Immunogenetics 38: 47–53, 1993

Immuno-genetics

A novel coding sequence belonging to a new multicopy gene family mapping within the human MHC class I region

Corine Vernet[1], Marie-Thérèse Ribouchon[1], Giovanna Chimini[2], Anne-Marie Jouanolle[3], Issa Sidibé[1], Pierre Pontarotti[1]

[1] CNRS CRPG-UPR 8291, CHR Purpan, F-31300 Toulouse, France
[2] CIML, Case 906, F-13009 Marseille, France
[3] CHU de Rennes, Laboratoire de Biochimie Médicale B, rue Henri Le Guilloux, F-35033 Rennes Cedex, France

Received November 6, 1992/Revised version received November 26, 1992

Abstract. The human major histocompatibility complex (MHC) region is a genomic region spanning about 4000 kilobases (kb) including the class I, class II, and class III subregions. The class I subregion is larger than the two others but with fewer genes described to date. It includes a) classical human leucocyte antigen (*HLA*) class I genes (*HLA-A, HLA-B, HLA-C*) which are highly polymorphic and encode products presenting the endogenous antigenic peptides to the T-cell receptors, and b) non-classical class I genes (*HLA-E, HLA-F, HLA-G*) whose function is still unknown. In this study, we describe the first coding sequence which is not structurally related to the class I genes, although it is localized within the MHC class I region. This novel gene, *P5-1*, belongs to a multiple copy family, all members of which map within the MHC. Although the *P5-1* sequence showed no similarity to sequences in different databanks, its transcription, which is restricted to lymphoid tissues, argues for an immunological function of its product.

Introduction

The human MHC is divided into three subregions (Chimini and Pontarotti 1991). The most centromeric subregion is the class II region (about 1000 kb) containing the class II genes *HLA-DP, -DQ, -DR* [human leucocyte antigen (HLA), and other recently described genes such as *LMP2, LMP7* (Glynne et al. 1991), *TAP1*, and *TAP2* (Spies et al. 1990; Trowsdale et al. 1990). Class II proteins present exogenous antigenic peptides to T-cell receptors. LMP2 an LMP7 products are involved in the processing of antigenic proteins. TAP1 and TAP2 products are likely to be involved in transporting antigenic peptides from the cytoplasm to the endoplasmic reticulum where they bind the HLA heavy and light chains. Other genes called *RING* have no elucidated function (Hanson et al. 1991).

The central class III subregion (about 1000 kb) includes genes encoding for components of the complement system (like *C4*); (Carroll et al. 1985) and genes implicated in the immune response such as *TNF* α, β (Spies et al. 1986; Carroll et al. 1987), and *HSP70* (Sargent et al. 1989a). However, some genes of this subregion, such as the *t-valyl* synthetase gene (Hsieh and Campbell 1991), have no evident relationship with the immune response. In addition, there are 18 *G* genes and two *BAT* genes whose function has not been elucidated (Sargent et al. 1989b; Kendall et al. 1990; Spies et al. 1985).

The third subregion of the MHC is the class I region. It spans about 2000 kb and no genes other than class I genes and pseudogenes have so far been described. These can be separated into the two following groups: classical HLA class I (*HLA-A, HLA-B, HLA-C*) genes whose products present endogenous antigenic peptides to T-cell receptors in association with the β-2 microglobulin β_2m chain, and non-classical class I (*HLA-E, HLA-F, HLA-G*) genes whose function is still unknown (Chimini and Pontarotti 1991). The number of MHC genes and pseudogenes in this region can be explained by the large series of duplications and deletions (Klein et al. 1991), and in this article, we describe a novel coding sequence, *P5-1*, mapped within the MHC class I region.

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned the accession number L06175.

Correspondence to: C. Vernet.

Materials and methods

Cells and strains. The human B-lymphoblastoid cell line HHK (B-LCL) was used for mapping, and it is homozygous for the whole MHC region and typed A3, B7, Cw7, DR6, DQw6, DP4. The yeast artificial chromosome (YAC) 6F6 has already been described (Chimini et al. 1990). The yeast B30, which contains a YAC of 310 kb, was obtained from D. Chaplin (Geraghty et al. 1992a). The insert contains *HLA-A* as well as three other class I loci. Cosmid 503 has an insert with *HLA-A*.

DNA probes. These were: human placental DNA, *α3cw3* (a general HLA class I probe) (Chimini et al. 1988), the genomic *P5* probe which is the insert of the *P5* subclone (cloned at the *Pst* I site of pAT 153 vector), EKP5 which is the 5′ part of the total cDNA *P5-1-1*. This 770 base pairs (bp) fragment is liberated by *Kpn* I digestion of the total cDNA. Labeling of gel-purified fragments was performed by the random priming method to a specific activity greater than 2×10^9 cpm/μg (Feinberg and Vogelstein 1983).

