Supporting Information

Huang et al. 10.1073/pnas.0903196106

SI Materials and Methods

Cloning and Expression of Bovine *Mr1* cDNA. Bovine *Mr1* cDNA was amplified from the total *Bos taurus* RNA (Zyagen) by RT-PCR. The cDNA was cloned with the TOPO TA cloning vector (Invitrogen) and sequenced by Washington University's Protein & Nucleic Acid Chemistry Laboratory. The full-length cDNA was cloned with the vector pMX.IRES.Puromycin or pMX.IRES.Blasticidin modified from the pMX retroviral expression system. Retrovirus-containing supernatant was generated and used for transduction. The sequence analyses and alignment were performed with Vector NTI Suite software (Invitrogen).

Analyses of Synonymous and Nonsynonymous Substitution. Encoding sequences for MR1 and class Ia molecules were aligned and used for the analysis of synonymous and nonsynonymous substitution. The 7 *Mr1* sequences (with GenBank accession number) were from human (U22963), chimpanzee (AJ275982), orangutan (CR858109), Rhesus monkey (XM_001110940), bovine (EU792881), rat (Y13972), and mouse (AF010448). The 29 class Ia alleles included HLA-A0205, HLA-A0302, HLA-B37, and HLA-Cw0303 from human; Patr-A_0101, Patr-A_0501, Patr-B_0301, and Patr-C_0301 from chimpanzee; Popy-A_01, Popy-B_0401, Popy-B_0801, and Popy-C_0203 from orangutan; Mamu-A2_0101, Mamu-A6_0102, Mamu-B_8001, and Mamu-B_1902 from Rhesus monkey; BoLA-N_00401, BoLA-N_01502, BoLA-N_01801, and BoLA-N_03101 from bovine; RT1-A1_f, RT1-$A1^q$, RT1-$A2^b$, and RT1-$A2^q$ from rat; and H-$2D^b$, H-$2L^d$, H-$2D^k$, H-$2K^b$, and H-$2K^d$ from mouse. The proportion of synonymous (*Ps*) and nonsynonymous substitution (*Pn*) was calculated using the Nei–Gojobori algorithm (1) and a sliding-window model for each continuous 30-codon set with PSWIN software (2). The ARS and non-ARS regions of MHC class Ia and MR1 molecules also were tested for synonymous and nonsynonymous substitution by pairwise comparisons based on a modified Nei–Gojobori algorithm (1, 3) using the SNAP program (http://www.hiv.lanl.gov) (4).

Expression and Purification of Recombinant MR1 Protein. Recombinant mMR1 protein was generated using baculovirus coexpressing GFP and a bicistronic construct encoding the honey bee melittin signal peptide, the ectodomain of the mMR1(α1–2)/L^d(α3), thrombin site, His tag, foot-and-mouth disease virus–encoded 2A peptide, and mouse β2m (allele b) protein. The soluble mMR1/L^d/β2m complex was purified from the supernatant of Hi-Five cells infected with the recombinant baculovirus using a Ni-NTA affinity column, followed by size-exclusion chromatography. The expression level was confirmed with SDS/PAGE and Western blot analysis using anti-MR1 mAb 4E3.

