

## Nucleotide sequence of a dog class I cDNA clone

**Ulla M. Sarmiento\* and Rainer Storb** 

Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98104, USA

Received March 13, 1990

The major histocompatibility complex (MHC) of the dog, termed DLA, can be divided into three serologically defined loci, designated DLA-A, DLA-B, and DLA-C (Bull et al. 1987), and a fourth region (termed DLA-D) defined by mixed leucocyte culture reactivity (Deeg et al. 1986). The DLA-A antigens are characterized as class I molecules by their association with  $\beta_2$ -microglobulin (Krumbacher et al. 1986) and by the correlation of serologic DLA-A specificities with protein polymorphisms (van der Feltz and Ploegh 1984; Neefjes et al. 1986). The DLA-B locus codes for class II antigens which are constitutively expressed on almost all circulating lymphocytes in the dog (Doxiadis et al. 1989). The DLA-C antigens are thought to be weakly expressed class I antigens (van der Feltz and Ploegh 1984). Recently, allelic polymorphism within the class I and class II regions of the DLA complex has been studied by restriction fragment length polymorphism (RFLP) analysis (Sarmiento and Storb 1988a, b, 1989). Based on the RFLP studies, the DLA class I region has been presumed to have at least eight class I genes, including the canine homologues of HLA-A, -B and HLA-E genes (Sarmiento and Storb 1989). Moreover, the DNA polymorphism detected by the human HLA-B7 probe corresponded to the serologically defined DLA-A allelic series (Sarmiento and Storb 1989). In this report, we describe the nucleotide sequence of a cDNA clone which encodes a class I gene in the dog.

To isolate DLA class I cDNA clones, a lambda ZAP phage cDNA library (Stratagene, La Jolla, California) constructed from the spleen of a dog (serotyped as *DLA-A9*, *B6*, *C11*) was screened by plaque hybridization

(Benton and Davis 1977) using a <sup>32</sup>P-labeled human HLA-B7 cDNA clone (Sood et al. 1981) as a probe. The hybridization was performed in a buffer containing 50% (v/v) formamide, 0.1% Denhardt's solution, 5× salinesodium phosphate-EDTA (SSPE; 0.9 M NaCl, 0.05 M NaH<sub>2</sub>PO<sub>4</sub>, 5 mM ethylenediaminetetraacetate), 5% dextran sulfate, 1% sodium dodecyl sulfate (SDS), and 200 µg/ml salmon sperm DNA overnight at 42 °C. The filters were washed twice at room temperature in  $2 \times SSPE$ , once in  $2 \times SSPE$ , 0.5% SDS, and once in  $0.5 \times SSPE$ . Twenty clones were isolated and plaquepurified after screening over 600 000 plaques. The cDNA inserts were contained within a phagemid, pBluescript SK(<sup>-</sup>), which was excised from the lambda ZAP phage in vivo by helper virus (Short et al. 1988). Single- and double-stranded DNA preparations were made for sequencing. The sequencing primers included T3 and T7 promoters flanking the pBluescript polylinker and three synthetic oligonucleotides termed A1 (spanning codons 1-9 in the sense strand, 5'-CGAATTCCACTCCCTG-AGGTATTTCTAC-3'), A2 (spanning codons 279-284 in the complementary strand, 5'-GACAATGGTGGACA-GAGG-3'), and A3 (spanning codons 201-209 in the complementary strand, 5'-AGAATTCCAGCGCCCAG-CACCTCAGGGT-3'). Four additional synthetic oligonucleotides (kindly provided by Y. Choo) based on the human HLA-B27 gene sequence were used for sequencing and were termed SP83-88 (spanning codons 83-88 in the sense strand), SP171-176 (spanning codons 171-176 in the sense strand), SP246-251 (spanning codons 246-251 in the sense strand), and SPC221-226 (spanning codons 221-226 in the complementary strand). The complete sequences of six full-length clones were determined on both strands using the dideoxy chain termination procedure (Sanger et al. 1977) and Sequenase (US Biochemicals, Cleveland, Ohio).

