

Linkage map of three *HLA-DR* β -chain genes: Evidence for a recent duplication event

(class II antigens/*HLA-DR* pseudogene/genetic complexity/haplotype comparison)

PIERRE ROLLINI, BERNARD MACH, AND JACK GORSKI

Department of Microbiology, University of Geneva School of Medicine, Geneva, Switzerland

Communicated by Werner Arber, July 1, 1985

ABSTRACT The predominant class II, or Ia, antigen of the human major histocompatibility complex is *HLA-DR*. It consists of an α and a β chain, the latter being responsible for the remarkable polymorphism of these Ia antigens. Studies with cloned genes had shown the existence of more than one *DR* β -chain locus. We have isolated about 100 kilobases of the *HLA-DR* β -chain gene region from a cosmid library generated from a consanguineous homozygous B-cell line of the *DR3* haplotype. Three *HLA-DR* β -chain genes have been characterized. They are arranged in a head-to-tail orientation. One of the genes lacks the region encoding the first domain of the *DR* β chain. The two other genes are transcribed, as shown by RNA blot hybridization analysis. A striking restriction site homology has been found within the *DR* β -chain gene cluster, suggesting a recent duplication event involving at least 25 kilobases of DNA. Moreover, the molecular map of *DR* β chain genes cloned from B-cell lines of two other *HLA-DR* haplotypes shows extensive homology between alleles of a given *DR* β -chain locus.

The class II antigens of the major histocompatibility complex are highly polymorphic transmembrane glycoproteins, consisting of an α subunit and a β subunit. These molecules are located predominantly on the surface of macrophages and B cells. They play a key role in the control of the immune response, functioning in cell-cell interactions and antigen presentation to regulatory T lymphocytes. In the human major histocompatibility complex (*HLA*), the class II molecules have been mapped to the *HLA-D* region on chromosome 6. Three subregions have been defined, *DP*, *DQ*, and *DR*. The *HLA-DR* antigen is the predominant surface product and it is the β chain that is responsible for the *DR* polymorphism (for reviews, see refs. 1 and 2).

The molecular organization of the *HLA-D* region has not yet been elucidated. The restriction map of the *DP* subregion is known. There are two *DP* α - and β -chain loci (3) found as α - β pairs with the α and β chain genes in a head-to-head orientation (4-6). The organization of the *DQ* and *DR* subregions is not known, although the genetic complexity of these subregions has been documented. There are two *DQ* α loci (7, 8) and two *DQ* β loci (9, 10). The *HLA-DR* subregion has been shown to contain only one nonpolymorphic α -chain gene (11, 12). Analysis of cDNA clones (13, 14) and genomic clones (10), as well as direct studies of cellular DNA by Southern blot hybridization (15), has allowed us to establish the existence of multiple *DR* β -chain loci.

In this study, we report the characterization of overlapping cosmid clones containing *DR* β -chain genes isolated from a genomic library prepared from DNA of a *HLA*-homozygous individual. Three *DR* β -chain genes in a head-to-tail conformation have been aligned. Two of these genes are transcribed, while the third one appears to be truncated, lacking

the first domain. Extensive restriction site homology between *DR* β -chain genes within the same haplotype has been observed, suggesting a recent duplication event. In addition, we observe a striking structural homology between this *DR* β -chain gene region and that of two other related haplotypes. The possible evolutionary implications of these observations are discussed.

MATERIALS AND METHODS

Construction of the Genomic Libraries. Details will be published elsewhere. Briefly, DNA from the consanguineous homozygous cell line AVL (*DR3,3*; kindly provided by M. Giphart) was partially digested with *Sau3AI* and fractionated on sucrose gradients. Fractions containing DNA fragments 30-45 kilobases (kb) long were collected and ligated with arms of cosmid pTCF (16). The ligation mixture was packaged *in vitro* (17) and transfected into *Escherichia coli* 490 A. Cosmids were spread, replicated (18), and screened by colony filter hybridization (19), using conditions described by Steinmetz *et al.* (20). Clones hybridizing to the specific probes after three rounds of screening were grown and their DNA was prepared. As screening probes, two DNA fragments free of repetitive sequences were isolated from a *DR*- β -chain genomic clone (unpublished data). The *DRw6,6* cosmid and *DR4,w6* phage genomic libraries have already been described (3, 10). The clones from the heterozygous phage library that correspond to the *DRw6* haplotype were identified by Southern blot hybridization (21).

