Genomics, Isoforms, Expression, and Phylogeny of the MHC Class I-Related MR1 Gene

Patricia Riegert, Valérie Wanner and Seiamak Bahram

*J Immunol* 1998; 161:4066-4077; ;
http://www.jimmunol.org/content/161/8/4066

References
This article cites 63 articles, 16 of which you can access for free at:
http://www.jimmunol.org/content/161/8/4066.full#ref-list-1

Subscription
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Genomics, Isoforms, Expression, and Phylogeny of the MHC Class I-Related MR1 Gene\textsuperscript{1,2}

Patricia Riegert,* Valérie Wanner,† and Seiamak Bahram\textsuperscript{3*†}

A growing number of non-MHC-encoded class I-related molecules have been shown to perform diverse, yet essential, functions. These include T cell presentation of bacterially derived glycolipidic Ags by CD1, transcytosis of maternal IgG by the neonatal Fc receptor, enriched presence and plausible function within exocrine fluids of the Zn-\(\alpha_2\)-glycoprotein, subversion of NK cytolytic activity by the CMV UL18 gene product, and, finally, crucial involvement in iron homeostasis of the HFE gene. A recently described member of this family is the MHC class-I related (MR1) gene. The most notable feature of MR1 is undoubtedly its relatively high degree of sequence similarity to the MHC-encoded classical class I genes. The human chromosome 1q25.3 MR1 locus gives rise not only to the originally reported 1,263-bp cDNA clone encoding a putative 341-amino acid polypeptide chain, but to many additional transcripts in various tissues as well. Here we define the molecular identity of all human and murine MR1 isoforms generated through a complex scenario of alternative splicing, some encoding secretory variants lacking the Ig-like \(\alpha_3\) domain. Moreover, we show ubiquitous transcription of these MR1 variants in several major cell lineages. We additionally report the complete 18,769-bp genomic structure of the MR1 locus, localize the murine orthologue to a syntenic segment of chromosome 1, and provide evidence for conservation of a single-copy MR1 gene throughout mammalian evolution. The 90% sequence identity between the human and mouse MR1 putative ligand binding domains together with the ubiquitous expression of this gene favor broad immunobiologic relevance. \textit{The Journal of Immunology}, 1998, 161: 4066–4077.

\textsuperscript{1} Basel Institute for Immunology, Basel, Switzerland; and \textsuperscript{2} Centre de Recherche d’Immunologie et d’Hématologie, Strasbourg, France

Received for publication January 30, 1998. Accepted for publication June 8, 1998.

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

\textsuperscript{*} Work in Strasbourg was supported by Association pour la Recherche sur le Cancer, the MRC, the Fonds de la Recherche Médicale Francaise, and the CNRS.

\textsuperscript{†} Address correspondence and reprint requests to Dr. Seiamak Bahram, Centre de Recherche d’Immunologie et d’Hématologie, 4 Rue Kirschleger, 67085 Strasbourg, France. E-mail address: siamak@hemato-ulp.u-strasbg.fr

\textsuperscript{3} Address correspondence and reprint requests to Dr. Seiamak Bahram, Centre de Recherche d’Immunologie et d’Hématologie, 4 Rue Kirschleger, 67085 Strasbourg, France. E-mail address: siamak@hemato-ulp.u-strasbg.fr

\textsuperscript{4} Abbreviations used in this paper: MHC-I, MHC class I gene or protein; MHC-II, MHC class II gene or protein; Fc, neonatal Fc receptor; MIC, MHC class I chain related; MR1, MHC class I related; FISH, fluorescence in situ hybridization.

\textsuperscript{5} Copyright © 1998 by The American Association of Immunologists 0022-1767/98/$02.00
The neonatal Fc receptor (FcRn) is the second well-characterized member of non-MHC-I-like structures. The human chromosome 19q13.3 (murine chromosome 7) FCGRT gene encodes a highly divergent 48-kDa heavy chain bound to \( \beta_2 \) m at the gut epithelial brush border where it acts as a transepithelial shuttle delivering maternal IgG from the intestinal lumen into the bloodstream, providing passive neonatal immunity (26, 27).

