Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6234–6238, September 1987
Immunology

Allelic variation in the DR subregion of the human major histocompatibility complex

(HLA/DR$_\beta$/gene conversion/evolution)

JOHN I. BELL, DAN DENNEY, JR., LINDA FOSTER, TERTIA BELT, JOHN A. TODD, AND HUGH O. MCDEVITT

Departments of Medical Microbiology and Medicine, Stanford University, Stanford, CA 94305

Contributed by Hugh O. McDevitt, March 6, 1987

ABSTRACT **Allelic variation in the DR subregion of the human major histocompatibility complex has been analyzed by nucleic acid sequencing of cDNA clones obtained from cell lines homozygous by consanguinity for all the common serological types DR1–9. Two expressed loci were identified in the haplotypes DR2, -3, -4, -7, and -9; one locus being present at a much lower frequency (4–7%) than the other. The low-frequency allele was highly conserved between each of the DRw53 (DR4, -7, -9) and the DRw52 (DR3, -5, -6) haplotypes. Analysis of the variation between alleles confirms the presence of three allelic hypervariable regions. At each variable residue, a limited range of amino acid substitutions are found, distinguishing them from immunoglobulin hypervariable regions. Dinucleotide substitutions are extremely common. Individual hypervariable regions are often shared between haplotypes. Much of the variation in these alleles can be attributed to the shuffling of these regions between haplotypes, possibly by the mechanism of gene conversion.**

The polymorphism encoded within the class II region of the murine and human major histocompatibility complex provides the structural basis for genetic control of the immune response (1). In man, this region has been divided into three major subregions, DP, DQ, and DR, each encoding both α and β chains (2). These chains associate as heterodimers on the surface of B cells, antigen presenting cells, and activated T cells and, in the presence of antigen, act as restriction elements for both humoral and cellular immune responses (2). The DR subregion maps at the telomeric end of the class II region (3) and, in at least two haplotypes, contains one α- and three β-chain loci (4, 5). The polymorphism in the expressed products of this subregion was originally defined with the mixed lymphocyte reaction (6). Serological reagents have provided extensive additional data on the population distribution of these alleles and their association with a wide variety of autoimmune diseases (7). The serological polymorphism associated with the DR subregion is found in two distinct patterns. There exists both the standard polymorphic series DR1–14 and the supertypic specificities DRw52 and DRw53 superimposed on the DR3, -5, and -w6 haplotypes (DRw52) and the DR4, -7, and -w9 haplotypes (DRw53).

At a molecular level, DR polymorphism has been determined by both restriction fragment length polymorphism (8, 9) and nucleic acid sequencing of cDNA and genomic clones. Three distinct allelic hypervariable regions are found in the β1 domain of both the murine E$_\beta$ and DR$_\beta$ molecules, and these seem to be primarily responsible for both functional (10) and serological polymorphism (11). The determination of this molecular polymorphism in all the standard DR haplotypes is an essential step in understanding the precise structural correlates of the serological reactivities and also provides an accurate impression of the major variations present in these molecules in the population at large. In addition, this information provides insights into both the mechanism by which the polymorphism was generated and the evolutionary association between different haplotypes. We report here the structural basis in the DR subregion for the major serological specificities DR1–9, including the supertypic specificities DRw52 and DRw53. These sequences were all generated from cDNAs obtained from lymphoblastoid cell lines homozygous by consanguinity, making their attribution to serologically defined haplotypes unambiguous. These sequences provide a standard set of the DR-specific sequences that underlay the serological and functional polymorphism known to map to the HLA-DR subregion.

MATERIALS AND METHODS

cDNA was prepared from the human lymphoblastoid cell lines PGF (DR2), WT49 (DR3), WT51 (DR4), JVM (DR5), APD (DR6), LBF (DR7), Madura (DR8), and DKB (DR9). For Madura and DKB, the cDNA was made by the method of Gubler and Hoffman (12), was amplified by the technique of J.A.T., J.I.B., and H.O.M. (unpublished technique) and was sequenced in both orientations after cloning directly into the vector M13mp10. PGF, WT49, WT51, JVM, APD, and LBF cDNA was made from poly(A)$^+$-selected RNA by a modification of homopolymer tailing. Briefly, the first strand was extended using terminal transferase and dGTP, and an oligonucleotide containing a poly(dC) stretch was then annealed onto the first-strand product. This served as the primer for the second-strand synthesis using DNA polymerase I. The cDNA was subsequently filled in with DNA polymerase I and treated with the *Eco*RI methylase. After ligating *Eco*RI linkers, the product was cut with *Eco*RI, gel-purified, and ligated into the vector λgt10. The libraries produced ranged from 1 to 5 × 10^6 independent clones. The libraries were screened using standard techniques (30) and the full-length DR$_\beta$ chain cDNA clone 2918.4 (11). Inserts from the positive clones were subcloned into the vectors pUC8 or M13mp8. With the exception of DR$_\beta$3a, only cDNA clones that included the initiating ATG were studied.

