ORIGINAL PAPER

Shirley A. Ellis · W. Ivan Morrison · Niall D. MacHugh · James Birch · Alison Burrells · Michael J. Stear

Serological and molecular diversity in the cattle MHC class I region

Received: 21 April 2005 / Accepted: 13 July 2005 / Published online: 6 August 2005 © Springer-Verlag 2005

Abstract Information on major histocompatibility complex (MHC) diversity in cattle is important to aid our understanding of immune responses and may contribute to maintenance of healthy cattle populations. Equally, understanding the mechanisms involved in generating this diversity may shed light on the complex nature of mammalian MHC evolution. The aim of this study was to assess molecular and serological variation within cattle MHC class I molecules and to study the mechanisms generating diversity. To address this aim, sequence variation was examined in 12 serologically assigned alleles from three putative loci and correlated with monoclonal antibody (mAb) binding data. The results demonstrate that both alloantisera and mAbs often fail to distinguish gene products that differ by a significant number of amino acids. Conversely, some mAbs could distinguish alleles differing by only one or two amino acids. Examination of the sequences demonstrates sharing of motifs between alleles, some encoded at distinct loci, supporting the occurrence of interlocus recombination within the cattle MHC class I region. The implications of this for MHC sequence diversity, and functional capability, are discussed.

S. A. Ellis (⊠) · J. Birch Institute for Animal Health, Compton, Newbury, RG20 7NN, UK e-mail: Shirley.ellis@bbsrc.ac.uk Tel.: +44-1635-577265 Fax: +44-1635-577263

W. I. Morrison · N. D. MacHugh · A. Burrells The Centre for Tropical Veterinary Medicine, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Roslin, Midlothian, EH25 9RG, UK

M. J. Stear

Division of Animal Production and Comparative Medicine, Institute of Comparative Medicine, Faculty of Veterinary Medicine, Glasgow University, Bearsden Road, Glasgow, G61 1QH, UK Keywords MHC class I · Cattle · Polymorphism

Introduction

Major histocompatibility complex (MHC) class I molecules are cell-surface glycoproteins whose primary function is to elicit an immune response by presentation of antigenic pathogen-derived peptides to cytotoxic T lymphocytes. In humans, there are currently more than 1,100 validated class I alleles encoded at three loci, HLA-A, -B, -C (http://www.ebi.ac.uk/imgt/hla). This extremely high level of polymorphism is probably maintained by natural selection, and most authorities would agree that low levels of diversity could be disadvantageous (Parham and Ohta 1996; Hughes and Yeager 1998).

Despite the obvious public health and commercial importance of healthy cattle, the extent of genetic variation in the cattle MHC class I region has still to be quantified. Cattle are found worldwide and occur in two species (*Bos taurus* and *Bos indicus*), each of which has a variety of morphologically distinguishable breeds. The origins of modern cattle, however, remain controversial but have clearly resulted from multiple domestication events using genetically diverse populations. The clearest distinction, between Indian cattle (*B. indicus*) and all others (*B. indicus* and *B. taurus*), has been demonstrated by mitochondrial DNA analysis (Loftus et al. 1994). In addition, many existing *B. indicus* breeds have been shown to carry taurine genes, consequently for MHC studies all cattle breeds are best considered as belonging to a single species.

International workshops were run between 1979 and 1996 to determine cattle MHC (BoLA) diversity, primarily using serological reagents to analyse class I. These studies identified over 50 distinct class I serological specificities, classified as two series, with most specificities behaving as members of one series (Davies et al. 1994). No previous attempt has been made to assess allelic diversity using molecular methods, and at present there are only 35 fulllength validated class I sequences (http://www.ebi.ac.uk/ ipd/mhc/bola/). Cattle appear to have multiple classical class I loci, as defined by phylogenetic and haplotype analysis, but all known haplotypes express only three or fewer of these loci (Ellis et al. 1999). Some of the loci are located very close to one another (Di Palma et al. 2002), which may mean that recombination is less frequent than is seen in humans. The serological specificities essentially function as haplo-type markers, with many alleles apparently serologically 'blank'. Thus, a haplotype expressing two or three class I genes may have only one serological specificity associated with it, for example, A17 (Table 1, Ellis et al. 1998), whereas in rare cases a haplotype expressing two genes may be associated with two serological specificities as in A10 and KN104 (Table 1, Bensaid et al. 1991).