Northern blot analysis. RNAs from the following human cell lines were used: HepG2 (hepatocellular carcinoma), NGP (HLA-negative neuroblastoma), A549 (lung carcinoma), JAR (choriocarcinoma), HPB-ALL (T-cell leukemia), YT (NK like cell line), MOLT4, CEM, KE37, 1301 (T-cell lines with different sets of surface markers), U937 (histiocytic lymphoma), IM9 (B lymphoblasts), phytohemagglutinin (PHA)-activated lymphocytes (PBL), LOVO (colic adenocarcinoma), and HHK (B-LCL). Poly A$^+$RNA was prepared according to the Fast Track kit procedure (Invitrogen, San Diego, CA). We also tested total RNA from human tissues extracted by the LiCl-Urea method (Auffray and Rougeon 1980): tumoral and healthy lung, colon, spleen, liver, thyroid, intestine, and T cell lines (J79, J1B5, B1.10, B1.8, all of them issued from Jurkat cell line) which were obtained from B. Rubin. 5 μg of each poly A$^+$RNA or 20 μg of total RNA were size-selected in a 1.2% agarose gel, blotted to nylon filters, and hybridized with the probes mentioned above.

Mapping of the P5-1 to four genes in the YAC 6F6, YAC B30 and cosmid 503. This was performed by pulsed field gel electrophoresis (PFGE) and normal electrophoresis gel analysis of genomic DNA, YAC 6F6, YAC B30, and cosmid 503 DNA. Preparation of high relative mass DNA in agarose blocks from cultured cells, tissues or yeasts, restriction digests, conditions used for PFGE and normal electrophoreses, transfer onto nylon membranes, and hybridization were all performed as previously described (Chimini et al. 1990).

Screening of cDNA libraries to isolate P5 cDNA clones. PHA-stimulated lymphocyte (Cat HL1031b) and human spleen cDNA library 5′ stretch (Cat HL1134a) were obtained from Clontech Laboratories (Palo Alto, CA). 8×10^5 plaques forming units of these libraries were doubledlifted on NEN membranes (NEN Research Products, Boston, MA). The membranes were treated according to the manufacter's instructions. The filters were hybridized with the genomic probe of *P5* and twelve positive plaques were purified. Bacteriophage DNA, prepared with the Qiagen-plasmid kit (Diagen, Düsseldorf, Germany) for λgt10, was digested by *Eco* RI, ligated, and subcloned (Sambrook et al. 1989).

Sequencing of cDNA inserts. Comparisons with databanks. Sequencing was performed using the T7 Sequencing kit (Pharmacia, Uppsala, Sweden), and is dependent on the base-specific termination of enzyme-catalyzed primer extension reactions (Sanger et al. 1977). Comparisons with databanks were carried out by using FASTP (Pearson and Lipman 1988) and BLAST (Altschul et al. 1990).

Results

Isolation of the P5-1 cDNA. The YAC 6F6 lies within the HLA class I region and has been mapped at 550 kb telomeric to *HLA-C* and at least 800 kb centromeric to *HLA-A*. The human insert is 150 kb long and contains the *HLA-X* class I gene which is located 50 kb from the *Not* I cloning site (El Kahloun et al. 1992); (Fig. 1 A).

In order to develop genomic probes, the YAC 6F6 was purified by PFGE, digested, and subcloned in a plasmid vector. A total of 1700 subclones was obtained and hybridized with radiolabeled human DNA. Inserts of subclones which do not hybridize with radiolabeled total human DNA should not contain repetitive sequences and were used systematically as probes on Northern blots.

One of these subclones, *P5* (*Pst* I-*Pst* I 1.2 kb genomic insert) detected a messenger RNA of 2.5 kb on different tissues and cell lines (data not shown). This positive signal was detected on RNA (either total or poly A$^+$) from B cell lines (HHK, IM 9, and 8317), YT (natural killer-like cell line), PHA-blasts, and healthy spleen. There was no detectable signal on hepatocellular carcinoma, neuroblastoma, choriocarcinoma, T leukaemia, T-cell lines, histiocytic lymphoma (U937), tumoral and healthy lung, or on colon, liver, thyroid, and intestine RNAs.

When the *P5* probe was used under stringent hybridization conditions [washing with 0.1× standard sodium citrate (SSC), 0.1% sodium dodecyl sulfate (SDS)] on digested genomic DNA, it revealed the presence of a single band identical to that detected on the YAC 6F6. With lower stringency (2× SSC, 0.1% SDS), no other band was demonstrated with DNA from YAC 6F6, but with total human DNA several bands appeared. The single *P5* locus on YAC 6F6 was named *P5-1*. The *P5* genomic fragment was used to screen a human PHA-stimulated lymphocyte and a human spleen cDNA libraries constructed in λgt11 and λgt10 vectors. Twelve of 8×10^5 plaques were found to be positive. The three longest were chosen for the determination of the nucleotide sequence. In order to determine whether they issued from the transcription of the *P5-1* gene or from genes related to *P5-1*, we also sequenced the genomic *P5* insert. The genomic *P5* insert was found to be identical in sequence to each of these three clones, suggesting that all the cDNA issued from the transcription of the *P5-1* gene. The longest insert obtained was named *P5-1-1* cDNA.