The Zn\(^{2+} \)a\(^{2+} \)gp (AZGP1) is another MHC-I like structure. A soluble single-chain 41-kDa molecule of as yet unknown function, Zn\(^{2+} \)a\(^{2+} \)gp was first purified from plasma, although it is more prevalent in exocrine secretions, especially those of the mammary gland. The gene has been mapped to human chromosome 7q22.1 and mouse chromosome 5 (28–30).

Highly divergent MHC-I genes are not all encoded outside the MHC; in particular, a novel highly divergent multigene class I family termed MIC (MHC class I chain related) is located within the MHC class I region (31). The MICA gene in this family is unusually polymorphic (32) (compared with mono-, oligomorphic MHC-Ib genes) and encodes a highly glycosylated single-chain, membrane-bound protein almost exclusively expressed in epithelia (33). HFE is yet another example of a drastically divergent MHC-I gene, adjacent to the MHC itself (approximately 4 megabases telomeric to HLA-F) (34). Surprisingly, the murine HFE orthologue does not map to the H2 complex, but is located within a paralogous locus on chromosome 13 (35, 36). A typical class I gene by domain structure, HFE displays only 30% overall homology to typical MHC-I molecules, but requires dimerization with \( \beta_2 \) m before surface expression (37).

In summary, a restricted collection of diverse, yet structurally related, MHC-I genes located on various chromosomes perform highly disparate, yet essential, functions. The latest addition to the group of non-MHC-encoded class I genes is the human MHC class I-related (MR1) gene (43). Discovered by degenerate PCR, MR1 was subsequently mapped to human chromosome 1q25.3. Northern blotting experiments using a number of human tissues revealed ubiquitous expression of multiple MR1 transcripts; a cDNA clone of 1263 bp encoding a putative polypeptide of 341-amino acid sequence (calculated \( M_r \) of 39 kDa) resembling an authentic MHC-I was originally reported (43). Perhaps the most interesting feature of MR1 is that despite being located outside the MHC, the molecule shows the highest degree of homology to the classical MHC-I molecules (43). This opens the possibility of analogous function for MR1 and raises enticing questions as to the scenario of genome-wide MHC-I dispersion.

Here we present the cDNA sequences of all human and murine MR1 isoforms and show evidence for their uniform expression in various cell lineages and tissues. Moreover, we define the genomic structure of MR1, localize the gene by fluorescence in situ hybridization (FISH) in the mouse genome, and study the extent of phylogenetic conservation of this molecule.

**Materials and Methods**

cDNA genomic cloning and sequencing

A full-length human MR1 cDNA clone was obtained by RT-PCR using oligonucleotides derived from published sequence (43) and human intestinal total RNA (Clontech, Palo Alto, CA) as template, employing standard procedures (44, 45). The oligonucleotides used were as follow: RT, 59-TGACTCTATGGGGGACAGAG-39; and PCR, 59-GGACTATGGGGAACTGATG-39 and 39-CTAACGTCTAGGGAGAAAAGG-59. The full-length cDNA clone of 1.3 kb was subsequently used as a probe to sequentially screen approximately 6 × 10\(^6\) phage clones of the following cDNA libraries: human fetal spleen and murine (C57/BL6) spleen (Stratagene, San Diego, CA). This led to isolation of several cDNA clones from each species, which were subsequently analyzed by sequence determination. The full-length human MR1 cDNA probe was also used to screen approximately 6 × 10\(^6\) phage clones of the 129/SvJ mouse genomic library (Stratagene). This yielded the isolation of several genomic clones that were subsequently characterized by blot hybridization of restriction-mapped...
FIGURE 2. The complete 18,769-bp MR1 genomic structure. The following exons are boxed: L (leader peptide), α1, α2, α3, TM (transmembrane), and cytoplasmic/3′ untranslated (Cyt/3′ UT).
fragments (isolation of the bacteriophage P1 containing the MR1 gene is detailed below). Cross-species screenings were conducted as follows. In brief, overnight hybridization was performed at 42°C in a solution containing 50% formamide, 5× Denhart’s solution, 5× SSPE, 0.1% SDS, and 100 μg/ml of denatured salmon sperm DNA (Sigma, St. Louis, MO). The nitrocellulose filters were subsequently treated in a final wash solution containing 2× SSC/0.1% SDS at 45°C. High stringency screening of the human cDNA library with an isogenic probe was performed using standard conditions (44, 45).