Restriction Enzyme Mapping. Cosmids were mapped by standard single and double digestions with restriction endonucleases followed by Southern blot hybridization using specific cDNA or genomic probes.

Hybridizations. Southern blot analysis (22) was as described (15). Dot hybridizations were performed as described (13). Filters were washed several times at 65°C in 2 \times standard saline citrate (NaCl/Cit; 1 \times NaCl/Cit is 150 mM NaCl/15 mM sodium citrate, pH 7.0) before a series of 30-min washes at 65°C at the indicated stringencies. Hybridizations with oligonucleotides were performed on cosmid DNA fragments or total RNA that had been electrophoresed on agarose gels and transferred to nitrocellulose or nylon membranes by standard procedures (22, 23). The conditions of hybridization and washes using the ³²P-labeled oligonucleotide probes (19-mers) were chosen to optimize the discrimination between perfect hybrids and hybrids with a single mismatch (unpublished results). The oligodeoxynucleotide probes correspond to a highly polymorphic region of the first domain of the *HLA-DR* β -chain loci and are as follows: I, CTCAGACGTAGAGTACTCC; III, CTCAGACTTACGCAGCTCC. In this region there is less than 50% homology with *DP* or *DQ* sequences.

RESULTS

Mapping of Cosmids Containing DR β -Chain Genes. A cosmid library was constructed with DNA from the consanguineous homozygous B-cell line AVL (DR3,3). The library was screened twice, first with a fragment corresponding to the first domain and adjacent sequences of an *HLA-DR* β -chain gene, and then with a probe containing the second domain and flanking sequences of a *DR* β -chain gene. Several cosmids were isolated after two rounds of screening. A third round was performed using 3' untranslated region probes specific for the three subregions *DP*, *DQ*, and *DR*. Twelve cosmids hybridizing to the 3' *DR* β -chain probe were mapped, and the restriction fragments corresponding to the signal sequence, first domain, second domain, and 3' untranslated region of the *DR* β molecule were identified by Southern blot hybridization with specific probes.

These twelve clones define a contiguous stretch of about 100 kb of DNA, containing three *DR* β -chain genes, which we call βI , βII , and βIII in the order of their transcription. The molecular map of the *DR* β region is shown in Fig. 1. As can be seen, the three genes are arranged in a head-to-tail orientation. The spacing between βI and βII is about 11 kb, and between βII and βIII , about 24 kb. A large intron (about 8 kb) separates the signal sequence from the first domain exons in βI and βIII .

The restriction map of the *DR* βI gene is quite different from that of the *DR* βII gene. However, there is extensive restriction site homology between the *DR* βI and *DR* βIII genes as well as the region downstream from each of these genes (indicated in Fig. 1 by boxes). This suggests a relatively recent duplication event involving at least 25 kb of DNA. An additional region of hybridization to a signal sequence probe has been found 10 kb downstream from the *DR* βIII gene (at map position 91 kb in Fig. 1).

Southern blot analysis of cellular DNA as well as cosmids containing the βI , βII , and βIII genes was performed to establish that the clones indeed represent genes found in the B-cell DNA and to determine the completeness of the library

