Immunogenetics 29: 297–307, 1989

Immuno-genetics

Allelic variation in HLA-B and HLA-C sequences and the evolution of the *HLA-B* alleles

Heike Pohla, Wolfgang Kuon, Piotr Tabaczewski, Christa Doerner, and Elisabeth H. Weiss

Institute for Immunology, University of Munich, Goethestraße 31, D-8000 München 2, Federal Republic of Germany

Abstract. Several new *HLA-B (B8, B51, Bw62)*- and *HLA-C (Cw6, Cw7)*-specific genes were isolated either as genomic cosmid or cDNA clones to study the diversity of HLA antigens. The allele specificities were identified by sequence analysis in comparison with published HLA-B and -C sequences, by transfection experiments, and Southern and northern blot analysis using oligonucleotide probes. Comparison of the classical HLA-A, -B, and -C sequences reveals that allele-specific substitutions seem to be rare events. *HLA-B51* codes only for one allele-specific residue: arginine at position 81 located on the $\alpha 1$ helix, pointing toward the antigen binding site. *HLA-B8* contains an acidic substitution in amino acid position 9 on the first central β sheet which might affect antigen binding capacity, perhaps in combination with the rare replacement at position 67 (F) on the $\alpha 1$ helix. *HLA-B8* shows greatest homology to HLA-Bw42, -Bw41, -B7, and -Bw60 antigens, all of which lack the conserved restriction sites *Pst* I at position 180 and *Sac* I at position 131. Both sites associated with amino acid replacements seem to be genetic markers of an evolutionary split of the *HLA-B* alleles, which is also observed in the leader sequences. *HLA-Cw7* shows 98% sequence identity to the *JY328* gene. In general, the *HLA-C* alleles display lower levels of variability in the highly polymorphic regions of the $\alpha 1$ and $\alpha 2$ domains, and have more distinct patterns of locus-specific residues in the transmembrane and cytoplasmic domains. Thus we propose a more recent origin for the *HLA-C* locus.

Introduction

HLA (human leucocyte antigen) molecules are membrane glycoproteins found on the surface of nearly all nucleated cells which are formed by a noncovalent association of a polymorphic heavy chain (M_r 40 000–45 000), encoded within the major histocompatibility complex (MHC) on the short arm of chromosome 6, and an invariant light chain (M_r 12 000), beta-2 microglobulin (B2m), encoded on chromosome 15 (Goodfellow et al. 1975, Grey et al. 1973, Coligan et al. 1981). Each heavy chain is composed of three extracellular domains ($\alpha 1$, $\alpha 2$, $\alpha 3$) of about 90 amino acids, a hydrophobic transmembrane segment, and a hydrophilic cytoplasmic region. The products encoded by three gene loci (*HLA-A, -B,* and *-C*) serve as restriction elements for virus-specific and allospecific cytotoxic T lymphocytes (CTL) and are therefore major determinants in directing tissue graft rejection (Zinkernagel and Doherty 1979). Additional non-*HLA-A, -B,* and *-C* genes of unknown function map between the *HLA-B* and/or around the *HLA-A* locus (Koller et al. 1987). A characteristic feature of the classical class I products is their very high degree of polymorphism, mainly in the $\alpha 1$ and $\alpha 2$ domain, which occurs within a highly organized framework of conserved sequences. The $\alpha 3$ domain and B2m are relatively conserved and show amino acid sequence homology to immunoglobulin constant domains (Orr et al. 1979, Peterson et al. 1972, Trägardh et al. 1979, Smithies and Poulik 1972). The transmembrane and cytoplasmic regions contain the majority of locus-specific residues and reveal a higher degree of intralocus conservation which reflects evolutionary ancestry rather than positive selection (Güssow et al. 1987).

We have isolated additional *HLA* class I genes and present here the molecular structures of *HLA-B8, -B51, -Bw62, -Cw6,* and *-Cw7* in comparison to the *HLA* genes published to date.

Materials and methods

Isolation of the HLA-B and HLA-C clones. The HLA-B-specific cd2.6 (HLA-B27) cosmid (Weiss et al. 1985) and the allelic cd3.3 cosmid were isolated from a genomic library (cd) constructed from peripheral white blood cells of a healthy donor (HLA-type: A2, B27/B51, Cw2/Cw3).

Address correspondence and offprint requests to: E. H. Weiss

Two other HLA-B-specific clones pMF18 and pMF28, and the HLA-C-specific clone pMF17 were isolated from a cDNA library (MF) kindly provided by A. Hahn, prepared from RNA of a chronic T-cell leukemia (HLA-type: A1/A2, B8/Bw62, Cw3/Cw7). The HLA-C-specific clones pMS1 and pMS24 were identified in a second cDNA library (MS) constructed from RNA of thymocytes (HLA-type: A2/A30, B13, Cw6). The libraries were screened by differential hybridization using a full-length class I probe derived from a B27 cDNA clone, an HLA-B-specific probe derived from the 3′ untranslated region, and a 200 bp *Pvu* II × *Sac* I HLA-C-specific fragment spanning the region of amino acids 246–312, part of the α3 domain, and the transmembrane segment of the HLA-Cw1 cDNA (Szöts et al. 1986). The probes were labeled by the random primer method (Feinberg and Vogelstein 1983) with α-^{32}P-dATP (3000 Ci/mmol).

DNA sequencing. The DNA sequence analysis of M13 subclones was carried out by the dideoxy chain termination method (Sanger et al. 1977).

Southern and northern blot analysis. Southern blot analysis of genomic DNA from various *HLA* haplotypes was done according to standard protocols (Southern 1975). A 1350 bp *Sac* I fragment from the 5′ end of the *HLA-B27* gene (Weiss et al. 1985) was selected for hybridization. Slot blot analysis with isolated RNA was performed with the Minifold II SRC 072/0 as described by the supplier (Schleicher and Schuell, Dassel, FRG). Purification of the oligonucleotides (synthesized by Dr. R. Mertz, Genzentrum, Munich, FRG) and hybridization was done as described (Geliebter et al. 1986) with the modification that the probe was labeled with terminal desoxynucleotidyl transferase and α-^{32}P-dATP (Collins and Hunsaker 1985).

HLA haplotype of the cells used for the studies.

EBV-transformed B-cell lines. AS (A1, B8, Cw7); CoL (A2/A3, B18/B51, Cw5); DS (A2/A3, B35/B37, Cw4); FrDa (A?, B27/?, C?); Juso (A1/A2, B7/B8, C?); KR (A2, B44, Cw5); LG2 (A2, B27, Cw1); LiCa (A3, B7, Cw7/Cw8); ML (A2/A3, B7/B27, Cw2); MWi (A2, B18/B35, Cw4/Cw7); Ni (A3/A29, B35/Bw65, Cw4/Cw8); RL (A2/A23, B13/B27, Cw3/Cw6); Sweig (A29, B40*, C?); TY (A11, B35, Cw4); VL (A2/A3, B13/B18, Cw5/Cw6); WDV (A3, B38, C!).

The T-cell leukemia line MF. (A1/A2, B8/Bw62, Cw3/Cw7).

The peripheral blood lymphocytes. CD (A2, B27/B51, Cw2/Cw3); CW (A24/A28, B44/Bw60, Cw3/Cw5); DB (A2, B44/Bw41, C?); Eb (A3/A32, B7/B51, Cw1); EW (A2/A30, B13/Bw62, Cw3/Cw6); GS (A2, B49/Bw62, Cw4); JZ (A25/A30, B18/Bw41); KL (A2/A3, B7/B35, Cw4); MS (A2/A30, B13, Cw6); PH (A2/A26, Bw62/?, Cw3); and in addition K562, a human chronic myelogenous leukemia cell line (HLA class I negative) and SW480, a human colorectal adenocarcinoma cell line (A2, B8/B17, C?).