Sequencing was performed using the chain-termination method of Sanger *et al.* (14) with dATP[^{35}S]. Two sequencing strategies were employed. In the first, complementary synthetic oligonucleotides were obtained for both strands of conserved sequences in the DR$_\beta$ chains. The oligonucleotide sequences were chosen to include sequences from the first and second domains and the 3′-untranslated region. The sequence and orientation of these oligonucleotides were as follows.

```
              114               131
Oligo A 5' GCCTGATGCCGAGTACTG 3'
        B 3' CGGACTACGGCTCATGAC 5'
```

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.

```
     240               256
C 5' GTGGAGAGGTTTACACC 3'
D 3' CACCTCTCCAAATGTGG 5'
     803                 821
E 5' GAGAGGGCTTTCTCAGGAC 3'
F 3' CTCTCCCGAAAGAGTCCTG 5'
```

(Overdots indicate the numbered bases.) These oligonucleotides were then used in combination with the standard oligonucleotide M13 primers to prime sequences of individual full-length cDNA clones. The second strategy employed was utilized for the clones DR_β2a, -2b, -7a, -4a, and -5a. This strategy was to sequence random clones generated by sonication, gel-purification, and cloning into M13mp8 (15). Sequence data were analyzed by using the computer program developed by Staden (16). Data from these clones provided multiple overlapping sequences on both strands, and each nucleotide was sequenced on average six times.

Southern blots were performed using established techniques (9). The probes used were a 300-base-pair fragment from the 3'-untranslated region of a DR1 DR_β chain (12) and a 200-base-pair 3'-untranslated region probe from the DR_β4a clone. Hybridization was performed overnight at 42°C in 50% (vol/vol) formamide, 2% (wt/vol) $NaDodSO_4$, 3× SSC (3 M NaCl/0.3 M Na citrate, pH 7.0.), salmon sperm DNA (20 μg/ml), and 3× Denhardt's solution (1× Denhardt's solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin). The filters were washed three times for 20 min at room temperature in 2× SSC/0.1% $NaDodSO_4$, and subsequently washed twice for 30 min in 0.1× SSC/0.1% $NaDodSO_4$ at 65°C.

RESULTS AND DISCUSSION

Two distinct DR_β cDNAs were identified from each of the haplotypes DR2, -3, -4, -7, and -9, confirming the presence in these haplotypes of at least two expressed β chains from the DR subregion. The absence of a second DR_β chain in the DR1 haplotype has been predicted from the less-complex Southern blot pattern obtained with this haplotype (9), while the absence of a second DR_β chain from the DR5 haplotype after analysis of 30 DR_β cDNA clones most probably reflects the low levels of expression of the second DR_β chain. Of the haplotypes that express two DR_β loci on a single haplotype, one DR_β chain was invariably found at a higher frequency than the other. cDNAs representing the common species were identified at 15 (DR3) to 25 (DR7) times the frequency of the rarer species. We have named the common species the "a" locus and the rarer species the "b" locus, because the lack of genomic organization data on all the haplotypes studied makes it impossible to designate a particular genomic locus for a particular cDNA product. Only where genomic organization and sequence data are available (DR3, DR4, and DR6) is it possible to attribute sequences to particular genomic loci (4, 5). In both these cases, the common species represents the product of the βI locus, and the rare DR_β cDNA is generated from the βIII locus. This βIII locus has been recognized (17) to encode at least some of the determinants recognized serologically as DRw52 (MT2) and DRw53 (MT3).

A pattern of conservation consistent with their supertypic associations is obvious among the products of the b locus. The 3b cDNA is identical in sequence to the MT2 sequence observed in the DR6 haplotype (18), whereas the 7b and 9b sequences are also identical to the MT3 sequence published from the βIII locus of the DR4 haplotype (4). This conservation of one locus within each of the DRw52 and DRw53 supertypic families supports the idea that the haplotypes DR3, -5, and -6 evolved together (19), as did the haplotypes DR4, -7, and -9 (20). This similarity extends to the a locus in the DRw52-related haplotypes. DR3, -5, -6 and -8 all share a similar first allelic hypervariable region, and the second is also virtually identical among DR3, -5, and -6. These data confirm that the DRw52 family has most probably derived from the same common haplotype (19) and supports the DR3 transfection data (17) that suggest that some of the structure elements of DRw52 antibody specificity may arise from the βI as well as the βIII locus. Evolutionary relationships among the a alleles of members of the DRw53 family are

