Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 9798–9802, December 1990
Immunology

Structure of miniature swine class II *DRB* genes: Conservation of hypervariable amino acid residues between distantly related mammalian species

(major histocompatibility complex/molecular evolution)

KENTH GUSTAFSSON*, SHARON GERMANA, FRANÇOIS HIRSCH†, KAREN PRATT‡, CHRISTIAN LEGUERN, AND DAVID H. SACHS§

Transplantation Biology Section, Immunology Branch, Division of Cancer Biology and Diagnosis, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by David R. Davies, September 12, 1990 (received for review June 22, 1990)

ABSTRACT **As part of our studies of the class II genes of miniature swine, we have isolated and characterized cDNA clones corresponding to *DRB* genes from two major histocompatibility complex homozygous strains. Comparison of the sequences of these clones to those of human *DRB* genes revealed a striking amino acid homology between the hypervariable residues of *SLA-DRB^c* and the human *DRB1-0101* allele. The percentage of differences in these residues between the pig *DRB^c* allele and the human *DRB1-0101* allele was significantly lower (29%) than that between the *DRB1-0101* allele and all other human alleles (average, 66.2%). This similarity was not seen in a comparison of the number of silent substitutions, by which the swine *DRB^c* and the human *DRB-0101* differed. Since phenotypic selection operates at the level of protein products rather than nucleotide sequences, these data suggest the existence of selective mechanisms that have resulted in similar hypervariable regions in certain alleles even in these widely disparate species. Consistent with this hypothesis, an examination of available murine and bovine class II sequences revealed a homology in hypervariable residues between the human *DRB1-1401* allele and the mouse E_{β^s} allele as well as a cow *DRB* allele. Consideration of these data along with intraspecies allelic sequence comparisons suggests that at least some of the interspecies similarities have emerged as the result of convergent evolution, possibly as the result of a need to react to common pathogens.**

The major histocompatibility complex (MHC) contains genes encoding cell-surface molecules playing a pivotal role in antigen presentation to T cells (1). The MHC of mammals is one of the most polymorphic genetic systems known, in terms of both number of alleles and structural differences between alleles. The MHC genes are divided into two classes, I and II, of which the class II genes encode two separate chains, α and β, forming a dimer on the cell surface. Almost all of the polymorphism in the class II β chains is concentrated in <20 amino acid residues, all of which have been hypothesized to be in direct contact with, or in close proximity to, the antigen and/or the T-cell receptor during antigen presentation (2).

The MHC of swine, termed SLA (3), has been shown to exhibit extensive similarities to its human counterpart in both structure and function (4). As a large animal model for studies of transplantation biology, our laboratory has developed a herd of partially inbred miniature swine in which three independent MHC haplotypes (*SLA^a*, *SLA^c*, and *SLA^d*) and three intra-SLA recombinant haplotypes (*SLA^f*, *SLA^g*, and *SLA^h*) are maintained (5, 6). Studies of vascular grafts in these animals have shown that long-term tolerance can be obtained without exogenous immunosuppression by matching for SLA class II loci only, despite differences at class I and numerous minor loci (7). Because these results suggest the possibility of modifying transplantation immunity by genetic engineering of SLA class II genes, we have recently begun a detailed analysis of porcine class II genes in miniature swine. Clones encoding class II genes both from an *SLA^c* genomic library (8, 9) and from *SLA^c* and *SLA^d* cDNA libraries (10–12) have been isolated and characterized. We report here the primary structure of two allelic miniature swine *DRB* genes and note a surprising similarity of both alleles to a single human *DRB1* allele (*DRB1-0101*).¶ These similarities may have implications for the evolutionary mechanisms leading to class II polymorphisms.

MATERIALS AND METHODS

Construction of cDNA Libraries. One cDNA library was constructed in the λZAP vector (Stratagene) from spleen mRNA of the *SLA^c* haplotype according to the instructions of the manufacturer (10). Another library was constructed by P. Kondaiah in the Okayama–Berg vector by using poly(A)-selected lymphocyte RNA from the *SLA^d* haplotype (12).