Results

Characterization of HLA-B- and -C-specific sequences. *HLA-B51* is contained in the cosmid cd3.3 isolated from the genomic library (cd). The restriction map of the cd 3.3 clone and the sequencing strategy are shown in Figure 1. Since the cosmid contains only one entire *HLA-*

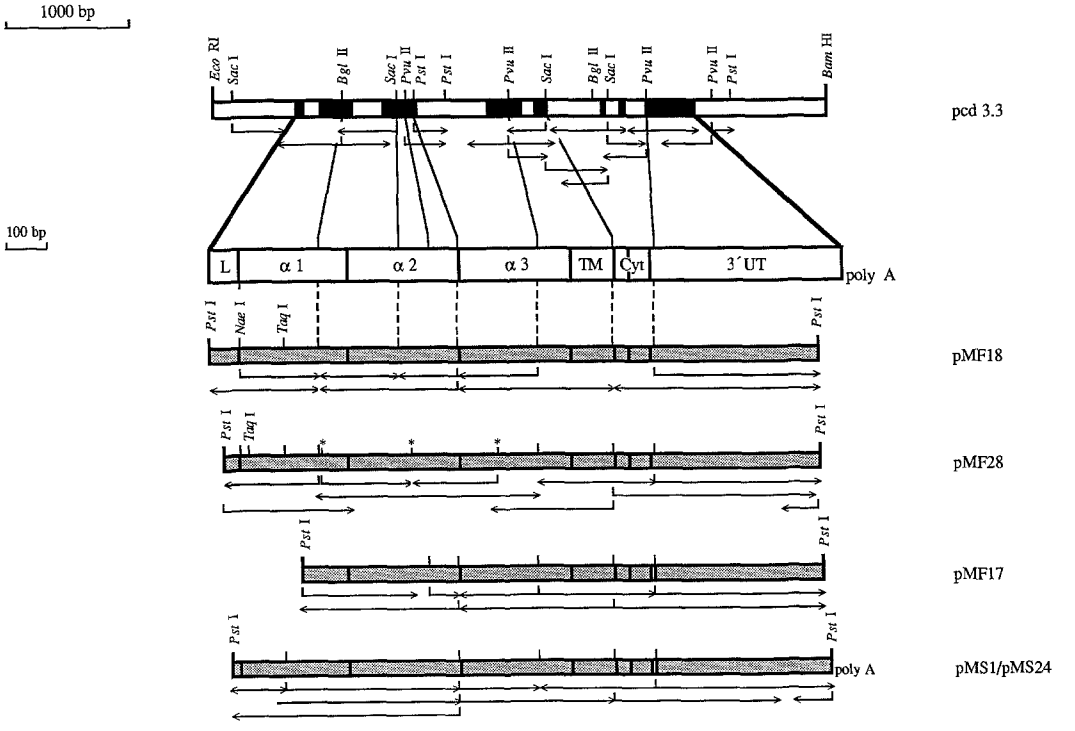


Fig. 1. Restriction map of the cosmid cd3.3 and sequencing strategy of the *HLA-B51* gene; the cDNA clones pMF18 (HLA-Bw62), pMF28 (HLA-B8), pMF17 (HLA-Cw7), and pMS1/pMS24 (HLA-Cw6). Restriction sites are shown in relation to the exons *(black boxes)*; the corresponding protein domains of a typical HLA-B heavy chain including the 3′ untranslated region are denoted by the following abbreviations: L, leader peptide; α1, α2, α3, extracellular domains; TM, transmembrane region; Cyt, cytoplasmic regions; 3′UT, 3′ untranslated region. Both HLA-C-specific clones contain a further segment representing the short, third cytoplasmic region (exon 8). The *Pst* I sites at the 5′and 3′ ends of the cDNA clones were generated during construction of the library. The *arrows* indicate the sequencing strategy. The *asterisks* indicate the restriction site *Sau* 3A needed for sequencing of additional fragments of pMF28. *Taq* I sites are denoted only for the cDNA clones

B-specific gene, this clone was used directly for DNA-mediated gene transfer. The analysis of the transfectant clones with the monoclonal antibody S4 (Johnson et al. 1987) demonstrated that cosmid cd3.3 encodes *HLA-B51* (data not shown).

HLA-B8 is encoded by the pMF28 clone, which hybridized with the HLA-B locus-specific probe. Figure 1 shows the restriction map of pMF28 and the sequencing strategy. Neither this clone nor the other cDNA clones possess the entire leader peptide and thus could not be used directly for transfection experiments to determine the HLA specificity. Partial sequence information of exons 5, 6, and 7 of the HLA-B-specific cDNA clone pHLA-1, isolated from the homozygous cell line LKT (HLA-A1/-B8; Ploegh et al. 1980), indicated that pMF28 encodes HLA-B8. The protein sequence of pMF28 is also identical in the three extracellular domains to that of the published B8 antigen cloned from the heterozygous lymphoblastoid cell line (LCL) 721 (Parham et al. 1988).

HLA-Bw62, the second allele of the heterozygous individual MF, must be encoded by the second clone, pMF18, identified with the HLA-B-specific probe. pMF18 reveals a restriction pattern distinct from the pMF 28 clone, as shown in Figure 1. To assign the pMF18 clone to *HLA-Bw62*, a specific oligonucleotide derived from the region of amino acids 42–48 in the α1 domain was synthesized; this sequence was subsequently found to be present in the HLA-B13 sequence (Zemmour et al. 1988). The Bw62/B13 oligomer was used to correlate the expression of the HLA-Bw62 antigen with the presence of mRNA shown by the slot blot analysis (Fig. 2). Specific bands were found only in those cells having Bw62 and/or B13.

Assignment of the HLA-Cw6 and Cw7 alleles. DNA sequence determination of the clone pMF17, pMS1, and pMS24, hybridizing with the HLA-C-specific probe, confirmed the typical transmembrane and cytoplasmic regions found only in HLA-C sequences (see Figs. 3 and 4). Furthermore, they showed the HLA-C-specific amino acid pattern in the α1 domain (amino acid 45, 66–76). pMF 17 differs from the published Cw3 sequence and thus should encode the HLA-Cw7 antigen, the second *HLA-C* allele of individual MF. The clones pMS1 and pMS24 are identical and very likely encode HLA-Cw6, since individual MS carries HLA-B13 and -Cw6 on both chromosomes. The parents of MS were HLA typed to confirm that MS inherited HLA-B13 and -Cw6 from both parents.

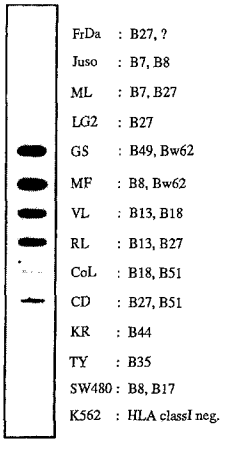


Fig. 2. RNA-slot blot analysis with a pMF18 (HLA-Bw62) oligonucleotide as the hybridization probe, derived from the sequence of amino acids 42–48 in the α1 domain: 5′ - GT CCG AGG ATG GCG CCC CGG −3′ (20 mer)

HLA class I sequence analysis. The amino acid sequences deduced from these HLA class I clones are compared with other published HLA antigen sequences (Fig. 4). The sequences of the known nonclassical class I loci JY328 (Srivastava et al. 1985), pHLA6.0 (Geraghty et al. 1987), HLA-E (LG2-C1), and the pseudogene pHLA12.4 (Malissen et al. 1982) are also included. In addition, the defective class I sequence, pLN-11A (Biro et al. 1983), was evaluated when locus-specific residues were assigned, although this sequence is not included in Figure 4. LG2-C1 is a full-length cDNA clone isolated from the LG2 cDNA library (Szöts et al. 1986, Pohla and Weiss 1986) corresponding to the RS5 sequence (Srivastava et al. 1987) and is identical to HLA-E (Koller et al. 1988) and JTW15 (Mizuno et al. 1988) except for one substitution in amino acid position 222 (R=AGG instead of E=GAG).

