Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 5433–5436, August 1985
Genetics

Exuberant restriction fragment length polymorphism associated with the DQ α-chain gene and the DX α-chain gene

(major histocompatibility complex)

ISABELLE LE GALL*, AGNES MARCADET*, MARIE-PIERRE FONT*, CHARLES AUFFRAY†, JACK STROMINGER‡, JEAN-MARC LALOUEL§, JEAN DAUSSET§¶, AND DANIEL COHEN*

*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; †Institut d'embryologie (Centre National de la Recherche Scientifique and Collège de France), 99 bis Avenue de la Belle Gabrielle, 94130 Nogent sur Marne, France; ‡Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138; and §Centre d'Etude du Polymorphisme Humain (C.E.P.H.), Collège de France, 3 rue d'Ulm, 75005 Paris, France

Contributed by Jean Dausset, April 9, 1985

ABSTRACT **Cellular DNAs from individuals of 23 families were digested with five restriction endonucleases (*Pvu* II, *Eco*RI, *Hin*dIII, *Bam*HI, and *Eco*RV) and then probed with a DX α-chain gene probe. Seventeen allogenotopes were observed, each of which could be assigned to a serologically defined haplotype by noting its segregation in families. Six sets of allogenotopes forming allelic series were noted. In comparison with restriction maps of the *DQα* and the *DXα* regions, each of these series has been assigned to the *DQα* or the *DXα* locus. Allogenotopes of the four *DQα* series constitute three clusters correlating with the supertypic groups of class II histocompatibility antigens DQw1 (DR1, DR2, and DRw6), DRw53 (DR4, DR7, and DR9), and DR3 plus DR5 plus DR8. These 13 *DQα* fragments constitute 22 different patterns. The two *DXα* series constitute two clusters, one of which is not found to be correlated strongly with DR specificities, whereas the other is correlated loosely ($r = 0.45$) with DR5 and DR7. This absence of strong linkage disequilibrium between the *DXα* series and the DR series contrasts with the *DQα* series and suggests a recombination point between *DQα* and *DXα* loci.**

The *HLA-D* region, the segment of human chromosome 6 extending from the *HLA-B* region to the enzyme locus glyoxylase, includes genes encoding α and β polypeptide chains for several class II histocompatibility products such as DP (SB), DQ (DC, DS, MB), and DR. The polymorphism of each of these loci can be detected both phenotypically, by immunological methods, and at the gene level, by Southern transfer techniques using appropriate DNA probes (1). A polymorphism occurring in the distribution of restriction sites revealed by such patterns of fragments is called a restriction fragment length polymorphism, or RFLP (2). We shall refer to a fragment of such a system, not invariably found in all individuals tested, as an allogenotope.

Two DQ α-chain genes have been found (3–6). These two genes, named *DQα* and *DXα*, are very similar (3). Each is linked to a single DQ β-chain gene, namely *DQβ* and *DXβ*, respectively (6). The distance between *DQα* and *DQβ* genes and *DXα* and *DXβ* genes is not yet known. RFLP associated with *DQα* and *DXα* has previously been reported (4, 5, 7, 8) for small samples of individuals. In this paper, we report an extensive study including 22 families analyzed with five restriction enzymes.

METHODS

Detection of Restriction Fragments by the Southern Technique. DNA was extracted from peripheral blood leukocytes as described (9). Digestions with restriction enzymes (*Eco*RI, *Eco*RV, *Hin*dIII, *Bam*HI, and *Pvu* II) were performed overnight using 5 units/μg of DNA. Samples then were electrophoresed in 0.7% agarose gel (40 hr at 30 volts in Tris acetate buffer) and transferred onto DBM paper (10). They then were hybridized for 40 hr in 50% (vol/vol) formamide/4× NaCl/Cit/0.2% Ficoll/0.2% polyvinylpyrrolidone/0.2% bovine serum albumin/5% (wt/vol) dextran sulfate containing the radiolabeled probe at 10 ng/ml and were washed at 60°C twice (15 min each) with 2× NaCl/Cit/0.5% $NaDodSO_4$ and once or twice with 0.2× NaCl/Cit/0.1% $NaDodSO_4$. (1× NaCl/Cit is 0.15 M NaCl/15 mM sodium citrate, pH 7.) Autoradiography was for 1–7 days. The DNA probe consists of a 1.6-kilobase-pair (kb) *Bam*HI–*Hin*dIII fragment of a DX α-chain gene, including exons encoding domains α1, α2, and a transmembrane domain (3).