The *P5-1-1* cDNA is 2531 nucleotides long (Fig. 2) with a poly (A) stretch of ten nucleotides. The estimated messenger size, by northern blot analysis, corresponded to this cDNA size, which suggests that the *P5-1-1* insert is a full-length cDNA. It contains a 675-bp open reading frame, with the initiation site at position 305 and the termination codon TAA at position 962, corresponding to a polypeptide of 219 amino

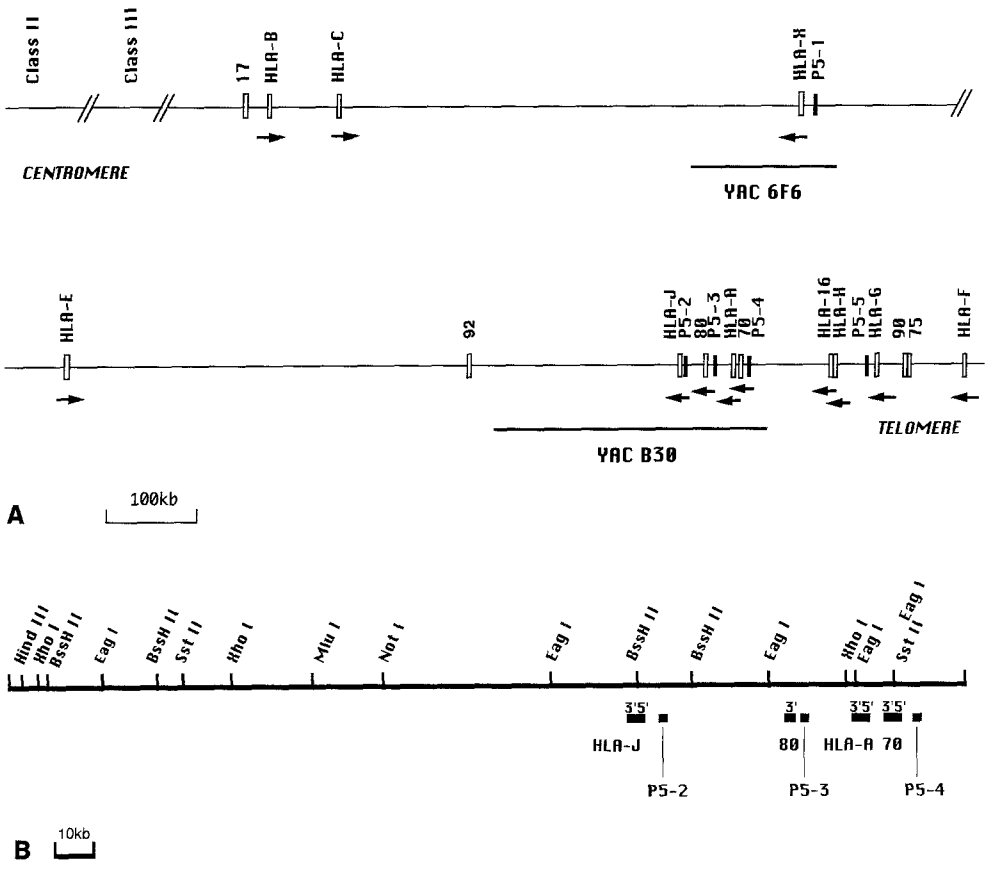

Fig. 1 A, B. **A** Map of human MHC class I region; **B** Map of YAC B30. Distances and class I gene orientations described (Vernet et al. 1991; El Kahloun et al. 1992; Geraghty et al. 1992a). *White* and *black boxes* correspond to class I and *P5*-related sequences, respectively. Nomenclature described (Bodmer et al. 1992).

acids. Four inframe stop codons are found at nucleotides 157, 230, 259, and 283 in the 5′ untranslated (UT) region. The sequence around the initiation site (TCCTCATGT) matched the consensus sequence for the initiation of translation in vertebrates ($GCC^{A}/_{G}CCATGG$, Kozak 1991). There is no other open reading frame which can encode a long polypeptide (>90 amino acids). A polyadenylation signal was found (AATAAA) 129 bp upstream to the poly (A) stretch.

