Primordial Linkage of β2-Microglobulin to the MHC

Yuko Ohta, Takashi Shiina, Rebecca L. Lohr, Kazuyoshi Hosomichi, Toni I. Pollin, Edward J. Heist, Shingo Suzuki, Hidetoshi Inoko and Martin F. Flajnik

*J Immunol* 2011; 186:3563-3571; Prepublished online 14 February 2011;
doi: 10.4049/jimmunol.1003933
http://www.jimmunol.org/content/186/6/3563

**Supplementary Material**
http://www.jimmunol.org/content-suppl/2011/02/14/jimmunol.1003933

**References**
This article cites 47 articles, 15 of which you can access for free at:
http://www.jimmunol.org/content/186/6/3563.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
Primordial Linkage of β2-Microglobulin to the MHC

Yuko Ohta,*†1 Takashi Shiina,*†1 Rebecca L. Lohr,* Kazuyoshi Hosomichi,†‡1 Toni I. Pollin,§ Edward J. Heist,§ Shingo Suzuki,† Hidetoshi Inoko,†‡1 and Martin F. Flajnik**,†1

β2-Microglobulin (β2M) is believed to have arisen in a basal jaw vertebrate (gnathostome) and is the essential L chain that associates with most MHC class I molecules. It contains a distinctive molecular structure called a constant-1 Ig superfamily domain, which is shared with other adaptive immune molecules including MHC class I and class II. Despite its structural similarity to class I and class II and its conserved function, β2M is encoded outside the MHC in all examined species from bony fish to mammals, but it is assumed to have translocated from its original location within the MHC early in gnathostome evolution. We screened a nurse shark bacterial artificial chromosome library and isolated clones containing β2M genes. A gene present in the MHC of all other vertebrates (ring3) was found in the bacterial artificial chromosome clone, and the close linkage of ring3 and β2M to MHC class I and class II genes was determined by single-strand conformational polymorphism and allele-specific PCR. This study satisfies the long-held conjecture that β2M was linked to the primordial MHC (Ur MHC); furthermore, the apparent stability of the shark genome may yield other genes predicted to have had a primordial association with the MHC specifically and with immunity in general. The Journal of Immunology, 2011, 186: 3563–3571.

*Department of Microbiology and Immunology, University of Maryland, Baltimore, MD 21201; †Division of Molecular Life Science, Department of Genetic Information, Tokai University School of Medicine, Itochukarana, Kanagawa 243-1143, Japan; ‡School of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan; §Department of Medicine, University of Maryland, Baltimore, MD 21201; and †Department of Zoology, Fisheries and Illinois Aquaculture Center, Southern Illinois University Carbondale, Carbondale, IL 62901

Y.O., T.S., H.I., and M.F.F. contributed equally to this work.

Received for publication November 30, 2010. Accepted for publication January 6, 2011.

This work was supported by National Institutes of Health Grant AI27877 (to Y.O., R.L.L., and M.F.F.) and by Scientific Research on Priority Areas “Comparative Genomics” (20017023) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to T.S., K.H., S.S., and H.I.). E.H. was supported by the Department of Zoology at Southern Illinois University Carbondale and by the W.W. Diehl Endowed Professorship of Biology to J. Carrier at Albion College.

The sequences presented in this article have been submitted to the DNA Data Bank of Japan under accession numbers AB571627 and to GenBank under accession numbers HM626581 and HM626583.

Address correspondence and reprint requests to Yuko Ohta, Department of Microbiology and Immunology, University of Maryland, 685 West Baltimore Street, Baltimore, MD 21201. E-mail address: yota@som.umaryland.edu

The online version of this article contains supplemental material.

Abbreviations used in this article: BAC, bacterial artificial chromosome; C1, constant-1; IgSF, Ig superfamily; LOD, log of the odds; β2M, β2-microglobulin; NJ, neighbor-joining; PBR, peptide-binding region; ssCP, single-strand conformational polymorphism; ZFP, zinc finger protein.

Copyright © 2011 by The American Association of Immunologists, Inc. 0022-1767/11/S16.00
Materials and Methods

Animals
Genomic DNA was isolated from RBCs for mapping analysis from the nurse shark family as previously described (21). The procedure of animal use was reviewed and approved by the Institutional Animal Care and Use Committee at the University of Maryland.

