

Polymorphism of HLA-DR β chains in *DR4*, *-7*, and *-9* haplotypes: Implications for the mechanisms of allelic variation

(cDNA clones/major histocompatibility complex/selection)

PETER K. GREGERSEN*, TETSUYA MORIUCHI[†], ROBERT W. KARR[‡], FUMIYA OBATA*, JUNKO MORIUCHI[§], JOAN MACCARI*, DONNA GOLDBERG*, ROBERT J. WINCHESTER*, AND JACK SILVER*

*Molecular Biology Unit, Department of Rheumatic Diseases, The Hospital for Joint Disease, New York, NY 10003; Departments of [†]Cell Biology and [‡]Internal Medicine, Tokai University School of Medicine, Ischara, 259-11, Japan; and [§]Department of Internal Medicine, University of Iowa College of Medicine, Iowa City, IA 52240

Communicated by Anton Lang, August 18, 1986

ABSTRACT We have isolated and sequenced cDNA clones corresponding to the *DR β 1* and *DR β 2* loci from two homozygous B-cell lines typed as *DR7* (Burkhart) and *DR9* (ISK). These nucleotide sequences were compared to β 1 and β 2 chains of other *DR* haplotypes. The first-domain sequences of β 2 chains are identical in *DR4* and *DR7* haplotypes. In addition, there is strong sequence homology within the 3' untranslated regions of β 1 genes from *DR4*, *-7*, and *-9* haplotypes, thus confirming the close evolutionary relationship among these three haplotypes. In contrast, the first-domain sequences of β 1 molecules from these haplotypes are very different from each other and do not reflect the *DR4*, *-7*, *-9* family relationship. Two explanations for the differences in degree of diversity between β 1 and β 2 chains are suggested. The differences may be a consequence of selection pressures; this implies functional differences for products of the β 1 and β 2 loci. Alternatively, closely linked segments of the human class II region may differ in their underlying rates of variation, independent of selection pressures, and this may in part account for the extraordinary diversity found in the β 1 first domain.

The *HLA-D* region of the human major histocompatibility complex consists of multiple closely linked loci that encode the α and β subunits of class II molecules. These molecules play a central role in the recognition events that lead to T-cell activation and an effective immune response by the organism. One of the most striking features of the human class II region is the extensive polymorphism found at many of these loci. This polymorphism has been shown to be responsible for the enormous variability of immune response patterns among individuals.

In general, among unrelated haplotypes, allelic differences are observed in genes from all three major subregions—*DR*, *DQ*, and *DP*—of the *HLA-D* region (1). Within the *DR* subregion, at least two polymorphic β chains, designated *DR β 1* and *DR β 2*, are expressed in most haplotypes (2). A third β -chain gene is a pseudogene; remnants of other *DR β* genes have also been found in the *DR* region, suggesting previous duplication and deletion events (3). In contrast, the *DR α* chain is not polymorphic (1).

In a previous analysis of class II genes from several closely related haplotypes that type serologically as *DR4* (4), we observed that polymorphism was generally restricted to the *DR β 1* molecule; *DR β 2* genes were identical in all *DR4* haplotypes examined. In addition, the *DQ α* and *DQ β* genes were also highly conserved within the *DR4* haplotype family. Thus, a pattern of higher variability at *DR β 1* contrasted with the extreme conservation of the surrounding class II loci. We

have now extended these observations to a larger family of haplotypes, including *DR4*, *DR7*, and *DR9*. These haplotypes are believed to be related by virtue of sharing the *DRw53* serologic specificity (5). In this report we demonstrate their close evolutionary relatedness at the nucleotide level. More interestingly, the pattern of high variability confined to the *DR β 1* locus, which was observed among *DR4* haplotypes, is even more striking within the *DR4*, *-7*, and *-9* family. These observations indicate that closely linked segments of the class II region may differ dramatically in their degree of polymorphism and raise questions about the genetic mechanisms that may account for such differences.

METHODS

Construction and Screening of cDNA Libraries. cDNA libraries were constructed from the homozygous B-cell lines ISK (*DR9*) and Burkhart (*DR7*) as described (6, 7). The libraries were screened for *DR β* chain genes with the 0.5-kilobase *Pst* I fragment of a previously isolated *DR β* chain gene from a *DR4* cell line (4).

Sequencing. Sequencing was performed using the dideoxy method of Sanger *et al.* (8). The majority of sequencing reactions were performed directly in the cloning vector (9)—either pBR322 or pcDV1—using synthetic primers (OCS Industries, Denton, TX) corresponding to highly conserved regions of the *DR β* molecule (4). Appropriate fragments were subcloned in pUC18 for sequencing of the 3' untranslated regions.