The conceptual protein encoded by this messenger RNA has a relative mass of 25000. Hydrophobicity plot analysis (data not shown) shows no clear hydrophobic region which would have suggested a membrane-spanning portion. There is no preferential distribution of the amino acids. We compared the *P5-1-1* nucleotide sequence and its deduced amino acid sequence with other known sequences present in GenBank, EMBL, and SWISS-PROT databases and found no reported sequences with significant similarity. Nevertheless we found a similarity (80% of homologies over 100 nucleic acids) between the 3′ noncoding region of *HLA* class I genes and the 3′ non coding region of *P5-1*. This similarity could mean that *P5* and class I genes share some sequences and reflect the proximity of *P5* and class I genes. More sequencing is needed to clarify this.

Genomic description of the multicopy family P5. As mentioned above, the genomic *P5* probe under low-stringency conditions gave several bands on a Southern blot of genomic DNA digested by various enzymes (*Hin* dIII, *Eco* RI, *Eco* RV, and *Bam* HI). For instance, the pattern of hybridization with *Hin* dIII showed six bands (from 2.8 to 17 kb) with different intensities. The two-allele polymorphism of the *P5-1* gene has already been described (David et al. 1991; Fig. 3). Preliminary data show that one member of the *P5* family, *P5-3*, displays a 70% sequence similarity with *P5-1*.

Localization of the P5 multicopy family within the MHC class I region. Under low-stringency conditions, hybridization with the genomic *P5* probe on HHK cell line DNA digested by *Not* I and *Mlu* I and separated by PFGE reveals the same bands as a class I specific probe (Fig. 4). These results show a genomic co-localization of the *P5*-related sequences and *HLA* class I genes. Furthermore, hybridization experiments on digested YAC B30 and cosmid 503 DNA (both containing the *HLA-A* locus) showed that three copies (*P5-2–P5-4*) are located around the *HLA-A* gene (Fig. 1B). Moreover, this probe revealed one band (*P5-5*) on a cosmid insert containing a *HLA-G* class I gene (data not shown). The size of *P5-5* is undetermined because it is located at the end of the cosmid insert (Fig. 1A). The

```
ggcccagacgccaaggttgcgggtcatggagtcccgaaccctcctcctgctgttctcgggagccgtggccctgatccagacctgggcaga        90

ttacaattacaatcaaggcagaaatgatctcatttttacattacaactcctggaaaaggcaatagactgagatgcaagtgtgcccccaag       180

tgatgggcagaaggagagaaggatgttttggatgcattctagaacacaggtaatctaaggagagttgatccaaggcacgtaggaagagtc       270

ctccagccactattagGCCATCAAAGGAGTCCTCATGTGTCCCAGGAAATGGTCCTGCTTTGGTGTCCCTGGTGTGACCCATCACCCGCT       360
                                        M  C  P  *R  K  W  S  C  F  G  V  P  G  V  T  H  H  P  L   19
GGAACAGCCTGAGAGAAGTAGGGCCTCTGCACCAATGCTGCTGAGGATGTCAGAGCACAGGAACGAGGCCTTGGGAAATTACCTGGAAAT       450
 E  Q  P  E  R  S  R  A  S  A  P  M  L  L  R  M  S  E  H  R  N  E  A  L  G  N  Y  L  E  M   49
GCGACTGAAATCTTCCTTCCTGAGGGGTCTGGGCTCTTGGAAATCAAACCCTCTCAGGTTGGGTGGCTGGACGATTCTCCTCACACTTAC       540
 R  L  K  S  S  F  L  R  G  L  G  *S  W  K  S  N  P  L  R  L  G  G  W  T  I  L  L  T  L  T   79
AATGGGACAAGGGGAACCAGGAGGTCCAAGGATCCCTGGGTTCCACACGAACTCCTCCTACCCTCATTGTGTGACAGCAGCCATGCCTCC       630
 M  G  Q  G  E  P  G  G  P  R  I  P  G  F  H  T  N  S  S  Y  P  H  C  V  T  A  A  M  P  P  109
TCCTGGGGATCAGGATTCTATTACCTGTGCCTGGAGAGGAGGGGACTCCTCTTCTCACCCGCTGGTCTCTGGACACATACTGTCCAATTC       720
 P  G  D  Q  D  S  I  T  C  A  W  R  G  G  D  S  S  S  H  P  L  V  S  G  H  I  L  S  N  S  139
CCCTGTGGCAGCTGTAATGTGTAGTTCAATGGGCACTCATTTGTCYCCTTTTAAGGTACCCTCCTTTAGAATCAGGACCTTCTACCCTGC       810
 P  V  A  A  V  M  C  S  S  M  G  T  H  L  S  P  F  K  V  P  *S  F  R  I  R  T  F  Y  P  A  169
AGAGTGTGGTTTTGGGAGAGAAGTGCAAAATCCCACGACAGGTGAGTTGAAGGAATGGGATATGGAGCCACATCCACTTCCACCCCTTGG       900
 E  C  G  F  G  R  E  V  Q  N  P  *T  T  G  E  L  K  E  W  D  M  E  P  H  P  L  P  P  L  G  199
TATCTGGACCCACGTGTTCTTCCTACTGAGATTACAGAACTGTAGAGATGTCTTTGATTTTtaaaatgcaccatgtcctgaaagatggca       990
 I  W  T  H  V  F  F  L  L  R  L  Q  N  C  R  D  V  F  D  F                                   219
ccctcccacccgcagagtgcttcctgcaagctggcgttgagctgtgcctatagaagctcttttcaacattctttatggccaggagccctt      1080
ggttggtgcagatggtgataggaccagtgggtcccacagcatggccacactgcacctccttcgctgtcaagtgggtccccccacaagata      1170
ctgcacggagagcagtgccaagcctgtggatcaggaatatcaacagcccccagagagtggtgctggctgagggtctgagagcaggacagg      1260
aaaaccacctatggaataggtgcctatccctgtgaagatgaacctctggcccttccaggatggaaggagtgcaatgtagtcaactcctca      1350
cttagggtctggttggtcacctaaagaaatagagccctaccagggaagatcattgggttcaaatgctgatgagtaggacatttagaggtg      1440
gcagtgtctggatctaccttggtaggagggagtcagtactgttggacccataggtagcctcatccctgccactgtgtttgctccatttat      1530
gtacccacctacccggcctgggctgaccatgggaaggctggctaatttcagtgcttgtgcttggttgttcagggccatttcaggtttggg      1620
tgttttctggggatgttaacatgggattcaggctcaactcacaagaaacttttcatctcatgatggatgctgttgggcatgtccaatgta      1710
tgacttcatgagttacacagatgctaattcgtaggggcacttggaatcacatggttgttttgtgtcccatggtcaagcattctatcttat      1800
cagggcctacagtaacatgccaaaagttgcttccaacatatttctctgctttggatggggcatatttctgtgctgtggatgacatggcct      1890
tactccagaatcccaggccctccactgtgactctcctactggtgcttggttcagctccaccccaaatcttaccccaccactggcactttc      1980
agcaccaggggtctgaaggatggtgactgcgccatggcctggatctgctgcagtgtcctttcctgtggaggctccactcaaagctggca      2070
tcctcctatgtcacctagagtgtgggtcaaagcaatacacctacatgtagaatgtgatgtcagaactcaaacaggctcaccaggcagtgt      2160
gcttcttccttgcatgaggatgcaagatgcaacagtttgtcttcacattggaaggacacccctggatgcccctaaccactagacctgtaa      2250
aacttcactgcagtggccacttctgaatctctgtaaggtttatttatcttcacccctctggagagaagatgttttaccaaagcctctagt      2340
gtaccgtcctcctcttactcatccatcccagtcaacatgatgttgtcaatgaataaaggaatttaatattctatagtatatccaggttc      2430
tccagatctcttaagactgtactatagaggcctggggaattataatagccctgaggcaaactatgaattaaagtgttgtggatcccacat      2520
gaaaaaaaaa                                                                                      2531
```