This sequence comparison reveals that *HLA-B51* contains one allele-specific amino acid: arginine, a basic residue at position 81 replaces the common nonpolar amino acids alanine or leucine of other genes, pointing toward the antigen binding site, according to the X-ray crystallographic structure analysis of the HLA-A2 molecule (Bjorkman et al. 1987a, b). The histidine present at position 171 instead of the highly conserved tyrosine is also found in *HLA-B14 (-Bw65)* and *-B18* (Parham et al. 1988). This histidine could also affect interactions between the MHC molecule and distinct processed antigens, for instance with the backbone of a peptide. As a member of the Bw4 subgroup, HLA-B51 shows the sequence 80(T), 82(L), and 83(R), postulated to be specific for the Bw4 epitope (Wan et al. 1986). The monoclonal antibody S4, which recognizes the HLA-B51 antigen, also reacts with additional *HLA-B* alleles *(Bw52, B13, B44, B45, Bw57)*. Analysis of these sequences discloses one epitope shared by the *HLA-B* alleles: asparagine at position 77 together with the Bw4 characteristic residues 80, 82, and 83, which may account for the antibody specificity.

The highest overall sequence identity of HLA-B8 was found with HLA-Bw42 and -Bw41 (Parham et al. 1988) and HLA-B7, -B40*, and -Bw60. With the exception of the *HLA-B40** mutant, all these antigens share identical substitutions at amino acids 131 (R) and 180 (E), which

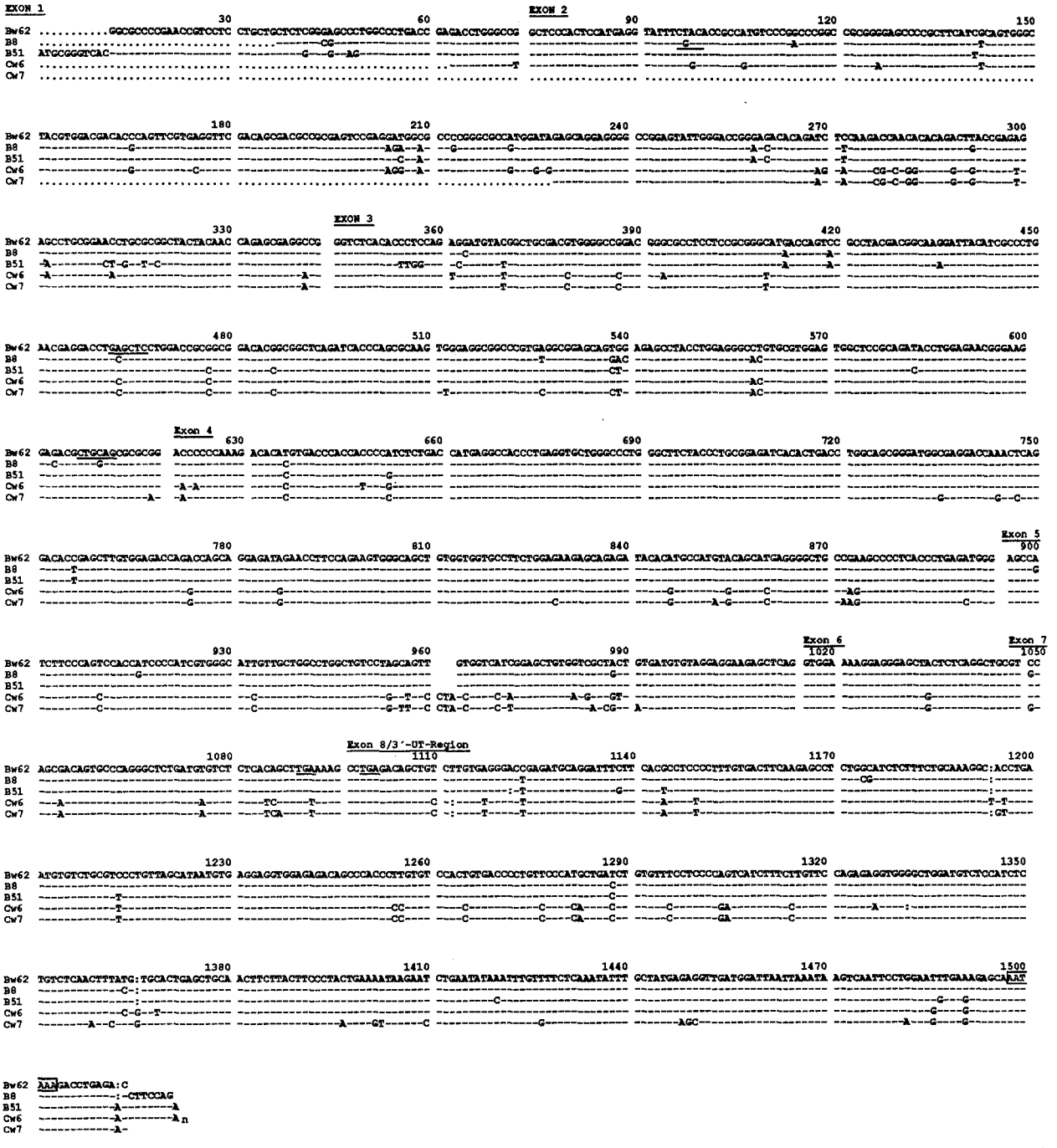


Fig. 3. Nucleotide sequence of HLA-Bw62 in comparison to HLA-B8, -B51, -Cw6 and -Cw7 sequences. Only the sequence of the exons and the 3′ untranslated region is shown. The translation termination codons and the restriction sites, *Taq* I, *Sac* I, and *Pst* I as mentioned in the Results and Discussion, are *underlined*; the polyadenylation signal is *boxed*

destroy the *Sac* I and *Pst* I restriction sites (see Fig. 3); they also all possess an identical replacement at position 177 (D). Southern blot analysis using a 1.35 kb *Sac* I fragment isolated from the 5′ end of the *HLA-B27* gene (500 bp 5′ of the start codon to amino acid 131; Weiss et al. 1985) as a hybridization probe, with *Sac* I-digested genomic DNA, revealed the 2.6 kb band in *HLA-B8*, *-B40**, and *Bw41* haplotypes and a 3.1 kb *Sac* I fragment in *HLA-B7* and *-Bw60* haplotypes which in addition lack a *Sac* I site at the end of exon 5 (Fig. 5). The latter two *HLA-B* alleles furthermore possess an identical substitution at position 178 (K). All *HLA-B* alleles coding for Ser at position 131 show the 1.35 kb band, identical to that of the probe (see lane LG2, for example). A weakly

COMPARISON OF THE AMINO ACID SEQUENCES

Leader

```
Bw62     M L V M A P R T V L L L L S G A L A L T E T W A
B8       . . . . . A - - - - - - - - -
B7       - - - - - A - - - - - - - - -

B27      - R - T - - - - L - - - - W - - V - - - - - - -
Bw47     - R - T - - - - L - - - - W - - V - - - - - - -
B44      . . . T - - - - L - - - - W - - V - - - - - - -
B13      - R - T - - - - L - - - - W - - V - - - - - - -
B51      - R - T - - - - - - - - - W - - V - - - - - - -
Bw58     - R - T - - - - - - - - - W - - V - - - - - - -

Cw1      - R - - - - - - L I - - - - - - - - - - - - - -
Cw2      - R - - - - - - L I - - - - - - - - - - - - - -
Cw3      - R - - - - - - L I - - - - - - - - - - - - - -
Cw6      . . . . . .
JY328    - - - - - - - - L - - - - L A P R - - - Q - - -

Aw24     - A - - G - - - L V - - - - - - - - - - Q - - -
A3     ♣ Q A - - - - - - L - - - - - - - - - - - Q - - -
Aw68     - A - - - - - - L V - - - - - - - - - - Q - - -
Aw69     - A - - - - - - L V - - - - - - - - - - Q - - -
A2       - A - - - - - - L V - - - - - - - - - - Q - - -

pHLA12.4 - V L - - - - - - L - - - - - - - - - - - Q - - -
HLA6.0   - V - - - - - - - L F - - - - - - - - T - - - - - -
LG2-C1   . . . . . - - - - - - - - - - E - - - - - Q - - -
```