RESULTS

***DR β 1* Sequence Analysis.** Fig. 1 shows the nucleotide sequence comparison of *DR β 1* molecules from *DR4*, *DR7*, and *DR9* haplotypes. The *DR4* sequence shown is from the cell line BIN40 and is representative of the closely related family of *DR4 DR β 1* chains (4). The cDNA clone of *DR β 1* from the *DR9* cell line ISK begins at codon 18 and therefore does not allow comparison at the first hypervariable region. As has been noted for other *DR β 1* alleles, most of the nucleotide polymorphism resides in the first domain and results in productive amino acid changes, as shown in Fig. 2.

Table 1 shows the degree of nucleotide divergence between these and other published *DR β 1* alleles, broken down into first domain, second domain, transmembrane/cytoplasmic tail, and 3' untranslated regions. The first domain nucleotide divergence between *DR4*, *-7*, and *-9* haplotypes ranges from 8.1% to 10.9%. This is similar to the degree of divergence among *DR β 1* chains of unrelated *DR* types. However, comparisons of the nucleotide sequences within the 3' un-

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.

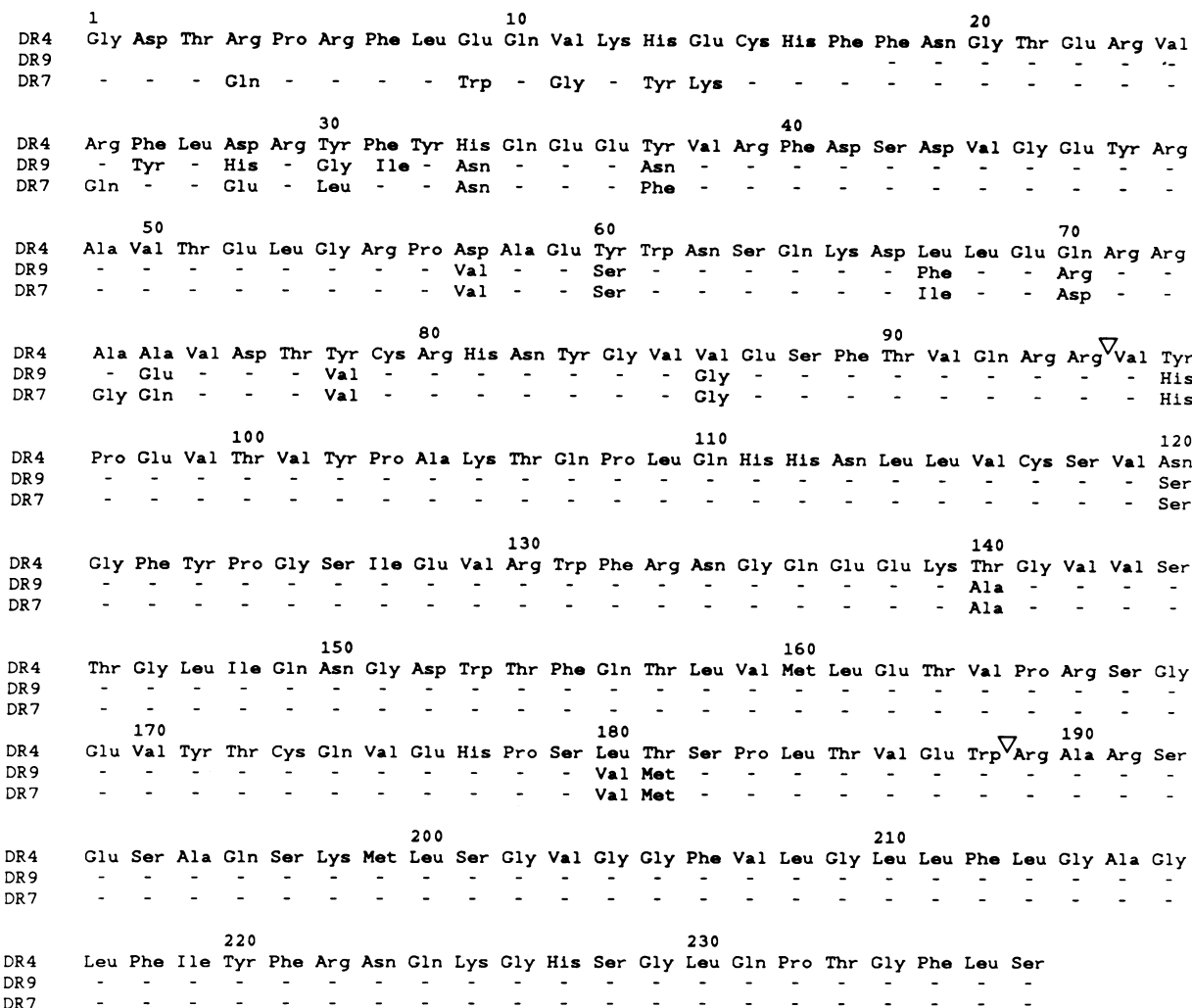


FIG. 2. Predicted amino acid sequence of DRβ1 chains from DR4, DR9, and DR7 haplotypes. Symbols as in Fig. 1.