The purpose of this study is to assess molecular and serological variation in alleles encoded at three putative loci, to shed light on the mechanisms generating diversity. The serological specificities are A10, and the two super-typic specificities are A6 (encompassing A17, A18 and A19) and A30 (encompassing A12 and A31).

Materials and Methods

Nomenclature

Twelve full-length cattle class I cDNA sequences were analysed in this study. New internationally agreed cattle MHC class I nomenclature is used, with former, local

 Table 1 Serological specificities associated with each allele

Broad specificity	Allele	Fine specificity	Breed	Other alleles expressed
A10	N*00101	A10 (KN104)	Boran	N*00301 ^a
A10	N*00102	A10	Boran	a
A10	N*00103	A10	N'Dama	N*00301 ^a
		(KN104)		
A10	N*00201	A10	Holstein	N*01201 ^a
A6	N*01301	A18	Holstein	None
A6	N*01302	A18	Holstein	None
A6	N*01401	A19	Holstein, Angus	N*01601
A6	N*01501	A17	Holstein	N*00801, N*00601 ^a
A30	N*02001	A12	Angus	N*00801 ^a
A30	N*01901	A12	Holstein	N*00801 ^a
A30	N*02101	A31	Holstein	N*02201
A30	N*00701	A30	Angus	various ^b

The serological specificity associated with each allele, the breed(s) in which it has been identified and any other alleles shown to be expressed on the same haplotype

^aSignifies a haplotype that has not been completely characterised at molecular level or a haplotype that varies significantly between individuals that are serologically indistinguishable

^bAllele N*00701 was identified in an animal typed serologically as A30/A2; a large number of additional alleles were expressed, but not definitively assigned to either haplotype (Garber et al. 1994)

names and accession numbers in parentheses. Details of the new nomenclature can be found in the cattle section of the Immuno Polymorphism Database (IPD; http://www.ebi.ac. uk/ipd/mhc/bola/; Robinson et al. 2005). In brief, class I nomenclature is based on the HLA nomenclature system. Allele names are based on amino acid sequence and consist of up to nine digits. The first three digits indicate the allele 'group', the second two indicate coding change, the next two indicate non-coding change and the last two indicate promoter/intron change-the last four digits are therefore rarely used, as non-coding changes in exons are rare, and there is little data available describing non-coding regions. Assignment of most alleles to loci is still problematic (Holmes et al. 2003), consequently, all alleles are currently prefixed 'N' to indicate 'not assigned' and numbered in a single series.

Origin of sequences

Table 1 provides full details on the origin of alleles, their serological specificities and lists other alleles found on the same haplotypes. Four serologically defined A10 alleles were studied: N*00101 (5.1, M69206, Bensaid et al. 1991), N*00201 (JSP.1, X92870, Pichowski et al. 1996), N*00102 (JSP.2, DQ001408) and N*00103 (JSP.3, DQ001409). Where not previously available, full-length clones were obtained by PCR from cDNA and sequenced as described in Ellis et al. (1999).

Four alleles defined by the supertypic serological specificity A6 were studied: N*01301 (HD6, X80934, Ellis et al. 1996), N*01302 (HD6.1, DQ001407), N*01401 (MAN2, AJ010861, Ellis et al. 1998) and N*01501 (3349.1, AJ010862, Ellis et al. 1998). N*01301 and N*01302 belong to the serological subtype A18, N*01501 belongs to A17 and N*01401 to A19. N*01301 was initially isolated from a cDNA library as described in Ellis et al. (1996), and the other three were derived by PCR as described in Ellis et al. (1999). Identical sequences have subsequently been identified in additional, unrelated animals carrying the same serological specificities.

The remaining four alleles were defined by the supertypic specificity A30: N*01901 (4221.1, AJ010865 Ellis et al. 1998), N*02001 (MAN3, AJ010864, Ellis et al. 1998), N*00701 (BSX, U01187, Garber et al. 1994) and N*02101 (HD1, X80933, Ellis et al. 1996). Only the product of allele N*02101 expressed the subtype A31, whereas N*02001 and N*01901 were defined as A12. N*02101 was initially isolated from a cDNA library as described in Ellis et al. (1996); N*01901, N*02001 (Ellis et al. 1999) and N*00701 (Garber et al.1994) were derived by PCR (Ellis et al. 1999).