Fig. 2. Nucleotide and amino acid sequence of *P5-1-1* cDNA. Sequence determined by method of Sanger and co-workers (1977). Consensus Kozak sequence and potential polyadenylation sites *underlined* and *bold*. *Underlined* stop codons in front of and at end of open reading frame. Position of open reading frame encoding the 219 amino acids indicated. Predicted amino acid sequence given in *one-letter* symbols below nucleotide sequence. *Asterisks* denote potential sites of cAMP-cGMP-dependent protein kinase, caseine kinase II, and protein kinase C phosphorylation sites.

sequence *P5-1* (14.5 or 17 kb with *Hin* dIII, depending on the allele) is mapped at 30 kb centromeric from a *Not* I restriction site of the YAC 6F6 and at 15 kb telomeric from the *HLA-X* gene. *P5-2* and *P5-3* (3.6 and 5.6 kb with *Hin* dIII) lie 5′ of *HLA-J* and 3′ of *HLA-A*, respectively. *P5-4* (2.8 kb with *Hin* dIII) is located next to the 5′ end of the *HLA-70* pseudogene. *P5-6* could be telomeric to the *HLA-A* or centromeric to the *HLA-X* locus. However, hybridization experiments on YACs and cosmids from the MHC class I region excluded the localization of *P5-6* near the *HLA-B*, *HLA-C*, and *HLA-E* loci.