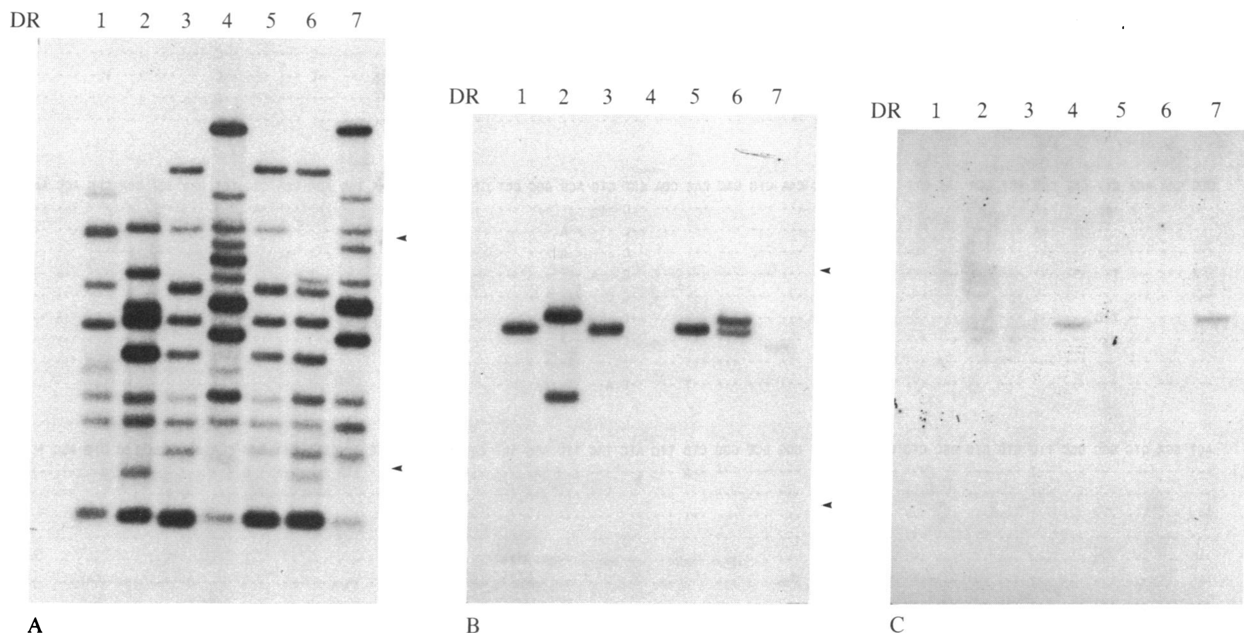

FIG. 1. Southern blots of DNA from homozygous consanguineous lymphoblastoid lines, representing DR types 1–7. (A) Southern blot of homozygous consanguineous cell lines probed with a full-length DR_β chain probe (11). (B) Same blot as in A probed with a 3'-untranslated region probe from a DR1 DR_β chain (11) and washed at 65°C in 0.1× SSC. (C) Autoradiograph of a Southern blot probed with a 3'-untranslated region probe from a DR_β4a cDNA clone and washed at the same stringency.

```
DR1   GGG GAC ACC CGA CCA CGT TTC TTG TGG CAG CTT AAG TTT GAA TGT CAT TTC TTC AAT GGG ACG GAG CGG GTG CGG TTG CTG GAA AGA TGC ATC TAT AAC CAA GAG GAG TCC GTG CGC TTC
DR2a  --- --- --- --- --- --- --- --- CA- --- GA- --- -A- --G --- --- --- --- --C --- --- --- --- --- --- --C --- C-C --- GA- --- --- --- --- --- --- GA- T-- --- ---
DR2b  --- --- --- --- --- --- --- C-- --- --- -C- --- AGG --G --- --- --- --- --- --- --- --- --- --- --- --C --- --C --- -A- T-- --- --- --G --- --- --- --- --- ---
DR3a  --- --- --- A-- --- --- --- --- GA- T-C TC- -C- -C- --G --- --- --- --- --- --- --- --- --- --- --- -AC --- --C --- -A- T-- C-- --- --G --- --- AA- --- --- ---
DR3b  --- --- --- --- --- --- --- --- GA- -T- --- --- -C- --G --- --- --- --- --- --- --- --- --- --- --- --C --- --G --- CA- T-- C-- --- --G --- --- -A- -C- --- ---
DR4a  --- --- --- --- --- --- --- --- GA- --- G-- --A CA- --G --- --- --- --- --C --- --- --- --- --- --- --C --- --C --- -A- T-- --- C-- --- --- --- -A- --- --- ---
DR5a  --- --- --- A-- --- --- --- --- GA- T-C TC- -C- -C- --G --- --- --- --- --- --- --- --- --- --- --- --C --- --C --- -A- T-- --- --- --- --- --- -A- --- --- ---
DR6a  --- --- --- A-- --- --- --- --- GA- T-C TC- -C- -C- --G --- --- --- --- --- --- --- --- --- --- --- --C --- --C --- -A- T-- C-- --- --G --- --- AA- --- --- ---
DR7a  --- --- --- -A- --- --- --- C-- --- --- GG- --- -A- A-G --- --- --- --- --C --- --- --- --- --- -A- --C --- --- --- CT- T-- --- --- --G --- --- -T- --- --- ---
DR7b  --- --- --- -A- --- --- --- --- GA- --- GC- --- -G- --G --- --- --- C-- --- --- --- --- --A --- T-- AAC --- ATC --- -A- --- --- --- --- --- --- -A- -C- --- -A-
DR8a  --- - - --- --- --- --- --- --- GA- T-C TC- -C- GG- --G --- T-- --- --- --- --- --- --- --- --- --- --C --- --C --- -A- T-- --- --- --- --- --- -A- --- --- ---
DR9a  --- --- --- -A- --- --- --- --- AA- --- GA- --- --- --G --- --- --- --- --C --- --- --- --- --- --- -AT --- C-C --- G-- --- --- --- --- --- --- AA- --- --- ---
DR9b  --- --- --- -A- --- --- --- --- GA- --- GC- --- -G- --G --- --- --- C-- --- --- --- --- --A --C T-- AAC --- ATC --- -A- --- --- --- --- --- --- -A- -C- --- -A-

DR1   GAC AGC GAC GTG GGG GAG TAC CGG GCG GTG ACG GAG CTG GGG CGG CCT GAT GCC GAG TAC TGG AAC AGC CAG AAG GAC CTC CTG GAG CAG AGG CGG GCC GCG GTG GAC ACC TAC TGC AGA
DR2a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --T --- --- --- --- --- --- --- --- T-- --- --A G-C --- --C --- --- --- --- --- --- --- ---
DR2b  --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --C --T --- --- --- --- --- --- --- --- A-- --- --- --- GC- --- --- --- --- --- --- --- --- ---
DR3a  --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -A- --- -G- CG- --- --- -A- --- --- ---
DR3b  --- --- --- --- --- --- --- --- --- --- -G- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -A- --- -G- CAG --- --- -AT --- --- ---
DR4a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -A- --- --- --- --- --- --- --- --- ---
DR5a  --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- -AG --- --- --- --- --- --- --- --- A-- --- --A G-C GA- --- --- --- --- --- --- --- --- ---
DR6a  --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- A-- --- --A G-C GA- --- --- --- --- --- --- --- --- ---
DR7a  --- --- --- --- --- --- --- --- --- --- --- --- --A --- --- --- -TC --- --- -C- --- --- --- --- --- --- A-- --- --- G-C --- --- -G- CA- --- --- --- GTG --- ---
DR7b  A-- --T --- C-- --- --- --- -A- --- --- --- --- --- --- --- --- --C --T --- --- --- --- --- --- --- --- --- --- --- -G- --- --- --- -A- --- --- --- --- --- ---
DR8a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- AGC --- --- --- --- --- --- --- --- --- T-- --- --A G-C --- --- --- CT- --- --- --- --- --- ---
DR9a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -TC --- --- -C- --- --- --- --- --- --- T-- --- --- -G- --- --- --- -A- --- --- --- GTG --- ---
DR9b  A-- --T --- C-- --- --- --- -A- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- --- --- --- --- --- -G- --- --- --- -A- --- --- --- --- --- ---

DR1   CAC AAC TAC GGG GTT GGT GAG AGC TTC ACA GTG CAG CGG CGA GTT GAG CCT AAG GTG ACT GTG TAT CCT TCA AAG ACC CAG CCC CTG CAG CAC CAC AAC CTC CTG GTC TGC TCT GTG AGT