Monoclonal antibodies and alloantisera

The monoclonal antibodies (mAbs) used were ILA7, ILA13, ILA31, ILA33, ILA35, ILA38, ILA40, ILA89 and FJ101. ILA31, ILA33, ILA35, ILA38 and ILA40 were all raised against *B. indicus* cells expressing A10 and KN18

serological specificities. All react with cells expressing the A10 specificity apart from ILA33 which reacts with cells expressing the KN18 specificity. ILA7 was raised against *B. taurus* cells expressing the A10 and A7 specificities and reacts with cells expressing the A10 specificity. ILA89 was raised against *B. indicus* cells expressing the A25 specificity, ILA13 was raised against *B. indicus* cells expressing the A10 and w4 specificities, and FJ101 was raised against

B. taurus cells expressing the A6 specificity. All mAbs were raised and characterised at the International Livestock Research Institute, Nairobi (Goddeeris 1990; Taracha et al. 1995), except FJ101 which was a gift from Dr. H. Leveziel, Institute National de la Recherche Agronomique, Jouy-en-Josas, France. We have no data regarding expressed alleles associated with the specificities w4, A7 and KN18. We

	Alpha 1				
N+00101 .	1 * * * 20 * * *	30 40	* 50 +~*	+~* +**++* +**	+* +*+ * 90
N*00101 :	: GSHSMRIFSIAVSRPGLGEPRILEVGIV	DDIQFVRFDSDAPNPR	MEPRARWVEQEGPEI	WDREIQRAKGNAQFF	RVSLINNLRGIINQSEA
N*00102 :	:FIJ :FIT				
N*00201 :	 :LFF			0RK	
				~	
N*01301 :	:LHRLFIT	RD	TQP-M-K	IS-ELWY	-EA
N*01302 :	:LHRLFIT	RD	TQP-M-K	LWY	-EA
N*01401 :	:FLHRLFIT	RD	KQP-M-K	IS-ELWY	-EA
N*01501 :	:LY-GRLFIT	RD	KQP-M-K	LWY	(-EA
N#00001		D			
N*02001 :	·DKHDFII	D	KP-I-K	RIS-E-ILVI	-G
N*02101 ·	·I	WD	KTP-T-K	RIS-E-TLVY	/-G
N*00701 :	:LKHDFIT	D	KP-I-K	RIS-E-TLVY	-G
	Alpha 2				
	* * * 110 * *	120 130	* +~*	+++~ +~* +* +~	· +~ +~* 179
N*00101 :	: GSHTLQWMSGCYVGPDGRPPRGFMQFG	YDGRDYLALNEDLRSW	TAVETMAQISKRKME	AAGEAEVQRNYLEGR	CVEWLRRYLENGKDTL
N*00102 :					
N*00103 :	. T.				
N 00201 .			v		
N*01301 :	: E-YDSLRYE-Y-		AD-A	ARF1	
N*01302 :	: L-YDSLRYE-Y-		AD-A	ARFT	
N*01401 :	: Y-R-YDLLST	I	AD-ATW-	V	GA-
N*01501 :	: L-YDLLT		AD-A	DR	G
N*02001 :	: F-L-YDLLDA	SI	-VAD-ATW-	VRFV	A-
N*01901 :	: Y-D-YDLLG-YD-Y-		AD-ATW-	VRFV	GA-
N*02101 :	: F-Q-FDLLG-YR-YA	l	AD-ATW-	DRFVE	GG
N*00/01 :	: F-Q-YDLLKA	SIE	AD-AIW-	AIWE	
	Alpha 3				
	Alpha 3 190 200	210 220	230	240 250	260
N*00101 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00103 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00103 : N*00201 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00103 : N*00201 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00103 : N*00201 : N*01301 : N*01302 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSC VV	260 EEQKYTCQVQHEGLQE RR
N*00101 : N*00102 : N*00103 : N*00201 : N*01301 : N*01302 : N*01302 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG VV	260 EEQKYTCQVQHEGLQE RR
N*00101 : N*00102 : N*00103 : N*00201 : N*01301 : N*01302 : N*01401 : N*01501 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG VV	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00201 : N*01301 : N*01302 : N*01401 : N*01501 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00201 : N*01301 : N*01302 : N*01401 : N*01501 : N*02001 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG VV	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00103 : N*01301 : N*01302 : N*01401 : N*01501 : N*01501 : N*02001 : N*01901 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSC 	260 SEEQKYTCQVQHEGLQE RR
N*00101 : N*00102 : N*00103 : N*01301 : N*01301 : N*01301 : N*01401 : N*01501 : N*02001 : N*02101 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG VV	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00201 : N*01301 : N*01302 : N*01401 : N*01501 : N*01501 : N*02001 : N*01901 : N*02101 : N*00701 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE RR RR