Isolation and Mapping of cDNA Clones. Both libraries were screened with a probe corresponding to the translated region of a human *DRB* cDNA clone (pII-b-3) obtained from the laboratory of Per Peterson (Uppsala, Sweden) (13). Approximately 20 positive clones from each library were isolated, rescreened, and partially characterized by using suitable restriction endonucleases. One clone from each library was chosen and mapped with a variety of restriction endonucleases in double digests.

Nucleotide Sequence Analysis. Fragments were subcloned into M13mp18 and M13mp19 phage and sequenced by the dideoxynucleotide chain-termination method (14) using the Sequenase kit (United States Biochemical). Consensus sequences were obtained with MicroGenie software (Beckman). The sequences of both strands were determined.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; ARS, antigen recognition site.

*Present address: Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala Biochemical Center, S75124 Uppsala, Sweden.
†Present address: Institut National de la Santé et de la Recherche Medicale U28, Hopital Broussais, 96 rue Didot 75014 Paris Cedex 14, France.
‡Present address: Department of Microbiology, University of Vermont, Burlington, VT 05403.
§To whom reprint requests should be addressed.

¶The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M55165 and M55166).

Computation of Nucleotide Substitutions. Silent and replacement substitutions were computed by using a program kindly provided by Masatoshi Nei (Houston, TX) (15). In this analysis, the potential substitutions of each type constitute the average of the potential substitutions in the two sequences in each pairwise comparison. The calculated substitution frequencies were not corrected for multiple substitutions since the data were not used to compute evolutionary distances but to analyze differences in frequencies of silent vs. replacement substitutions.

RESULTS

Nucleotide Sequences of the Two Allelic *SLA-DRB* cDNA Clones. As shown in Fig. 1, sequences of the two cDNA clones showed them to be full length, containing all of the translated portions of two corresponding allelic miniature swine *DRB* genes. By analogy to human *DRB* genes (13), a putative initiation codon was found in amino acid position −29. Several distinguishing features of human *DRB* genes were also found in the pig *DRB* sequences, including (*i*) identical size of the predicted cytoplasmic portions (amino acid positions 220–237, (*ii*) sizes of extracellular domains, (*iii*) location of the putative attachment site for oligosaccharide (13), and (*iv*) homologous 5′ and 3′ untranslated regions (data not shown). The two allelic pig *DRB* cDNA clones differed by a total of 42 nucleotide substitutions, of which 12 were situated in untranslated regions. Twelve of the remaining 30 substitutions were in the region encoding the signal sequence (residues −29 to −1), 12 in the region encoding the first extracellular domain (residues 1–94), 3 in the region encoding the second extracellular domain (residues 95–188), and 1 in the region encoding the transmembrane portion (residues 198–220). Ten of the nucleotide substitutions in the region encoding the first domain represented amino acid replacements, and 2 were silent substitutions, while all of the nucleotide substitutions in the second domain were silent. Six of the 10 predicted replacement substitutions in the first domain (60%) were situated in positions corresponding to the putative antigen recognition site (ARS) (2), which is similar to the percentage of replacement substitutions (67%) reported for first domains among several human *DRB* alleles (16). Comparisons of the nucleotide sequence of the two pig *DRB* genes to the sequences corresponding to the first domain of 41 different human *DRB* genes is presented in Fig. 2. Human alleles labeled with asterisks were chosen to represent allelic groups of *DRB* sequences (17). Homology varied between 78% and 82% with no human *DRB* sequence appearing to be significantly more similar to the pig sequence than any other.