SI Results and Discussion

Fig. S3 shows that synonymous substitution was predominant in *Mr1*. To study the evolutionary constraints on *Mr1* gene products, we compared the synonymous and nonsynonymous substitutions between *Mr1* and class Ia gene orthologs. Higher nonsynonymous substitution (*Pn-Ps* >0) implicates positive Darwinian selection of a favored phenotype. Higher synonymous substitution (*Pn-Ps* <0) suggests purifying or negative selection to maintain a structural and functional constraint throughout evolution (5, 6). As shown in Fig. S3, comparisons of 7 *Mr1* orthologs revealed that nonsynonymous substitution was minimal and synonymous substitution was predominant throughout the α1–α3 domains of the *Mr1* gene. This indicates that only mutations with a silent effect on its phenotype were permitted, and that *Mr1* genes underwent purifying selection. In contrast to *Mr1* orthologs, comparison of 29 class Ia alleles (Fig. S3*A*, *Right*) from the same species showed an expected excess of nonsynonymous substitution in the α1/α2 domains (*Pn-Ps* >0) (4, 7), especially from codons 62–89 in the α1/α2 domains involved in antigen or TCR interaction (8). Extending these findings, an averaged pairwise comparison was performed for the ARS containing 57 residues and the non-ARS regions in the α1–α3 domains for class Ia molecules (8) and their aligned regions in MR1 molecules. The resulting synonymous and nonsynonymous substitutions were relatively well balanced within *Mr1* orthologs and non-ARS of class Ia genes. In contrast, a dramatic excess of nonsynonymous substitutions was found in the ARS of class Ia genes, as reported previously (Fig. S3*B*) (7). The positive *Pn-Ps* value in class Ia molecules has been interpreted as evidence that polymorphisms in class Ia provide a selective advantage against infections by generating pathogen-specific immune responses (7). More specifically, expression of multiple polymorphic class I gene products by an *Mhc* haplotype results in an increase in the total peptide repertoire that an individual can present to its T cells, thereby increasing the likelihood of specific pathogen recognition. In contrast, polymorphism of *Mr1* appears to be disadvantageous, suggesting that if MR1 does bind a ligand, it likely does so in a manner distinct from that of class Ia molecules. Consistent with this conclusion, activation of MAIT cells by MR1-overexpressing cells is not dependent on the proteasome or TAP (9), 2 critical components of peptide presentation by class I to conventional $CD8^+$ T cells.

Fig. S4 compares variable residues in the MR1-threading model, CD1d, and H-2Kb crystal structures. Because cross-species antigen presentation by CD1d to *i*NKT cells also has been reported using mouse and human components (10), comparing the locations of amino acid differences in the α1/α2 domains of MR1 versus CD1d or class Ia proteins from mouse, rat, bovine, and human is of interest. As shown in the figure, the residues lining the presumptive ligand-binding groove of MR1 are highly conserved, albeit completely distinct from those used by other classical and nonclassical MHC proteins (11). Most of the 42 variable residues in MR1 proteins are located distal from the platform groove, with the exception of a small cluster located around the presumptive F-pocket, where the C terminus of peptide ligands are anchored in classical MHC proteins. In comparison, CD1d or class Ia molecules have substantially more (98 and 78, respectively) variable residues among these 4 species, and they are widely distributed throughout its ligand-binding platform. Relevant to T cell activation, CD1d or class Ia molecules have a greater number of variable residues contributing to ligand and/or TCR interaction; thus, MR1 likely binds its evolutionarily conserved ligands and presents them to MAIT cells in a manner quite distinct from CD1d presentation of lipids to *i*NKT cells. More specifically, a recent study of co-crystal structures found that the invariant TCRα subunit of *i*NKT cells engages the central region of the CD1d-α-GalCer complex, with its TCR β subunit mostly hanging over the side (12). In light of this finding, it perhaps is surprising that the area of the *i*NKT TCR footprint on CD1d is not more evolutionarily conserved. In contrast, it is attractive to speculate that the structurally conserved surface of the MR1 molecule is the area engaged by the MAIT

cell–invariant TCR α chain. Indeed, such an engagement would predict that MR1 interacts with the MAIT cell TCR similar to the way in which classical MHC molecules interact with conventional T cells. Regardless, it seems reasonable to speculate that MR1 uses a different conserved mechanism than the CD1d molecule to bind a family of chemically related ligands.

1. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
2. Reid SD, Selander RK, Whittam TS (1999) Sequence diversity of flagellin (fliC) alleles in pathogenic *Escherichia coli*. *J Bacteriol* 181:153–160.
3. Ota T, Nei M (1994) Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. *Mol Biol Evol* 11:613–619.
4. Doyle CK, et al. (2003) Hyperconservation of the N-formyl peptide binding site of M3: evidence that M3 is an old eutherian molecule with conserved recognition of a pathogen-associated molecular pattern. *J Immunol* 171:836–844.
5. Nei M (2005) Selectionism and neutralism in molecular evolution. *Mol Biol Evol* 22:2318–2342.
6. King JL, Jukes TH (1969) Non-Darwinian evolution. *Science* 164:788–798.
7. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167–170.
8. Bjorkman PJ, et al. (1987) The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature* 329:512–518.
9. Huang S, et al. (2008) MR1 uses an endocytic pathway to activate mucosal-associated invariant T cells. *J Exp Med* 205:1201–1211.
10. Godfrey DI, McCluskey J, Rossjohn J (2005) CD1d antigen presentation: Treats for NKT cells. *Nat Immunol* 6:754–756.
11. Hansen TH, Huang S, Arnold PL, Fremont DH (2007) Patterns of nonclassical MHC antigen presentation. *Nat Immunol* 8:563–568.
12. Borg NA, et al. (2007) CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature* 448:44–49.
13. Garboczi DN, et al. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* 384:134–141.
14. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. *J Mol Biol* 238:777–793.