Assignment of Allogenotopes in Each *HLA* Haplotype. Twenty-two HLA-typed families were investigated. In each of these families, it was possible to deduce the four *HLA* parental haplotypes. In informative families, each allogenotope segregated with one or two parental chromosomes. In noninformative families (i.e., where a given allogenotope was found in all the members), it was still possible to assign the allogenotopes on the basis of the two different relative intensities observed for the same band in different individuals. We made the assumption that the stronger intensity most probably corresponds to homozygosity and the weaker intensity to heterozygosity (Fig. 1). Thus it was always possible to assign a given allogenotope to one, two, three, or four parental chromosomes. In each case, this assignment coincided with the *HLA* assignment. It was possible to make this correlation since the 22 selected families always had at least three children, of which two were HLA nonidentical and one was haplo-identical with the two.

Computer Analysis. The association (positive or negative) between serological types and fragments, as well as among fragments themselves, was investigated by computing χ^2 measures of association in 2 × 2 tables. In addition to correlations, one-sided probabilities of positive association are also reported. The distribution of fragments among haplotypes was investigated by cluster analysis, grouping fragments exhibiting similar distribution in haplotypes by use of the average pair group method applied to correlation. The ordering of the fragments thus created was used to display the data so as to best reveal this clustering. A threshold value can be chosen so as to retain only those fragments clustering above that level.

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: RFLP, restriction fragment length polymorphism; kb, kilobase pair(s).

¶To whom reprint requests should be addressed.

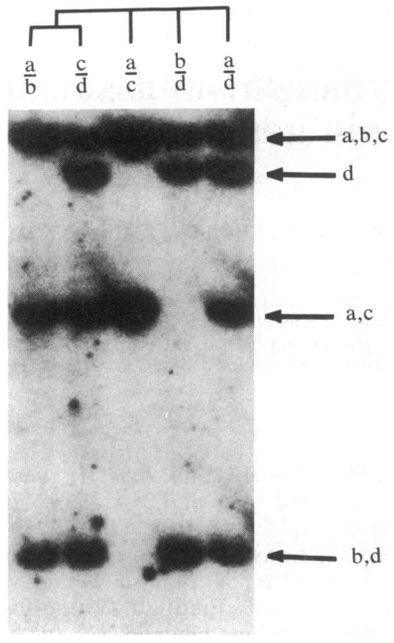

FIG. 1. Example of *Pvu* II restriction fragments detected with *DXα* probe. In the family, serologically determined parental *HLA* haplotypes are indicated a/b (father) and c/d (mother). Fragments segregate with serologically determined haplotypes. Fragments for which a single parent is heterozygous will be either present or absent in each offspring but always give bands of the same intensity (second band from the top). Fragments for which both parents are heterozygous may be present or absent in offspring, and, if present, give bands of either the same or double the intensity (third and fourth bands). Fragments for which one parent is homozygous and the other heterozygous are found in all offspring at single or double intensity (band at top).

RESULTS AND DISCUSSION

Of the five restriction enzymes used in this study (*Pvu* II, *Bam*HI, *Hin*dIII, *Eco*RI, and *Eco*RV) only *Eco*RV did not determine allogenotopes but did determine two constant fragments, of 12.0 kb and 8.8 kb, respectively. The other four restriction enzymes determined 17 allogenotopes (Fig. 2) and four constant fragments, one for each enzyme (*Eco*RI, 5.0 kb; *Pvu* II 1.5 kb; *Bam*HI, 11.9 kb; *Hin*dIII, 2.6 kb).