Comparisons at the DNA and Protein Levels in First Domain Hypervariable Residues Between Pig and Human *DRB* Alleles. A comparison of the 17 most polymorphic residues among 41 human *DRB* sequences (i.e., positions having three or more

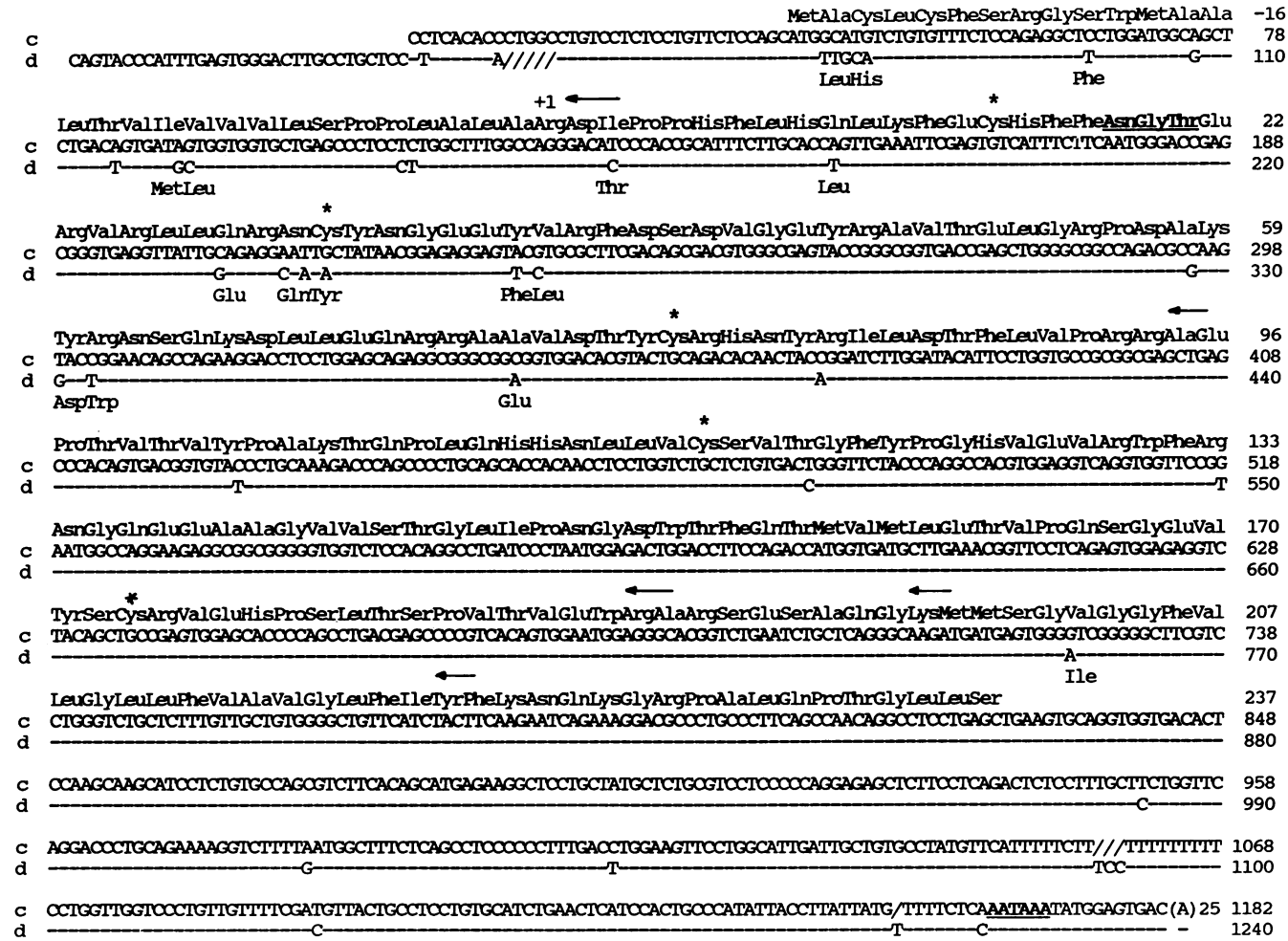

FIG. 1. Comparison of nucleotide and predicted amino acid sequences of the two miniature swine cDNA clones encoding *SLA-DR β* chains of the *SLA*^{c} and *SLA*^{d} haplotypes. Identities between the two sequences are indicated with hyphens, and slashes show deletions or insertions in the respective sequences. The putative N-linked glycosylation site is underlined, as is a candidate polyadenylylation site in the 3′ untranslated region. Asterisks indicate cysteine residues, and arrows delineate the boundaries of the protein domains.