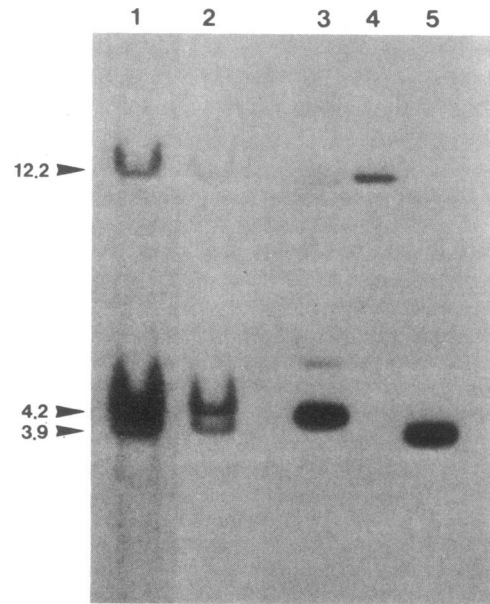


FIG. 2. Southern blot analysis of cosmids and of cellular DNA digested with *Bam*HI and hybridized with a fragment containing the second domain and adjacent sequences of a *DR* β chain gene. Lanes 1 and 2, cellular DNA (AVL); lane 1 is a longer exposure. Lanes 3-5, recombinant cosmids 4-1 (*DR* βI), 6-2 (*DR* βII), and 10-4 (*DR* βIII), respectively (see Fig. 1). Faint bands in lane 3 result from partial digestion of the cosmid. Fragment sizes are indicated in kb.

(Fig. 2). Digestion with *Bam*HI and hybridization with a second domain probe revealed only three bands in cellular DNA (lanes 1 and 2), each band corresponding to one of the three cloned genes (lanes 3-5).

βII Is a Truncated *DR* Gene. In the course of the restriction mapping, it was observed that the *DR* βII gene did not hybridize to probes specific for the first domain of the *DR* β chain. This gene was explored in more detail. DNA dot

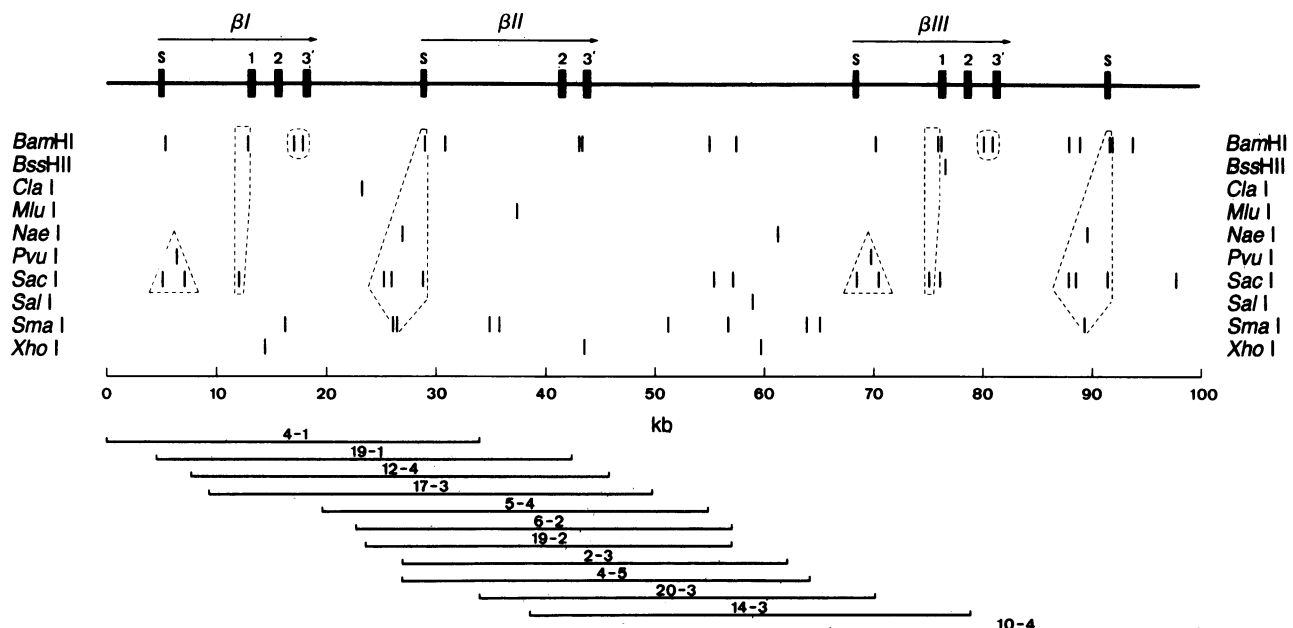


FIG. 1. Molecular map of the *DR* β region. The location of exons, as described from higher-resolution maps of clones or subclones, is shown by filled boxes. S, signal sequence; 1, first domain; 2, second domain; 3', 3' untranslated region. The exons for the transmembrane region and the cytoplasmic tail are not indicated. Arrows above the map indicate the 5' \rightarrow 3' orientation of the genes. Positions of restriction sites are shown by vertical bars. No site for *Nru*I has been found in any of the cosmids. The fragments cloned in individual cosmids are indicated by horizontal bars with their respective isolation number. Repetitive patterns in the restriction map are indicated by boxes of different shapes (broken lines). The three *DR* β loci are numbered I to III following the direction of transcription.