Allelic Behavior of *DQα* and *DXα* Restriction Fragments. The relative distribution of these fragments among the 88 haplotypes was studied. The 17 allogenotopes form six allelic series comprising from 2 to 4 allogenotopes (Fig. 2). Fragments within a series are mutually exclusive (i.e., they are never found together in the same haplotype). The *Bam*HI series, in contrast to the four others, does not cover all 88 haplotypes: it cannot be ruled out that some allogenotopes comigrate with the constant fragment.

Allogenotopes in a given series are determined by a single enzyme. However, a given enzyme might determine two series, as do *Pvu* II and *Eco*RI. It was possible to assign the two *Eco*RI series to *DQα* and *DXα* loci; cosmids containing these two genes have been isolated from an *HLA* homozygous cell line typed HLA-DR4 (6). In these cosmids, the *DQα* gene is associated with a 6.3-kb *Eco*RI fragment, whereas one of the *Eco*RI series comprises a 6.2-kb fragment probably representing the same fragment. As will be discussed, this allogenotope is specifically found in *DR4*, *DR7*, and *DR9* haplotypes. Therefore, this series is very probably associated with the *DQα* gene. The *DXα* gene comprises a 5.1-, a 2.2-, and a 1.2-kb *Eco*RI fragment. In our study, a constant 5.0-kb *Eco*RI fragment was found, corresponding probably to the 5.1-kb EcoRI *DXα* fragment. The second *Eco*RI series comprises a 1.0-kb fragment which might correspond to the 1.0-kb *DXα* fragment. Therefore, this second series is probably related to the *DXα* locus. In the same vein, the *Bam*HI and *Hin*dIII series can be assigned to the *DQα* locus.

Unfortunately, *Pvu* II-restricted sites have not been studied in these cosmid restriction maps. However, the two *Pvu* II series are probably assignable, one to the *DXα* locus and the other to the *DQα* locus, since the allogenotopes of these series are found to be in strong linkage disequilibrium with allogenotopes of either series assigned to the *DQα* locus or series assigned to the *DXα* locus. This is illustrated in Fig. 3, which shows the correlation between allogenotopes. Thus, we refer to fragments associated with the *DQα* locus as *DQα* allogenotopes and to fragments associated with the *DXα* locus as *DXα* allogenotopes.

The 13 *DQα* Allogenotopes Constitute Three Clusters, Loosely Describing an Allelic Series. Analysis of the distribu-