♣ : M A R G D

α1 Domain

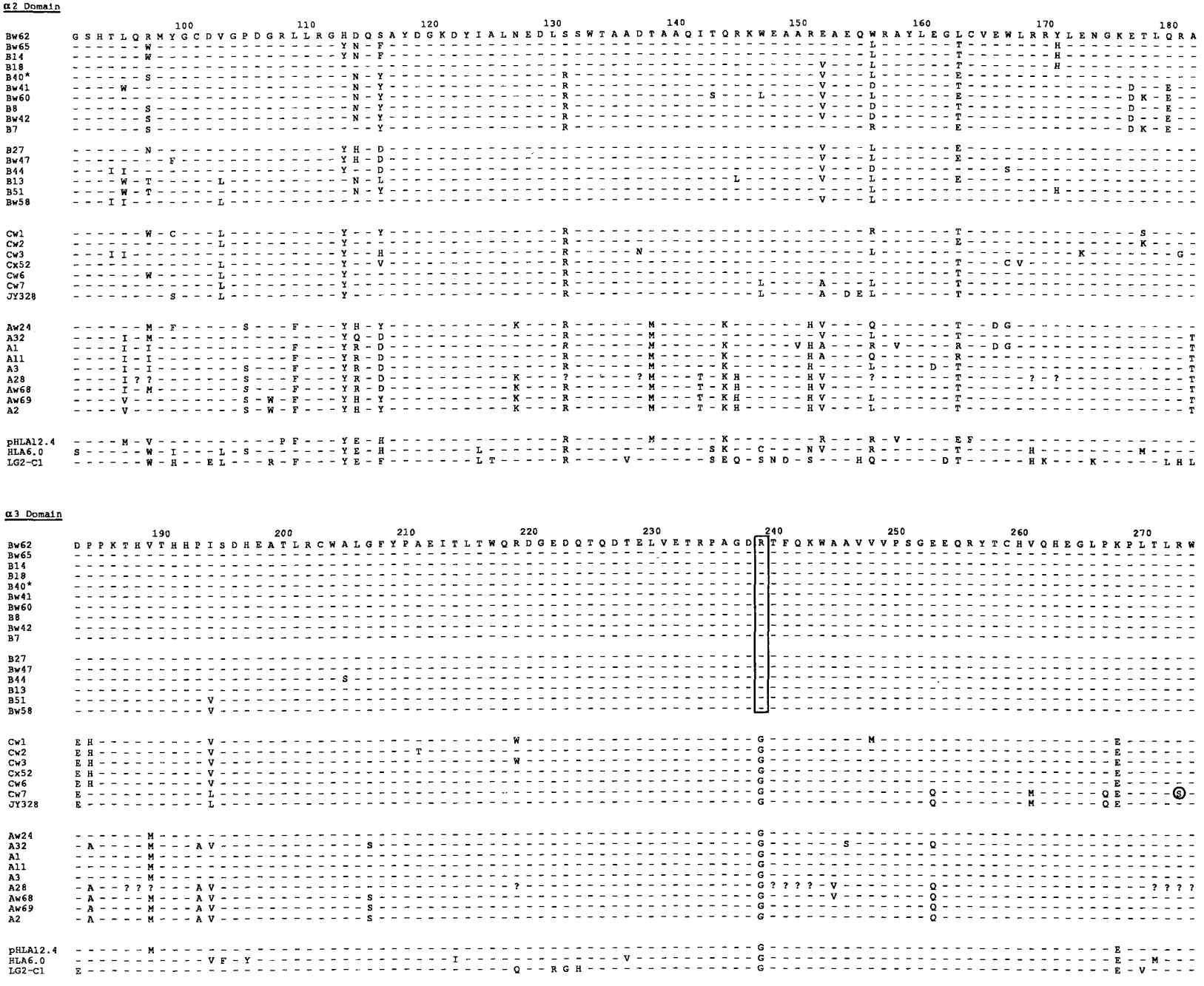
```
                    10                  20                  30                  40                  50                  60                  70                  80                  90
Bw62     G S H S M R Y F Y T A M S R P G R G E P R F I A V G Y V D D T Q F V R F D S D A A S P R M A P R A P W I E Q E G P E Y W D R E T Q I S K T N T Q T Y R E S L R N L R G Y Y N Q S E A
Bw65     - - - - - - - - - - V - - - - - - - - - S - - - - - - - - - - - - - - - - - - E E - - - - - - - - - - - - N - - - C - - - - - - D - - - - - - - - - - - - - - - - - -
B14      - - - - - - - - - S V - - - - - - - - - S - - - - - - - - - - - - - - - - - - E E - - - - - - - - - - - - N - - - C - - - - - - D - - - - - - - - - - - - - - - - - -
B18      - - - - - - - - H - S V - - - - - - - - - S - - - - - - G - - - - - - - - - - T E - - - - - - - - - - - - N - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
B40*     - - - - - - - - H - S V - - - - - - - - - T - - - - - - L - - - - - - T - - - K E - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Bw41     - - - - - - - - H - - - - - - - - - - - - T - - - - - - L - - - - - - T - - - K E - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Bw60     - - - - - - - - H - - - - - - - - - - - - T - - - - - - L - - - - - - T - - - K E - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
B8       - - - - - - - - D - - - - - - - - - - - - S - - - - - - - - - - - - - - - - - - E E - - - - - - - - - - - - N - - - F - - - - D - - - - - - - - - - - - - - - - - -
Bw42     - - - - - - - - - S V - - - - - - - - - - S - - - - - - - - - - - - - - - - - - E E - - - - - - - - - - - - N - - - Y - A Q A - - D - - - - - - - - - - - - - - - - - -
B7       - - - - - - - - - S V - - - - - - - - - - S - - - - - - - - - - - - - - - - - - E E - - - - - - - - - - - - N - - - Y - A Q A - - D - - - - - - - - - - - - - - - - - -

B27      - - - - - - - - H - S V - - - - - - - - - T - - - - - - - L - - - - - - - - - - E E - - - - - - - - - - - - - - - - C - A K A - - D - - D - - T - L R - - - - - - -
Bw47     - - - - - - - - - - - - - - - - - - - - - T - - - - - - - L - - - - - - T - - - K E - - - - - - - - - - - - - - - - - - - - - - - - D - - T - L R - - - - - - -
B44      - - - - - - - - - - - - - - - - - - - - - T - - - - - - - L - - - - - - T - - - K E - - - - - - - - - - - - - - - - - - - - - - - - N - - T A A R - - - - - - -
B13      - - - - - - - - - - - - - - - - - - - - - T - - - - - - - - - - - - - - T - - - - - - - - - - - - - - - - - - - - - - - - - - - - N - - T A L R - - - - - - -
B51      - - - - - - - - - - - - - - - - - - - - - T - - - - - - - - - - - - - - - - - - T E - - - - - - - - - - - - N - - - F - - - - - - - - - N - - T R L R - - - - - - -
Bw58     - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - T E - - - - - - - - - - - - G - - R N M - A S A - - - - - N - - I A L R - - - - - - -

Cw1      C - - - - K - - F - S V - - - - - - - - - - S - - - - - - - - - - - - - - - - - - G E - - - - V - - - - - - - - - - - - K Y - R Q A - - D - V - - - - - - - - - - - - -
Cw2      C - - - - - - - - - - V - - - S - - - - H - - - - - - - - - - - - - - - - - - - G E - - G R - V - - - - - - - - - - - - K Y - R Q A - - D - V N - - K - - - - - - - - - -
Cw3      - - - - - - - - C - - V - - - - - - - - H - - - - - - - - - - - - - - - - - - - G E - - - - - V - R K - - - - - - - - - K Y - P Q A - - D - V - - - - - - - - - - - - -
Cx52     . . . . . .
Cw6      C - - - - - - - C - - V - - - - - - - - - S - - - - - - - - - - - - - - - - - - G E - - - - - V - - - - - - - - - - - - K Y - R Q A - A D - V N - - K - - - - - - - - - D D
Cw7      . . . . . .
JY328    - - - - - - - - D - - V - - - - - - - - - S - - - - - - - - - - - - - - - - - - G E - - - - - V - - - - - - - - - - - - K Y - R Q A - A D - V - - - - - - - - - - - - D D

Aw24     - - - - - - - - S - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - E - - G K V - A H S - - D - - N - - I A L R - - - - - - -
A32      - - - - - - - - F - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - Q - - R N V - A H S - - D - - - - - I A L R - - - - - - -
A1       - - - - - - - - - - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q K - E - - - - - - - - - - - - - - Q - - R N M - A H S - - D - A N - G T - - - - - - - - - D
A11      - - - - - - - - - - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - Q - - R N V - A Q S - - D - V D - G T - - - - - - - - - D D
A3       - - - - - - - - F - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - Q - - R N V - A Q S - - D - V D - G T - - - - - - - - - -
A28      ? - - - - - - - - - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - N - - R N V - A Q S - - D - V D - G T - - - - - - - - - -
Aw68     - - - - - - - - - - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - N - - R N V - - Q S - - D - V D - G T - - - - - - - - - -
Aw69     - - - - - - - - - - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - N - - R N V - - Q S - - D - V D - G T - - - - - - - - - -
A2       - - - - - - - - F - S V - - - - - - - - - - - - - - - - - - - - - - - - - - - - Q - - E - - - - - - - - - - - - - - G - - R K V - A H S - - H - V D - G T - - - - - - - - - -

pHLA12.4 R - - - - - - - - - T - - - - - A - - - - - - - S - - - - - - - - - - - - - - - - - - D - - - - E E - - - - - - M - R - - - - K - - - - - N - - - C - A Q A - - E - - N - - I A L R - - - - - - G
HLA6.0   - - - - - - - - - - - - - - S A - V - - - - - - - - - - - - - - - - - - M - - - - - - - - - - S - C - - - - E - - - - - - V - - - - - - - - - E E - - - R N T - A H A - - D - M N - Q T - - - - - - - - - -
LG2-C1   - - - - L K - - H - S V - - - - - - - - - - - - - - S - - - - - - - - - - - - - - N - - - - - - - V - - - - - - M - - - - - S - - - - - - - - - - - R S A R D T A - I F - V N - - T - - R - - - - - - - - -
```