hybridizations were performed with various probes (Fig. 3). A series of 3' untranslated region probes, known to be specific for each of the subregions *DP*, *DQ*, and *DR*, was used first to show that β_{II} is really a *DR* gene (Fig. 3A). Indeed, the β_{II} gene hybridizes to the *DR* 3' probe and not to the *DP* or *DQ* probes. Fig. 3B shows the results with a probe for the second domain (cDNA) and a probe encoding the second domain plus flanking sequences (genomic). As can be seen, the *DR* β_{II} gene hybridizes to these sequences. However, when a probe containing the first domain and flanking sequences was used no hybridization was observed with the β_{II} gene, even at low stringencies (Fig. 3C). When a Southern blot of cosmid 6-2 (β_{II}) was hybridized at low stringency with a cDNA probe specific for the signal sequence and the first domain, a probe that gives two bands with the genes β_I and β_{III} , only the band corresponding to the signal sequence was seen (Fig. 3C).

β_I and β_{III} Are Transcribed. The transcription of the two complete *DR* β -chain genes (β_I and β_{III}) was studied by RNA blotting using oligonucleotide probes. Specific oligonucleo-

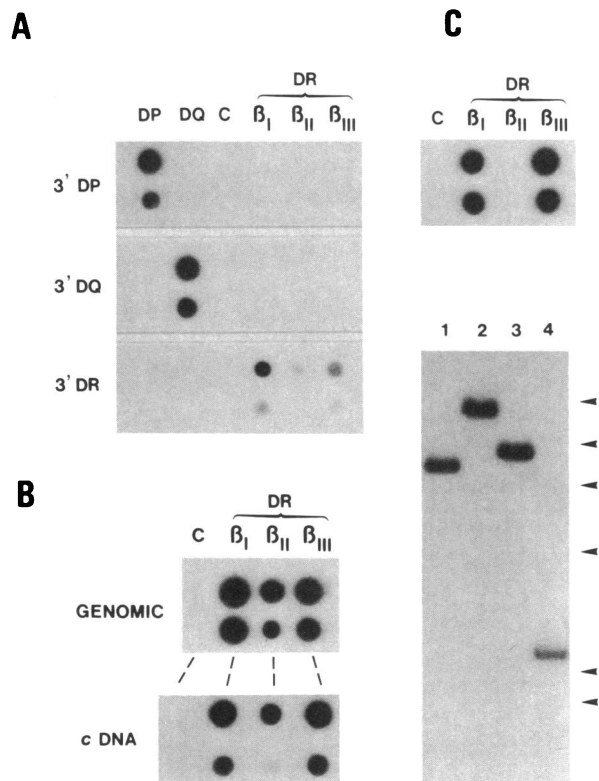


FIG. 3. Dot hybridizations of recombinant clones with various probes. (A) Subregion specificity. DNA of the cosmids 4-1 (*DR* β_I), 6-2 (*DR* β_{II}), and 10-4 (*DR* β_{III}) was used. DQ is a phage containing a *DQ* β -chain gene (10). DP is a cosmid containing a *DP* β -chain gene (3). C is the vector pTCF control. The probes all represented the 3' untranslated portion of cDNA clones encoding DP (3), DQ, and DR (14) β chains. The final wash in each case was $0.5\times$ NaCl/Cit for the upper row, $0.1\times$ NaCl/Cit for the lower. (B) Hybridizations with a *DR* β -chain second-domain (cDNA) probe (14) and a second domain plus flanking region probe (genomic). Identification and stringencies as in A. (C) Hybridizations with *DR* β first-domain probes. Dot hybridization is with a fragment containing the first domain plus flanking region of a *DR* β -chain gene. The final wash was $2\times$ NaCl/Cit for the upper row, $0.1\times$ NaCl/Cit for the lower. The Southern blot shows cosmid 6-2 (*DR* β_{II}) hybridized with a *Pst* I fragment from a cDNA (14) containing the signal sequence and the first domain of the *DR* β molecule. Digestions were as follows: lane 1, *Mlu* I + *Bam*HI; lane 2, *Mlu* I + *Xho* I; lane 3, *Sma* I + *Mlu* I; lane 4, *Sma* I + *Bam*HI. The final wash was $2\times$ NaCl/Cit. The arrowheads denote the migration of phage λ DNA digested with *Hind*III.

