Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6289–6292, October 1983
Genetics

Analysis of HLA class I genes with restriction endonuclease fragments: Implications for polymorphism of the human major histocompatibility complex

(histocompatibility antigens/gene family/gene conversion)

DANIEL COHEN, PASCALE PAUL, MARIE-PIERRE FONT, ODILE COHEN, BRIGITTE SAYAGH, AGNES MARCADET, MARC BUSSON, GUY MAHOUY, HOWARD M. CANN*, AND JEAN DAUSSET†

Unité de Recherches sur l'Immuno-Génétique de la Transplantation Humaine (Institut National de la Santé et de la Recherche Médicale U.93), Institut de Recherches sur les Maladies du Sang, Hôpital Saint Louis, 75010 Paris, France

Contributed by Jean Dausset, July 11, 1983

ABSTRACT **Cellular DNA from HLA-typed individuals was digested with the restriction endonucleases *Hin*dIII, *Eco*RV, and *Eco*RI. The separated restriction endonuclease fragments were hybridized with a HLA class I cDNA probe by using the Southern transfer technique. Digestion of cellular DNA with *Hin*dIII generated 22 restriction endonuclease fragments, 11 of which showed polymorphism for presence or absence in a population sample. With *Eco*RV, 13 fragments were identified; 6 showed polymorphism. *Eco*RI generated 11 fragments, of which 1 was polymorphic. Of these 18 polymorphic fragments generated by the three restriction endonucleases, each of 5 was found to be positively associated with one allele of the HLA-A or -B allelic series (HLA-Aw24, -B8, -B15, -Bw35, and -B40). One fragment was positively associated with two HLA-A series alleles (HLA-A1 and -A11). Another fragment was positively associated with five HLA-B series alleles (HLA-B5, -B7, -B14, -Bw16, and -Bw35) and one fragment was positively associated with alleles at two loci (HLA-B14 and -Cw5). The serologically defined allele HLA-Aw24 was associated with two polymorphic fragments, one association showing a positive correlation and the other a negative correlation. Each informative family studied thus far has shown segregation of the restriction fragment with the associated serologically defined allele. The fragments associated with serologically defined alleles occurred in the population sample studied at low or moderate frequencies. The remaining polymorphic fragments occur at high frequency, suggesting that class I genes not serologially detected show less polymorphism than serologically defined class I genes.**

The HLA chromosomal region, the human major histocompatibility complex (MHC), located in a 2–3 × 10^3-kilobase-pair (kb) segment of DNA on the short arm of chromosome 6, is a cluster of at least three gene families coding for cell surface proteins (class I and class II loci) and several complement factors (class III loci) (1, 2). Serologic and biochemical (phenotypic) analysis of the gene products of class I loci has demonstrated at least three polymorphic loci, *HLA-A*, *-B*, and *-C*, each with a series of alleles coding for a series of M_r 45,000 alloantigen polypeptide chains (2). Restriction endonuclease analysis of human cellular DNA by using the Southern transfer technique has demonstrated a complex pattern of restriction fragments that hybridize with a cloned cDNA class I gene, compatible with the interpretation that considerably more class I loci exist than are found with phenotypic analysis (3–5).

We have found considerable polymorphism of class I restriction endonuclease fragments with various restriction enzymes. We call a polymorphic restriction fragment—i.e., one that is generated from the cellular DNA of some individuals but not from that of others—an *allogenotope*. We have reported correlation and segregation of an allogenotope, generated by the restriction enzyme *Eco*RV, with the serologically defined allele HLA-B8, suggesting that restriction fragment-length polymorphism (RFLP) of class I sequences will coincide with phenotypically defined class I polymorphisms (4). Here we report additional correlations between class I allogenotopes and serologically defined class I HLA alleles.