The 18,769-bp full-length MR1 gene sequence was assembled as follows. Almost 13 kb was derived from the available λ clones (directly or after subcloning); the remaining 5' 6 kb encoding exon 1 and a large part of exon 1 were obtained employing the following pair of oligonucleotides, 5'-GGTTGATGATGCTCCTGTTACC-3' (derived from leader peptide sequence of murine cDNAs) and 3'-CGGCAATTTCTCGGTCAGAGAC-5' (pan-ultimate sequence within by then available λ DNA), in a PCR reaction using the Expand Long Template kit (Boehringer Mannheim, Mannheim, Germany) and P1 DNA as template following the manufacturer’s recommendations (see below for P1 isolation). Finally, upstream sequence from exon 1 was obtained using the Human GenomeWalker kit (Clontech) following the manufacturer’s recommendations and employing the following primer:

- 3'-GTAACACAATTACCACTGTCGTGTCGCTA-5'
- 3'-CTTGGCTACCGCAAGACATGGAGATT-5'

Double-strand sequencing of cloned DNA fragments (plasmid, λ, or P1) was performed using flanking vector sites and custom-synthesized oligonucleotides (primer walking). The chain termination method was conducted using either the Thermo Sequenase cycle sequencing kit (Amersham, Aylesbury, U.K.) and [35S]dATP (Amersham, UK) or the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq DNA polymerase, FS (Perkin-Elmer, Foster City, CA) according to the manufacturer’s protocol. In the latter technique, the reactions were run on an ABI 373XL DNA Sequencer STRETCH (Perkin-Elmer-Applied Biosystems, Foster City, CA), and the results were analyzed using ABI PRISM Sequencing Analysis, Factura, and Sequence Navigator software (all from Perkin-Elmer-Applied Biosystems). Further sequence analysis employed the EditSeq and MegAlign (default parameters) programs of the LaserGene Navigator (DNastar, Madison, WI).

**RNA blots**

All cell lines were purchased from American Type Culture Collection (Manassas, VA) and grown in the recommended medium at 37°C in an atmosphere supplemented with 5% CO₂. Total RNA was isolated from approximately 10⁷ cells after lysis in a solution containing 4.0 M guanidinium thiocyanate, 0.1 M Tris (pH 7.5), and 1% β-mercaptoethanol, followed by ultracentrifugation (33,000 rpm, 16 h) on a cushion of 5.7 M CsCl/0.01 M EDTA (pH 7.5). Twenty micrograms of total RNA was subjected to electrophoresis on a 1% agarose/formaldehyde gel (44, 45). Upon transfer onto Zeta-Probe GT (Bio-Rad, Richmond, CA), the membrane was hybridized to a randomly [α-³²P]dCTP-labeled MR1 cDNA probe (Ready To Go DNA Labeling Kit, Pharmacia, Uppsala, Sweden) in a solution containing 10% dextran sulfate, 1.5× SSPE, and 0.5% SDS overnight at 65°C. The membrane was subsequently washed at 65°C in a solution containing 0.1× SSC/0.1% SDS. The wet membrane was exposed for 4 days at −80°C to X-OMAT AR film (Eastman Kodak, Rochester, NY) opposing an intensifying screen. The mouse poly(A)⁺ multiple tissue Northern blot was purchased from Clontech and hybridized to a full-length murine MR1 cDNA probe under stringent conditions according to the manufacturer’s specifications. The blot was subsequently exposed for 72 h to an intensifying screen at −70°C. The loading control probe consisted in both cases of a 2-kb human β-actin cDNA fragment (Clontech) and required 2 h of exposure at room temperature.