The nucleotide and predicted amino acid sequence of the longest [1511 base pairs (bp)] DLA class I cDNA clone (pBT-I16, available to interested investigators) is illus-

<sup>\*</sup> Present address: University of Pennsylvania, School of Veterinary Medicine, Department of Pathobiology, 3800 Spruce Street, Philadelphia, PA 19104-6008, USA.

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession number M32283.

Address correspondence and offprint requests to: U.M. Sarmiento.

Met Glu Val Val Met Pro Arg Ala Leu Leu Val Leu Leu Ser Ala Ala Leu Ala Leu Thr - 5 ATG GAG GTG GTG ATG CCG CGA GCC CTC CTC GTG CTG CTG TCG GCG GCC CTG GCC CTG ACC 60 Pro Thr Arg Ala|Gly Ser His Ser Leu Arg Tyr Phe Tyr Thr Ser Val Ser Arg Pro Gly 16 CCG ACC CGG GCG|GGC TCC CAC TCC CTG AGG TAT TTC TAC ACC TCC GTG TCC CGG CCC GGC 120 Ala Gly Asp Pro Arg Phe Ile Ala Val Gly Tyr Val Asp Asp Thr Gln Phe Val Arg Phe 36 GCG GGG GAC CCC CGC TTC ATC GCC GTC GGC TAC GTG GAC GAC ACG CAG TTC GTG CGG TTC 180 Asp Ser Asp Ala Ala Thr Gly Arg Met Glu Pro Arg Ala Pro Trp Val Glu Gln Glu Gly 56 GAC AGC GAC GCG GCC ACT GGG AGG ATG GAG CCG CGG GCG CCG TGG GTG GAG CAG GAG GGG 240 Pro Glu Tyr Trp Asp Arg Gln Thr Arg Thr Ile Lys Glu Thr Ala Arg Thr Phe Arg Val 7б CCG GAG TAT TGG GAC CGG CAG ACG CGG ACC ATC AAG GAG ACC GCA CGG ACT TTC CGA GTG 300 Asp Leu Asp Thr Leu Arg Gly Tyr Tyr Asn Gln Ser Glu Ala Gly Ser His Thr Arg Gln 96 GAC CTG GAC ACC CTG CGC GGC TAC TAC AAC CAG AGC GAG GCC GGG TCT CAC ACC CGC CAG 360 Thr Met Tyr Gly Cys Asp Leu Gly Pro Asp Gly Arg Leu Leu Arg Gly Tyr Ser Gln Asp 116 ACC ATG TAC GGC TGT GAC CTG GGG CCC GAC GGG CGC CTT CTC CGC GGG TAC AGT CAG GAC 420 Ala Tyr Asp Gly Ala Asp Tyr Ile Ala Leu Asn Glu Asp Leu Arg Ser Trp Thr Ala Ala 136 GCC TAC GAC GGC GCC GAT TAC ATC GCC CTG AAC GAG GAC CTG CGC TCC TGG ACC GCG GCG 480 Asp Thr Ala Ala Gln Ile Thr Gln Arg Lys Trp Glu Ala Ala Gly Val Ala Glu Leu Gln 156 GAC ACG GCG GCG CAG ATC ACC CAG CGC AAG TGG GAA GCG GCA GGT GTA GCA GAG CTA CAA 540 Trp Arg Asn Tyr Leu Glu Thr Thr Cys Val Glu Trp Leu Arg Arg Tyr Leu Glu Met Gly 176 TGG AGG AAC TAC CTG GAG ACG ACG TGC GTG GAG TGG CTG CGG AGG TAC CTG GAG ATG GGG 600 Lys Glu Thr Leu Leu Arg Ala Asp Pro Pro Ser Thr Arg Val Thr His His Pro Val Ser 196 AAG GAG ACG CTG CTG CGC GCA GAC CCC CCA AGC ACA CGT GTG ACC CAC CAC CCC GTC TCT 660 Asp His Glu Val Thr Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Glu Ile Thr Leu 216 GAC CAT GAG GTC ACC CTG AGG TGC TGG GCG CTG GGC TTC TAC CCT GCG GAG ATC ACC CTG 720 Thr Trp Gln Arg Asp Gly Glu Asp Gln Thr Gln Asp Thr Glu Val Val Asp Thr Arg Pro 236 ACC TGG CAG CGG GAT GGG GAG GAC CAG ACC CAG GAC ACA GAG GTT GTG GAC ACA AGG CCT 780 Ala Gly Asp Gly Thr Phe Gln Lys Trp Ala Ala Val Val Val Pro Ser Gly Gln Glu Gln 256 GCA GGA GAT GGG ACC TTC CAG AAG TGG GCG GCC GTG GTG GTG CCT TCT GGA CAG GAG CAG 840 Arg Tyr Thr Cys His Val Gln His Glu Gly Leu Pro Glu Pro Ile Thr Arg Arg Trp|Glu 276 AGA TAC ACG TGC CAC GTC CAG CAT GAG GGG CTG CCG GAG CCT ATC ACG CGG AGA TGG GAG 900 Pro Ser Pro Leu Ser Thr Ile Val Ile Val Ser Ile Ala Ala Leu Val Leu Val Val 296 CCT TCC CCT CTG TCC ACC ATT GTC ATC GTC AGC ATT GCT GCT CTG GTT CTC CTC GTG GTC 960 Ala Gly Val Ile Gly Ala Val Ile Trp Arg Lys Gln Arg Ser!Gly Gly Lys Gly Pro Gly 316 GCT GGG GTG ATT GGA GCT GTG ATC TGG AGG AAG CAG CGC TCA GGA GGA AAA GGA CCA GGC 1020 Tyr Ser His Ala Ala Arg Asp Asp Ser Ala Gln Gly Ser Asp Val Ser Leu Thr Ala Pro 336 TAC TCT CAT GCT GCA|CGT GAT GAC AGT GCC CAG GGC TCT GAT GTG TCT CTG ACA GCT CCT 1080 Arg Val \*\*\* 338 AGA|GTG TGA GACCAGCTGCCTGTGGGACTGACGGATGCAAGATGTGTTCACATCTCACGTGATGACATCAACAACC 1156  ${\tt CTGGCTTGTCTGCAAACAGTGTCAGGATGTGCCTGTGTCCCTAGGAGCATAATGTGAGGAGGTGGGGAGATTGGCCC}$ 1235 ACCCTGCCCACCATGACCTGTCCCTAATCTGATGTGCGCTCTCCTCTGATGTGCTTTCCTGTCCAGGAGAGGCAGGG 1314 CTGGACCATCTCCATCCCTGTCTTTGTTTCATGTTGAGTACTAATCTCTTACTATCCGATTGAAAATAAGAATCCAGAT 1393 ATGAGTTTGTGTTTCCTGAGTCTTGGGATGTGGGGGCTGATGAGGT<u>AATAAA</u>AGGAGATTTGTGAAGTTGAGAGAGCA<u>AA</u> 1472 1511