MATERIALS AND METHODS

Techniques for extraction and restriction endonuclease digestion of human cellular DNA as well as electrophoresis, transfer, and hybridization of restriction endonuclease fragments have been described (4). After hybridization, the membranes were washed at 60°C with 0.3 M sodium chloride/0.03 M sodium citrate, pH 7 (four times, 45 min each), with 30 mM sodium chloride/3 mM sodium citrate (once, 45 min), and finally with 15 mM sodium chloride/1.5 mM sodium citrate (once, 45 min). The restriction endonucleases—*Hin*dIII (Boehringer Mannheim and Amersham), *Eco*RV, and *Eco*RI (Boehringer Mannheim)—were used according to manufacturers' recommendations. The probe used for these studies was the 1.4-kb cDNA clone containing most of the coding sequence for a HLA-B gene, most likely HLA-B7 isolated by Sood *et al.* (6).

RESULTS AND DISCUSSION

Allogenotopes That Correlate and Segregate with Serologically Defined HLA-A Alleles. Cellular DNAs from 68 unrelated HLA-typed individuals, whose complete haplotypes were known, were digested with *Hin*dIII, and the restriction endonuclease fragments carrying HLA class I sequences were visualized by using the Southern technique (7). The autoradiograms of the patterns of restriction fragments that hybridized with the HLA class I probe (stringent conditions of washing) showed 20–22 bands, ranging in size from approximately 36 to 1.7 kb (Fig. 1). Of these bands, 11 were polymorphic fragments—i.e., 11 allogenotopes were noted.

A positive correlation was noted between a 4.8-kb fragment (Fig. 1) and the antigens HLA-A1 and HLA-A11. This allo-

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

Abbreviations: kb, kilobase pair(s); RFLP, restriction fragment-length polymorphism; MHC, major histocompatibility complex.

* Present address: Departments of Pediatrics and Genetics, Stanford University Medical Center, Stanford, CA 94305.
† To whom reprint requests should be addressed.

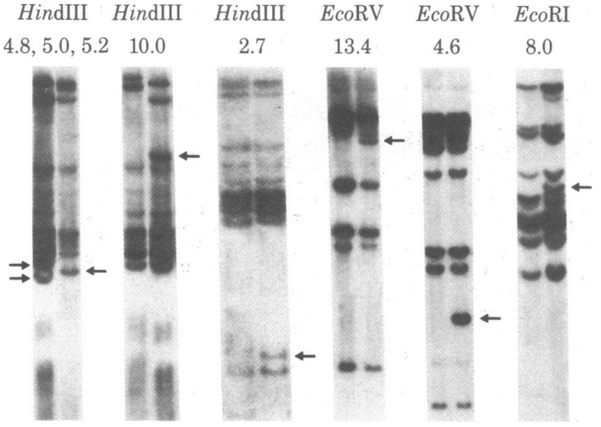

FIG. 1. For each allogenotope (shown by an arrow), a positive and negative pattern is shown. The restriction enzyme used and the size (in kb) of the allogenotope are indicated.

genotope, noted for 23 individuals, was generated from DNA from all 15 HLA-A1 and 7 HLA-A11 individuals tested and from 1 individual who carried neither HLA-A1 nor -A11 ($r = 0.97$) (Fig. 2 *Left*, allogenotope a).

A 5.0-kb fragment (Fig. 1) was associated with the antigen HLA-Aw24. DNA from 14 of the 15 HLA-Aw24 individuals generated this fragment, and it was also noted for one -A1, -Aw23 individual ($r = 0.91$) (Fig. 2 *Left*, allogenotope b). These two *Hin*dIII fragments (4.8 and 5.0 kb) might either be derived from the genes encoding the correlated antigens or from genes in linkage disequilibrium. The exceptions (Fig. 2) found for each correlation favor the second hypothesis. However, these exceptions could reflect structural variations not (or not yet) detectable immunologically. It is of interest that HLA-A1 and -A11, strongly crossreacting antigens (8), are correlated with the same allogenotope. This fragment is lost from DNA of deletion mutant cell lines that have lost the expression of HLA-A1 (3). Thus, the population and cell line data suggest that the 4.8-kb fragment is derived from the genes encoding HLA-A1 and -A11, rather than from a gene in linkage disequilibrium.