N*00101 : N*00102 : N*0103 : N*01301 : N*01301 : N*01401 : N*01401 : N*02001 : N*02001 : N*02101 : N*02701 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSC VV	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00103 : N*01301 : N*01301 : N*01401 : N*01401 : N*01501 : N*02001 : N*02101 : N*02101 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00201 : N*01301 : N*01301 : N*01301 : N*01401 : N*01501 : N*02001 : N*02101 : N*02101 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00201 : N*01301 : N*01302 : N*01401 : N*01501 : N*01501 : N*02001 : N*02001 : N*02011 : N*00701 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*01201 : N*01301 : N*01301 : N*01301 : N*01401 : N*01501 : N*02001 : N*02001 : N*02011 : N*00701 : N*00101 : N*00102 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*010301 : N*01301 : N*01301 : N*01401 : N*01401 : N*01901 : N*02001 : N*02101 : N*00701 : N*00701 : N*00102 : N*00102 : N*00103 :	Alpha 3 190 200 : LRA DPPKAHVTRHPISGREVTLRCWAL :	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG VV	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*0103 : N*01301 : N*01301 : N*01301 : N*01401 : N*01401 : N*02001 : N*02101 : N*00701 : N*00101 : N*00101 : N*00103 : N*00201 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*00201 : N*01301 : N*01302 : N*01401 : N*01501 : N*01501 : N*02001 : N*02001 : N*00701 : N*00701 : N*00102 : N*00102 : N*00103 : N*00103 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*0103 : N*01301 : N*01301 : N*01301 : N*01401 : N*01501 : N*02001 : N*02001 : N*02011 : N*00701 : N*00101 : N*00102 : N*00103 : N*00101 : N*00101 : N*00101 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE RR RR
N*00101 : N*00102 : N*01201 : N*01302 : N*01401 : N*01401 : N*01901 : N*02001 : N*02101 : N*02101 : N*00701 : N*00102 : N*00103 : N*00103 : N*00201 : N*01301 : N*01302 : N*01302 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG V	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*01003 : N*01301 : N*01302 : N*01401 : N*01501 : N*02001 : N*02101 : N*02101 : N*00101 : N*00101 : N*00103 : N*00103 : N*00103 : N*01301 : N*01301 : N*01301 : N*01301 : N*01301 : N*01401 : N*01401 : N*01401 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE RR RR
N*00101 : N*00102 : N*0103 : N*01301 : N*01302 : N*01401 : N*01501 : N*02001 : N*02001 : N*02001 : N*02011 : N*00101 : N*00102 : N*00102 : N*00101 : N*00101 : N*00101 : N*01201 : N*01201 : N*01201 : N*01201 : N*01201 : N*01201 : N*01201 : N*01201 : N*01201 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEQKYTCQVQHEGLQE RR RR RR
N*00101 : N*00102 : N*01201 : N*01301 : N*01302 : N*01401 : N*01501 : N*02001 : N*02001 : N*02001 : N*00101 : N*00102 : N*00103 : N*00101 : N*00101 : N*00101 : N*00101 : N*01301 : N*01301 : N*01501 : N*01501 : N*02001 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG V	260 EEEQKYTCQVQHEGLQE RR RRR RR
N*00101 : N*00102 : N*01201 : N*01201 : N*01302 : N*01401 : N*01501 : N*02001 : N*02001 : N*01901 : N*00101 : N*00102 : N*00101 : N*01301 : N*0130	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE
N*00101 : N*00102 : N*01003 : N*01301 : N*01302 : N*01401 : N*01501 : N*01501 : N*02101 : N*02101 : N*00101 : N*00101 : N*00103 : N*01001 : N*01301 : N*01301 : N*01301 : N*01301 : N*01301 : N*01401 : N*01501 : N*01901 : N*02001 :	Alpha 3 190 200 LRA DPPKAHVTRHPISGREVTLRCWAL 	210 220 GFYPEEISLTWQRNGE 	230 DQTQDMELVETRPSG	240 250 DGNFQKWAALLVPSG 	260 EEEQKYTCQVQHEGLQE RR RR RR