**Zoo blots**

Eight micrograms of various high m.w. genomic DNA were digested with HindIII (HindIII/BglII for mouse DNA) and analyzed by standard Southern blotting employing the full-length human MR1 cDNA as a probe. DNA from the following species was used in this study: human (a gift from Dr. A. Meyer, Strasbourg, France); primates: bonobo (Pan paniscus), gorilla (Gorilla gorilla), and gibbon (Hylobates lar; all provided by Dr. M. Chorney, Hershey, PA); dog, bovine, porcine, rat (all purchased from Clontech); and finally DBA/2J-C57/BL6 F1 mouse (Basel Institute for Immunology animal facility, Basel, Switzerland). Hybridization and low stringency wash of the blot were performed as described above. The membrane was subsequently exposed for 1 wk to Kodak X-OMAT AR films opposing an intensifying screen.

**FIGURE 3.** Phylogeny. A, Evolutionary conservation of the MR1 locus across mammalian evolution. Note the relative simplicity of bands given the large genomic length. The mouse genomic DNA pattern is the result of cross-species evolution. The mouse genomic DNA pattern is the result of stringency screening of the human cDNA library with an isogenic probe was performed using standard conditions (44, 45).

- **Mouse**
- **Rat**
- **Porcine**
- **Bovine**
- **Dog**
- **Gorilla**
- **Bonobo**
- **Human**

The identity of chromosome 1 is ascertained via double labeling employing a chromosome 1-specific marker (in **B2** and **B3** the sample chromosomes are oriented head-to-head).
**Bacteriophage P1 cloning and FISH**

The following two oligonucleotides, 5'-CAGACCTGTGGTCGTC-3' and 5'-CTAATACGCAGTGTTAG-3', amplifying a unique 275-bp genomic fragment within exon 3 of the murine MR1 gene, were used to PCR screen a 129/SvJ mouse embryonic stem cell genomic P1 library (Genome Systems, St. Louis, MO) following published protocols (45, 46, 47). This yielded isolation of two P1 clones. The identity of one of these was ascertained through amplification and sequence analysis employing the same primer pair mentioned above. This clone, called F298, was subsequently used to FISH-localize the MR1 locus (Genome Systems, St. Louis, MO). In brief, DNA from clone F298 was labeled with digoxigenin dUTP by nick translation. Labeled probe was combined with sheared mouse DNA and hybridized to normal metaphase chromosomes derived from mouse embryoblast cells in a solution containing 50% formamide, 10% dextran sulfate, and 2× SSC. Specific hybridization signals were detected by incubating the hybridized slides in fluorescentlylabeled anti-digoxigenin Abs followed by counterstaining with 4',6-diamidino-2-phenylindole, dihydrochloride. The initial experiment resulted in specific labeling of the distal portion of the largest chromosome pair. A second experiment was conducted in which a probe specific for the centromeric region of chromosome 1 was cohybridized with clone F298. This experiment resulted in the specific labeling of the centromere and the distal portion of chromosome 1. Measurements of 10 specifically hybridized chromosomes 1 demonstrated that F298 is located at a position that is 75% of the distance from the heterochromatic-euchromatic boundary to the telomere of chromosome 1, an area that corresponds to band IG1–2. A total of 80 metaphase cells were analyzed, with 76 exhibiting specific labeling.

**Results**

The human MR1 gene is the latest addition to the restricted family of the non-MHC-encoded class I genes. Beside cDNA cloning, the original report of Hashimoto et al. (43) established the localization of the human gene to chromosome 1q25.3 and ubiquitous transcription of the molecule in various tissues. The present report aims to define the MR1 transcriptional unit; isolate, map, and characterize the murine orthologue; investigate the extent of the phylogenetic conservation of the gene; as well as outline key architectural residues of this molecule.

