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

Corine Vernet¹, Marie-Thérèse Ribouchon¹, Giovanna Chimini², Anne-Marie Jouanolle³, Issa Sidibè¹, Pierre Pontarotti¹

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 and 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 β2m 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, α3cω3 (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 S' 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 2×108 cpm/μg (Feinberg and Vogelstein 1983).

Northern blot analysis. RNAs from the following human cell lines were used: HepG2 (hepatocellular carcinoma), NBP (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 (colonic 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 (J77, 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 cosm 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 S' stretch (Cat HL1.134a) were obtained from Clontech Laboratories (Palo Alto, CA). 8×105 plaques forming units of these libraries were doubledlifted on NEN membranes (NEN Research Products, Boston, MA). The membranes were treated according to the manufacturer’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 Agt10, 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 Nor I cloning site (El Kahouni 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 leukemia, 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 Agt11 and Agt10 vectors. Twelve of 8×105 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