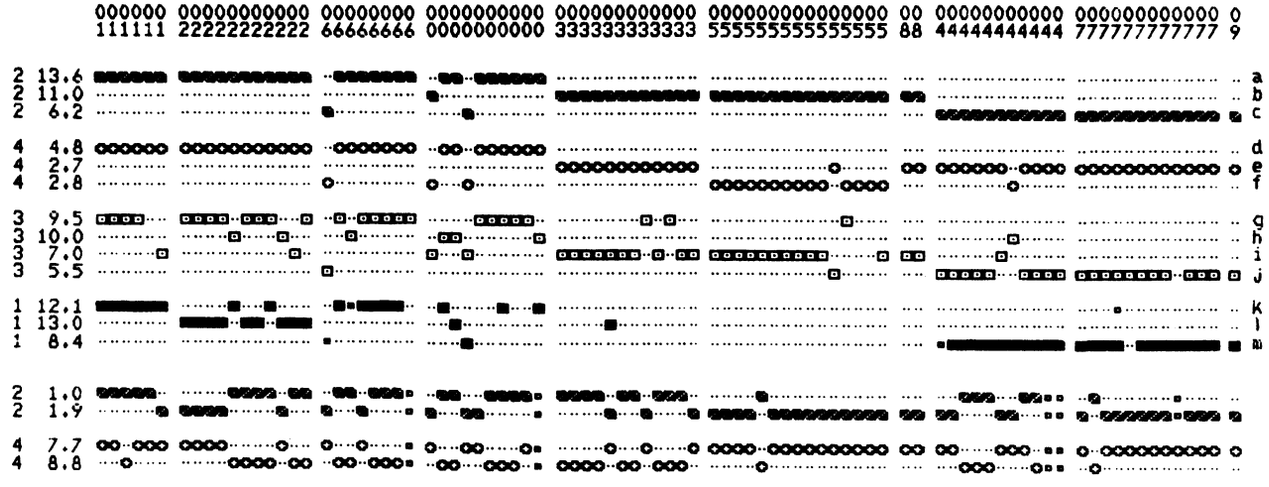

FIG. 2. Distribution of 17 DQα allogenotopes among 88 haplotypes. Rows indicate allogenotopes identified by two numbers. The first number refers to the restriction enzyme used (1, *Bam*HI; 2, *Eco*RI; 3, *Hin*dIII; 4, *Pvu* II) and the second number indicates the size of the fragment in kb. Columns correspond to haplotypes; numbers at the top indicate the DR type of each haplotype. The presence of a fragment in a given haplotype is indicated by one of four symbols, which represents the enzyme used (■, *Bam*HI; ●, *Eco*RI; ⊡, *Hin*dIII; ⊙, *Pvu* II). Questionable assignment of a fragment to a haplotype is indicated by a small open square. The four fragments at the bottom correspond to the *DXα* locus; the upper 13 fragments correspond to the *DQα* locus, and each is designated by a small letter at right, from a to m. These 17 fragments constitute six allelic series separated here by blank rows.

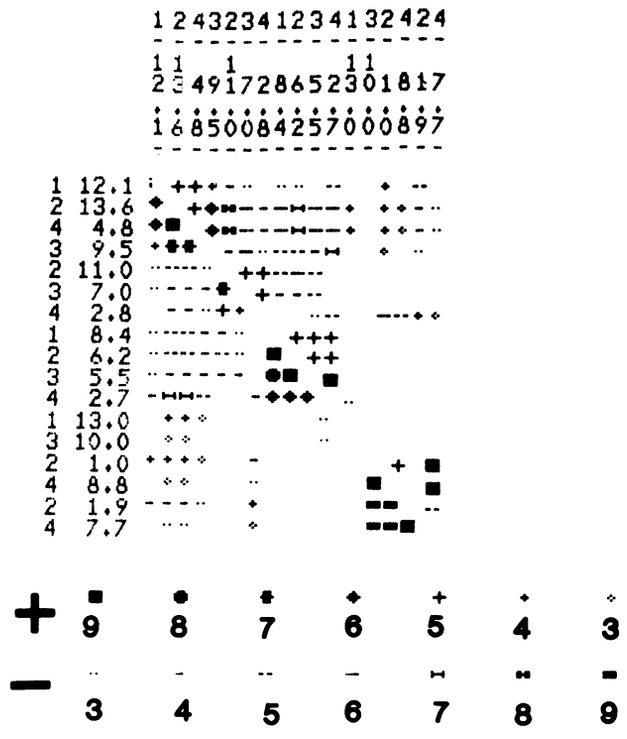

FIG. 3. Positive and negative correlations between restriction fragments. Upper right triangle represents analysis using the coefficient of correlation (r). The positive (+) or negative (−) range of rounded r or P values is indicated by symbols at the bottom of the figure. For r values, each number corresponds to $r \times 10$, and for P values each number corresponds to $-\log P$. The fragments are arranged in the same order horizontally and vertically. The first number refers to the restriction enzyme used (1, BamHI; 2, *Eco*RI; 3, *Hin*dIII; 4, *Pvu* II) and the second number indicates the size of the fragment in kb. Only r values >0.3 or <−0.3 and only P values $<10^{-3}$ are shown. Along the diagonal, one can see several triangular groupings indicating strong positive correlations, thereby defining clusters of fragments. Negative correlations between clusters can also be seen.