Fig. 1 Shows an alignment of the predicted amino acid sequences derived from 12 full-length cattle class I cDNA sequences (see text for details of sequences). The A10 group contains N*00101, N*00101, N*00103 and N*00201; the A6 group contains N*01301, N*01302, N*01401 and N*01501 and the A30 group contains N*02001, N*01901, N*02101 and N*00701. *Dashes* indicate se-

quence identity, *dots* represent gaps introduced to maximise alignment, * indicates residues predicted in human to contact peptide, + indicates residues predicted in human to contact T cell receptor, ~ indicates residues potentially contacting both. Data based on HLA-A*0201 structure as summarised in Marsh et al. 2000

have previously identified two alleles associated with the A25 specificity (AY188804, AY188805).

A panel of 116 alloantisera was used which defines 32 workshop and 10 locally defined specificities (Stear et al. 1990). All cDNAs apart from N*00701 (which was not available) were transfected into mouse P815 cells to assess recognition by monoclonal antibodies using standard flow cytometry methodology, essentially as described in Ellis et al. (1999).

Results

Figure 1 shows an alignment of the full-length amino acid sequences derived from the 12 cDNAs. Within both the A30 group and the A6 group, the sequences are almost identical in alpha 1 (4 variable positions in each case), but very different in alpha 2 (17–19 variable positions). There is little variation in the 3' portion of the molecule [alpha 3, transmembrane (TM) and cytoplasmic domains] in these two sets of sequences (two to four variable positions in total).

Within the A6 group, N*1301 and N*1302 (both encoding the A18 specificity) differ by a single amino acid at position 97 caused by two nucleotide changes. Several other alleles (including N*01501 in this group) also have this substitution. The Y-G motif around position 10 in N*01501 is interesting because it is seen in several other alleles encoded at two additional loci. Three sequences in the A10 group are from African cattle breeds (N*00101, N*00102, N*00103); these seem to have been generated by a small recombination event in alpha 1. The differences between the European allele N*00201 and the other three A10 alleles are spread evenly across the coding region; these are discussed in detail in Pichowski et al. (1996). Most of the substitutions seen in N*00201 are commonly seen in other alleles apart from a unique V at position 139 in alpha 2 and a V at position 245 in alpha 3. Within the A30 group, N*02101 has three substitutions in alpha 1 where the other three sequences are identical. There are between 9 and 17 amino acid differences between the four sequences in alpha 2, and these are evenly spread across the region.

There are distinct differences between the three sets of sequences in the transmembrane and cytoplasmic domains. It is difficult to identify locus-specific motifs because there are relatively few sequences putatively assigned to each locus; however, the best candidates appear in this region. Apart from the difference in length of the TM region, which distinguishes the A10 set of alleles (and is shared only with other alleles thought to be encoded at the same locus), the main area of diversity stretches from position 304 to part way through exon 7 (position 327).

Table 2 shows reactivity of a panel of mAbs on selected P815 transfectants, as assessed by flow cytometry. Despite the fact that seven of the mAbs were originally raised against cells carrying the A10 specificity, several show cross-reactivity against apparently unrelated alleles in both the A6 and A30 groups. The overall similarity between and within the three sets is similar ($\sim 86\%$ identity between groups, ~94–99% identity within groups). Within the A6 group, ILA7, ILA31, ILA35 and ILA40 cross-react with N*01301; ILA7 and ILA35 also recognise N*01302, whereas ILA31 does not, and ILA40 shows only a very weak reaction. These cross-reactions are somewhat surprising, given the lack of sequence similarity between the A10 group of alleles and the A6 alleles and the fact that these two groups of alleles are encoded at different loci. Even more surprising is the fact that two of the mAbs appear to be able to distinguish N*01301 and N*01302, which differ by a single amino acid at position 97 in the alpha 2 domain. Although this amino acid is not directly accessible to mAb binding, it is in close contact with bound peptide, and could therefore modify the exposed face of the molecule.