Bacterial artificial chromosome library screening
The 17 bacterial artificial chromosome (BAC) filters with 11-fold genomic coverage (22) were screened with radiolabeled full-length b2M or ring3 probes under high-stringency conditions (23). Membranes were exposed to x-ray film for various lengths of time to obtain positive signals and the desired background. Putative positive clones were then re- spotted on nylon membranes for colony hybridization and tested by Southern blotting to confirm true positives. BAC insert DNA was isolated using the PhasePrep BAC DNA kit (Sigma-Aldrich), and the sequence was determined by shotgun sequencing at the sequencing facility at Tokai University with 7.5% coverage.

Sequence alignment and phylogenetic tree
Amino acid sequences of constant-1 (C1)-IgSF domains were aligned using the ClustalW program with minor adjustments. A rooted neighbor-joining (NJ) bootstrapped (1000 runs) phylogenetic tree (24) was constructed, and the consensus tree was then viewed with the TreeView program (25).

Database searches
Genome synteny in various species was retrieved and analyzed from publicly available Web sites as noted. Genes from mouse, chicken, human, opossum, and zebrafish were retrieved from GenBank (http://www.ncbi.nlm.nih.gov), and information on other genomes was retrieved from the following Web sites: elephant shark genome (http://blast.fugu.sg.org/); Anolis genome (http://genome.ucsc.edu/cgi-bin/hgGateway?db=anoCar1); Xenopus genome (http://genome.jgi-psf.org/Xennu/Xennu.home.html); and Fuq genome (http://genome.jgi-psf.org/Takru4/Takru4.home.html).

In-house EST collection
We constructed the cDNA library using the Gateway System (Invitrogen) from adult nurse shark pancreas. To eliminate Ig H and L chain probes under high-stringency conditions. Negative colonies (∼8000) were then manually picked and sequenced from the vector end. All draft sequences were blastx searched against GenBank databases, and rare obtained ∼1150 sequences not specific to the pancreatic enzymes (Y. Ohta and M.F. Flajnik, personal observations).

Single-strand conformation polymorphism analysis
Nurse shark ring3 primers were designed based on the sequence obtained from BAC GC614H19 clone. Multiple primers were tried, and we selected the primer set anchoring exons 4 and 5 for the single-strand conformation polymorphism (ssCP) analysis. The primers were exon 4 forward, 5′-GTAAACACTGGACAAATT-3′; and exon 5 reverse, 5′-ATTGGGAAGCTGGACAGAT-3′. PCR amplification was performed for 4 min at 94°C, followed by 35 cycles of 94°C for 1 min, 56°C for 1–2 min, 72°C for 1 min, and a final extension at 72°C for 10 min. An ∼550-bp fragment amplicon was cloned into the pCRII TA cloning vector (Invitrogen), and individual clones were sequenced. Nurse shark families 2 and 3 were genotyped using 12 DNA microsatellite markers and assigned sires (E.J. Heist, J.C. Carrier, H.L. Pratt, and T.C. Pratt, submitted for publication).

Statistical analysis of linkage
We used parametric linkage analysis to formally assess the evidence for linkage of b2M to the MHC region in the offspring of deletion-carrying sires. This approach assesses the odds of the likelihood of obtaining the observed data set if the two loci are linked versus if the loci are not linked, showing as a log of the odds (LOD) score. The paternal sibships were determined based on consolidated data from combination of Southern blotting, sequencing of MHC class Ia alleles, and microsatellite analyses (shown in Table I).

The LOD score is calculated as follows when parental phase (linkage status) is known: LOD = log10\{[(10^{−b} + 1 − θb)/(0.5 + NR)] − θR\}, where θ is the recombination fraction, NR is the number of nonrecombinant offspring, and R is the number of recombinant offspring.

Because the parental phase was unknown in the current study due to a lack of grandparental genotypes, a phase ambiguous LOD score was first calculated for each family by taking the log of the average odds for the two possible phases (1 and 2 in Table I), and the resulting LOD scores were then summed over the two families to obtain the LOD score at a given recombination fraction. LOD scores were calculated at recombination fractions between 0 and 0.5 to obtain the recombination fraction where the LOD score was maximized (26). The corresponding p value was calculated using a one-sided χ2 test of LOD × 2 log(10) (27).