A 5.2-kb fragment (Fig. 1) was negatively correlated with both HLA-Aw24 ($r = -0.64$) and the two antigens HLA-A1 and -A11 ($r = -0.43$). This fragment, which occurred frequently (78%), was not found in DNA from the individuals heterozygous for HLA-Aw24/A1 or HLA-Aw24/A11 ($r = -0.61$) (Fig. 2 *Left*, allogenotope c). These results suggest that the alleles HLA-Aw24, -A1, and -A11, or genes in linkage disequilibrium, could be derived from a gene associated with the 5.2-kb *Hin*dIII fragment. Thus, this allogenotope, because it occurs at high frequency, might be related to several other HLA-A alleles.

Family studies have demonstrated that each of these three *Hin*dIII allogenotopes segregates with HLA.

Allogenotopes That Correlate and Segregate with Serologically Defined HLA-B Alleles. *Hin*dIII generated two allogenotopes, 10.0- and 2.7-kb fragments (Fig. 1), which were found to be associated with the serologically defined HLA-B series alleles HLA-B40 and HLA-B15, respectively. The 10.0-kb fragment was associated with HLA-B40 ($r = 0.77$); most HLA-B40 haplotypes that carry this fragment (four of five) also carry the allele for HLA-Cw3 (Fig. 2 *Center*, allogenotope h). The 2.7-kb fragment was noted for all nine of the HLA-B15 individuals ($r = 0.74$) and for six additional individuals (Fig. 2 *Center*, allogenotope f).

Cellular DNA from 66 unrelated individuals was digested by the restriction endonuclease *Eco*RV and probed for class I sequences. Of the 12 or 13 bands detected, 6 were allogenotopes, and, of these, 3 correlated with alleles of the *HLA-B* locus. We have already described (4) the 8.6-kb *Eco*RV fragment that correlated with HLA-B8 (Fig. 2 *Center*, allogenotope g). In this study, this allogenotope, found in all 13 of the HLA-B8 individuals, was also noted in one of two HLA-B17 individuals ($r = 0.95$) in the population sample analyzed.

A 13.4-kb allogenotope (Fig. 1) was associated with the serologically defined HLA-B alleles -B5 (-Bw51), -B7, -B14, -Bw16 (-Bw38, -Bw39), and -Bw35 ($r = 0.91$). Two individuals carrying HLA-Bw51 and HLA-Bw39, respectively, did not carry this fragment, but it was noted in one individual with the phenotype HLA-B18, -B40 (Fig. 2 *Center*, allogenotope d). The antigen HLA-Bw35 was also correlated with another *Eco*RV fragment. This 4.6-kb fragment (Fig. 1), noted for all nine of the HLA-Bw35 individuals ($r = 0.94$), was also noted for the only HLA-Bw52 individual in the population sample (Fig. 2 *Center*, allogenotope e). The 13.4-kb fragment and the 4.6-kb fragment could both be derived from the HLA-Bw35 gene or a gene in linkage disequilibrium. This would require the presence of an intrageneic *Eco*RV site. Alternatively, this double association may also reflect the existence of two closely linked loci in the *B* region. The latter hypothesis is supported by the existence in the mouse of clusters of genes associated with *K*, *D*, and *L* loci in the H-2 complex (9). Such clusters of genes in humans are difficult to detect by cloning experiments because class I genes appear to be widely spaced in the MHC (10).

A previously published family study demonstrated segregation of the 8.6-kb fragment with HLA-B8 (4). Each of the other allogenotopes that correlated with one of the HLA-B series alleles segregated with the associated allele(s) (except for the *Hin*dIII 10.0-kb fragment correlating with HLA-B40; informative families are not yet available).