α2 Domain
α3 Domain
Bw62
Bw65
B14
B18
B40*
Bw41
Bw60
B8
Bw42
B7
B27
Bw47
B44
B13
B51
Bw58
Cw1
Cw2
Cw3
Cx52
Cw6
Cw7
JY328
Aw24
A32
A1
A11
A3
A28
Aw68
Aw69
A2
pHLA12.4
HLA6.0
LG2-C1

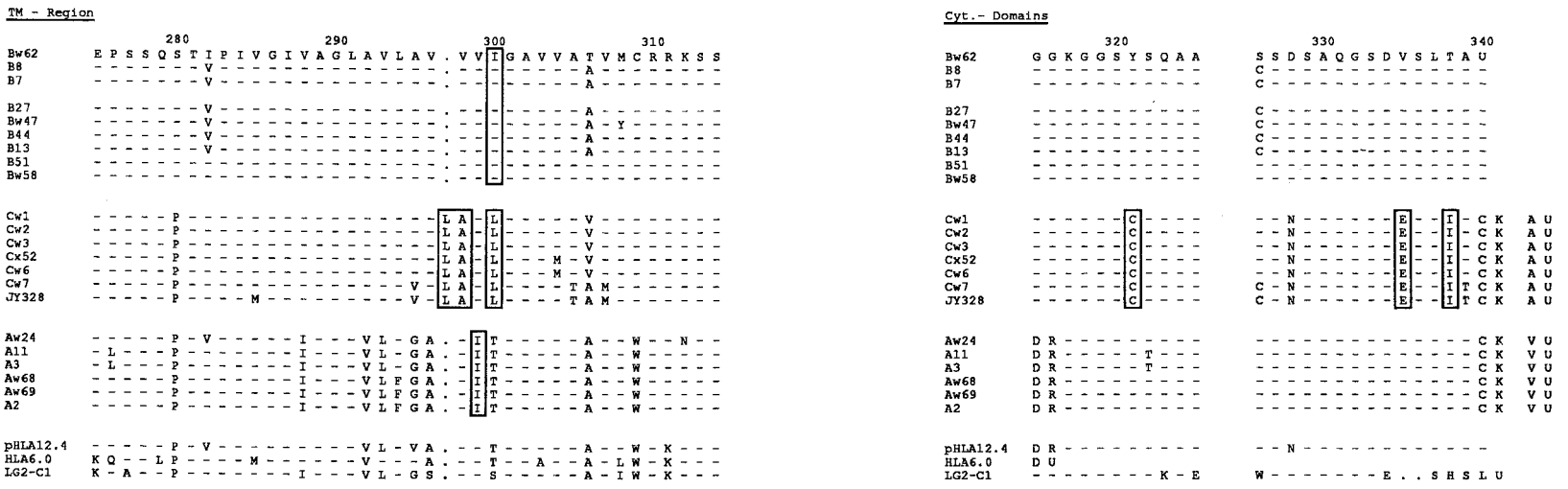


Fig. 4. Amino acid sequence of HLA-Bw62 in comparison to the other HLA antigens. The standard one-letter amino acid code is used. A *dash* indicates a residue identical to the Bw62 sequence, and unknown residues or deletions are represented by *question marks* or *points*. The first three amino acids of the HLA-Bw62 sequence are deduced from the HLA-B7 molecule. Discussed residues of HLA-B8, -B51, -Cw6, and -Cw7 are indicated by a *circle*, and locus-specific residues are *boxed*. The remaining sequences were obtained from the following references: Bw65, B14, B18, Bw41, Bw42 (Parham et al. 1988); B40*, Bw60 (Ways et al. 1987); B7 (Biro et al. 1983); B27 (Weiss et al. 1985, Szöts et al. 1986); B8 (Pohla and Weiss 1986; Parham et al. 1988); B13, Bw47 (Zemmour et al. 1988); B44 (Kottman et al. 1986); Bw58 (Ways et al. 1985); Cw1, Cw2 (Güssow et al. 1987); Cw3 (Sodoyer et al. 1984); Cx52 (Takata et al. 1988); JY328 (Srivastava et al. 1985); Aw24 (N'Guyen et al. 1985); A32 (Wan et al. 1986); A1 (Parham et al. 1988); A11 (Cowan et al. 1987); A3 (Strachan et al. 1984); A28 (Lopez de Castro et al. 1982); Aw68, Aw69 (Holmes et al. 1987); A2 (Koller and Orr 1985); pHLA12.4 (Malissen et al. 1982); HLA6.0 (Geraghty et al. 1987) and LG2-C1 (Pohla and Weiss 1986)

hybridizing band of 3.1 kb is also present in the WDV DNA. Since a 1.35 kb *Sac* I fragment gives a strong hybridization signal, this most likely represents B38, while the former fragment must be derived from another class I sequence in the WDV genome. In the case of *Pst* I × *Eco* RI double-digested DNA (data not shown), the 1.35 kb *Sac* I probe detects a strongly hybridizing 1.9 kb *Pst* I fragment specific for the *HLA-B7, -B8, -Bw41,* and *-Bw60* haplotypes lacking the *Pst* I site at the end of exon 3 instead of a 1.65 kb *Pst* I fragment, which is present if glutamine replaces glutamic acid at position 180. HLA-B8 contains an aspartic acid on the first β strand of the α1 domain at position 9, which has been defined as a potential ligand for an antigenic peptide. Interestingly, this rare acidic substitution is located on the same side of the groove as the highly polymorphic residue 67 (HLA-B8: F) and might join in antigen binding. At the nucleotide level, position 9 is associated with an additional *Taq* I restriction site (see Figs. 1 and 3).