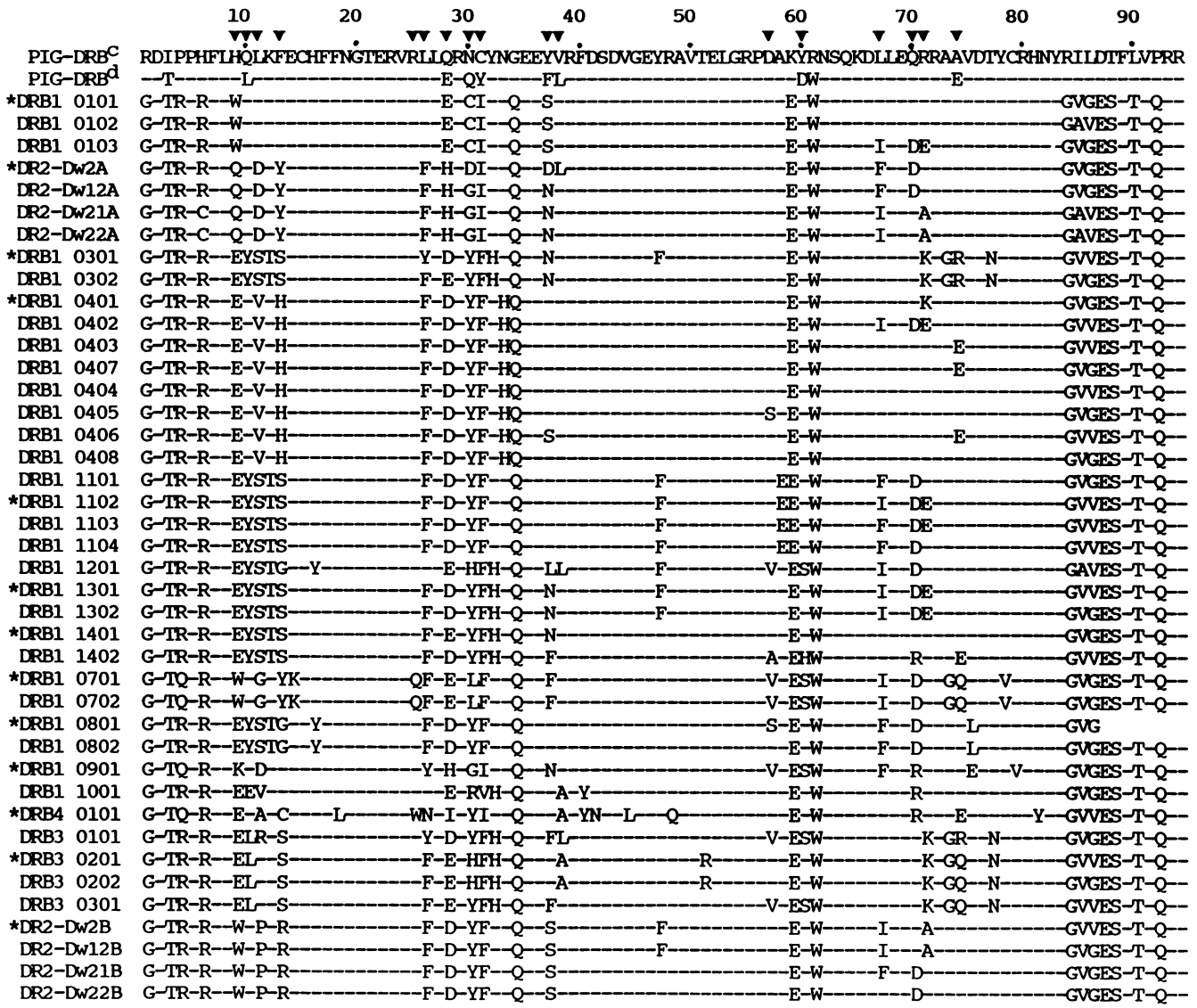

FIG. 2. Comparison of amino acid sequences of the two pig *DRB* alleles to 41 human *DRB* alleles. All human sequences are from ref. 16 and references therein. Arrowheads at the top show the 17 most polymorphic residues as compared between the different human alleles (17). Dashes denote identities with the pig *DRB^c* allele. Asterisks indicate the human alleles used for comparisons in Table 1.

different amino acids among these alleles) (17) with the corresponding positions of the pig *DRB* alleles revealed a surprising finding (Fig. 2). Both pig *DRB* alleles showed marked similarities to the human *DRB1-01* alleles with the pig *DRB^c* allele differing by only 29% of 17 residues from the human *DRB1-0101* and *DRB1-0102* alleles, whereas the average percentage of differing residues between the pig *DRB^c* and the remaining human *DRB* alleles was 66.2% ± 10.8% (Table 1). In addition, no other human allele (except other members of the *DRB1-01* subtype) was as similar to *DRB1-0101* as was the pig *DRB^c* allele (66.6% ± 12.4% differences). Both of these comparisons were statistically significant (Student's *t* test).

It seemed possible to us that phenotypic selection might have been responsible for this observed conservation between DRB hypervariable region residues. To test this hy-

Table 1. Percentages and statistical significance of silent substitutions and amino acid replacements in polymorphic and nonpolymorphic regions of human and porcine *DRB* sequences

Residues	Comparison	% of silent substitutions		% of amino acid replacements	