The most striking similarity between these groups of sequences is the M substitution at position 147 (also directly involved in peptide binding) found in the A10 alleles and in three of the A6 alleles. Almost all other classical class I alleles (including human) have a W at this

Table 2 Binding of mAbs to class I transfectants

	ILA7	ILA31	ILA35	ILA38	ILA40	FJ101	ILA13	ILA33	ILA89
N*00101 (5.1)	+	+	+	+	+	_	+	_	+(vw)
N*00102 (JSP.2)	+	+	+	+	+	-	+	_	+
N*00103 (JSP.3)	+	+	+	+	+	-	-	_	_
N*00201 (JSP.1)	+	+	+	+	+	-	+	_	_
N*01301 (HD6)	+	+	+	-	+	-	-	_	_
N*01302 (HD6.1)	+(vw)	-	+	-	+(vw)	-	-	_	_
N*01401 (MAN2)	_	-	-	-	_	-	-	_	_
N*01501 (3349.1)	_	-	-	-	_	+	-	_	_
N*02001 (MAN3)	_	-	-	+	_	-	-	+	+
N*01901 (4221.1)	_	-	-	+	_	-	-	+	+
N*02101 (HD1)	-	_	_	+	_	_	_	-	_

Binding of a panel of nine monoclonal antibodies to stable P815 transfectants each expressing 1 of 11 cattle MHC class I alleles. Former local names for each allele are shown in parentheses. + indicates positive binding, +(vw) indicates very weak binding, - indicates no binding position (apart from a very small number of HLA-B and -C alleles which have L). The mAbs in question do not recognise N*01501 although this has the M substitution; a possible explanation is that other differences found in that sequence alter the conformation such that the mAbs cannot bind. The mAb ILA38 recognises all three of the A30 alleles, but none of the A6 alleles, suggesting that it is recognising a different epitope on the A10 alleles from the other four mAbs.

FJ101 is the only mAb demonstrating allele-specific binding within this panel. Two variants of N*01501 have been partially sequenced, one from an African breed of cattle whose peripheral blood mononuclear cell type serologically as A6, and that show reactivity with FJ101, and the other from European cattle exhibiting the same reactivity. The partial sequences identify three variable positions in alpha 2 when compared to N*01501 (97, 99 and 156).

ILA13 reacts with three alleles in the A10 group, but does not bind to N*00103, which only differs from N*00102 by a single substitution, S to I at position 24. This amino acid is predicted to be buried and not directly accessible to mAb binding, as position 97 in N*01302. These binding patterns demonstrate the exquisite level of specificity achievable with mAbs raised against class I alleles.

ILA33, which was raised against an uncharacterised African haplotype (KN18), fails to recognise N*02101, but sees the other two A30 alleles. Significantly, this allele can also be distinguished from the other A30 sequences by typing with alloantisera, being defined by the A31 specificity.

ILA89 shows a similar pattern of reactivity to ILA33, strongly recognising the same two alleles in the A30 set, but showing additional reactivity with two unrelated alleles. This mAb was raised against the A25 specificity, and one of the two alleles associated with this (N*02901) is believed to be encoded at the same locus as the A30-related alleles.

Although mAb reactivity in these experiments has only been assessed using mouse cells transfected with cattle class I genes, we have previously shown identical mAb reactivities (using some of the same mAbs) on both transfectants and cattle cells (Ellis et al. 1999). In addition, we have shown that association between cattle heavy chain and mouse $\beta_2 m$ is not involved in the observed mAb reactivities with transfectants because cattle $\beta_2 m$ (from FCS in the culture medium) is rapidly substituted when cattle heavy chains reach the cell surface (Ellis and MacHugh, unpublished data).