An Allogenotope That Correlates with Alleles of the *HLA-B* and *-C* Loci. Analysis of DNA from 54 individuals revealed only one *Eco*RI allogenotope among the 11 bands detected with the conditions of hybridization and washing used in this study. This 8.0-kb fragment (Fig. 1) was found to correlate with HLA-B14 and HLA-Cw5 ($r = 0.75$). This allogenotope was noted for 11 of 15 HLA-Cw5 individuals; it was also noted in all 7 HLA-B14 individuals in the population sample. This allogenotope was found for two individuals who were neither HLA-Cw5 nor HLA-B14 (Fig. 2 *Right*, allogenotope i).

Although these alleles, HLA-Cw5 and -B14, have not been consistently noted to be associated (i.e., in linkage disequilibrium), it is possible that the 8.0-kb fragment might represent a segment from a gene located between the *HLA-B* and *-C* loci, in linkage disequilibrium with these two alleles. With the small map distance between the *HLA-B* and *-C* loci, we would expect alleles at these loci to show linkage disequilibrium if each were associated with a gene between them on the chromosome. Thus, we favor an alternative hypothesis, which suggests that this allogenotope reflects a sequence common to both of these serologically defined alleles. This 8.0-kb fragment segregates with HLA in family studies.

Other Allogenotopes Are Not Correlated with Classical HLA Antigens. This study has found 18 allogenotopes generated by *Hin*dIII, *Eco*RV, and *Eco*RI and detected by a cDNA probe containing an almost full-length coding sequence for a HLA class I gene. Eight of these allogenotopes were positively correlated with serologically defined class I antigens. Of the other 10 fragments, only 1 (*Hin*dIII, 5.2 kb) was negatively correlated with a HLA antigen. The other 9 allogenotopes were not correlated positively or negatively with any serologically defined class I alleles, and they occurred at relatively high frequency (60–90%). The population sample of this study is rather small, and negative as well as positive correlations might be missed. Never-