To define the genomic structure of MR1, a PCR-cloned human cDNA encompassing the entire coding sequence was used to isolate orthologous genomic clones from a murine bacteriophage λ library. However, because of the lack of the entire transcriptional unit within the obtained λ clones, a murine bacteriophage P1 library (average insert size, 90–100 kb) was PCR-screened employing MR1-specific primers. Double-stranded sequence analysis identified MR1 within 18,769 bp of mouse genomic DNA (Figs. 1 and 2; GenBank accession no. AF039526). The structure of the gene is that of a typical MHC-I gene, with respective α1, α2, and α3 extracellular domains, and transmembrane and cytoplasmic sequences encoded by separate exons. Particularities include an unusually large first intron of 8,857 bp. As in other MHC-I and -II genes, each domain starts with a conserved motif (type I splicing) (44) and is flanked by intronic sequences harboring the canonical GT/AG splice junctions. Finally, sequence analysis of 1384 bp putative promoter sequence of human MR1 gene did not reveal any obvious similarity to other MHC-I promoters (see GenBank accession no. AF039526).

A comparative genomic Southern blot (zoo blot) including human, mouse, nonhuman primates, and more distant species revealed the conservation of what appears as a single-copy MR1 gene across mammalian evolution (Fig. 3A). To map the gene within the mouse genome, a bacteriophage P1 clone was used to FISH-localize a single MR1 locus within bands G1–2 of mouse chromosome 1, syntenic to human 1q25.3, the location of the orthologous locus (Fig. 3B).

![Figure 4](http://www.jimmunol.org/) **FIGURE 4.** Ubiquitous expression of several MR1 transcripts in both man and mouse. **A**. Diverse human cell lineages, i.e., Jurkat (T cell leukemia), Ramos (B cell, Burkitt lymphoma), Molt-4 (T cell, acute lymphoblastic leukemia), Raji (B cell; Burkitt lymphoma), HT-1080 (fibrosarcoma), SK-LMS-1 (vulval sarcoma), HeLa (cervical epithelial carcinoma), U-373 (glioblastoma), and HT-29 (colon adenocarcinoma) express MR1 isoforms. **B.** All major murine tissues express a number of MR1 transcripts. β-Actin intensity served as the loading control in both experiments (lower panels). See Figure 5 for the molecular identities of these transcripts.

The original Northern blot experiment reported by Hashimoto and colleagues (43) revealed ubiquitous expression of what appeared as multiple human MR1 transcripts in various tissues examined. However, it was not known whether this reflects universal transcription in all cell lineages or solely within a broadly present single cell type, e.g., endothelium, epithelium, etc. Hence, an equivalent experiment was performed, this time using several human lineage-specific cell lines. The observation of several transcripts in all of these cells confirms truly ubiquitous MR1 expression, resembling the pattern of MHC-Ia transcription (Fig. 4A). Finally, to ascertain a similar situation in the mouse, a multiple tissue Northern blot revealed, as predicted, a broad range of expression for the murine gene, once again present in multiple isoforms (Fig. 4B).

To fully understand the molecular identity of multiple MR1 transcripts, a saturation cDNA cloning strategy was clearly relevant. Extensive screening of human and mouse spleen cDNA libraries coupled with RT-PCR experiments allowed us to completely characterize the complex transcriptional activity of this gene (Figs. 5, 6, and 7). In humans, we identified one RT-PCR-cloned transcript identical with the original 1263-bp cDNA clone complete match with STS_U22963 (locus G26705) positioned at 838.4 cR from the top of the chromosome 1 linkage group.
reported by Hashimoto and colleagues (43), hereafter called hMR1A, as well as three other cDNA species, respectively designated hMR1B, hMR1C, and hMR1D (Figs 5 and 6). Interestingly, all newly described transcripts lack the Ig-like \(\alpha_3\) domain. Moreover, the 1578-bp hMR1C transcript bears a stop codon a few amino acids after the end of the \(\alpha_2\) domain and therefore encodes a putative soluble isoform, whereas the 776-bp hMR1B and the 2046-bp hMR1D remain putative integral transmembrane proteins (Figs. 5 and 6). In the mouse, six distinct cDNA species were delineated (Fig. 7). The mMR1A cDNA of 2406 bp encodes a bona fide MHC-I molecule organized similar to the human orthologue. However, due to the presence of a cryptic splice donor site 77 bp inside the \(\alpha_1\) domain and through a scheme best described as partial exon skipping, several transcripts containing aborted open reading frames were generated through splicing of this initial exon 2 fragment directly to the beginning of exon 3. These include the mMR1B, -C, -E, and -F of 2130, 1641, 1115, and 2854 bp, respectively. The mMR1D is a 3219-bp truncated, partially spliced transcript carrying part of intron 3 (Figs. 5 and 7). Finally, the extensive length heterogeneity observed among both human and murine MR1 transcripts is due to the respective sizes of their 3' untranslated regions, a fact readily visible on Northern blot experiments (see above).