DR2a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- G-- -G- --- --- A-- --- --- --- --- --- --- --- --- --- --- --- -A-
DR2b  --- --- --- --- --- -TG --- --- --- --- --- --- --- --- --C C-A --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR3a  --- --- --- --- --- -TG --- --- --- --- --- --- --- --- --C C-T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --T --- --- ---
DR3b  --- --- --- --- --- -TG --- --- --- --- --- --- --- --- --C C-T --- C-- --- --- --- --- --- G-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR4a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C T-T --- G-- --- --- --- --- --- G-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -A-
DR5a  --- --- --- --- --- -TG --- --- --- --- --- --- --- --- --C C-T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- ---
DR6a  --- --- --- --- --- -TG --- --- --- --- --- --- --- --- --C C-T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- ---
DR7a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C C-T --- G-- --- --- --- --- --- G-C --- --T -G- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR7b  T-- --- --- --- --- -TG --- --- --- --- --- --- --- --- --C C-A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -A-
DR8a  --- --- --- --- --- ---
DR9a  --- --- --- --- --- ---
DR9b  T-- --- --- --- --- -TG

DR1   GGT TTC TAT CCA GCC AGC ATT GAA GTC AGG TGG TTC CGG AAC GGC CAG GAA GAG AAG GCT GGG GTG GTG TCC ACA GGC CTG ATC CAG AAT GGA GAT TGG ACC TTC CAG ACC CTG GTG ATG
DR2a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- A-- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --C --- --- --- --- --- --- --- ---
DR2b  --- --- --- --- --- --- --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- A-- --- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- ---
DR3a  --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --- A-- --- --- --- --- --- --- --- --- --C --- --- --C --- --- --- --- --- --- --- ---
DR3b  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --G --- --- --- --- --- --- --C --- --- --- --- --- --- --- ---
DR4a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- A-- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- ---
DR5a  --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --- A-- --- --- --- --- --- --- --- --- --C --- --- --C --- --- --- --- --- --- --- ---
DR6a  --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --- A-- --- --- --- --- --- --- --- --- --C --- --- --C --- --- --- --- --- --- --- ---
DR7a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- ---
DR7b  --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- ---

DR1   CTG GAA ACA GTT CCT CGG AGT GGA GAG GTT TAC ACC TGC CAA GTG GAG CAC CCA AGT GTG ACG AGC CCT CTC ACA GTG GAA TGG AGA GCA CGG TCT GAA TCT GCA CAG AGC AAG ATG CTG
DR2a  --- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --- --- --- --- --- --- --- --- -A- --- --- --- --- --- --- --- --- ---
DR2b  --- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --- --C --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR3a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR3b  --A --- --- T-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --A --- --- --- --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- ---
DR4a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C C-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR5a  --- --- --- T-C --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR6a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR7a  --- --- --- --- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR7b  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --C A-- -T- --- --- --- --G --- C-- --- --T --- --- --- --- --- --- --- --- --- --- ---

DR1   AGT GGA GTC GGG GGC TTC GTG CTG GGC CTG CTC TTC CTT GGG GCC GGG CTG TTC ATC TAC TTC AGG AAT CAG AAA GGA CAC TCT GGA CTT CAG CCA ACA GGA TTC CTG AGC TGA
DR2a  --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --A --- --- --- --- -A- --- --- --- --G --- --- --- --- --C --- --- --- C-- G-- --- ---
DR2b  --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR3a  --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -G- --- --- --- --- ---
DR3b  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR4a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR5a  --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -G- --- --- --- --- ---
DR6a  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -G- --- --- --- --- ---
DR7a  --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- T-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
DR7b  --- --- --- --- --- --T --- --- --- --- --- --- --- --- A-A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- C-- T-- --- ---
```