HLA-Bw62 is characterized by a high degree of similarity in the first domain with the HLA-B13 sequence. The major differences lie in the cluster of amino acids 77–83 encompassing the Bw6 epitope. HLA-Bw62, in the first half of the α1 domain, is identical to the sequence of the HLA-B51 and the HLA-Bw58 molecules. In this case, the antigens share serological crossreactivity. The sequence comparison also shows that the HLA-Bw62 antigen has sequence identity with HLA-B40*, -Bw60, -Bw41, -B13, -B44, and -Bw47 between amino acids 62 and 76. Zemmour and co-workers (1988) proposed this linear sequence to be responsible for the serological crossreactivity between HLA-B13 and -Bw47 and as a possible epitope for the crossreactivity observed with molecules of the HLA-B12- and HLA-B40*-related groups. Since the same sequence is also preserved in the HLA-Bw62 antigen which does not display any crossreactivity with the above-mentioned specificities, this region alone cannot explain the serological patterns.

The Cw7 sequence contains one allele-specific residue at position 273 (S) and is virtually identical (98%) to the JY328 sequence (Srivastava et al. 1985), but differs from the other *HLA-C* alleles (90–93%) as a result of several unique substitutions in the conserved α3 domain, transmembrane and cytoplasmic regions (amino acids 194, 261, 267, 305, 307, 326, 339; shown in Fig. 4). The high sequence similarity explains the Cw7 reactivity of the JY328 transfectants (Duceman et al. 1986).

Discussion

To study the diversity of HLA antigens, several new *HLA-B-* and *HLA-C*-specific genes were isolated and compared with previously published sequences. When non-HLA-A/-B/-C class I sequences are included in such a comparison, only a few locus-specific residues can be

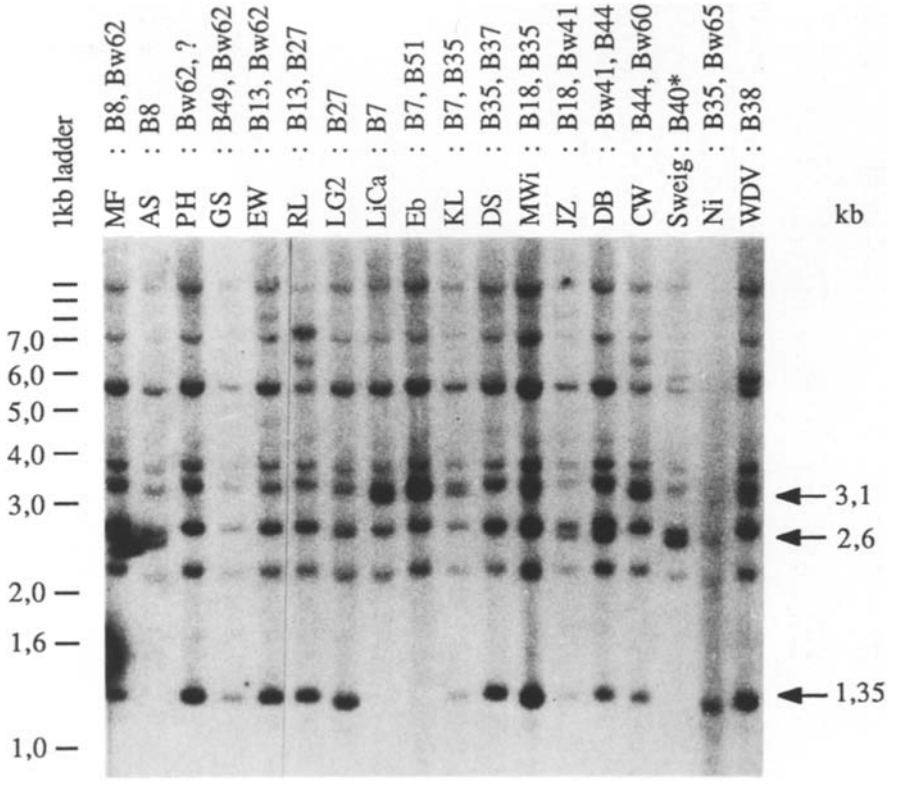


Fig. 5. Southern blot analysis of *Sac* I-digested genomic DNA isolated from donors of various haplotypes (see Materials and methods for HLA types). A 1.35 kb *Sac* I fragment (500 bp 5′ of the start codon to amino acid 131 of the *HLA-B27* gene; Weiss et al. 1985) was used as the hybridization probe. The genomic DNA of lane Ni shows slight degradation

identified. These include not only the HLA-A-specific residues, amino acids 71 (S) and 299 (I), and the HLA-B-specific residues, amino acids 239 (R) and 300 (I), but also several HLA-C locus-specific residues: 45 (G), 297 (L), 298 (A), 300 (L), 321 (C), 335 (E), and 338 (I), and in addition the eighth exon encoding an alanine residue.

This sequence comparison also confirms that allele-specific substitutions are rare events and probably the result of point mutations. Half of the amino acid replacements are localized within a framework of conserved sequences and not in the polymorphic regions. Thus, individual alleles are characterized by a cluster of amino acid substitutions, and a unique constellation of polymorphic residues affects the recognition of antibodies and the T-cell receptor.

As was pointed out previously for the HLA-A and HLA-B molecules, polymorphic amino acids display locus-specific distributions (Parham et al. 1988). Figure 6 analyzes the variability of HLA-C molecules, as revealed for eight known HLA-C sequences, including Cw2.2 (Parham et al. 1988), JY328 (Srivastava et al. 1985), and Cx52 (Takata et al. 1988). Although the number of sequences is still limited, this analysis reveals that only two positions (amino acids 9, 116) for HLA-C exhibit more than three different amino acids. Thus, the *HLA-C* alleles are characterized by a somewhat lower level of variability in the regions of the $\alpha 1$ and $\alpha 2$ domain which are highly polymorphic in *HLA-A* and *-B*. The highly polymorphic region at the end of the $\alpha 1$ domain in *HLA-A* and *-B* is rather invariant in the *HLA-C* alleles. Together with the more distinct pattern of locus-specific residues in the transmembrane and cytoplasmic domains, one could propose a more recent origin for the *HLA-C* locus. The variability seen in $\alpha 3$ is higher for the *HLA-C* locus than for the HLA-A and HLA-B molecules. This is accounted for mainly by the HLA-Cw7 and JY328 sequences which differ at five positions from the other *HLA-C* alleles.

In an attempt to find restriction sites and sequence similarities which might reflect evolutionary relationships among the *HLA-B* alleles, we used several single-copy probes derived from the *HLA-B* locus to investigate a large number of haplotypes (Weiss et al. 1988). Using gene flanking probes, polymorphisms in the gene flanking regions which correspond with sequence similarities in the coding regions were generally not found. However, some patterns might be interpreted in this direction. For HLA-Bw62, -B51, and -Bw58 we found identity in the transmembrane and cytoplasmic regions even at the nucleotide level; only one nucleotide substitution was found in the first half of exon 2, and minor differences are present in the 3′ untranslated region.