Discussion

Sequence and function. In this study, we report the first gene, *P5-1*, which is not structurally related to class I genes and which is located within the class I region. The *P5-1* locus maps between *HLA-B* and *HLA-E*, 15 kb telomeric to the *HLA-X* gene. Its 2.5 kb messenger RNA encodes an open reading frame of 219 amino acids from which we can deduce a protein with a relative mass of 25 000 and without a membrane-spanning domain. Both the nucleotide sequence and its deduced amino acid sequence showed no similarity to other known sequences in the GenBank, EMBL, and SWISS-PROT databanks. However, its pattern of transcription, restricted to normal spleen, PHA-activated lymphocytes, and B and YT (NK like) cell lines, argues that the product of this gene may be involved in the immune response. The presence of this lymphoid tissue-specific gene in the MHC region, otherwise rich in genes whose products are implicated in the immune function, could be an example of the teleological clustering of genes.

How can the organization of the MHC class I region be explained? The organization of *HLA* class I genes

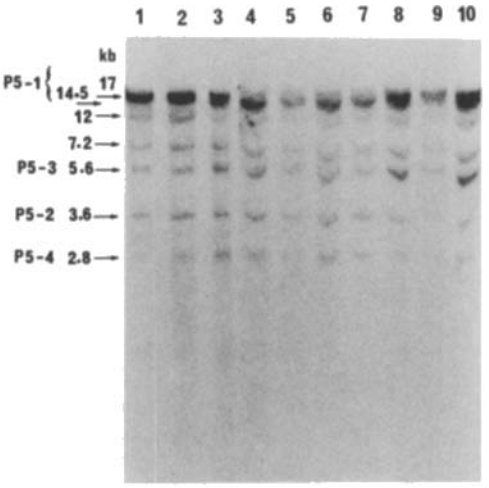

Fig. 3. DNA from different human donors digested by the *Hin* dIII restriction enzyme, analyzed by electrophoresis in 0.8% agarose gel, transferred onto nylon, and hybridized with ^{32}P-labeled 1.2 kb *P5* genomic insert. Six fragments revealed with *P5* probe. *Hin* dIII restriction polymorphism (17 kb: allele 1; 14.5 kb: allele 2) noted.

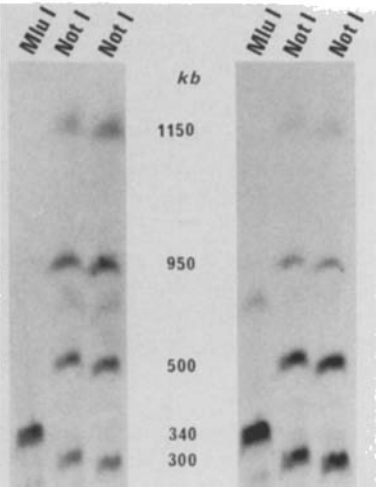

Fig. 4. PFGE blot analysis. Genomic DNA from HHK (human B lymphoblastoid cell line) digested with *Not* I and *Mlu* I. DNA digests electrophoresed by pulsed field in 0.8% agarose gel (Chimini et al. 1990), blotted onto nylon membrane, and hybridized with following ^{32}P-labeled probe: a) the general class I probe, *α3cw3*, b) the 1.2 kb genomic *P5* insert. HHK-strain DNA digest shows hybridization of same fragments with two probes.

within the MHC class I region reflects several chromosomal rearrangements which have occurred during the evolution of this region. These genomics rearrangements can be illustrated by several examples: for instance, *HLA-80, HLA-90*, and *HLA-16* (which form a class I subfamily, since these closely related sequences show the lowest homology with respect to any of the other class I sequences, Geraghty et al. 1992b) are linked to the closely related *HLA-A, HLA-G*, and *HLA-H*, respectively. One could suggest large duplication events. However, these hypothetical duplications must have been followed by large additionnal rearrangements in the case of *HLA-A/HLA-80* and *HLA-G/HLA-90* couples: *HLA-A* lies 7 kb from *HLA-80*, while the distance between *HLA-80* and *HLA-G* is around 30 kb; *HLA-90* lies 5′ of *HLA-G* while *HLA-80* lies 3′ of *HLA-A*. Thus, any large duplication would have to have been followed by an inversion of *HLA-80* or *HLA-90*. Two independent duplication events may also explain the existence of the two gene pairs, and the proximity of each pair members to one another may be coincidal (Geraghty and al. 1992b, 1992c). In the case of the *HLA-H/HLA-16* and *HLA-A/HLA-80* couples, tandem duplication is the most likely possibility, as the two regions have broadly the same overall structure (Chorney et al. 1990).

Another example of the plasticity of the MHC class I region is given by Nei and Hughes (1989). They have shown that exons 4 and 5 from *HLA-A* locus alleles are much more similar to *HLA-E* than to *HLA-B* and *HLA-C* alleles, whereas in exons 2 and 3, alleles from the three classical class I genes (i.e., *HLA-A, HLA-B*, and *HLA-C*) are much more similar to each other than any one of them is to *HLA-E*. The most likely hypothesis is that interlocus recombination led to the formation of a hybrid gene in which exons 2 and 3 were derived from the original *HLA-A* locus and exons 4 and 5 from the *HLA-E* locus. A second example of such intergenic exchange is given by Geragthy and co-workers (1992c) for the *HLA-92* and *HLA-E* genes.