Discussion

Many of the variable amino acids in all of the sequences are found in positions predicted (from human and mouse studies) to be directly involved in peptide binding. The sequence alignments show some areas of sequence conservation within the groups, particularly within the alpha 1, transmembrane and cytoplasmic domains, and areas of relative conservation across all three groups, particularly in the alpha 3 domain. The alpha 2 domain shows a significant level of variation within the A6 and A30 groups. This probably reflects the fact that the A6 and A30 groups are defined by supertypic specificities and all (apart from N*01302) represent discrete allele groups, whereas three of the four A10 sequences are allelic variants (i.e. less than eight amino acid differences in total, see http://www.ebi. ac.uk/ipd/mhc).

There is still a level of uncertainty around the number of classical MHC class I loci in cattle. The wide range of haplotype configurations, leading to the supposition that any of the genes may be 'silenced' or deleted, simply adds to the problem. Two separate mapping studies involving different haplotype configurations indicate that there are at least five loci, and subsequent haplotype and phylogenetic analyses support this and provide evidence that this number could be even larger (Ellis, 2004). A recent study of expressed class I genes in sheep has generated similar data (Miltiadou et al. 2005). Although the sequences involved in the present study are not those that have been mapped, they have nevertheless been extensively analysed previously (Holmes et al. 2003). These studies have clearly shown that these three groups of alleles fall within different phylogenetic lineages, suggesting the products of three discrete loci.

There are several areas that demonstrate sharing of sequence between putative loci, for example amino acids 22–24 (all three groups), the beginning of alpha 2 (the A6 and A30 groups), the beginning (A6 and A30) and end (all three groups) of alpha 3. We have previously described this phenomenon using additional examples and provided evidence that it is the result of interlocus recombination (Holmes et al. 2003). The data shown here support our previous conclusions.

Within the A10 group, differences between N*00201 and the other three alleles probably reflect the fact that the latter three alleles were all derived from African cattle breeds, which are known to be genetically divergent from European cattle. These three alleles appear to have undergone more recent divergence as a consequence of small recombination events.

Within the A30 group, N*02101 has three substitutions in alpha 1 where the other three sequences are identical. Two of these substitutions (W at position 35, and T at position 49) appear unique when compared to all full length cattle class I sequences. However, because these sequences are likely to represent a fraction of the true level of diversity, it is still possible that the substitutions in N*02101 arose by recombination rather than point mutation.

It has been known for some time that alloantisera used for cattle class I typing are often unable to distinguish between similar alleles. What is perhaps more surprising is the degree to which mAbs that were raised to be allele specific show very specific and limited cross-reactions on very different alleles that are in several cases encoded at distinct loci. Plausible explanations can, in some cases, be derived from close examination of the sequences, which clearly show the sharing of small sequence motifs between alleles/loci.

The mechanisms that generate diversity in human MHC class I genes are likely to be the same in cattle. These are point mutations and intralocus recombination involving small sequence motifs (Parham et al. 1995). Comparison of the sequences reported in this study support our previous suggestion that interlocus recombination is occurring in cattle MHC genes (Holmes et al. 2003). The consequences of such interlocus recombination are likely to be higher overall diversity and fewer locus-specific areas of sequence. Although this may be true in alpha 1 and alpha 2, there appear to be potential locus-specific motifs within the transmembrane and cytoplasmic domains, which in general seem rather more variable than in other species, e.g. human, pig. Another consequence may be greater overall similarity between products of different loci in structure and perhaps function. Therefore, it is unlikely that any will have specialised function, for example, interacting with NK receptors.

At present, insufficient data exist to make many predictions regarding overall cattle MHC class I diversity. Preliminary molecular typing data in cattle indicate that the loci represented here by the A10 and A30 groups, together with an additional locus, are the most polymorphic and the most commonly represented in different haplotypes, whereas the locus represented by the A6 alleles is less commonly present and is represented by few alleles (Ellis and MacHugh, unpublished data). However, the existence of several allelic 'variants' (defined as containing fewer than eight amino acid differences, see http://www.ebi.ac. uk/ipd/mhc/bola/) within this small data set, and an additional three identified so far among the 31 published full length cattle class I sequences, indicate that there are likely to be many more examples.

An awareness of the scope and limitations of existing reagents and knowledge of variant alleles generated by a variety of mechanisms will lead to more accurate assessments of cattle MHC diversity in the future. This is important for the maintenance of healthy cattle populations and to aid our understanding of cattle immune responses to both pathogens and vaccination.