Comparing individual domains of human and murine MR1 polypeptides reveals an unusually high degree of sequence similarity (i.e., \(\alpha_1\), 90%; \(\alpha_2\), 89%; and \(\alpha_3\), 73%; see Figs. 8 and 9). The same comparison including representatives of other MHC-I molecules reveals, as expected, conservation of almost all so-called invariant residues, mostly thought to maintain the class I scaffold (48, 49). Moreover, this analysis defines at least three novel pan-MHC-I conserved amino acids (underlined hereafter). These residues are P20, G26, V28, D29, S38, P47, W51, W60, E/D61, T64, and the N86XS glycosylation site within the \(\alpha_1\) domain; G91, H93, Q96, M98, G100, C101, G112, Y118, D119, G120, W133, C164, L168, R170, L172, and G175 throughout the \(\alpha_2\) domain (Fig. 8); and P185, C203, F208, Y209, P210, P235, ...
FIGURE 6. Human cDNA clones and sequences. The first sequence is the original cDNA sequence identified by Hashimoto et al. (43); all other sequences were uncovered during the present study and are respectively called hMR1A-D.
D238, C259, V261, and H263 in the Ig-like α3 domain. Among the eight residues defined by high resolution crystallographic analysis to contact peptide termini in MHC-Ia molecules (50), only three are conserved in MR1, i.e., Y7, T143, and W147 (Figs. 8 and 9). Finally, homologous residues to positions 223 to 229, 245, and 247 of the MHC-Ia α3 domain, indispensable for interaction with the CD8 coreceptor and present in both human and mouse MHC-encoded class I molecules (51, 52) (Figs. 8 and 9). Finally, homologous residues to positions 223 to 229, 245, and 247 of the MHC-Ia α3 domain, indispensable for interaction with the CD8 coreceptor and present in both human and mouse MHC-encoded class I molecules (51, 52).
are clearly absent from human and mouse MR1 (Fig. 8).

Exploring the available human and murine MR1 sequences potentially originating from three and six distinct chromosomes, respectively (Refs. 43 and 53 and sequences reported here), we conclude that unlike MHC-Ia or MIC genes (2, 32) MR1 does not seem to be overtly polymorphic, as not a single nucleotide difference was detected within respective coding sequences.

**Discussion**

The 90% sequence identity between human and mouse MR1 molecules is truly remarkable, not seen either in the non-MHC-I genes, where the level of identity does not exceed 60(α1)/62(α2)% for CD1, 59/74% for Znα2gp, 69/66% for FeRn, and 78/76% for HFE, or in the MHC-encoded class Ia and Ib molecules, which have average human-mouse similarities of 69/70% and 51/41%, respectively (Fig. 9). This outstanding degree of trans-species sequence identity in MR1 should naturally include preservation of key structural/functional residues. Keeping this in mind, examination of such positions reveals interesting information. As previously noted, most if not all structural residues conserved across all vertebrate MHC-I molecules are also maintained in MR1, predicting folding of the putative MR1 polypeptide chain in an analogous three-dimensional configuration. Moreover, the N86XS glycosylation site present in all human and mouse MHC-Ia and -ib, HFE, and human Znα2gp, but absent from all CD1 isotypes as well as from MICA/MICB molecules, is present in both human and murine MR1 (Fig. 8).