The localization of the *P5* copies within the MHC class I region also reflects the large rearrangements that have occurred within this region. Indeed, the *P5* members are associated with a class I gene subset: *HLA-J/P5-2, HLA-A/P5-3, HLA-70/P5-4*, and *HLA-G/P5-5* (Fig. 1). This could be the result of tandem duplications but, as shown for the *HLA-A/HLA-80* and the *HLA-G/HLA-90* couples, the different *P5* copies do not always lie on the same side of their *HLA* class I companion. For example, *P5-3* lies 3′ of *HLA-A* while *P5-2* lies 5′ of *HLA-J*. Thus, either duplications were followed by other rearrangements or *P5* copies, and *HLA-A, HLA-J, HLA-70*, and *HLA-G* occurred by independent duplications. Furthermore, some of the class I genes which do not have a *P5* companion correspond to the genes which are not in the *HLA-A* subfamily: for example, the *HLA-B, HLA-C*, and *HLA-E* genes. This could be explained by a putative translocation event that placed a *P5* ancestor within the class I region after the separation of *HLA-A* from an *HLA-B, HLA-C*, and *HLA-E* gene ancestor (about 40 million years ago; Klein and Figueroa 1986). It is therefore also probable that the *P5/HLA* class I block predated the duplication events that gave rise to a *HLA-A, HLA-B, HLA-C*, and *HLA-E*

gene ancestor. The absence of *P5* copies near these latter genes suggests that some of the duplication events did not copy the entire block, or that *P5* copies have been subsequently lost. As noted above, it is also possible that the *P5* ancestral sequence duplicated independently in this region. It could also be postulated that an *HLA-A* ancestor has been duplicated in the *P5* region.

HLA class I and P5 phylogenic trees, based on synonymous mutations, and the localization of *P5* members in regard to class I genes in other mammalian species, will permit a better understanding of the setting up of this region.

Acknowledgments. We are grateful to John Clayton, Bent Rubin, and Tanja Crnogorac-Jurcevic for revision of our manuscript. We thank Michael Chorney and David Chaplin for the YAC B30. We thank Merry Ferrer for the artwork. This work is supported by grants to P. P. from AFM (Association Française contre les Myopathies), CNRS, and MRT.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. Basic local alignment search tool. *J Mol Biol 215:* 403–410, 1990

Auffray, C. and Rougeon, F. Purification of mouse immunoglobulin heavy chain messenger RNAs from total myeloma tumor RNA. *Eur J Biochem 107:* 303, 1980

Bodmer, J. G., Marsh, S. G. E., Albert, E. D., Boomer, W. F., DuPont, B., Erlich, H. A., Marsh, B., Mayr, W. R., Parham, P., Sasazuki, T. et al.: Nomenclature of factors of the HLA system. *Tissue Antigens 39:* 161–173, 1992

Carroll, M. C., Belt, K. T., Palsdottir, A., and Yu, Y. Molecular genetics of the fourth component of human complement and steroid 21-hydroxylase. *Immunol Rev 87:* 39–45, 1985

Carroll; M. C., Kartzman, P., Alicot, E. M., Koller, B. H., Geraghty, D. E., Orr, H. T., Strominger, J. L., and Spies, T. Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. *Proc Natl Acad Sci USA 84:* 8535–8539, 1987

Chimini, G., Pontarotti, P., Nguyen, C., Toubert, A., Boretto, J., and Jordan, B. R. The chromosome region containing the highly polymorphic HLA class I genes displays limited large scale variability in the human population. *EMBO J 7:* 395–400, 1988

Chimini, C., Boretto, J., Marguet, D., Lanau, F., Lauquin, G., and Pontarotti, P. Molecular analysis of the human MHC class I region using yeast artificial chromosome clones. *Immunogenetics 32:* 419–426, 1990

Chimini, G. and Pontarotti, P. Molecular organization and evolution of the chromosomal region containing the human Major histocompatibility complex. *In* R. Srivastava, B. P. Ram, and P. Tyle (eds): *Immunogenetics of the Major Histocompatibility Complex,* pp. 65–99, V. C. H Publishers, New York, 1991

Chorney, M. J., Sawada, I., Gillespie, G. A., Srivatava, R., Pan, J., and Weissman, S. M. Transcription analysis, physical mapping, and molecular characterization of a nonclassical human leukocyte antigen class I gene. *Mol Cell Biol 10:* 243–253, 1990

David, V., Boretto, J., Jouanolle, A. M., Mauvieux, V., El Khaloun, A., Perichon, M., Blayau, M., and Pontarotti, P. Two polymorphisms at the locus D698 defined by YAC. *Nucleic Acids Res 18:* 5582, 1991