Interestingly, two restriction sites were found, which are present in more conserved positions when all class I sequences were compared, and which divide the *HLA-B* genes into two groups. The larger group of alleles contains the restriction sites *Sac* I and *Pst* I as genetic markers for serine at position 131 and glutamine at position 180. The alleles *HLA-B7, -B8, Bw41, -Bw42,* and *-Bw60* are members of the second group. They lack these restriction sites and encode an additional identical residue (D) at position 177. One exception is the HLA-B40* antigen which was postulated to have lost the epitope encompassing amino acids 177 and 180 by gene conversion, a result which supports this evolutionary relationship of *HLA-B* alleles. The relationship is also observed when comparing the leader sequences of the *HLA-B7* and *-B8* genes (only

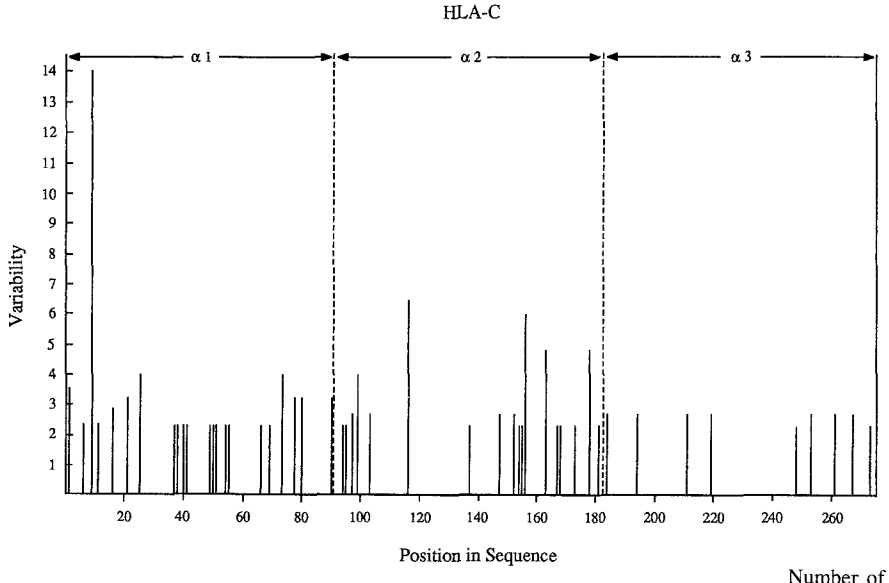


Fig. 6. Variability at different amino acid positions for the three extracellular domains of the eight HLA-C molecules including Cw2.2 (Parham et al. 1988), JY328 (Srivastava et al. 1985), and Cx52 (Takata et al. 1988). Cw2.2 only differs from Cw2.1 at position 49/50 (A/P in place of G/R). The protein sequences were analyzed as described by Wu and Kabat (1970).

$$\text{Variability} = \frac{\text{Number of different amino acids at a given position}}{\text{Frequency of the most common amino acid at that position}}$$

protein sequences of the mature antigens have been published for the other alleles of this group, see Fig. 4). Similarly, the leader peptides of the HLA-A and HLA-C molecules share a characteristic amino acid pattern. Surprisingly, the two recently published HLA-B homologous sequences from chimpanzee also contain the same set of amino acid substitutions as the uncommon serine at position 131, glutamic acid at 177, and glutamine at 180, present in the one group of *HLA-B* alleles (Lawlor et al. 1988, Mayer et al. 1988). This pattern of similarity leads to the conclusion that this divergence in an ancestral -B sequence arose before separation of the human and chimpanzee species (Klein 1987, Mayer et al. 1988).

Acknowledgments. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 217). We are grateful to Drs. E. D. Albert (Kinderpoliklinik, Munich) and R. Wank for HLA typing. We acknowledge the help of Dr. Mertz (Genzentrum, Munich) in synthesizing the oligonucleotides. Furthermore, we thank Dr. A. Hahn for providing the T-CLL cDNA library, Dr. J. Johnson for the mAb S4, Dr. D. Schendel for helpful discussions, M. Lang for technical assistance, and S. Förster for typing the manuscript.

References

Biro, P. A., Pan, J., Sood, A. K., Kole, R., Reddy, V. B., and Weissman, S. M.: Sequences of human repetitive DNA, non-α-globin genes, and major histocompatibility locus genes. Part III: The major Histocompatibility complex. *Cold Spring Harbor Symp Quant Biol 42:* 1082–1086, 1983

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., and Wiley, D. C.: Structure of the human class I histocompatibility antigen, HLA-A2. *Nature 329:* 506–512, 1987a

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett,, W. S., Strominger, J. L., and Wiley, D. C.: The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature 329:* 512–518, 1987b

Coligan, J. E., Kindt, T. J., Uehara, H., Martinko, J., and Nathenson, S. G.: Primary structure of a murine transplantation antigen. *Nature 291:* 35–39, 1981

Collins, M. L. and Hunsaker, W. R.: Improved hybridization assays employing tailed oligonucleotide probes: a direct comparison which 5′-end-labeled oligonucleotide probes and nick-translated plasmid probes. *Anal Biochem 151:* 211–224, 1985

Cowan, E. P., Jelachich, M. L., Biddison, W. E., and Coligan, J. E.: DNA sequence of HLA-A11: remarkable homology with HLA-A3 allows identification of residues involved in epitopes recognized by antibodies and T cells. *Immunogenetics 25:* 241–250, 1987

Duceman, B. W., Ness, D., Rende, R., Chorney, M. J., Srivastava, R., Greenspan, D. S., Pan, J., Weissman, S. M., and Grumet, F. C.: *HLA-JY328*: Mapping studies and expression of a polymorphic HLA class I gene. *Immunogenetics 23:* 90–99, 1986

Feinberg, A. P. and Vogelstein, B.:A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem 132:* 6–13, 1983

Geliebter, J., Zeff, R. A., Schulze, D. H., Pease, L. R., Weiss, E. H., Mellor, A. L., and Nathenson, S. G.: Interaction between $K^b$and Q4 gene sequences generates the K^{bm6} mutation. *Mol Cell Biol 6:* 645–652, 1986

Geraghty, D. E., Koller, B. H., and Orr, H. T.: A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. *Proc Natl Acad Sci USA 84:* 9145–9149, 1987

Goodfellow, P. N., Jones, E. A., van Heyningen, V., Solomon, E., Bobrow, M., Miggiano, V., and Bodmer, W. F.: The β_2-microglobulin gene is in chromosome 15 and not in the *HLA* region. *Nature 254:* 267–269, 1975

Grey, H. M., Kubo, R. T., Colon, S. M., Poulik, M. D., Cresswell, P., Springer, T., Turner, M. J., and Strominger, J. L.: The small subunit of HLA antigens is β_2-microglobulin. *J Exp Med 138:* 1608–1612, 1973

Güssow, D., Rein, R. S., Meijer, I., de Hoog, W., Seemann, G. H. A., Hochstenbach, F. M., and Ploegh, H. L.: Isolation, expression and the primary structure of *HLA-Cw1* and *HLA-Cw2* genes: Evolutionary aspects. *Immunogenetics 25:* 313–322, 1987

Holmes, N., Ennis, P., Wan, A. M., Denney, D. W., and Parham, P.: Multiple genetic mechanisms have contributed to the generation of the HLA-A2/A28 family of class I MHC molecules. *J Immunol 139:* 936–941, 1987

Johnson, J. P., Contag, I., and Wank, R.: The isolation and characterization of murine monoclonal antibodies directed to polymorphic epitopes on HLA antigens. *Hybridoma 6:* 17–27, 1987

Klein, J.: Origin of major histocompatibility complex polymorphisms: the trans-species-hypothesis. *Hum Immunol 19:* 155–162, 1987