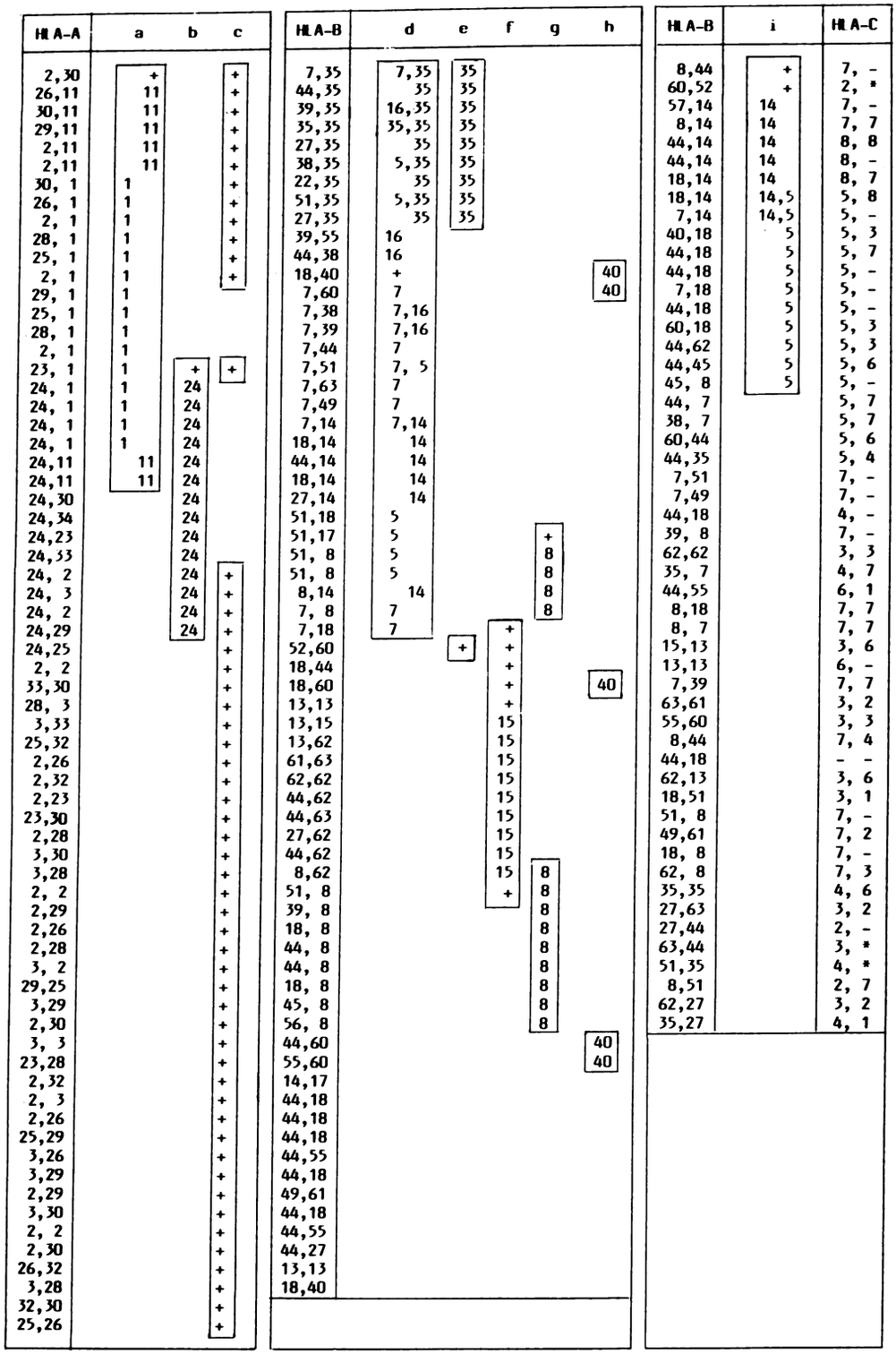

FIG. 2. Correlation between HLA-A, -B, and -C alleles and various allogenotopes. Unrelated genotyped individuals are listed vertically and their HLA-A, -B, or -C types are given: − means a "blank" allele not yet serologically defined; * means a possibility of homozygosity—i.e., homozygous or heterozygous for a blank allele; an allogenotope noted for an individual is indicated in the box, either by + or by the HLA allele(s) number with which the allogenotope is positively correlated. In the case of a split of a HLA antigen, the number corresponding to the split is given in the HLA type, and the number of the original antigen is given in the box. (*Left*) Locus *HLA-A*: allogenotope a (*Hin*dIII, 4.8 kb) correlated with HLA-A1 and -A11 ($r = 0.97$); allogenotope b (*Hin*dIII, 5.0 kb) correlated with HLA-Aw24 ($r = 0.91$); allogenotope c (*Hin*dIII, 5.2 kb) correlated negatively with HLA-Aw24 ($r = -0.64$). (*Center*) Locus *HLA-B*: allogenotope d (*Eco*RV, 13.4 kb) correlated with HLA-B5, -7, -14, -w16, and -w35 ($r = 0.91$); allogenotope e (*Eco*RV, 4.6 kb) correlated with HLA-Bw35 ($r = 0.94$) [note that the individuals (except one) with this allogenotope also show the allogenotope d]; allogenotope f (*Hin*dIII, 2.7 kb) correlated with HLA-B15 ($r = 0.74$); allogenotope g (*Eco*RV; 8.6 kb) correlated with HLA-B8 ($r = 0.95$); allogenotope h (*Hin*dIII, 10.0 kb) correlated with HLA-B40 ($r = 0.77$) (four of five individuals carry HLA-Cw3). (*Right*) Loci *HLA-B* and *-C*: allogenotope i (*Eco*RI, 8.0 kb) correlated with HLA-B14 and -Cw5 ($r = 0.75$).

theless, some of these allogenotopes could be associated with nonserologically defined class I genes. The number of class I human genes has been estimated to be ≈20, which is considerably more than the three serologically defined (*HLA-A*, *-B*, and *-C*) loci. Pseudogenes (11, 12) and the gene equivalent of the murine Qa or T1a cluster could account for this difference. If these genes had the extensive polymorphism of the serologically defined genes, we would expect to find some class I allogenotopes occurring at low frequencies that do not correlate with HLA antigens. However, class I allogenotopes with these characteristics were not observed in this study, suggesting that these other class I loci are less polymorphic than the serologically defined HLA loci. This could explain why these genes are not defined by immunogenetic techniques involving alloantibodies.