The genomic organization of MHC class I genes parallels their multidomain polypeptide chain configuration. Indeed, each extracellular domain is encoded by a separate exon, preceded by the leader sequence and followed by transmembrane, cytoplasmic, and 3′-untranslated sequences (one or more depending on the locus; Fig. 1) (44). Examination of the large number of genomic structures available reveals, with the exception of CD1, a clear dichotomy between the structurally homogeneous HLA-A-G genes (mouse H2-K, L, D, Q, T, and M) and those encoding the divergent (both functionally and structurally) MICA/B, HFE, FeRn, Znα2gp, and MR1 molecules; this is regardless of their physical location in the genome, i.e., whether they are located within the MHC proper (Fig. 1). Indeed, the former group averages a total gene length of 3 to 5 kb, whereas the latter easily exceeds 6 kb, with 6.8 kb for HFE, 10 kb for Znα2gp, 11 kb for FeRn, 12 kb for MICA/B, and close to 19 kb for the MR1 gene reported here (Fig. 1). With the exception of FeRn, where this is due to a lengthy fourth intron (between exons 3 and 4), the extra length is due to an unusually large first intron preceding the α1 domain, reaching a peak stretch of 8857 bp in MR1 (Fig. 2). Whether this unusual length has any functional repercussions, possibly at the transcriptional level, remains to be seen. Finally, it is noteworthy that large MHC-I genes are not exclusive to mammals, as, for instance, several carp genes have been documented with first introns up to 14 kb (54).

Alternative splicing has long been documented for MHC class I genes (55–57) and involves mainly transmembrane and cytoplasmic sequences. The true physiologic contribution of various isoforms, however, is not known, especially given the fact that they are severely under-represented compared with the canonical transcript; the majority are only detectable upon PCR amplification. Thus, high expression of a plethora of MR1 isoforms in both human and mouse, all readily visible on Northern blots, is unique
FIGURE 9. Sequence distance among members of the MHC-I family. Domain-by-domain sequence analysis of representatives of human and murine classical and nonclassical MHC-I genes was performed, based upon sequence alignment employing default parameters of the Clustal method, MegAlign program, Lasergene Navigator (DNAstar). Relevant comparisons are highlighted (box and arrow) and additionally bold-italicized with regard to MR1.
among all MHC-I (Fig. 4). In man, the apparent over-representation of transcripts lacking the Ig-like α3 domain may have functional relevance, although this can only be verified once specific Abs to various extracellular domains are available.