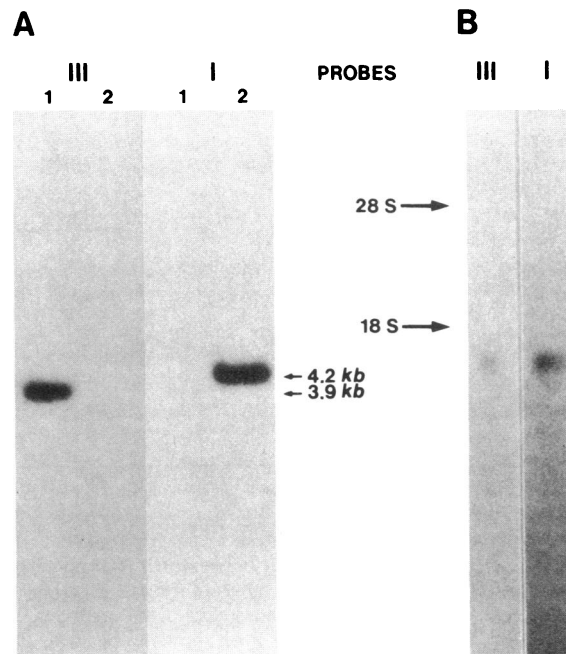


FIG. 4. Hybridizations with oligonucleotides. (A) Southern blot analysis of cosmids 10-4 (*DR* β_{III} , lanes 1) and 4-1 (*DR* β_I , lanes 2). One microgram of cosmid digested with *Bam*HI was loaded in each case. Oligonucleotide III is specific for *DR* β_{III} , oligonucleotide I, for *DR* β_I . (B) Blots of total AVL RNA (20 μ g) hybridized with the specific oligonucleotides I and III. Migration positions of rRNAs are indicated.

ides (19-mers; see *Materials and Methods*) against a polymorphic region in the first domain of β_I and β_{III} were used under conditions in which only perfectly matched sequences are stable. Fig. 4A shows the β_I and β_{III} specificities of the two probes as shown by Southern blotting of cosmids containing β_I or β_{III} genes. It is evident that each probe is specific for a given gene. The same specificity has been observed with cellular DNA (AVL) analyzed by Southern blotting, in which each probe hybridizes to only one band corresponding to β_I or β_{III} , respectively (unpublished data).

Blots of total RNA from the AVL line, hybridized with the two oligonucleotide probes, reveal in both cases a band of the size of *DR* β -chain mRNAs (Fig. 4B; refs. 13 and 23). *DR* β_I and β_{III} are therefore both transcribed.

Comparison of *HLA-DR* β -Chain Genes from Different Haplotypes. Genomic libraries have also been constructed with DNA from two other B-cell lines, HHK (*DRw6, w6*) and a heterozygous *DR4, w6* cell. Several *DR* β -chain genes were isolated from these libraries. By Southern blot hybridization it is possible to assign a clone isolated from a heterozygous cell line to a given haplotype (21). The restriction maps of the clones containing *DRw6* genes were compared to those of the *DR3* genes shown in Fig. 1. Extensive homology was found, allowing an alignment of the *DR* β -chain genes of these three haplotypes (Fig. 5). The overall organization of the *DR* β -chain gene cluster is therefore very similar in these *DR3* and *DRw6* haplotypes.