Intrageneic or Extrageneic Location of Class I Polymorphic Sequences Detected by Restriction Fragment-Length Variation. A restriction site can be present or absent at a given position in allelic genes or in their flanking sequences. A "conserved" site is present at the same position in all alleles of a series. "Unconserved" sites determine allogenotopes. In an allelic series, these restriction fragments can be flanked by 2, 1, or even 0 unconserved sites. In the last case, an allogenotope flanked by two conserved sites is created when, at least in one allele of the series, an unconserved site is located between the two conserved sites. The presence of an intrageneic unconserved site flanked by two conserved sites will determine two allogenotopes. For class I sequences, these two allogenotopes would be detected by the nearly complete cDNA probe used in this study and would have the same distribution in the population—i.e., they would be completely positively correlated. Such associated fragments were not observed in this study. Thus, the unconserved sites detected here might more often be extrageneic than intrageneic, suggesting that the flanking sequences of human class I alleles are more divergent than intrageneic sequences.

A gene conversion mechanism has been proposed for generating polymorphism in a multigene family (13). Two duplicated parental gene sequences form a heteroduplex that is randomly corrected, leading to the generation of a new allele. Thus, the more divergent the duplicated genes are, the more the new allele diverges from the parental alleles. All of the restriction enzymes used in this study and others (4, 5) generate numerous allogenotopes. This probably reflects a high degree of allelic divergence and is compatible with a gene conversion-like mechanism for generating polymorphism in the MHC (14–16).

We are grateful to F. Rougeon and R. Ollo for their helpful discussion. This work was supported by grants from the Institut National de la Santé et de la Recherche Médicale and from the Direction des Recherches d'Etudes et Techniques du Ministère des Armées (81/166).

1. Dausset, J. (1981) *Science* **213**, 1469–1474.
2. Ploegh, H. L., Orr, H. T. & Strominger, J. L. (1981) *Cell* **24**, 287–299.
3. Orr, H. T., Bach, F. H., Ploegh, H. L., Strominger, J. L., Kavathas, P. & Demars, R. (1982) *Nature (London)* **296**, 454–456.
4. Cann, H. M., Ascanio, L., Paul, P., Marcadet, A., Dausset, J. & Cohen D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1665–1668.
5. Ascanio, L., Paul, P., Marcadet, A., Mahouy, G., Fradelizi, D., Cohen, D. & Dausset, J. (1982) *C. R. Hebd. Seances Acad. Sci. Paris* **295**, 433–437.
6. Sood, A. K., Pereira, D. & Weissman, S. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 616–620.
7. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
8. Colombani, J., Colombani, M. & Dausset, J. (1970) in *Histocompatibility Testing 1970*, ed. Terasaki, P. I. (Munksgaard, Copenhagen), pp. 79–92.
9. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28**, 489–498.
10. Malissen, M., Damrotte, M., Birnbaum, D., Trucy, J. & Jordon, B. R. (1982) *Gene* **20**, 485–489.
11. Steinmetz, M., Moore, K. W., Frelinger, J. G., Sher, B. T., Shen, F., Boyse, E. & Hood, L. (1981) *Cell* **25**, 683–692.
12. Malissen, M., Malissen, B. & Jordon, B. R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 893–897.
13. Ollo, R. & Rougeon, F. (1983) *Cell* **32**, 515–523.
14. Lopez de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3813–3817.
15. Pease, L. R., Schulze, D. H., Pfaffenbach, G. M. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 242–246.
16. Weiss, E. H., Mellor, A., Golden, L., Fahrner, K., Simpson, E., Hurst, J. & Flavell, R. A. (1983) *Nature (London)* **301**, 671–674.