Results
Characterization of nurse shark b2M
Cartilaginous fish are the oldest living vertebrates having an adaptive immune system centered upon Ig, TCR, and MHC (1). When it was suggested that class I and class II genes may have evolved in separate linkage groups from studies of teleost fish (28), we demonstrated in family studies that the two MHC classes were closely linked in two shark species, nurse shark and houndshark (21). To gain further insight into the primordial MHC organization, we have isolated many shark genes associated with adaptive immunity, including b2M. The full-length b2M clone was found in an in-house EST collection (GenBank accession number HM625831), as well as from a previously published genomic sequence (GenBank accession number GQ865623) (29), and the deduced amino acid sequence was aligned with b2M from other species (S1). As was noted in previous studies, evolutionarily
 conserved residues are either found in all C1-IgSF (or just IgSF) domains (29, 30) or are predicted to be at class Ia-chain interaction sites (31). Some cartilaginous fish β2M have potential N-glycosylation sites that are rare in tetrapods but present in several bony fish species (32). Consistent with previous studies (33, 34), phylogenetic tree analysis revealed that cartilaginous fish β2M clustered with the orthologous proteins and to the IgSF domains of MHC class IIA/DMA, suggesting that they share the most recent common ancestor (Fig. 1A). Also consistent with previous studies (33), the IgSF domains of class IIB and class Ia shared the most recent common ancestor. β2M expression pattern seems to coincide with MHC class I expression (Fig. 1B).

**Mapping of β2M to the MHC in family studies**

Two families of nurse sharks previously were used to map several genes to the MHC (21, 36, 37). All of these families showed multiple paternity, at least five fathers in family 1 and seven in family 2. Southern blotting analysis using many restriction enzymes demonstrated that β2M is a single-copy gene (five representative digestions are shown in Fig. 1C); unfortunately, no RFLPs were obtained to test the linkage status, and thus we sequenced the gene from animals with different MHC haplotypes, hoping to find polymorphisms. A two-nucleotide deletion was detected in one of the paternal β2M alleles “p3” from groups “i” (p3/m2) and “j,” (p3/m1) from family 2 with 39 members (Fig. 1).

![FIGURE 1.](http://www.jimmunol.org/)

**Mapping of β2M to the MHC in family studies**

Two families of nurse sharks previously were used to map several genes to the MHC (21, 36, 37). All of these families showed multiple paternity, at least five fathers in family 1 and seven in family 2. Southern blotting analysis using many restriction enzymes demonstrated that β2M is a single-copy gene (five representative digestions are shown in Fig. 1C); unfortunately, no RFLPs were obtained to test the linkage status, and thus we sequenced the gene from animals with different MHC haplotypes, hoping to find polymorphisms. A two-nucleotide deletion was detected in one of the paternal β2M alleles “p3” from groups “i” (p3/m2) and “j,” (p3/m1) from family 2 with 39 members (Fig. 1).
FIGURE 2. The shark β2M is linked to the MHC. A, The two-nucleotide (CC) deletion polymorphism was found in intron 2 of β2m sequences in "p3" paternal allele from siblings belonging to the groups "i" and "j." Thus, allele-specific primers were designed based on this polymorphism. All primers are underlined. The ends of coding regions are boxed. The (AG) at the end of intron 2 is underlined.

B, PCR was carried out with a combination of allele-specific and universal NSB2mEx3Rev reverse primers. Presence or absence of the amplicon using the "p3"-specific primers was used for typing (top gel) the family 2 with 39 offsprings. Maternal primers were used for the positive control (bottom gel). Forward primers are indicated on the left side of the gels, and mother and sibling numbers are indicated above the gel along with MHC groups (36).

C, Allele-specific PCR in the families 1 and 3. Only two animals belonging to the MHC groups "h" possessed the "CC-deletion" allele, and two animals belonging to the "g" groups had this allele in family 3. We partially typed family 3 based on the MHC groups by sequencing of the PBR of the class Ia alleles (maternal and paternal alleles are designated as numbers above the gel) and by Southern blotting with a probe containing MHC class Ia leader and α1 domains (small dot, band for maternal haplotype 1; large dot, maternal haplotype 2). The "p2" allele of the "g" group is the only haplotype possessing the "CC-deletion" allele of β2M. D, Plot of LOD scores at corresponding recombination fractions. The sums of the two families were used (Supplemental Table I).
and “j” groups; thus this father had the MHC haplotypes shown to have been sired by the same father as offspring in the “i” obtained a maximum LOD score of 3.14 [1378:1 odds of linkage analysis (26) to evaluate the evidence for linkage. We identified other BAC clones that were either β2M- or ring3-single-positive; unfortunately, none of them was positive for other MHC genes, again consistent with larger intergenic distances in sharks compared with those of other species (36). Chen et al. (29) drew a premature conclusion of non-MHC linkage; however, determining the linkage status of β2M (or almost any gene) based on a single BAC sequence is not sufficient for the shark genome, where there are large intragenic and intergenic distances. Several nurse shark BAC clones (22) were isolated with the ring3 and β2M probes, and some of them were positive for both genes. As previously reported (29), the β2M gene contains at least three exons, having a similar genomic organization and size to other species. The shark ring3 gene spans ∼20 kb and contains 12 exons, which is approximately twice as large as mammalian ring3 genes (e.g., 12.8 kb and 9.7 kb for human and mouse, respectively), consistent with a larger gene size found in most shark MHC genes (36). Sequencing through an entire BAC clone (GC_614H19) confirmed that the β2M and ring3 genes were adjacent to each other ∼45 kb apart (Fig. 4).