DISCUSSION

The highly polymorphic *HLA-DR* β -chain genes play an essential role in the immune response. Analysis of cDNA clones (13, 14) and genomic clones (10), as well as studies of cellular DNA by Southern blotting (15), had allowed us to establish the existence of multiple *DR* β -chain loci. The study of the function of these genes in antigen presentation and the understanding of the molecular basis of their polymorphism requires the determination of the number of loci, of their

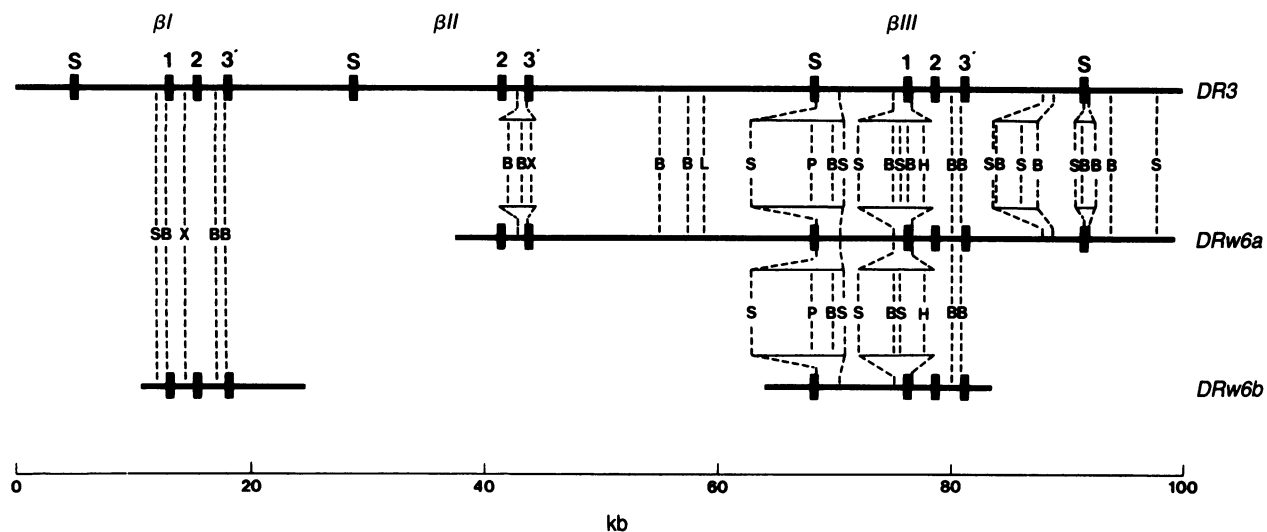


FIG. 5. Restriction site homology between *DR* β -chain genes from different haplotypes. The three linked genes from the *DR3* haplotype are depicted as in Fig. 1 (*DR3*). About 60 kb of the *DR* β region have been characterized in overlapping clones from a *DRw6,w6* cosmid library (*DRw6a*; ref. 3). Two *DR* β -chain genes have been isolated in overlapping phages from a library made with DNA from a *DR4,w6* heterozygous cell line (10). These two genes have been shown to be of the *DRw6* haplotype by Southern blotting (*DRw6b*). Genes are aligned by homology of restriction sites. In some areas the map has been enlarged 4-fold for clarity. An identical site in two related genes is indicated by a broken line. The code for restriction enzymes is as follows: B, *Bam*HI; H, *Bss*HII; L, *Sal* I; P, *Pvu* I; S, *Sac* I; and X, *Xho* I.

organization, and of their structure within a single haplotype. This paper presents the linkage map of three *HLA-DR* β -chain genes in the *DR3* haplotype.

Overlapping cosmid clones have allowed us to define a contiguous stretch of 100 kb of DNA in the *HLA-DR* β region. Three *DR* β -chain genes, arranged in a head-to-tail orientation, have been characterized. We suggest that they should be called *DR* β I, *DR* β II, and *DR* β III, the roman numerals assigned in the order of transcription. The spacing between these genes is 11 and 24 kb. The middle gene, β II, has a restriction map which differs from that of *DR* β I or *DR* β III. Hybridization studies confirm this observation and also show that this gene does not hybridize to *DR* β chain probes specific for the first domain exon and adjacent intron sequences (Fig. 3). One can postulate the existence of another exon whose sequence is so different that it no longer hybridizes to any human *DR* β first domain probe, even at low stringency. An alternative interpretation is that the *DR* β II gene has lost its first domain and adjacent sequences after a deletion event. In either case, the *DR* β II gene is not expected to code for a normal *DR* β chain, and it can thus be considered as a *DR* β pseudogene.