Fig. 1. Nucleotide sequence and predicted amino acid sequence of a DLA class I cDNA clone. The exon borders are indicated by vertical lines. The two putative polyadenylation signals in the 3' untranslated region are underlined. The numbers indicate the amino acid (top line) and nucleotide (bottom line) sequence positions. The asterisks represent the stop codon (TGA).

trated in Figure 1. The longest open-reading frame contains 362 codons flanked by 425 bp of the 3' untranslated region. The first of the two putative polyadenylation signals (5'-AATAAA-3') is located 349 nucleotides downstream from the termination codon (TGA) and the second polyadenylation signal is found 22 bp upstream from the poly(A) tail (Fig. 1). The signal peptide, which contains two in-frame methionine (ATG) codons at positions -24 and -20, is composed of 24 N-terminal residues. The remaining 338 amino acids represent the extracellular portion of the alpha chain ( $\alpha$ 1 domain, 90 residues;  $\alpha$ 2 domain, 93 residues;  $\alpha$ 3 domain, 92 residues), the transmembrane region (35 residues), and the cytoplasmic tail (28 residues). The dog sequence is characterized by an insertion of 3 bp in exon 3 (at the position designated Leu-155) and by a deletion of four codons from exon 5. Four cysteines involved in disulfide bonds are present in the extracellular portion at positions 101, 165, 204, and 260. The conserved site for N-linked glycosylation, Asn-Gln-Ser, is found at positions 86-88 and the serine phosphorylation site (SD/EXSL) is present between residues 329 and 333 of the cytoplasmic domains of the dog clone. The *DLA* class I sequence also contains the Arg-Phe-Asp-Ser sequence at positions 35–38 which is found in other class I antigen  $\alpha$ 1 domains and in MHC class II antigen  $\beta$ 1 domains (Young et al. 1987).