17 polymorphic*	*SLA-DRB^c vs. HLA DRB1-0101*	73		29	$P = 0.007$
	SLA-DRB^c vs. other 12 *HLA DRB* alleles	46.1 ± 11.0	$P < 10^{-6}$	66.2 ± 10.8	$P = 0.014$
	HLA DRB-0101 vs. other 12 *HLA DRB* alleles	6.1 ± 5.6		66.6 ± 12.4	
Non-polymorphic	*SLA-DRB^c* vs. *HLA DRB1-0101*	ND		14	$P = 0.17$
	SLA-DRB^c vs. other 12 *HLA DRB* alleles	27.3 ± 3.0	$P < 10^{-6}$	16.4 ± 1.4	$P = 0.001$
	HLA DRB-0101 vs. other 12 *HLA DRB* alleles	8.1 ± 1.4		3.6 ± 2.2	

Number of substitutions is expressed as percent of potential sites. All sequences compared are from Fig. 2. *P* values were calculated by Student's *t* test.

*Same residues as those defined in Fig. 2.

pothesis, we examined the amount of divergence between the relevant sequences. Table 1 shows the percentages of silent substitutions and amino acid replacements obtained from a triangular comparison between sequences corresponding to the DRB hypervariable residues (as indicated in Fig. 2) of the human *DRB1-0101*, the other human *DRB* alleles, and the pig DRB^c allele. The number of silent substitutions expressed as percent of potential silent sites (hereafter referred to as the frequency of silent substitutions) between the 17 polymorphic residues of the pig DRB^c allele and the corresponding regions of the human *DRB1-0101* allele was 73%. Similar comparisons between the human *DRB1-0101* or pig allele with the other 12 human *DRB* alleles revealed that the frequency of silent substitutions in intraspecies comparisons is markedly lower (6.1%) than that obtained for the cross-species comparison (46.1%).