FIG. 2. Nucleic acid sequence of the coding region of HLA-DR$_\beta$ cDNA clones from the haplotypes DR1 (LG2) (11), DR2 (PGF), DR3 (WT49), DR4 (WT51), DR5 (JVM), DR6 (APD), DR7 (LBF), DR8 (Madura), and DR9 (DKB).

much less obvious. It is clear that the evolutionary histories of these two families are different.

A further index of the relative evolutionary divergence of the different DR haplotypes has been obtained from studying the 3′-untranslated regions of these molecules. Nucleic acid hybridization data from Southern blots confirm the evolutionary pattern suggested by the DRw52 and DRw53 sequence data (Fig. 1). When a 3′-untranslated region probe obtained from the DR1 DR_β chain was used to hybridize to a Southern blot containing DNA from homozygous cell lines of DR types 1–7 followed by high stringency washes, this probe hybridized to the haplotypes DR1, -2, -3, -5, and -6 but failed to hybridize to the DR4 and DR7 (DRw53) haplotypes (Fig. 1*B*). This indicates that these five haplotypes are more similar to each other than they are to the DRw53-related haplotypes. The opposite experiment using a 3′-untranslated region probe for the DR_β 4a cDNA resulted in hybridization at high stringency to a single band in only the DRw53-related haplotypes DR4 and DR7 and no hybridization to the DRw52 or the related DR1 or DR2 haplotypes (Fig. 1*C*).