Koller, B. H. and Orr, H.-T.: Cloning and complete sequence of an *HLA-A2* gene: Analysis of two *HLA-A* alleles at the nucleotide level. *J Immunol 134:* 2727–2733, 1985

Koller, B. H., Geraghty, D., Orr, H. T., Shimizu, Y., and DeMars, R.: Organization of the human class I major histocompatibility complex genes. *Immunol Res 6:* 1–10, 1987

Koller, B. H., Geraghty, D. E., DeMars, R., and Orr, H. T.: *HLA-E:* A novel HLA class I gene expressed in resting T lymphocytes. *J Immunol 141:* 897–904, 1988

Kottman, A. H., Seemann, G. H. A., Güssow, D. H., and Roos, M. H.: DNA sequence of the coding region of the *HLA-B44* gene. *Immunogenetics 23:* 396–400, 1986

Lawlor, D. A., Ward, F. E., Ennis, P. H., Jackson, A. P., and Parham, P.: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature 335:* 268–271, 1988

Lopez de Castro, J. A., Strominger, J. L., Strong, D. M., and Orr, H. T.: Structure of crossreactive human histocompatibility antigens HLA-A28 and HLA-A2: Possible implications for the generation of HLA polymorphism. *Proc Natl Acad Sci USA 79:* 3813–3817, 1982

Malissen, M., Malissen, B., and Jordan, B. R.: Exon/intron organization and complete nucleotide sequence of an *HLA* gene. *Proc Natl Acad Sci USA 79:* 893–897, 1982

Mayer, W. E., Jonker, M., Klein, D., Ivanyi, P., van Seventer, G., and Klein, J.: Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J 7:* 2765–2774, 1988

Mizuno, S., Trapani, J. A., Koller, B. H., Dupont, B., and Yang, S. Y.: Isolation and nucleotide sequence of a cDNA clone encoding a novel *HLA* class I gene. *J Immunol 140:* 4024–4030, 1988

N'Guyen, C., Sodoyer, R., Trucy, J., Strachan, T., and Jordan, B. R.: The *HLA-Aw24* gene: Sequence, surroundings and comparison with the *HLA-A2* and *HLA-A3* genes. *Immunogenetics 21:* 479–489, 1985

Orr, H. T., Lancet, D., Robb, R., Lopez de Castro, J., and Strominger, J. L.: The heavy chain of human histocompatibility antigen HLA-B7 contains an immunoglobulin-like region. *Nature 282:* 266–270, 1979

Parham, P., Lomen, C. E., Lawlor, D. A., Ways, J. P., Holmes, N., Coppin, H. L., Salter, R. D., Wan, A. M., and Ennis, P. D.: Nature of polymorphism in HLA-A, -B and -C molecules. *Proc Natl Acad Sci USA 85:* 4005–4009, 1988

Peterson, P. A., Cunningham, B. A., Berggard, I., and Edelman,

G. M.: β_2-microglobulin: A free immunoglobulin domain. *Proc Natl Acad Sci USA 69:* 1697–1701, 1972

Ploegh, H. L., Orr, H. T., and Strominger, J. L.: Molecular cloning of a human histocompatibility antigen cDNA fragment. *Proc Natl Acad Sci USA 77:* 6081–6085, 1980

Pohla, H. and Weiss, E. H.: The organization of *HLA* class I genes: Sequencing of *HLA-B8, -Bw62* genes, a new HLA-C specific clone and identification of functional non-*HLA-A, -B -C* class I genes (Abstract) *Immunobiology 173:* 273–274, 1986

Sanger, F., Nicklen, S., and Coulson, A. R.: DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA 74:* 5463–5469, 1977

Smithies, O. and Poulik, M. D.: Initiation of protein synthesis at an unusual position in an immunoglobulin gene? *Science 175:* 187–189, 1972

Sodoyer, R., Damotte, M., Delovitch, T. L., Trucy, J., Jordan, B. R., and Strachan, T.: Complete nucleotide sequence of a gene encoding a functional human class I histocompatibility antigen (*HLA-Cw3*). *EMBO J 3:* 879–885, 1984

Southern, E.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol 98:* 503–517, 1975

Srivastava, R., Duceman, B. W., Biro, P. A., Sood, A. K., and Weissman, S. M.: Molecular organization of the class I genes of human major histocompatibility complex. *Immunol Rev 84:* 93–113, 1985

Srivastava, R., Chorney, M. J., Lawrence, S. K., Pan, J., Smith, Z., Smith, C. L., and Weissman, S. M. N.: Structure, expression and molecular mapping of a divergent member of the class I HLA gene family. *Proc Natl Acad Sci USA 84:* 4224–4228, 1987

Strachan, T., Sodoyer, R., Damotte, M., and Jordan, B. R.: Complete nucleotide sequence of a functional class I HLA gene, *HLA-A3*: Implications for the evolution of *HLA* genes. *EMBO J 3:* 887–894, 1984

Szöts, H., Riethmüller, G., Weiss, E. H., and Meo, T.: Complete sequence of HLA-B27 cDNA identified through the characterization of structural markers unique to the HLA-A, -B and -C allelic series. *Proc Natl Acad Sci USA 83:* 1428–1432, 1986

Takata, H., Inoko, H., Ando, A., Haranaka, M., Watanabe, B., Tsuji, K., and Iri, H.: Cloning and analysis of HLA class I cDNA encoding a new HLA-C specificity Cx52. *Immunogenetics 28:* 265–270, 1988

Trägardh, L., Rask, L., Wiman, K., Fohlman, J., and Peterson, P. A.: Amino acid sequence of an immunoglobulin-like HLA antigen heavy chain domain. *Proc Natl Acad Sci USA 76:* 5839–5842, 1979

Wan, A. M., Ennis, P. D., Parham, P., and Holmes, N.: The primary structure of *HLA-A32* suggests a region involved in formation of the Bw4/Bw6 epitopes. *J Immunol 137:* 3671–3674, 1986

Ways, J. P., Coppin, H. L., and Parham, P.: The complete primary structure of *HLA-Bw58. J Biol Chem 260:* 11924–11933, 1985

Ways, J. P., Lawlor, D. A., Wan, A. M., and Parham, P.: A transposable epitope of HLA-B7, -B40 molecules. *Immunogenetics 25:* 323–328, 1987

Weiss, E. H., Kuon, W., Dörner, C., Lang, M., and Riethmüller, G.: Organization, sequence and expression of the *HLA-B27* gene: A molecular approach to analyze HLA and disease associations. *Immunobiology 170:* 367–380, 1985

Weiss, E. H., Blömer, K., Dörner, C., Kuon, W., Lang, M., Pohla, H., Schattenkirchner, M., and Riethmüller, G.: Molecular biology of the *HLA-B7* locus. *Br J Rheumatol 27* (suppl II): 12–18, 1988

Wu, T. T. and Kabat, E. A.: An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med 132:* 211–250, 1970

Zemmour, J., Ennis, P. D., Parham, P., and Dupont, B.: Comparison of the structure of *HLA-Bw47* to *HLA-B13* and its relationship to 21-hydroxylase deficiency. *Immunogenetics 27:* 281–287, 1988

Zinkernagel, R. M. and Doherty, P. C.: MHC-restricted cytotoxic T cells: Studies on the biological role of polymorphic major transplantation antigens determining T-cell restriction-specificity, function and responsiveness. *Adv Immunol 27:* 51–77, 1979

Received November 15, 1988; revised version received January 12, 1989

Note added in proof: Recently two Cw7 subtypes in linkage disequilibrium with either B7 or B8 have been identified (Hajek-Rosenmayr et al. 1988, Meeting Abstract). Therefore, pMF17 should encode the Cw7.2 variant, whereas JY328 might correspond to the Cw7.1 variant.