Divergence at the DNA and Protein Level in Nonhypervariable Residues. Table 1 also shows the averages of individual computations of divergence at the DNA and protein levels between nonhypervariable positions corresponding to various *DRB* alleles. The frequency of silent substitution between the pig DRB^c and the human alleles was still higher (27.3% ± 3%) than between the human *DRB1-0101* allele and the other human alleles (8.1% ± 1.4%). In contrast to the hypervariable amino acid residues, the nonhypervariable residues of the pig DRB^c allele did not differ significantly in the percentages of divergence from the *DRB1-0101* allele (14%) vs. the other 12 human *DRB* alleles (16.4% ± 1.4%).

Sharing of Polymorphic Residues Among Other Class II β Chains from Other Species. After finding this striking homology between the pig *DRB* hypervariable region sequences and those of one human *DRB* allele, we attempted to find similar homologies among available *DRB* sequences from other species. Examination of the GenBank data base revealed the following two additional examples of such sharing (Fig. 3): (*i*) the polymorphic residues of human *DRB1-1401* were much more similar to the mouse E_{β^s} (18) than were all other human alleles; and (*ii*) the same human allele, *DRB1-1401*, differed by only 7 residues from a cow *DRB* allele (19). The silent substitution frequencies of intraspecies and interspecies comparisons for these species were similar to those for the pig vs. human comparisons (data not shown).

DISCUSSION

A comparison of the primary structures predicted from the sequences of the two isolated cDNAs encoding miniature swine *DR* β chains shows that ≈60% of the variability between them occurs in positions that constitute the hypervariable ARS as defined in human class II β chains, for which 67% of the amino acid variability occurs in the ARS. Both of the pig alleles show a surprising similarity to the human *DRB1-0101* allele (12 of 17 polymorphic residues for DRB^c and 9 of 17 for DRB^d) (Fig. 2). No other human allele is as similar to the *DRB1-0101* allele in the 17 hypervariable residues as is the pig DRB^c allele. However, the remainder of the first domain shows no cross-species allele-specific similarities on the amino acid level or on the nucleotide level, and for these residues the intraspecies homologies are markedly higher than any cross-species comparison (Table 1).

The high degree of polymorphism observed for class II genes appears to be governed mainly by positive selection for polymorphism in the ARS while the remainder of the first domain as well as the second domain are subjected to a conservative selection against polymorphism (13, 16, 20). The differences in the ARS are chiefly responsible for intraspecies alleles. Sharing of class II alleles between distantly related species has previously been suggested on the basis of serological cross-reactions, although the structural basis of such cross-reactions was not determined (21). Shared serologic MHC specificities have also been identified in primate species closely related to humans (22), and structural studies have confirmed the existence of conserved class II alleles in closely related species, suggesting a trans-species mode of allelic inheritance (23–26). However, because one would expect high degrees of homology throughout the related genes of such closely related species, restricted divergence of the entire gene sequences seems more likely than selection for specific hypervariable regions as an explanation for the shared alleles. To the best of our knowledge, structural similarities between individual alleles over relatively long evolutionary distances have not previously been reported. It has, however, been hypothesized that some of the polymorphism present today developed very early during mammalian evolution (27).

We can envision two possible explanations for the hypervariable region similarities we have observed between these distant species: (*i*) the cross-species allelic similarities existed prior to mammalian speciation and have been conserved, even though gene conversion-like events may have caused partial homogenization of sequences from related loci; and (*ii*) some or all of the conserved residues have emerged through convergent evolution. The silent substitution frequencies between the *DRB* alleles within each species are markedly lower than the silent substitution frequencies for the homologous genes between the two species. Thus, the pig *DRB* differs from the human *DRB* alleles by an average of 46% of silent mutations in the hypervariable codons and 27% silent substitutions in the remainder of the second exon. In contrast, among human alleles, the comparable percent silent substitutions are 6% within the hypervariable codons and 8% in the nonhypervariable codons. Therefore, if the cross-species allelic similarities existed at a time prior to mammalian speciation one would have to postulate that all human alleles have diverged from each other since that time, and that the *DRB1* (and subtypes) represents the ancestral allele from which such divergence occurred. The same reasoning, however, could be applied to the similarity between the human *DRB1-1401* and the corresponding murine and bovine alleles (Fig. 3), rendering these alleles ancestral to all others. The inconsistency in this reasoning is further magnified by the fact that the human alleles *DRB1-0101* and