Acknowledgements This work was supported by the Biological Sciences and Biotechnology Research Council, UK. We would like to thank Karen Staines for technical assistance. Experiments carried out in this study comply with the UK law.

References

- Bensaid A, Kaushal A, Baldwin CL, Clevers H, Young JR, Kemp SJ, MacHugh ND, Toye PG, Teale AJ (1991) Identification of expressed bovine class I MHC genes at 2 loci and demonstration of physical linkage. Immunogenetics 33:247–254
- Davies CJ, Joosten I, Bernoco D, Arriens M, Bester J, Ceriotti G, Ellis SA, Hensen EJ, Hines HC, Horin P, Kristensen B, Lewin HA, Morgan AL, Morita M, Nilsson P, Oliver RA, Orlova A, Ostergard H, Park CA, Simon M, Spooner RL, Stewart JA (1994) Polymorphism of bovine MHC class I genes. Joint report of the fifth international BoLA workshop. Eur J Immunogenet 21:239–258
- Di Palma F, Archibald SD, Young JR, Ellis SA (2002) A BAC contig of approximately 400 kb contains the classical class I MHC genes of cattle. Eur J Immunogenet 29:65–68
- Ellis SA (2004) The cattle MHC: is it unique? Vet Pathol and Immunopathol 102:1-8
- Ellis SA, Staines KA, Morrison WI (1996) cDNA sequence of cattle MHC class I genes transcribed in serologically defined haplotypes A18 and A31. Immunogenetics 43:156–159
- Ellis SA, Staines KA, Stear MJ, Hensen EJ, Morrison WI (1998) DNA typing for BoLA class I using PCR-SSP. Eur J Immunogenet 25:365–370
- Ellis SA, Holmes EC, Staines KA, Smith KB, Stear MJ, McKeever DJ, MacHugh ND, Morrison WI (1999) Variation in the number of expressed MHC genes in different cattle class I haplotypes. Immunogenetics 50:319–328
- Garber TL, Hughes AL, Watkins DI, Templeton JW (1994) Evidence for at least three transcribed BoLA class I loci. Immunogenetics 39:257–263
- Goddeeris BMA (1990) A bovine class I MHC molecule plus a Theileria parva antigen resembles mouse H-2Kd. Anim Genet 21:323–328
- Holmes EC, Roberts AFC, Staines KA, Ellis SA (2003) Evolution of MHC class I genes in Cetartiodactyls. Immunogenetics 55: 193–202
- Hughes AL, Yeager M (1998) Natural selection at MHC loci of vertebrates. Annu Rev Genet 32:415–435
- Loftus RT, MacHugh DE, Bradley DE, Sharp PM, Cunningham P (1994) Evidence for 2 independent domestications of domestic cattle. Proc Natl Acad Sci U S A 91:2757–2761
- Marsh SGE, Parham P, Barber LD (2000) In: the HLA FactsBook pp 62–63, Academic
- Miltiadou D, Ballingall KT, Ellis SA, Russel GC, McKeever DJ (2005) Haplotype characterization of transcribed ovine major histocompatibility complex (MHC) class I genes. Immunogenetics DOI 10.1007/s00251-005-0008-y
- Parham P, Adams EJ, Arnett KL (1995) The origins of HLA-A, B, C polymorphism. Immunol Rev 143:141–180
- Parham P, Ohta T (1996) Population biology of antigen presentation by MHC class I molecules. Science 272:67–74
- Pichowski JS, Ellis SA, Morrison WI (1996) Sequence of 2 cattle MHC class I cDNAs associated with BoLA specificity A10. Immunogenetics 43:253–254
- Robinson J, Waller MJ, Stoehr P, Marsh SGE (2005) IPD—the Immuno Polymorphism Database. Nucleic Acids Res 33:523– 526
- Stear MJ, Pokorny TS, Fryda-Bradley S, Lie O, Bull RW (1990) Genetic analysis of the antigens defined at the third international BoLA workshop. J Immunogenet 17:21–28
- Taracha EL, Goddeeris BM, Teale AJ, Kemp SJ, Morrison WI (1995) Parasite strain specificity of bovine cytotoxic T cell responses to Theileria parva is determined primarily by immunodominance. J Immunol 155:4854–4860