MHC-I genes constitute diverse multigenic families in various vertebrate species (58–60). On the other hand, non-MHC-encoded class I genes are, in general, more simply arranged, although the human Znα2gp locus contains a pseudogene in addition to the functional locus (28), and the CD1 loci are present in various configurations in different species (20). Within this context, both the low complexity of genomic bands seen in our zoo blot experiments and especially observation of single hybridization spots in both human and mouse FISH favor the existence of a single MR1 locus throughout mammalian evolution, in clear contrast therefore to most MHC-encoded and to some non-MHC-I-encoded genes. Given the strong sequence conservation between such phylogenetically distant species as *Homo sapiens* and *Mus musculus*, these data are largely in favor of strong selective pressure maintaining MR1 structure. Finally, the location of MR1 on human chromosome 1 might not be fortuitous, as the relative proximity of the CD1 locus is intriguing. Indeed, capitalizing on theories first forwarded by Ohno, who proposed creation of the present day vertebrate genomes via tetraploidization (61), a process recently verified through sequence analysis of the complete yeast genome (62), Kasahara and colleagues identified another MHC-like gene cluster outside chromosome 6 in man and chromosome 17 in mouse (9q33–34 in man and chromosome 2 in mouse) harboring elements resembling those found in the MHC itself, e.g., ATP-binding cassette transporters (such as the MHC-encoded TAP molecules), proteasome Z subunit (such as MHC-encoded LMP2 and -7), Grp78 stress protein (such as the MHC-encoded 70-kDa heat shock protein genes), complement C5 gene (such as the MHC-encoded C2/Bf/C4 molecules), etc. (63). Through similar observations, this time regarding some nonimmunee genes (NOTCH and PBX), Katsanis and co-workers added the long arm of chromosome 1 as yet another paralogous locus (64). However, the true MHC signature, i.e., MHC-I and -II genes, has not been found or noted on either 9q34 or 1q. In this light, the relatively close proximity of human CD1 and MR1 loci might parallel for the first time the human 6p21.3 (mouse 17) MHC-II/III topology and therefore define an authentic MHC paralogous locus. In this respect, it is interesting to note that through many functional/structural aspects (see above) CD1 is more closely related to MHC-II than to MHC-I. Given these facts, CD1 could be considered the chromosome 1-encoded MHC-II parologue and MR1 an authentic MHC-I equivalent. Moreover, several other key genes chart this genomic segment and therefore contribute to its emergence as yet another MHC paralogous locus. These are 1) the retinoid X receptor γ located between MR1 and CD1; interestingly, the two other RXRs genes are located, respectively, in the MHC (retinoid X receptor β) and chromosome 9q34 (retinoid X receptor α) (65); 2) the recently mapped proteasome β subunit PSMB4 to 1q21 (66), in comparison to other β subunits encoded within 9q34 for Z (PSMB7) (67) and the MHC for LMP2/7 (PSMB9/8) (3); 3) the 1q25-3q22-located tenasin-R gene (67) paralogous to MHC-encoded tenasin-X molecule (3); and 4) the so-called regulator of the complement activation (RCA) gene cluster comprising complement receptors type 1 and 2, decay-accelerating factor, and C4 binding protein, closely linked on 1q32 (65); reminiscent perhaps of a somewhat similar organization within the MHC, where the complement genes C2, BF, and C4 are sandwiched between MHC-I and -II loci (3) and 9q34, the location of yet another complement subunit, C5 (65). Finally, a close inspection of the other loci neighboring MR1 and CD1 reveals the existence of an unusually large number of immune-related loci. These are IL6RA, Ly9, CD48, CD3z, CD64, LYAM1, ELAM1, OX40L, IL10, CD34, CRP, and a constellation of Fc receptors, i.e., FCGRI1, FCR1A, FCR1G, FCGR, and FCGR3A and -B (65).

The relatively high and truly ubiquitous transcription of MR1 is unique among the so-called nonclassical MHC-I genes (Fig. 4). Human CD1a, -b, and -c, for instance, are mainly expressed on cortical thymocytes and Langerhans cells, whereas both human and mouse CD1d seem to be mainly transcribed in intestinal epithelial cells (20). The expression pattern of neonatal FcRn is selective to the intestinal brush border and placental syncytiotrophoblast (26). Znα2gp is highly prevalent in breast fluids (28), and HFE expression appears to be confined to the gut (36). Even among the MHC-encoded nonclassical genes and with the exception of human HLA-E (5), there is an evident degree of compartmentalization, e.g., epithelia for MICA/B (31), trophoblasts for HLA-G (14), intestine or thymocytes for most TL genes (15), and so on. Within this context the wide cellular expression of MR1 is clearly reminiscent of ubiquitous classical MHC-I molecules and is therefore in favor of a general physiologic function.

MHC-Ia molecules show an extraordinary degree of polymorphism, with more than 300 alleles collectively defined for human HLA-A/B/C genes (2). In clear contrast, MHC-Ib genes, with the important exception of the MIC loci (32), are not polymorphic, a situation that has led some authors to postulate their presentation of invariant ligands (68), for instance glycolipidic moieties by CD1 molecules to CD4, CD8 double-negative αβ, or γδ T lymphocytes (20). Following this rationale, the apparent monomorphism of MR1 favors possible interaction with a putative invariant ligand.

In summary, MR1 is a broadly expressed, phylogenetically conserved, invariant class I molecule located within an MHC-paralogous locus. Although at this point any suggestions regarding MR1 function within the immune system are highly speculative, possible interaction with a growing family of orphan leukocyte Ig superfamily receptors, whose founding members interact with a diverse set of MHC-I molecules, warrants further investigation (69).

Acknowledgments

S.B. thanks Georges Hauptmann for encouragement. We thank Nassima Fodil for help with genomic cloning, and Drs Marco Colonna, Louis Du Pasquier, and Susan Gilliland for critical reading of this manuscript.

References