	9	10	11	13	25	26	28	30	31	37	38	57	60	67	70	71	74
DRB1-0101	W	Q	L	F	R	L	E	C	I	S	V	D	Y	L	Q	R	A
Pig DRBc	H	-	-	-	-	-	Q	N	C	Y	-	-	-	-	-	-	-
DRB1-1401	E	Y	S	S	R	F	E	Y	F	N	V	D	Y	L	Q	R	A
Mouse Eβs	-	-	-	-	-	L	-	-	-	-	L	-	N	F	-	-	-
Cow DRB	-	-	-	-	-	-	D	-	Y	T	-	-	-	F	E	K	E

FIG. 3. Cross-species comparisons of similar allelic sequences in the 17 most polymorphic residues in human *DRB* alleles (17). Identities to human alleles *DRB1-0101* and *DRB1-1401*, respectively, are shown with dashes. Numbers above comparisons indicate the amino acid residue numbers in human class II *DR* β-chain first domains. Human and murine sequences are from ref. 17, and the cow sequence is from ref. 19.

DRB1-1401, when compared to each other, show no silent substitutions, suggesting recent divergence. Neither would gene conversion-like events, in our view, explain the similarity observed between swine and human *DRB* hypervariable regions, since the resulting homogenization would affect both silent and replacement sites equally. We suggest, therefore, that convergent evolution provides a more plausible explanation for the similarities we have observed between allelic hypervariable residues in these interspecies comparisons.

The hypervariable regions of MHC class II β chains have been implicated as major contributors to the binding of antigenic peptides during cellular interactions of the immune response, as determined by restriction patterns and T-cell receptor specificity (2, 28, 29). It is therefore intriguing to speculate that presentation of common pathogens could provide one possible mechanism for the appearance of similar functional portions of allelic MHC class II molecules in different species. In this case, two distantly related MHC class II molecules that share homologous hypervariable regions might have been selected for ability to present the same processed antigenic peptide to their respective immune systems. Precedent for this hypothesis can be found in the shared usage of identical amino acids between different heavy-chain variable region immunoglobulin molecules with common antigen specificity (30). In addition, different T-cell receptor β chains with a common antigen specificity have been shown to be under selection for identical amino acids in the region analogous to the third complementarity region of immunoglobulins (31). Studies of T-cell activation by class II plus known antigenic peptides may help to substantiate this hypothesis.

We thank Mona El-Gamil for expert technical assistance and Dr. Paduruh Kondaiah for generously providing us with a cDNA library from the SLA^d haplotype. K.G. was the recipient of a grant from the Swedish Medical Research Council. This study was supported in part by the Swedish Cancer Society.