of DQα chain sequence analysis (6). It is of particular interest that sequence comparisons within the first domain, as shown in Table 1, do not reflect this family relationship.

cDNA clone DRβ#4 from the DR7 cell line Burkhardt contains an insertion of approximately 150 base pairs in the terminal portion of the 3' untranslated region. This insertion

Table 1. Nucleotide sequence differences of DRβ1 chains

	First domain						Second domain				
	DR7	DR9	DR1	pIIB3	DR5		DR7	DR9	DR1	pIIB3	DR5
DR4	10.9	8.1	6.7	7.0	7.0	DR4	3.1	3.1	3.9	3.5	3.1
DR7		8.6	10.9	13.1	11.3	DR7		0	4.2	4.6	4.2
DR9			9.0	10.7	8.1	DR9			4.2	4.6	4.2
DR1				10.2	8.8	DR1				4.6	4.2
pIIB3					6.3	pIIB3					0

	Transmembrane/cytoplasmic region						3' untranslated region				
	DR7	DR9	DR1	pIIB3	DR5		DR7	DR9	DR1	pIIB3	DR5
DR4	1.3	0.6	0	1.3	1.3	DR4	2.7	3.7	14.1	15.3	14.8
DR7		0.6	1.3	1.3	1.3	DR7		1.8	12.1	13.1	12.1
DR9			0.6	0.6	0.6	DR9			13.8	16.1	15.9
DR1				1.3	1.3	DR1				8.2	7.8
pIIB3					0	pIIB3					0.3

Numbers represent percent difference. The 3' untranslated sequences were aligned by using a modification of the method of Needleman and Wunsch (10, 11). Insertions of one nucleotide or greater were counted as a single event. DR1, DR5, and pIIB3 sequences are from refs. 12, 13, and 14, respectively.

shows no significant homology to any sequence in the current Genbank data base.[†] In particular, it is not homologous to the *Alu I* sequence that we (4) and others (15) have described at the 3' end of incompletely or alternatively spliced *DRβ1* transcripts. For the purposes of the nucleotide divergence calculations shown in Table 1, this insertion was counted as a single event.

DRβ2 Sequence Analysis. The first domain sequence of the *DRβ2* molecule from the *DR7* cell line Burkhardt is identical to that obtained for the *DRβ2* molecule from *DR4* cell lines (4).

DISCUSSION

Our data firmly establish the close evolutionary relationship among *HLA-D* region haplotypes that encode the serologic determinants *DR4*, -7, and -9. The special relatedness of this family of haplotypes had been suspected because of their sharing of the serologic specificity DRw53 (5, 16), and more recently by the finding of sequence homology among *DQα* genes from these haplotypes (6). The remarkable degree of sequence conservation in the *DRβ1* 3' untranslated regions of these haplotypes, as compared with other *DRβ1* alleles, constitutes the strongest evidence for a close evolutionary relationship among *DR4*, -7, and -9 haplotypes. This is summarized in Table 1. When comparing 3' untranslated regions within the *DR4*, -7, and -9 family group, sequence divergence is between 1.8% and 3.7%. However, comparisons with *DRβ1* alleles outside the *DR4*, -7, -9 haplotype family show a much greater sequence divergence in the 3' untranslated region, in the range of 12.1–16.1%. Second-domain sequence comparisons are not informative with respect to this family relationship, presumably because selection pressures prevent a large degree of divergence in this exon. However, *DRβ1* molecules in *DR7* and *DR9* haplotypes have identical second-domain sequences, indicating that they may be more closely related to each other than to *DR4*. In addition to the strong sequence homologies just mentioned for *DRβ1*, we have also found the *DRβ2* molecule to be identical in *DR4* and *DR7* haplotypes; protein analysis indicates that the *DRβ2* molecules in *DR9* cell lines share this identity (17, 18). The *DRβ2* gene is largely responsible for the DRw53 serologic specificity, which is common to *DR4*, -7, and -9 haplotypes (ref. 16 and unpublished observations). The absolute conservation of *DRβ2* constitutes further support for a close genetic relationship among DRw53-bearing haplotypes.