**Genetic descent of β2M**
The chromosomal location of the β2M gene varies greatly among vertebrate species (Fig. 5). Genomic synteny is well conserved in...
the region of chicken β2M relative to humans except for deletions of certain genes (43), and the same seems to be true for the Anolis lizard in which the synteny near the β2M gene (GenBank accession number FG703784, etc.) is conserved (genomic scaffold-670, 634,364 bp) (Supplemental Table II). Mouse β2M is linked to the so-called minor histocompatibility complex on chromosome 2 (16) and is located within a small region syntenic to human chromosome 15 (43). Notably, a smaller syntenic block is embedded with genes mapping to human chromosome 14q11.2 in a marsupial, the opossum. Although these regions can be accounted for by block translocations or syntenic breakpoints, synteny is not conserved in species from lower vertebrate classes as β2M is surrounded by genes mapping to various human chromosomes. The amphibian Xenopus β2M is linked to the genes mapping to human chromosomes 16 and 17 (genomic scaffold-673). In zebrafish, β2M (chromosome 4) is surrounded by genes mapping to human chromosome 12p12, and various locations in the human genome have syntenic regions on the Fugu scaffold-171 (638,182 bp). As mentioned above, the teleost fish experienced a recent genome-wide duplication (“3R”), and there is another β2M locus in the zebrafish genome that is ~60% similar to its paralogue at the amino acid level. Notably, the second β2M locus is found at the telomeric region of chromosome 8 and is distantly linked to a class IIA gene and two class Ib genes of the L-lineage (44) (Supplemental Table II). Although the β2M linkage is not very close (i.e., 6.5 Mbp apart) in this chromosomal region (considering the rapid reorganization of syntenic regions in the teleost fish), this linkage group of class II/class I/β2M is likely a vestige of the primordial synteny. Combining all of the evidence, our study in nurse shark demonstrates that β2M was originally encoded in the MHC, and from extensive database analysis in many taxa, this gene underwent multiple translocations in gnathostomes, either stepwise or independently from the MHC (Fig. 5).

**Discussion**

Compared with other vertebrate models (e.g., chicken or teleost fish), the shark genome seems to be stable, first demonstrated with the linkage of MHC class I and II genes (21), which was lost in bony fish (28), and later with linkage conservation of genes found in the mammalian MHC class III region (37). These MHC linkage data are consistent with global genomic studies in the elephant shark suggesting that cartilaginous fish have greater preservation of synteny than is found in any teleost model (45, 46). The β2M linkage to the shark MHC demonstrated here is likely the primordial condition, thus further supporting the conservation of the cartilaginous fish genome. Furthermore, the close proximity of class I, class II, and β2M is consistent with the theory that they were derived from a common ancestor by tandem (cis) duplication. The close linkage of β2M and class I may have regulated their original coordinated expression and upregulation. Class I and β2M expression is nearly identical in the nurse shark (Fig. 1B), but in other vertebrates β2M is made in excess (47). Furthermore, the number of β2M loci is expanded in rainbow trout (48) and polyploid Xenopus species (18).

Unlike class II genes, class I genes are extraordinarily plastic. Besides the MHC-linked classical class Ia genes, there are also many nonclassical class Ib genes with varied functions, some encoded in the MHC and others not. The majority of class Ib proteins associates with β2M as well, and it has been speculated that there was an advantage of translocation of β2M out of the MHC so that it would not be subject to duplications and deletions (19), like class I genes in many vertebrates. Consistent with the idea of maintaining genomic stability, but in contrast to class I and class II genes, both β2M and ring3 genes are in a very stable part of the shark MHC, with very few polymorphisms and transposable elements (Fig. 4); there was no polymorphism detected by using restriction enzymes/Southern blotting with either the ring3 or β2M probe. Although there are a few bony fish species in which
the number of β2M loci has been expanded (49), and there are two loci in the tetraploid *Xenopus laevis* (18), generally these species are exceptions. There seems to be only one β2M locus in the nurse shark genome, because genomic Southern blotting with many restriction enzymes yielded a single band with an exon-specific probe (Fig. 1C).