Figure 2 summarizes the comparison of the predicted amino acid sequences of class I genes from the dog, cat, ox, and human. In the coding region, the DLA class I clone appears to be most similar to the cat class I gene with 84.5% overall nucleotide and 77.5% amino acid sequence similarity with the pFLA24 sequence (Yuhki et al. 1989). The dog class I sequence also has 84.4% and 82.9% nucleotide similarity with the bovine BL3-7 clone (Ennis et al. 1988) and the human HLA-A11 allele (Cowan et al. 1987), respectively (Fig. 2). Pairwise comparison of the individual domains reveals that the  $\alpha 1$  and  $\alpha 3$  domains of the dog class I clone have 81% and 88% amino acid similarity with those of the human HLA-A11 allele, respectively (Fig. 2). The DLA  $\alpha$ 2 domain has 84.9% amino acid similarity with the BoLA BL3-7 clone, while the transmembrane and cytoplasmic domains of the DLA clone are most closely related to those of the cat class I clones with 60% and 75% protein sequence similarity,

αl	10	20	30	40	50	60	70	80	90	
DLA-A	GSHSLRYPYT	SVSRPGAGDP	RFIAVGYVDD	TQFVRFDSDA	ATGRMEPRAP	WVEQEGPEYW	DRQTRTIKET	ARTFRVDLDT	LRGYYNQSEA	
FLA	F	AL-E-	s		PNP-E	- M	NIYLD-	-QISN-N-	F-RS	
BoLA	F	GL-E-			- PNPR · V -	- M	NIY-D-	-QIN-	T	
HLA-A11	M F	R-E-			- SQ	-I	- QE NV - AQ	SQ-DG-		
α2	100	110	120	130	140	150	160	170	180	183
DLA-A	GSHTRQTMYG	CDLGPDGRLL	RGYSQDAYDG	ADYIALNEDL	RSWTAADTAA	QITQRKWEAA	GVAELQWRNY	LETTCVEWLR	RYLEMGKETL	LRA
FLA	NI-R	VD R - F -	s	К		R	/-E	GA	К·-D	۷
Bola	NI-A	V	FW-FG	R E -		· · · K · · · · · ·	-A/T	GE	N D	
HLA-A11	I-I	V-SF-	R	К	M	••••K••••••	HA/-Q-A-	GR	N	Q-T
α3	193	203	213	223	233	243	253	363	273	275
DLA-A	DPPSTRVTHH	PVSDHEVTLR	CWALGFYPAE	ITLTWQRDGE	DQTQDTEVVD	TRPAGDGTFQ	KWAAVVVPSG	QEQRYTCHVQ	HEGLPEPITR	RW
FLA	ES-NR-	-1R		Q	- H - · · L - E			E	- K NT	• -
BoLA	KAH	SIR	E-	·S····E··	M-L-E	s	L	E R	QL-L	•••
HLA-A11	K-HM	-IA			£-£			E	K-L-L	
тм	285	294	304	310	<b>CYT</b> 320	330	338	<u>% IDENT</u>	TY WITH I	DLA-A
DLA-A	EPSPLSTIVI	VSIAA/LVLL	VVAGVIGAVI	////WRKQRS	GGKGPGYSHA	ARDDSAQGSD	VSLTAPRV	NUCLEOTI	DE AMINO	ACID
FLA	···S-PF-T-	LG-I-GVAV-	TV-V	////KC-	IQP	T	S••M••K-	84.5	78	.1
BoLA	P/QTSFL-	MG-IVG	L-A	////K	-ERI-TQ-	- SS	V - K -	84.4	77	.8
			(C-VTTV	ANW DVS-	DRGS-TO-	-65	CK-	82.9	76	.9