The most intriguing result of these studies is that sequence comparisons within the first domain of *DRβ1* do not reflect the *DR4*, -7, -9 family relationship. As shown in Table 1, first-domain sequence divergence ranges from 8.1% to 10.9% within the *DR4*, -7, -9 haplotype family. This is approximately the same degree of divergence that is seen when comparing any two unrelated *DRβ1* alleles, regardless of *DR* type. This is a similar but more dramatic example of the pattern of variability we have observed previously between *DR4* subtypes (4). In a recent study we found that *DR4* haplotypes of differing *HLA-D* types displayed variability that was restricted to the first domain of *DRβ1*; other linked loci such as *DRβ2*, *DQα*, and *DQβ* were identical within the *DR4* family. Thus, a pattern emerges for both *DR4* and the larger *DR4*, -7, -9 family in which the *DRβ1* gene diverges more rapidly than genes at other loci. How can such a large degree of variability in the first domain of *DRβ1* be explained in the face of the highly conserved nature of other class II genes, such as *DRβ2* and *DQα*, which are closely linked to *DRβ1*?

The traditional arguments from selection assume that the underlying rates of variation for *DRβ1*, *DRβ2*, and *DQα* are identical and that differing selection pressures at these loci account for their different degrees of polymorphism. If this reasoning is correct, strong selective advantage to the population must result from polymorphism at the *DRβ1* locus but not at the *DRβ2* locus. This would suggest that some functional difference exists for products of *DRβ1* and *DRβ2* genes. *DRβ2* molecules are generally expressed in lower quantities in these cell lines (4, 16); such quantitative differences may influence the degree of selection pressure for variation at this locus.

An alternative explanation for the greater polymorphism of *DRβ1* alleles as compared with *DRβ2* alleles is suggested by studies of the murine class II region. Steinmetz *et al.* (19) have shown that the *I-E* and *I-A* regions differ in their degree of variability between inbred strains as well as in the outbred population. Of particular interest is the fact that variability in the *I-A* region is found in both coding and noncoding regions. Likewise, the conserved nature of the *I-E* region is not restricted to coding regions. These studies suggest that the underlying rate of variability is quite different in the *I-A* and *I-E* regions. The boundary between tracts of high variability and low variability could be placed somewhere between the first domain of the *I-Eβ* gene and its 3' flanking region (19). Intriguingly, this is the same region where a recombinational hot spot is thought to exist, marking the boundary between the *I-E* and *I-A* regions. A recent analysis of murine recombinant inbred strains shows that recombination can take place within the *I-Eβ* gene itself (20). It is provocative to speculate that the site of recombinational hot spots and the boundary between conserved and variable portions of the class II region may be functionally related. In this connection, our finding that the 3' untranslated region of the *DRβ1* genes is highly conserved in the *DR4*, -7, -9 family would place such a boundary within the *DRβ1* gene in humans. Whether or not such speculations are correct, the underlying concept that different regions of the class II genome may differ in their rates of mutation does offer an alternative explanation for differences in allelic variability at *DRβ1* and *DRβ2*. Support for this view might be obtained by examining allelic variability of intron sequences at these two loci, since noncoding sequences are presumably not under selection pressure and therefore reflect more accurately the underlying rate of mutation.

An examination of published data indicates that family relationships similar to that of *DR4*, -7, and -9 exist among other *DR* haplotypes. As shown in Table 1, *DRβ1* molecules from a previously published *DR3,6* cell line (14) and a *DR5* haplotype (13) show strong nucleotide homology in their 3' untranslated regions—they differ by only 0.3%. These alleles are also identical in their second-domain sequence. Since both of these haplotypes encode the DRw52 serologic specificity, this suggests that DRw52 may also define an evolutionarily related haplotype group. As more sequence data become available, it will be of interest to see if the pattern of conservation of *DRβ2* and polymorphism of *DRβ1* is maintained in these and other haplotype families.

The data presented here indicate that at least two different mechanisms may be operating in generating the diversity of the class II region and in the evolution of various *HLA-D* haplotypes. One of these appears to be the gradual accumulation of mutations over evolutionary time; the small number of differences observed among the *DQα* and *DRβ2* genes and the 3' untranslated regions of the *DRβ1* genes from *DR4*, -7, and -9 haplotypes are presumably a consequence of this mechanism. In contrast, the large number of differences observed among the first domains of the *DRβ1* genes indicate the existence of an additional mechanism leading to rapid divergence of these genes. Elucidation of this mechanism,

[†]National Institutes of Health (1986) *Genetic Sequence Databank: Genbank* (Research Systems Div., Bolt, Beranek, and Newman, Inc., Boston), Tape Release 44.0.

and its role among the forces generating the polymorphism of the *HLA-D* region, should provide a better understanding of the dynamics of the major histocompatibility complex.

