BoLA-DRA* is closest to the ovine gene. Indeed, ovine *DRA* shares the greatest overall similarity in terms of predicted amino acid sequence to MR1 (95.7%), as shown in Table 1. Human, swine and equine *DRA* genes can all be included on one branch and they exhibit approximately 80% identity in terms of being encoded to MR1. In addition, rat and mouse *DRA* genes are located on the other branch, and these are the least similar to MR1 (69.6% and 68.8%, respectively). Thus, the *BoLA-DRA* gene is more different from the *DRA* genes of rat and mouse than it is from the *DRA* genes of the human, pig and horse. These results suggest that bovine *DRA* and ovine *DRA* branched off from the equine/human/swine lineage after the bifurcation of the mouse/rat and equine/human/swine/ovine /cattle lineages.

MR1 is the first cDNA clone of a bovine *DRA* gene. The absence of variations in our clone at residues of which amino acids are highly conserved among products of mammalian class II *A* genes supports the notion that this gene may be expressed and functional. Further study is

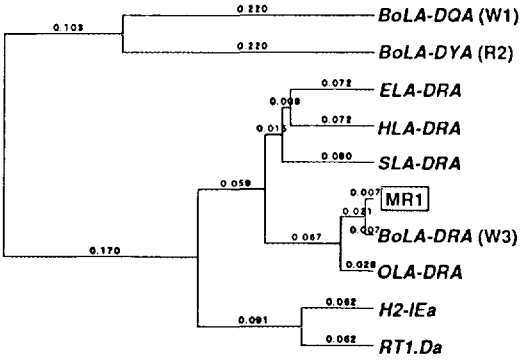

Figure 3. Phylogenetic tree showing evolutional relationships among MHC class II α genes, constructed from predicted amino acid sequences of the α1, α2, CP, TM, and CY regions. The sequences used are those shown in Figure 2. Numbers above the horizontal lines indicate the evolutional distance between one protein and another.

required to characterize the expression and function of this clone. Recently, as a first step towards an understanding of these events, we isolated and characterized full-length cDNAs for bovine DRB. Further, the *DRB* clone and the *DRA* cDNA clone described here have been used to transfect fibroblasts and attempts are being made to determine whether the BoLA-DRAB dimer is expressed on the surface of these cells. Thus, the availability of this cDNA clone will facilitate investigations of issues such as genomic organization, regulation of gene expression, and structure/function relationships.

ACKNOWLEDGMENTS: We thank Dr. H. Amanuma for encouragement, Dr. K. Yokoyama for critical reading of the manuscript; Drs. H. Inoko and A. Ando for the gift of pDAa1-4 cDNA, Drs. T. Sudo, Y. Ogawa, and T. Itoh for preparation of cDNA library, and Dr. H. Koyama for the gift of BLSC-KU1 cells. This study was supported by Special Coordination Funds from the Science and Technology Agency of the Japanese Government.

REFERENCES

1. Klein, J. (1986) Natural History of the Major Histocompatibility Complex, John Wiley & Sons, New York.
2. Usinger, W. R., Curie-Cohen, M., and Stone, W. H. (1977) *Science* 196,1017-1018.
3. Anderson, L., Bohme, J., Rask, L., and Peterson, P. A. (1986) *Anim. Genet.* 17, 95-112.
4. Anderson, L., Bohme, J., Peterson, P. A., and Rask, L. (1986) *Anim. Genet.* 17, 295-304.
5. Groenen, M. A. M., van der Poel, J. J., Dijkhof, R. J. M., and Giphart, M. J. (1990) *Immunogenetics* 31, 37-44.
6. Anderson, L., and Rask, L. (1988) *Immunogenetics* 27, 110-120.
7. Muggli-Cockett, N. E., and Stone, R. T. (1988) *Anim. Genet.* 19: 213-225.
8. Muggli-Cockett, N. E., and Stone, R. T. (1989) *Anim. Genet.* 20, 361-369.
9. Stone, R. T., and Muggli-Cockett, N. E. (1990) *Anim. Genet.* 21, 353-360.
10. van der Poel, J. J., Groenen, M. A. M., Dijkhof, R. J. M., Ruyter, D., and Giphart, M. J. (1990) *Immunogenetics* 31, 29-36.
11. Xu, A., McKenna, K., and Lewin, H. A. (1993) *Immunogenetics* 37, 231-234.
12. Onuma, M., Koyama, H., Aida, Y., Okada, K., Ogawa, Y., Kirisawa, R., and Kawakami, Y. (1986) *Leukemia Res.* 10, 689-695.
13. Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
14. Gubler, U., and Hoffman, B. J. (1983) *Gene* 25, 263-269.
15. Seed, B. (1987) *Nature* 329, 840-842.
16. Seed, B., and Aruffo, A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 3365-3369.
17. Ando, A., Inoko, H., Kimura, M., Ogata, S., and Tsuji, K. (1986) *Human Immunol.* 17, 355-367.
18. Sanger, F., Nicklen, S., and Coulson A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
19. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444-2448.
20. Kimura, M. (1980) *J. Mol. Evol.* 16, 111-120.
21. Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.* 4, 406-425.
22. Fabb, S. A., Maddox, J. F., Gogolin-Ewens, K. J., Baker, L., Wu, M-J., and Brandon, M. R. (1993) *Anim. Genet.* 24, 249-255.
23. Korman, A. J., Auffray, C., Schamboeck, A., and Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6013-6017.
24. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A., and Strominger, J. L. (1979) *Nature* 282, 266-270.
25. Koppelman, B., and Cresswell, P. (1990) *J. Immunol.* 145: 2730-2736.

26. Hirsch, F. Germana, S., Gustafsson, K., Pratt, K., Sachs, D. H., and Leguern, C. (1992) *J. Immunol.* 149, 841-846.
27. Holowachuk, E. W., Greer, M-K., and Martin, D. R. (1987) *Nucleic Acids Res.* 15,10551-10567.
28. Hyldig-Nielsen, J. J., Schenning, L., Hammerling, U., Widmark, E., Heldin, E., Lind, P., Servenius, B., Lund, T., Flavell, R., Lee, J. S., Trowsdale, J., Schreier, P. H. Zablitzky, F., Larhammar, D. Peterson, P. A., and Rask, L. (1986) *Nucleic Acids Res.* 11, 5055-5071.
29. Albright, D., Bailey, E., and Woodward, J. G. (1991) *Immunogenetics* 34, 136-138.