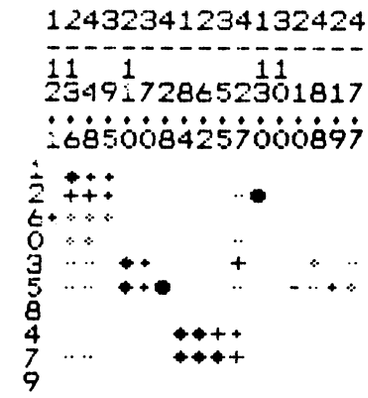

FIG. 5. r values for the correlations between DR types and the 17 allogenotopes. Allogenotopes are listed across the top of the figure (indicated as in Fig. 3); DR types are listed at left. Symbols used for the range of rounded r values are as in Fig. 3.

tion of allogenotopes among haplotypes reveals the existence of clusters of allogenotopes (i.e., fragments that tend to occur jointly in the same haplotypes). The distribution of such clusters among the haplotypes themselves, classified according to serological specificities, allowed us to investigate the relationship between polymorphism defined at the genomic level and serologically defined polymorphism (Figs. 4 and 5).

The first cluster, comprising four allogenotopes, is found in haplotypes carrying DR1,2,w6 and *null* (DQw1-like). It can be seen that one fragment in this cluster (*Bam*HI 12.1 kb) correlates with DR1 plus DRw6 only. However, another *Bam*HI fragment (13 kb), not associated with this cluster, correlates with DR2 alone ($P = 10^{-8}$).

The second cluster, comprising three fragments, correlates with DR3, DR5, and DR8. Within these clusters, a single *Pvu* II fragment correlates with DR5 only ($P = 10^{-8}$). It can be seen that the DR specificity DRw52 is associated not only with DR5, DR3, and DR8 but also with DRw6, which is not correlated with this cluster.

The third cluster includes four fragments, three of which correlate with DR4, DR7, and DR9, like the specificity DRw53. The fourth fragment correlates with DR4, DR3, DR7, DR8, and DR9.

The Four *DXα* Allogenotopes Form Two Clusters Which Describe an Allelic Series. The first cluster comprises a 1.9-kb *Eco*RI fragment and a 7.7-kb *Pvu* II fragment. The second comprises a 1.0-kb *Eco*RI fragment and an 8.8-kb *Pvu* II fragment. In six haplotypes (four of which were typed DR1, one DR null, and one DR4), these clusters disappear in the sense that the 1.0-kb *Eco*RI fragment is associated with the