El Kahloun, A., Vernet, C., Jouanolle, A. M., Boretto, J., Mauvieux, V., Le Gall, J-Y., David, V., and Pontarotti, P. A continuous restriction map from HLA-E to HLA-F. Structural comparison between different HLA-A haplotypes. *Immunogenetics 35:* 183–189, 1992

Feinberg, A. P. and Volgelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem 107:* 303, 1983

Geraghty, D. E., Pei, J., Lipsky, B., Hansen, J. A., Taillon-Miller, P., Bronson, S. K., and Chaplin, D. D. Cloning and physical mapping of the HLA class I region spanning the HLA-E-to-HLA-F interval by using yeast artificial chromosomes. *Proc Natl Acad Sci USA 89:* 2668–2673, 1992a

Geraghty, D. E., Koller, B. H., Hansen, J. A., and Orr, H. T. The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. *J Immunol 149:* 1934–1946, 1992b

Geraghty, D. E., Koller, B. H., Pei, J., and Hansen, J. A. Examination of four HLA class I pseudogenes. *J Immunol 149:* 1947–1956, 1992c

Glynne, R., Powis, S. H., Beck, S., Kelly, A., Kerr, L. A., and Trowsdale, J. A proteasome-related gene between the two ABC transporter loci in the class II region of the human MHC. *Nature 353:* 357–360, 1991

Hanson, I. M., Poustka, A., and Trowsdale, J. New genes in the class II region of the human major histocompatibility complex. *Genomics 10:* 417–424, 1991

Hsieh, S. L. and Campbell, R. D Evidence that genes G-7A in the human major histocompatibility complex encodes Valyl-transfer RNA-synthetase. *Biochem J 278:* 809–816, 1991

Hughes, A. L. and Nei, M. Ancient interlocus exchange in the history of the HLA-A locus. *Genetics 122:* 681–686, 1989

Kendall, E., Sargent, C. A., and Campbell, R. D. The human major histocompatibility complex contains a new cluster of genes between HLA-D and the complement C4 loci. *Nucleic Acids Res 18:* 7251–7257, 1990

Klein, J. and Figueroa, F. Evolution of the major histocompatibility complex. *Crit Rev Immunol 6:* 295–386, 1986

Klein, J., Zhu, Z., Gutknecht, J., Figueroa, F., and Kasahara, M. Mhc: Lessons in evolution. *In* R. Srivastava, B. P. Ram, and P. Tyle (eds.): *Immunogenetics of the Major Histocompatibility Complex,* pp. 18–38, V. C. H., New York, 1991

Kozak, M. An analysis of vertebrate mRNA sequences: intimations of transcriptional control. *J Cell Biol 115:* 887–903, 1991

Pearson, W. R. and Lipman D. J. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA 85:* 2444–2448, 1988

Sambrook, J., Fritsch, E. F., and Maniatis, T. *In* C. Nolan (ed.) *Molecular Cloning: A Laboratory Manual, 2nd edn.*, Cold Spring Harbor Laboratory, Cold Spring Harbor, 1989

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

Sargent, C. A., Dunham, I., Trowsdale, J., and Campbell, R. D. Human major histocompatibility complex contains genes for the major heat shock protein HSP 70. *Proc Natl Acad USA 86:* 1968–1972, 1989a

Sargent, C. A., Dunham, I., and Campbell, R. D. Identification of multiple HTF island associated genes in the human major histocompatibility complex class III region. *EMBO J 8:* 2305–2312, 1989b

Spies, T., Sorrentino, R., Boss, J. M., Okada, K., and Strominger, J. L. Structural organization of the human major histocompatibility complex. *Proc Natl Acad Sci USA 82:* 5165–5169, 1985

Spies, T., Morton, C. C., Nedospasov, S. A., Fiers, W., Pious, D., and Strominger, J. L. Genes for the tumor necrosis factors α and β are linked to the human major histocompatibility complex. *Proc Natl Acad Sci USA 83:* 8699–8702, 1986

Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blanck, G., Mellins, E., Pious, D., and DeMars, R. A gene in the human major histocompatibility complex class II region controlling the class I antigen pathway. *Nature 348:* 744–747, 1990

Trowsdale, J., Hanson, I., Mockridge, I., Beck, S., Townsend, A., and Kelly, A. Sequences encoded in the class II region of the MHC related to the "ABC" superfamily of transporters. *Nature 348:* 741–744, 1990

Vernet, C., Chimini, G., Boretto, J., Le Bouteiller, P., and Pontarotti, P.: Organization and evolution of the MHC chromosomal region: an overview. *In* J. Klein and D. Klein (eds.): *Molecular Evolution of the Major Histocompatibility Complex,* pp 1–11, Springer, Berlin Heidelberg, 1991