Comparison of sequences between the different DR_β alleles (Fig. 2) confirms the observation originally made in mouse class II β chains (21) and extended to human class II products that the majority of the polymorphism in these alleles is confined to the first domain and is localized to the amino acid residues 9–14, 25–38, and 67–74. These regions have been described as allelic hypervariable regions (21), to differentiate them from the degree of variation seen in immunoglobulin hypervariable regions (22). The hypervariable residues of immunoglobulins display much greater amounts of variation than is seen in HLA-DR_β residues. Variation at HLA-DR_β residues is in only four instances (residues 11, 13, 30, and 37) highly variable (i.e., more than four different amino acids) (Fig. 3). In addition, these regions are frequently shared between haplotypes (Fig. 3), further defining the characteristics of the allelic hypervariable regions.

Selection is certainly an important factor in fixing these polymorphisms in the population but the exact mechanisms by which such variation is established remains unclear. The allelic hypervariable regions may be the product of sequential single base-pair mutations followed by gene conversion (19) or homologous recombination events (23). Much of the polymorphism present among DR_β alleles within the hypervariable regions is seen as multiple base changes within any coding triplet. In fact, of the possible sequences seen at any triplet, half are multibase changes. This is four times the rate seen in the rest of the sequences.

It is difficult to provide conclusive evidence for true gene conversion events in man, but the sequences of these DR_β alleles support the suggestion that gene conversion-like events are responsible for the generation of much of class II polymorphism. The DR_β 5a allele provides one such example. The first allelic hypervariable region of this allele shares a sequence with the two other DRw52-associated alleles DR_β 6a and DR_β 3a. Its second allelic hypervariable region is identical to that found in the DR_β 2b allele, and its third allelic hypervariable region is shared by a member of yet another supertypic family, DR4 Dw10 (10). The DR_β 5a allele from the JVM line is, therefore, composed of a patchwork of sequences found in other alleles. The DR_β 6a allele is similarly of interest as it shares its first allelic hypervariable region with the DR_β 3a allele and its third allelic hypervariable region with DR_β 5a and DR4/Dw10. This particular Dw10 "cassette" seems to be extremely mobile and produces variation within DR types, which is recognized readily by T cells, but for which specific serological reagents have been difficult to obtain. The DR_β 8 sequence is also very similar to the DR_β 3a and DR_β 6a sequences in the first allelic hypervariable region making this portion of the molecule a potential candidate for DRw52 reactivity. The DR_β 8 sequence shares its second allelic hypervariable region with DR_β 2b and its third with DR_β 2a.