Fig. 2. Alignment of predicted amino acid sequences and domain structures of class I molecules from dog, cat, ox, and human. The standard one-letter amino acid code is used. The *numbers* indicate the DLA class I cDNA amino acid sequence which is given in full (*top line*). Dashes indicate sequence identity with the top line. Slashes indicate gaps introduced to maximize alignment. The sequences have been published elsewhere as follows: *pFLA24* (Yuhki et al. 1989), *BoLA-BL3-7* (Ennis et al. 1988), and *HLA-A11* (Cowan et al. 1987).

respectively. Thus, the dog class I clone appears to be closely related to particular class I alleles of other species, supporting the hypothesis of trans-species evolution of MHC genes (Klein 1987). Most of the amino acid differences are found at positions reported to have high variability in other species (Fig. 2; Parham et al. 1988). The *DLA* class I sequence contains 13 dog-specific residues at Asp-19, Thr-66, Ile-67, Glu-69, Arg-72, Leu-155 (3 bp insert), Thr-163, Asp-233, Val-284, Ala-289, Gly-298, Gln-308, and Arg-337. The dog sequence shares eight residues with one or both of the *FLA* class I alleles (Yuhki et al. 1989) at Ser-114, Met-175, Arg-189, Ile-271, Leu-280, Pro-315, His-319, and Pro-336. Some of the latter substitutions may be unique to the carnivore lineage.

Further comparison of the dog and human class I sequences (Parham et al. 1989) demonstrates that the dog class I cDNA clone shares a total of 31, 25, and 22 locusspecific nucleotides with the human HLA-A, -B, and -Cloci, respectively (data not shown). The latter finding, the percent nucleotide and amino acid similarity, and the length of the peptide encoded by the dog cDNA suggest that the DLA class I gene reported in this paper represents a canine homologue of the HLA-A locus. Based on linked polymorphisms at eight positions in exon 4, the HLA-A sequences have been divided into two groups: HLA-A2/A28 family and HLA-A1/A3/A11 family (Parham et al. 1989). The dog sequence shares six out of eight characteristic nucleotide substitutions in exon 4 with members of the HLA-A1/A3/A11 family, in contrast to sharing only two out of eight positions with the HLA-A2/A28 family (data not shown). This is in agreement with our previous finding of cross-reactivity between dog lymphocytes and human alloantisera to the HLA-A1/A3/A11 family of molecules (unpublished observations). Moreover, all chimpanzee ChLA-A alleles described thus far have been related to alleles of the HLA-A1/A3/A11 family (Parham et al. 1989).

In conclusion, it is worth noting that the sequence data presented here are representative of all the class I clones isolated from the dog spleen cDNA library. This finding is consistent with the expression of a single class I gene in the dog. It also suggests that the DLA class I clone reported here is functional despite the insertion of a codon in exon 3. The other class I genes detected by RFLP studies of dog genomic DNA may represent pseudogenes or genes that are transcribed at very low levels in the adult spleen. We propose that the dog class I gene reported in this paper be termed the *DLA-A* gene and that the sequence be designated the *DLA-A* gallele. Further sequence analysis of other *DLA-A* types will elucidate the degree of allelic variation in the class I region of the canine MHC.

Acknowledgments. We thank Dr. Yoon Choo for providing the sequencing primers and many helpful discussions, and Drs. John Hansen (for providing the *HLA-B7* cDNA probe), Pei Ji, and Dan Geraghty for technical advice. We also thank Susan DeRose and Rosemary Hering for their expert technical help. This work was supported by the Joseph Steiner Award and NIH grants CA31787 and CA15704 (R.F.S.) and by a fellowship from the Medical Research Council of Canada (U. M. S.).