Two of the *DR* β -chain genes (β I and β III) are transcribed. This correlates with our previous cDNA cloning results (14, 24), in which two *DRw6* and two *DR4* cDNA clones were found. The presence of only two active loci might be of significance with respect to the load imposed on the T-cell repertoire. It is not known yet whether *DR* β II is transcribed. However, it is interesting to note that so far there have been no reports of *DR* β cDNA clones that are truncated or contain an unrelated first domain.

A closer look at the map of the *DR* β cluster reveals blocks of restriction site homology (see boxes in Fig. 1). The arrangement of these homologous regions, even in the non-coding regions, suggests a recent event resulting in the duplication of a portion of the *DR* β gene. This event involved at least 25 kb of DNA and included an entire gene and downstream sequences, including the following signal sequence. Alternatively, a large-scale homogenization (by recent gene conversion) of previously duplicated genes could explain these two regions of homology.

The presence of a fourth signal sequence, found at the 3' end of the cluster, is interesting. It could indicate the

presence of a fourth gene, which would be the homologue of β II. We do not think this is likely in this particular haplotype because the Southern blot of the cell line reveals no fragments other than those of the cloned genes.

We suggest that the *DR* β -chain gene cluster has had a complicated history of duplications, deletions, and conversions. For example, the duplication of the entire β I and β II gene segment would result in four *DR* β chain loci, which could in some cases undergo further deletions involving one or another of the *DR* β loci. In other cases, instead of a recent duplication, extensive deletions might result in a single *DR* β -chain gene. In that respect, analysis of homozygous *DR1,1* cell lines, both by Southern blot (15) and genomic cloning (unpublished data), suggests that there are fewer *DR* β genes in *DR1*. This would make it likely that the number of *DR* β -chain genes might vary from one to four in different haplotypes. In addition, some of these genes might be pseudogenes and not result in a functional *DR* β -chain product.

It should be noted that the results presented here argue in favor of the hypothesis that the original number of *DR* β -chain genes was two (β I and β II). There are two mouse *I-E* β -chain genes (25, 26), the mouse equivalent of *DR*, which are also in head-to-tail orientations. Taken together, this suggests that the pair of β chains of this subregion of the major histocompatibility complex has existed before man-mouse divergence.

Several clones containing *DR* β chain genes were isolated from two other genomic libraries generated from a homozygous *DRw6,w6* cell line and a heterozygous *DR4,w6* cell line. All *DRw6* β chain genes were mapped and aligned with the cloned *DR3* genes (Fig. 5). It is clear that the overall organization of the region is the same in these three haplotypes. An interesting point to note is that the supertypic specificity called MT2 (27) includes *DR3* and *w6*, as well as *DR5,w8*. By comparison, the restriction map of a *DR* β pseudogene isolated from a *DR4* (non-MT2) homozygous cell line (28) is completely different from the map of any of the *DR3* or *w6* genes. This homology, at the level of restriction mapping of cloned genes, between haplotypes within a supertypic group reinforces the results of Southern blotting analysis, in which many identical bands are seen in DNAs of the same supertypic group. These results indicate that the

haplotypes within a supertypic group are evolutionarily related. This comparison also shows that, at least for genes of the same supertypic group, allelic sequence comparisons can be made by relating the map of a given gene to a given locus identified on the *DR* β linkage map (Fig. 5). This is an important point in the study of *DR* β chain polymorphism, as until now sequence comparisons have been made without the possibility of distinguishing between alleles or pseudoalleles.

Finally, the availability of these genes now allows the study of the expression of individual *DR* genes after DNA-mediated cell transformation (29). This should result in a correlation between the individual *DR* β -chain loci and the structure or functional specificities of their products, as measured by serological reactivities or specific interaction with T lymphocytes. Indeed, expression of the *DR* β III locus from the *DR3*, *DRw6a*, and *DRw6b* haplotypes (Fig. 5) in mouse L cells has allowed the identification of that locus as encoding the MT2 (w52) serological specificity (ref. 21 and unpublished results).