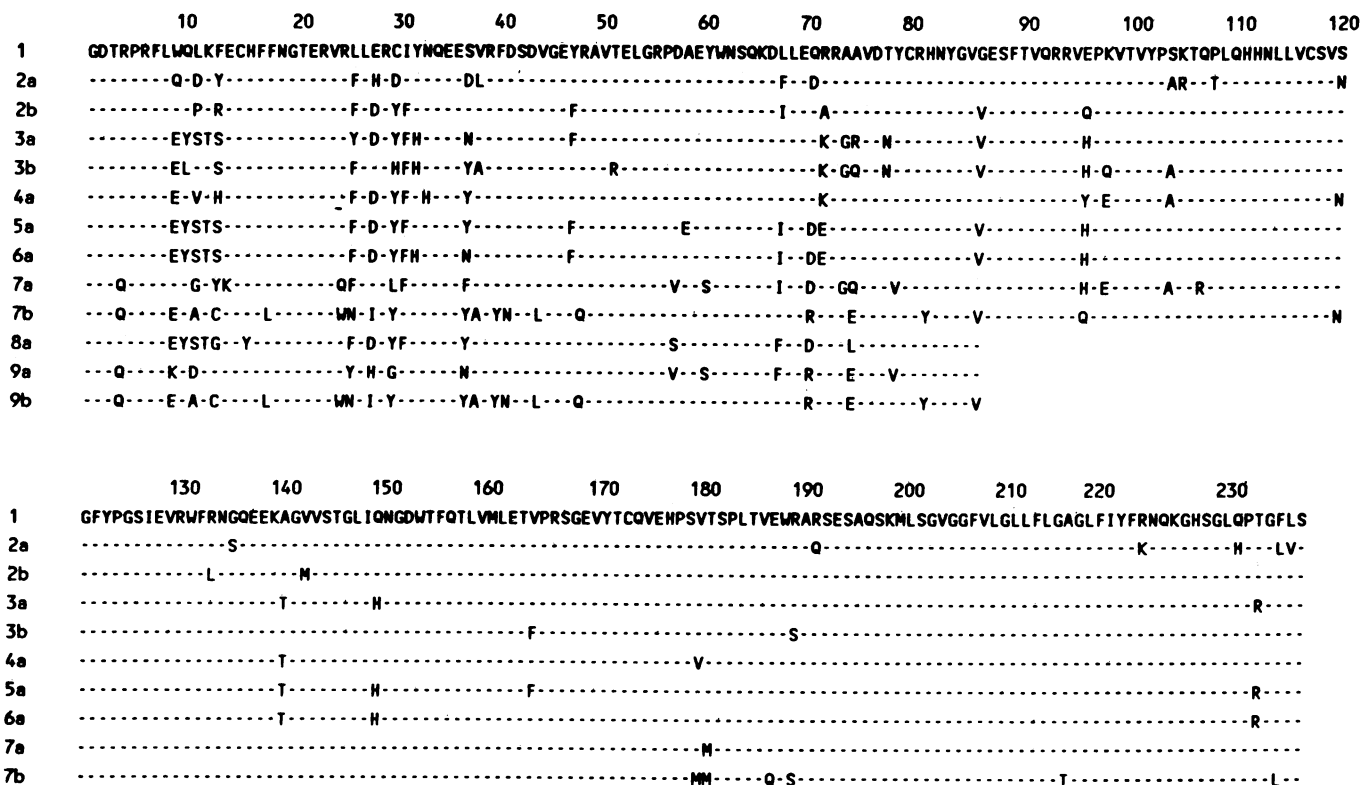

FIG. 3. Derived amino acid sequence (in the single-letter amino acid code) of the DR_β cDNA clones for the haplotypes DR1, -2, -3, -4, -5, -6, -7, -8, and -9 as determined by this study.

Other potential examples of gene conversion include the Arg-Arg-Arg-Ala-Glu sequence in the third hypervariable region (amino acids 70–74) that is shared by the DR_β 7b/9b (DRw53 related) alleles and the DR_β 9a allele, but no others. The DR_β 4a allele shares the sequence Phe-Leu-Asp-Arg-Tyr-Phe with DR_β 2b, DR_β 5a, DR_β 6a, and DR_β 8a in the second allelic hypervariable region (amino acids 26–31) and also shares almost the entire sequence of the third allelic hypervariable region with the DR1 allele. An identical match in this region is present between DR4/Dw14 (10) and DR1.

These data strongly imply that much of the variation among DR_β alleles arises from shuffling of these regions, both within and between the supertypic families. Although one can speculate that the most likely mechanism for such events is gene conversion, direct proof of this hypothesis is not available. More important than the mechanism is the conclusion from the analysis of these sequences that the total amount of variation among DR alleles is less than anticipated, since most serologically defined alleles share at least one and as many as three allelic hypervariable regions with other alleles.

Perhaps the most dramatic example of a potential gene conversion event does not involve shuffling of these regions, but only the substitution of the codon GTG for TAC in the DR_β7a and DR_β9a sequences at position 78. An identical GTG sequence can be found in the DQ_β chain at position 78, providing compelling evidence for a DQ_β to DR_β gene conversion event at this position (13). Of interest, apparent conversion events have occurred both among members of the same supertypic families and between members of different families. This example is one in which the genetic exchange seems to have occurred between two different subregions.

These sequences allow attribution of several sequences obtained from heterozygous cell lines to specific haplotypes. pIIβ3 (24) represents the DR_β 3a sequence from Raji cells, and DR_β1 (18) is the DR_β 6b sequence from a DR4,6 cell line. The DR5 sequence here is distinct from that published (25), and this is compatible with the atypical cellular typing obtained from the DR5 cell line JVM (26). The DR2 nucleic acid sequences are the first to be obtained and confirm DR2 protein sequence studies on the same line (27). The DR6 sequence from APD cells provides a full-length sequence from that haplotype, distinct from the DR_β 6b sequence (19) and sharing the first domain with the DR_β 6a allele (19). The DR_β 9a sequence completes the 5′-coding region from that allele (28) and the DR_β 7b allele is the first full-length sequence of the DRw53 (MT3)-associated allele. To our knowledge, the DR_β 8 sequence is the first to be published, whereas the DR_β 4a and the DR_β 7a sequences are identical to those described (28, 29).

These studies provide a base of sequence data from expressed genes that, because of their origin from homozygous consanguineous lines, can be attributed to single serologically defined haplotypes. The majority of the DR alleles found in the Caucasian population are included in this study. In addition to providing insights into the evolution and mechanisms of generation of polymorphism, these clones and sequences will permit the generation of allele-specific oligonucleotides and/or monoclonal antibody reagents. These will permit rapid and accurate analysis of the DR subregion on a population level. Finally, they provide a base-line set of sequences with which alleles from patients with the vast array of HLA-mediated diseases can be compared.