1. Klein, J. (1987) *Natural History of the Major Histocompatibility Complex* (Wiley, New York).
2. Brown, J. H., Jardetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J. & Wiley, D. C. (1988) *Nature (London)* **332**, 845–850.
3. Vaiman, M., Renard, C., Lafage, P., Ameteau, J. & Nizza, P. (1970) *Transplantation* **10**, 155–164.
4. Pescovitz, M. D. & Sachs, D. H. (1986) in *HLA Class II Antigens*, eds. Solheim, V. G., Moller, E. & Ferrone, S. (Springer, Berlin) pp. 541–556.
5. Sachs, D. H., Leight, G., Cone, J., Schwartz, S., Stuart, L. & Rosenberg, S. (1976) *Transplantation* **22**, 559–567.
6. Pennington, L. R., Lunney, J. K. & Sachs, D. H. (1981) *Transplantation* **31**, 66–71.
7. Pescovitz, M. D., Thistlethwaite, J. R., Jr., Auchincloss, H., Jr., Ildstad, S. T., Sharp, T. G., Terrill, R. & Sachs, D. H. (1984) *J. Exp. Med.* **160**, 1495–1505.
8. Sachs, D. H., Germana, S., El-Gamil, M., Gustafsson, K., Hirsch, F. & Pratt, K. (1988) *Immunogenetics* **28**, 22–29.
9. Pratt, K., Sachs, D. H., Germana, S., El-Gamil, M., Hirsch, F., Gustafsson, K. & LeGuern, C. (1989) *Immunogenetics* **31**, 1–6.
10. Gustafsson, K., LeGuern, C., Hirsch, F., Germana, S., Pratt, K. & Sachs, D. H. (1990) *J. Immunol* **145**, 1946–1951.
11. Hirsch, F., Sachs, D. H., Gustafsson, K., Pratt, K., Germana, S. & LeGuern, C. (1989) *Immunogenetics* **31**, 52–56.
12. Kondaiah, P., Van Obberghen, E., Ludwig, R. L., Dhar, R., Sporn, M. A. & Roberts, A. B. (1988) *J. Biol. Chem.* **263**, 18313–18317.
13. Gustafsson, K., Wiman, K., Emmoth, E., Larhammar, D., Bohme, J., Hyldig-Nielsen, J.-J., Ronne, H., Peterson, P. A. & Rask, L. (1984) *EMBO J.* **3**, 1655–1661.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
16. Jonsson, A.-K., Andersson, L. & Rask, L. (1989) *Scand. J. Immunol.* **30**, 409–417.
17. Marsh, S. & Bodmer, J. (1989) *Immunol. Today* **10**, 305–312.
18. Mengle-Gaw, L. & McDevitt, H. O. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 7621–7625.
19. Groenen, M. A., Van der Poel, J. J., Dijkhof, R. J. & Giphard, M. J. (1989) *Immunogenetics* **31**, 37–44.
20. Hughes, A. L. & Nei, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 958–962.
21. Sachs, D. H., El-Gamil, M., Kiszkiss, P., Lunney, J. K., Mann, D. L., Ozato, K. & Shinohara, N. (1979) in *T and B Lymphocytes: Recognition and Function XVI*, eds. Bach, F., Bonavida, B., Vitetta, E. & Fox, C. F. (Academic, New York), pp. 15–29.
22. Ward, F. E., Seigler, H. F., Metzgar, R. S. & Reid, D. M. (1973) *Tissue Antigens* **3**, 389–401.
23. McConnell, T. J., Talbot, W. S., McIndoe, R. A. & Wakeland, R. (1988) *Nature (London)* **332**, 651–654.
24. Figueroa, F., Sunther, E. & Klein, J. (1988) *Nature (London)* **335**, 265–267.
25. Lawlor, D. A., Ward, F. E., Eunis, B., Jackson, A. P. & Parham, P. (1988) *Nature (London)* **335**, 268–271.
26. Fan, W., Kasahara, M., Gutknecht, J., Klein, D., Mayer, W. E., Jonker, M. & Klein, J. (1989) *Hum. Immunol.* **26**, 107–122.
27. Klein, J. (1987) *Hum. Immunol.* **19**, 155–162.
28. Heber-Katz, E., Hansburg, D. & Schwartz, R. H. (1983) *J. Mol. Cell. Immunol.* **1**, 3–14.
29. Ronchese, F., Schwartz, R. H. & Germain, R. N. (1987) *Nature (London)* **329**, 254–256.
30. Akolkar, P. N., Sikder, S. K., Bhattacharya, S. B., Liao, J., Gruezo, F., Morrison, S. L. & Kabat, E. A. (1987) *J. Immunol.* **138**, 4472–4479.
31. Hedrick, S. M., Engel, I., McElligott, D. L., Fink, P. J., Hsu, M.-L., Hansburg, D. & Matis, L. A. (1988) *Science* **239**, 541–544.