Note Added in Proof. Recent DNA sequence analyses of *DR3* and *DR6* haplotypes (21) and *DR5* haplotypes (22) support the hypothesis that DRw52 defines a group of evolutionarily related haplotypes analogous to the DRw53 family describe here.

We thank Ms. Roslyn Berger and Ms. Ann Rupel for their help in the preparation of this manuscript. This work was supported in part by grants from the National Institutes of Health. Funds were also generously provided by the Milton Petrie Endowment and the Klein Endowment.

1. Korman, A. J., Boss, J. M., Spies, T., Sorrentino, R., Okada, K. & Strominger, J. L. (1985) *Immunol. Rev.* **85**, 45–86.
2. Trowsdale, J., Young, J. A. T., Kelly, A. P., Austin, P. J., Carson, S., Meunier, H., So, A., Erlich, H. A., Spielman, R. S., Bodmer, J. & Bodmer, W. (1985) *Immunol. Rev.* **85** 5–43.
3. Spies, T., Sorrentino, R., Boss, J. M., Okada, K. & Strominger, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5165–5169.
4. Gregersen, P. K., Shen, M., Song, Q., Merryman, P., Degar, S., Seki, T., Maccari, J., Goldberg, D., Murphy, H., Schwenzler, J., Wang, C. Y., Winchester, R. J., Nepom, G. T. & Silver, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2642–2646.
5. Lepage, V., Grumet, F. C., Busson, M., Colombani, J., Ness, D. & Degos, L. (1984) in *Histocompatibility Testing 1984*, eds. Albert, E. D., Baur, M. P. & Mayr, W. R. (Springer, New York), pp. 203–204.
6. Moriuchi, J., Moriuchi, T. & Silver, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3420–3434.
7. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161–170.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
9. Wallace, R. B., Johnson, M. J., Suggs, S. V., Miyoshi, K., Bhatt, R. & Itakura, K. (1981) *Gene* **16**, 21–26.
10. Needleman, S. B. & Wunsch, C. D. (1970) *J. Mol. Biol.* **48**, 443–453.
11. Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* **147**, 195–197.
12. Tonelle, C., DeMars, R. & Long, E. O. (1985) *EMBO J.* **4**, 2839–2847.
13. Tieber, V., Abruzzini, L., Didier, D. K., Schwartz, B. & Rotwein, P. (1986) *J. Biol. Chem.* **261**, 2738–2742.
14. Gustafsson, K., Wiman, K., Emmoth, E., Larhammar, D., Bohme, J., Hylding-Nielsen, J. J., Ronne, H., Peterson, P. & Rask, L. (1984) *EMBO J.* **3**, 1655–1661.
15. Segall, M., Cairns, J. S., Dahl, C. A., Curtsinger, J., Freeman, S., Nelson, P., Cohen, O., Wu, S., Nicklas, J. N., Noreen, H. J., Linner, K. M., Saunders, T. L., Choong, S. A., Ohta, N., Reinsmoen, N. L., Alter, B. J. & Bach, F. H. (1985) *Immunol. Rev.* **85**, 129–148.
16. Sorrentino, R., Lillie, J. & Strominger, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3794–3798.
17. Obata, F., Endo, T., Tate, G., Miyokawa, M., Katagiri, M. & Kashiwagi, N. (1986) *J. Immunol.* **136**, 2187–2190.
18. Matsuyama, T., Schwenzler, J., Silver, J. & Winchester, R. J. (1986) *J. Immunol.* **137**, 934–940.
19. Steinmetz, M., Malissen, M., Hood, L., Orn, A., Maki, R. A., Dastoornikoo, G. R., Stephan, D., Gibb, E. & Romaniuk, R. (1984) *EMBO J.* **3**, 2995–3003.
20. Steinmetz, M., Stephan, D. & Lindahl, K. F. (1986) *Cell* **44**, 895–904.
21. Gorski, J. & Mach, B. (1986) *Nature (London)* **322**, 67–70.
22. Didier, D. K., Schiffenbauer, J., Shuman, S., Abruzzini, L., Gorski, J., Watling, D., Tieber V. & Schwartz, B. D. (1986) *J. Immunol.* **137**, 2627–2631.