We are grateful to M. Giphart for the cell lines AVL and HHK, E. Kawashima for synthesizing oligonucleotides, M. Steinmetz for making available a preprint of his work, and C. de Pr eval for help with the RNA blot analysis and useful discussions. This research was supported by the Swiss National Fund for Scientific Research.

1. Steinmetz, M. & Hood, L. (1983) *Science* **222**, 727-733.
2. Kaufman, J., Auffray, C., Korman, A., Shackelford, D. & Strominger, J. (1984) *Cell* **36**, 1-13.
3. Gorski, J., Rollini, P., Long, E. & Mach, B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3934-3938.
4. Trowsdale, J., Kelly, A., Lee, J., Carson, S., Austin, P. & Travers, P. (1984) *Cell* **38**, 241-249.
5. Servenius, B., Gustaffson, K., Widmark, E., Emmoth, E., Andersson, G., Larhammar, D., Rask, L. & Peterson, P. A. (1984) *EMBO J.* **3**, 3209-3214.
6. Okada, K., Prentice, H. L., Boss, J. M., Levy, D. J., Kappes, D., Spies, T., Raghupathy, R., Mengler, R. A., Auffray, C. & Strominger, J. L. (1985) *EMBO J.* **4**, 739-748.
7. Auffray, C., Kuo, J., Demars, R. & Strominger, J. (1983) *Nature (London)* **304**, 174-177.
8. Spielman, R. S., Lee, J., Bodmer, W., Bodmer, J. G. & Trowsdale, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3461-3465.
9. Boss, J. & Strominger, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5199-5203.
10. Gorski, J., Rollini, P., Kawashima, E., Long, E. O. & Mach, B. (1984) in *UCLA Symposia: Regulation of the Immune System*, eds. Serkarz, E., Cantor, H. & Chess, L. (Liss, New York), pp. 47-56.
11. Lee, J., Trowsdale, J. & Bodmer, W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 545-549.
12. Wake, C. T., Long, E. O., Strubin, M., Gross, N., Accolla, R., Carrel, S. & Mach, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6979-6983.
13. Long, E. O., Wake, C. T., Strubin, M., Gross, N., Accolla, R. S., Carrel, S. & Mach, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7465-7469.
14. Long, E. O., Wake, C. T., Gorski, J. & Mach, B. (1983) *EMBO J.* **2**, 389-394.
15. Wake, C. T., Long, E. O. & Mach, B. (1982) *Nature (London)* **300**, 372-374.
16. Grosveld, F. G., Lund, T., Murray, E. J., Mellor, A. L., Dahl, H. H. M. & Flavell, R. A. (1982) *Nucleic Acids Res.* **10**, 6715-6732.
17. Hohn, B. & Murray, K. (1979) *Proc. Natl. Acad. Sci. USA* **74**, 3259-3263.
18. Hanahan, D. & Meselson, M. (1980) *Gene* **10**, 63-67.
19. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961-3965.
20. Steinmetz, M., Stephan, D., Dastoornikoo, G. R., Gibb, E. & Romaniuk, R. (1985) in *Immunological Methods*, eds. Lefkowitz, I. & Pernis, B. (Academic, New York), Vol. 3, in press.
21. Gorski, J., Tosi, R., Strubin, M., Rabourdin-Combe, C. & Mach, B. (1985) *J. Exp. Med.* **162**, 105-116.
22. Southern, E. (1975) *J. Mol. Biol.* **98**, 503-517.
23. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201-5205.
24. Long, E. O., Gorski, J., Rollini, P., Wake, C. T., Strubin, M., Rabourdin-Combe, C. & Mach, B. (1983) *Hum. Immunol.* **8**, 113-121.
25. Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Frelinger, J., Wake, C., Long, E., Mach, B. & Hood, L. (1982) *Nature (London)* **300**, 35-42.
26. Long, E. O., Gorski, J. & Mach, B. (1984) *Nature (London)* **310**, 233-235.
27. Tanigaki, N. & Tosi, R. (1982) *Immunol. Rev.* **66**, 5-37.
28. Larhammar, D., Servenius, B., Rask, L. & Peterson, P. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1475-1479.
29. Rabourdin-Combe, C. & Mach, B. (1983) *Nature (London)* **303**, 670-674.