We are grateful to Kristina Karagianes for typing this manuscript. J.I.B. is a Fellow of the Alberta Heritage Fund for Medical Research, and J.A.T. is supported by the Science and Engineering Research Council. This research was supported by Grants CA-39069 and GM-35635 from the National Institutes of Health.

1. McDevitt, H. O. & Chinitz, A. (1969) *Science* **163**, 1207–1208.
2. Bell, J. I., Denney, D. W. & McDevitt, H. O. (1985) *Immunol. Rev.* **84**, 51–71.
3. Hardy, D. A., Bell, J. I., Long, E. O., Lindsten, T. & McDevitt, H. O. (1986) *Nature (London)* **323**, 453–455.
4. Spies, T., Sorrentino, R., Boss, J. M., Okada, K. & Strominger, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5165–5169.
5. Rollini, P., Mach, B. & Gorski, K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7197–7201.
6. Bach, F. H. & Voynow, N. K. (1966) *Science* **153**, 545.
7. van Rood, J. J., van Leeuwen, A., Keuning, J. J. & Blusse van Oud Alblas, A. (1975) *Tissue Antigens* **5**, 73–79.
8. Cohen-Hagenauer, O., Robbins, E., Massart, C., Busson, M., Deschamps, I., Hors, J., Lalouel, J.-M., Dausset, J. & Cohen, D. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3335–3339.
9. Bell, J. I., Denney, D., MacMurray, A., Foster, L., Watling, D. & McDevitt, H. O. (1986) *J. Immunol.*, in press.
10. Cairns, S., Curtsinger, J. M., Dahl, C. A., Freeman, S., Alter, B. J. & Bach, F. H. (1985) *Nature (London)* **317**, 166.
11. Bell, J. I., Estess, P., St. John, T., Saik, R. S., Watling, D. L., Erlich, H. A. & McDevitt, H. O. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3405–3409.
12. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25**, 263–269.
13. Boss, J. & Strominger, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5199–5203.
14. Sanger, F., Nicklem, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Bankier, A. T. & Barrell, B. G. (1983) in *Techniques in the Life Sciences*, ed. Flavell, R. A. (Elsevier, Shannon, Ireland) Vol. B508, pp. 1–34.
16. Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951–2961.
17. Gorski, J., Tosi, R., Strubin, M., Rabourdin-Combe, C. & Mach, B. (1985) *J. Exp. Med.* **162**, 105–116.
18. Long, E. O., Wake, C. T., Gorski, J. & Mach, B. (1983) *EMBO J.* **2**, 389–394.
19. Gorski, J. & Mach, B. (1986) *Nature (London)* **322**, 67–70.
20. Moriuchi, J., Moriuchi, T. & Silver, J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3420–3424.
21. Mengle-Gaw, L. & McDevitt, H. O. (1985) *Annu. Rev. Immunol.* **3**, 367–396.
22. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. (1983) *Sequences of Proteins of Immunological Interest* (U.S. Dept. of Health and Human Services, Washington, DC).
23. Lee, B. S., Rust, N. A., McMichael, A. J. & McDevitt, H. O. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4591–4595.
24. Gustafsson, K., Wiman, K., Emmoth, E., Larhammer, D., Bohme, J., Hyldig-Nielsen, J. J., Ronne, H., Peterson, P. & Rask, L. (1984) *EMBO J.* **3**, 1655–1661.
25. Didier, D. K., Schiffenbauer, J., Shuman, S., Abruzzini, L. F., Gorski, J., Watling, D. L., Tieber, V. L. & Schwartz, B. D. (1986) *J. Immunol.* **137**, 2627–2631.
26. Termijtelen, A., Khan, P. M., Shaw, S. & van Rood, J. J. (1983) *Immunogenetics* **18**, 503–512.
27. Kratzin, H., Gotz, H., Thinnes, F. P., Kruse, T., Barnikol, H. U., Wernet, P. & Hilschmann, N. (1986) in *HLA Class II Antigens*, eds. Solheim, B. G., Moller, E. & Ferrone, S. (Springer, Berlin), p. 49.
28. Gregersen, P. K., Moriuchui, T., Karr, R. W., Obata, F., Moriuchi, J., Maccari, J., Goldberg, D., Winchester, R. J. & Silver, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9149–9153.
29. Gregersen, P. K., Shen, M., Song, Q. L., Merryman, P., Degar, S., Seki, T., Maccari, J., Goldberg, D., Murphy, H., Schwenzer, J., Wang, C. Y., Winchester, R. J., Nepom, G. T. & Silver, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2642–2646.
30. Davis, R. W., Botstein, D. & Roth, J. R., (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).