## References

- Benton, W. D. and Davis, R. W.: Screening lambda gt recombinant clones by hybridization to single plaques in situ. Science 196: 180-182, 1977
- Bull, R. W., Vriesendorp, H. M., Cech, R., Grosse-Wilde, H., Bijma, A. M., Ladiges, W. L., Krumbacher, K., Doxiadis, I., Ejima, H., Templeton, J., Albert, E. D., Storb, R., and Deeg, H. J.: Joint report of the third international workshop on canine immunogenetics. II. Analysis of the serological typing of cells. *Transplantation* 43: 154–161, 1987
- Cowan, E. P., Jelachich, M. L., Biddison, W. E., and Coligan, J. E.: DNA sequence of HLA-A11: remarkable homology with HLA-A3 allows identification of residues involved in epitopes recognized by antibodies and T cells. *Immunogenetics* 25: 241–250, 1987
- Deeg, H. J., Raff, R. F., Grosse-Wilde, H., Bijma, A. M., Buurman, W. A., Doxiadis, I., Kolb, H. J., Krumbacher, K., Ladiges, W., Losslein, K. L., Schoch, G., Westbroek, D. L., Bull, R. W., and Storb, R.: Joint report of the third international workshop on canine immunogenetics. I. Analysis of homozygous typing cells. *Transplantation 41*: 111–117, 1986
- Doxiadis, I., Krumbacher, K., Neefjes, J. J., Ploegh, H. L., and Grosse-Wilde, H.: Biochemical evidence that the DLA-B locus codes for a class II determinant expressed on all canine peripheral blood lymphocytes. *Exp Clin Immunogenet* 6: 219–224, 1989
- Ennis, P. D., Jackson, A. P., and Parham, P.: Molecular cloning of bovine class I MHC cDNA. J Immunol 141: 642-651, 1988
- Klein, J.: The origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol 19*: 155–162, 1987
- Krumbacher, K., van der Feltz, M. J. M., Happel, M., Gerlach, C., Losslein, L. K., and Grosse-Wilde, H.: Revised classification of the DLA loci by serological studies. *Tissue Antigens* 27: 262–268, 1986
- Neefjes, J. J., Doxiadis, I., Stam, N. J., Beckers, C. J., and Ploegh, H. L.: An analysis of class I antigens of man and other species by one-dimensional IEF and immunoblotting. *Immunogenetics* 23: 164–171, 1986
- Parham, P., Lomen, C. E., Lawlor, D. A., Ways, J. P., Holmes, N., Coppin, H. L., Salter, R. D., Wan, A. M., and Ennis, P. D.: Nature of polymorphism in HLA-A, -B, and -C molecules. *Proc Natl Acad Sci USA* 85: 4005–4009, 1988
- Parham, P., Lawlor, D. A., Lomen, C. E., and Ennis, P. D.: Diversity and diversification of HLA-A,B,C alleles. J Immunol 142: 3937–3950, 1989
- Sanger, F., Nicklen, S., and Coulson, A. R.: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467, 1977
- Sarmiento, U. M. and Storb, R.: Restriction fragment length polymorphism of the major histocompatibility complex of the dog. *Immunogenetics* 28: 117-124, 1988a
- Sarmiento, U. M. and Storb, R.: Characterization of class II alpha genes and DLA-D region allelic associations in the dog. *Tissue Antigens* 32: 224–234, 1988b
- Sarmiento, U. M. and Storb, R.: RFLP analysis of DLA class I genes in the dog. *Tissue Antigens* 34: 158-163, 1989
- Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D.: Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res 16*: 7583-7600, 1988

- Sood, A. K., Pereira, D., and Weissman, S. M.: Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigen HLA-B by use of an oligonucleotide primer. *Proc Natl Acad Sci USA* 78: 616–620, 1981
- Young, J. A. T., Wilkinson, D., Bodmer, W. F., and Trowsdale, J.: Sequence and evolution of HLA-DR7- and -DRw53-associated beta-chain genes. Proc Natl Acad Sci USA 84: 4929-4933, 1987
- Yuhki, N., Heidecker, G. F., and O'Brien, S. J.: Characterization of MHC cDNA clones in the domestic cat. J Immunol 142: 3676-3682, 1989
- van der Feltz, M. J. M. and Ploegh, H. L.: Immunochemical analysis of glycosylated and non-glycosylated DLA class I antigens. *Immunogenetics* 19: 95-107, 1984