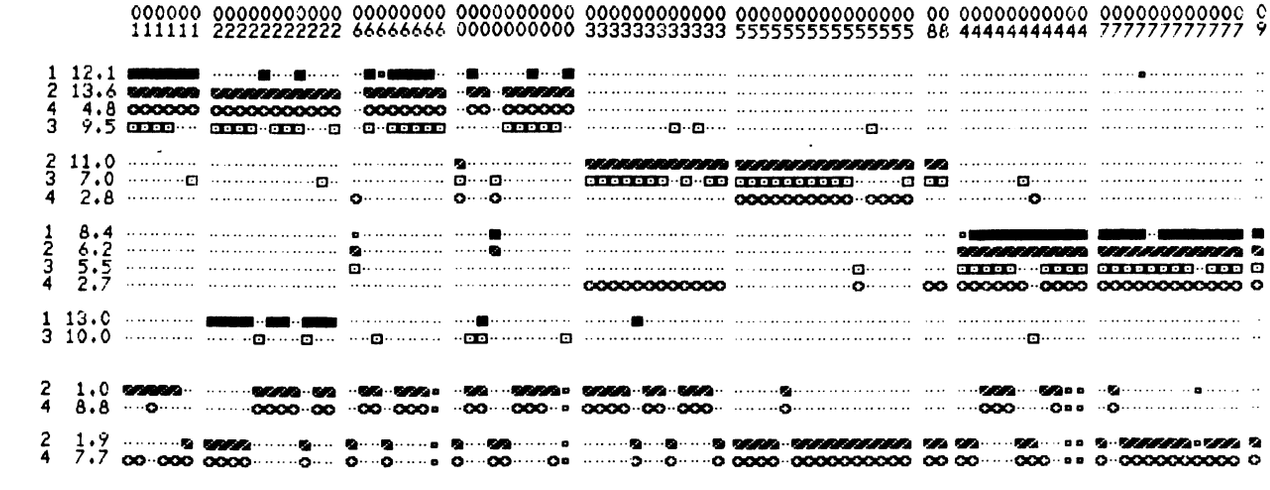

FIG. 4. Haplotypes and allogenotopes are represented as in Fig. 2. Five clusters are shown. The three top clusters (sets of rows) consist of *DQα* fragments. The next two lines each represent a *DQα* allogenotope, neither of which is associated with the three clusters. The two remaining clusters (pairs of rows) at the bottom of the figure are constituted by *DXα* allogenotopes.

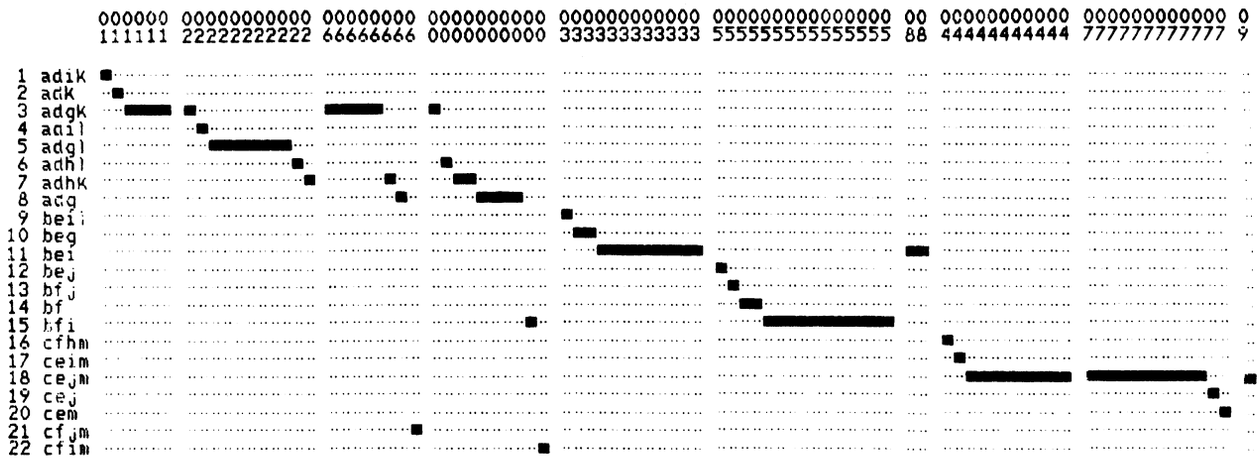

FIG. 6. Distribution of the 22 patterns of the 13 fragments found at the *DQα* locus. Haplotypes are indicated as in Fig. 2. Patterns are indicated at left by a number followed by a combination of letters from a to m, corresponding to DQα fragments as indicated in Fig. 2. The presence of a given pattern in a given haplotype is indicated by a filled square.

7.7-kb *Pvu* II fragment (Fig. 4). This probably reflects a recombination event within the *DXα* region.

It is remarkable that the *DXα* series, in contrast to the *DQα* series, does not exhibit strong linkage disequilibrium with serologically defined alleles. This suggests that the DQ α-chain genes are in strong linkage disequilibrium with the DR β-chain genes and that a recombination point occurs between a region comprising DR β-chain and DQ α-chain genes and the *DXα* locus.

However, such recombinations seem to occur only rarely between *DR5* or *DR7* haplotypes and the other haplotypes, since *DR5* and *DR7* haplotypes carry, in most cases, only one of these two *DXα* clusters (*Eco*RI 1.9 kb and *Pvu* II 7.7 kb; $r = 0.45$; Fig. 4). This apparent recombination inhibition could be due to a lack of pairing between these two classes of haplotypes, related perhaps to inversion, deletion, or duplication events.