The primordial linkage of β2M to the MHC does not contribute to the debate on which gene came first, class I or class II. Among the various IgSF domains, the C1-type is a rare form, found primarily in molecules associated with adaptive immunity (50). Therefore, it is reasonable to propose that C1-type IgSF-encoding genes like β2M were present in the “proto-MHC,” which then acquired the PBR from another gene family. Furthermore, it has been speculated that all molecules containing C1-type IgSF domains arose from a common ancestor, and thus an Ig/TCR precursor may have originated from the “proto-MHC” (20). Consistent with previous studies dating back almost 30 y (3, 5, 33, 34), our phylogenetic analysis demonstrated a common origin for the class IIA/DMB/β2M and the class Ia/DMA/class IIB lineages, and all of these genes share an ancestral C1 domain-encoding exon that emerged after the split between Ag receptors and MHC genes (Fig. 1B). Whereas class IIA, β2M, class IIB, and class Ia share an immediate common ancestor that arose by tandem duplication from the ancestral molecule, each DM gene was apparently generated by tandem duplications of class IIA and class IIB, perhaps early after the emergence of tetrapods, as no DM genes have been found in the teleost or cartilaginous fish; the maximum likelihood and Bayesian inference trees favor this scenario (S2). The NJ tree (Fig. 1B), however, suggests that shark class IIA and IIB genes cluster with class II genes from other species rather than at the basal position of class II/DM, suggesting that sharks may indeed possess DM.

An orthologous gene related to the ancestor of ring3 is present in the urochordate (e.g., amphioxus) “proto-MHC” (42), and thus the MHC-linkage of ring3 in sharks is not surprising. To determine the linkage status in other cartilaginous fish species, we examined the elephant shark genome. Current analyses of the elephant shark genome (46) has yielded only short (<1 kbp) scaffolds (AAVX01540028.1) in which we only identified the β2M C1 domain. Three scaffolds were found to contain some exons of the elephant shark ring3 gene [AAVX01538535 (754 bp), AAVX01069837 (5232 bp), AAVX01012433 (4324 bp)]; however, the assembly is still in its early stages. Further progress in this genome project will reveal the synteny around β2M and all of the other MHC genes and likely provide insight into the natural history of the adaptive immune system by revealing other genes that have been translocated out of the MHC during vertebrate evolution. For example, there is good evidence from various vertebrates that both IgSF- and C-type lectin-containing NK cell receptor genes (in humans, they are encoded in leukocyte receptor complex and NK complex, respectively) and...
MHC were genetically linked at an early point in vertebrate evolution (20, 51, 52), suggesting that NK receptors co-evolved with MHC proteins. We have found a fragment of a zinc finger protein (ZFP), ZFP112-like, in BAC clone GC_614H19, adjacent to βM (Fig. 5). ZFP112 is found on human chromosome 19q13.3 near FeRn (19q13.3), a nonclassical class Ib molecule, and the leukocyte receptor complex (19q13.4). This region had been suggested to be an MHC paralogous region by pericentric inversion of 19p13.1. Whether the nurse shark ZNF112 is a pseudogene or divergent from human/rodent ZFP112 genes, the linkage of ZFP112 suggests that the linkage of NK receptor(s) and MHC could be preserved in the shark genome. Furthermore, we found βM on the same chromosome as TCRα/δ in horse (chromosome 1), cow (chromosome 10), and both TCRα/β and Ig in the opossum genome (Fig. 5, Supplemental Table II). In addition, Ag receptor loci and other genes involved in immune defense (e.g., B7 ligands and Fc-like receptors) are linked to genes related to the Xenopus MHC (Y. Ohta and M.F. Flajnik, manuscripts in preparation), and cathepsins S and L are found on MHC paralogous regions in mammals (20). Such evidence is consistent with our hypothesis that Ag receptors (TCR, Ig), NK receptors, and other genes involved in Ag processing and generally in immune function might have been linked in a “pre-adaptive immune complex” in the ancestral configuration.

Acknowledgments
We thank Dr. Mike Crisitiello and Caitlin Doremus for critical reading.

Disclosures
The authors have no financial conflicts of interest.

References