What Is the Significance of Allogenotope Clusters and Their Correlation with Serologically Defined Epitopes? The existence of clusters suggests that some polymorphic restriction sites are in strong linkage disequilibrium. Such patterns define genomic segments whose structure, probably similar in a given set of haplotypes, drastically diverges from other sets of haplotypes. These genomic segments might encompass several genes. However, it is probable, as discussed above, that the *DQα* clusters define a genomic segment including the unique *DQα* gene; similarly, the *DXα* clusters define another genomic segment which includes the unique *DXα* gene. The genomic segment defined by *DQα* clusters might also comprise *DR* β-chain genes, since it has been reported that some DR β-chain gene restriction fragments form clusters correlated with DR4, DR7, and DR9, specificities which are also correlated with a *DQα* cluster (11).

Correlations between restriction fragments and serologically defined epitopes have been reported previously (1). It is very unlikely that these correlating restriction sites specifically affect epitopic sequences. In the few restriction maps established, polymorphic correlating sites are found in the noncoding part of the gene. Statistically, this is to be expected, since the ratio of coding versus noncoding sequences varies from 0.25 to 0.10 (1). It is interesting to note the correlation between two distant mutations: the epitope, in the coding part, and the restriction sites, generally in the noncoding part. It is also tempting to suggest that both polymorphic sequences are generated simultaneously by the same mutation event—for example, gene conversion by heteroduplex correction. Consequently, any polymorphic restriction site found even in the noncoding part could reflect a polymorphism of the coding region: this polymorphism may not have been detected or might not be detectable in the product. Indeed, clear serologically defined epitopes have not been reported for the DQ α-chain gene product, although polymorphism of the coding part has already been described (3). In our study, 22 different patterns were described for *DQα* restriction fragments (Fig. 6). These patterns might correspond to 22 allelic products. However, this generalization is probably rather extreme; some restriction sites in the noncoding segments might be independent of any polymorphism of the coding segments. However, the polymorphism of noncoding sequences could be of extreme importance if it is in fact related to the regulation of expression of these genes.

1. Cohen, D., Paul, P., Le Gall, I., Marcadet, A., Font, M. P., Cohen-Haguenauer, O., Sayagh, B., Cann, H., Lalouel, J. M. & Dausset, J. (1985) *Immunol. Rev.*, in press.
2. Botstein, D., White, R., Skolnick, M. & Davies, R. (1980) *Am. J. Hum. Genet.* **32,** 314–331.
3. Auffray, C., Lillie, J. W., Arnot, D., Grossberger, D., Kappes, D. & Strominger, J. L. (1984) *Nature (London)* **308,** 327–333.
4. Auffray, C., Ben-Nun, A., Roux-Dosseto, M., Germain, R. N., Seidman, J. G. & Strominger, J. L. (1983) *EMBO J.* **2,** 121–124.
5. Trowsdale, J., Lee, J., Carey, J., Grosveld, F., Bodmer, J. & Bodmer, W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1972–1976.
6. Okada, K., Boss, J. M., Prentice, H., Spies, T., Mengler, R., Auffray, Ch., Lillie, J., Grossberger, D. & Strominger, J. L. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3410–3414.
7. Cohen, D. & Dausset, J. (1983) in *Progress in Immunology V*, ed. Tada, Y. (Academic, London), pp. 1–12.
8. Spielman, R. S., Lee, J., Bodmer, W. F., Bodmer, J. L. & Trowsdale, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3461–3465.
9. Cann, H. M., Ascanio, L., Paul, P., Marcadet, A., Dausset, J. & Cohen, D. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1665–1668.
10. Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Stark, J. R. & Wahl, G. M. (1979) *Methods Enzymol.* **68,** 220–242.
11. Cohen, D., Le Gall, I., Marcadet, A., Font, M. P., Lalouel, J. M. & Dausset, J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7870–7874.