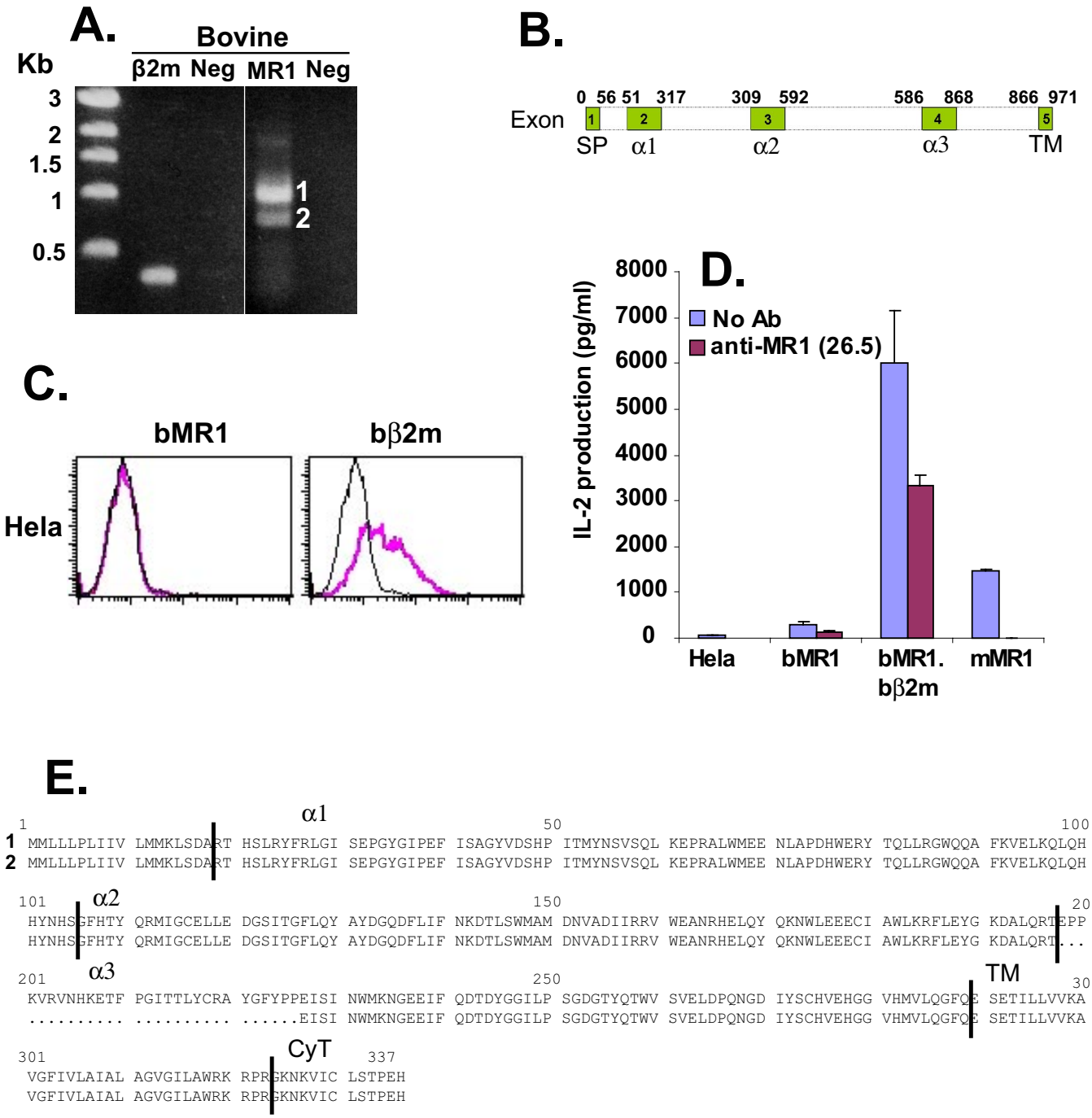
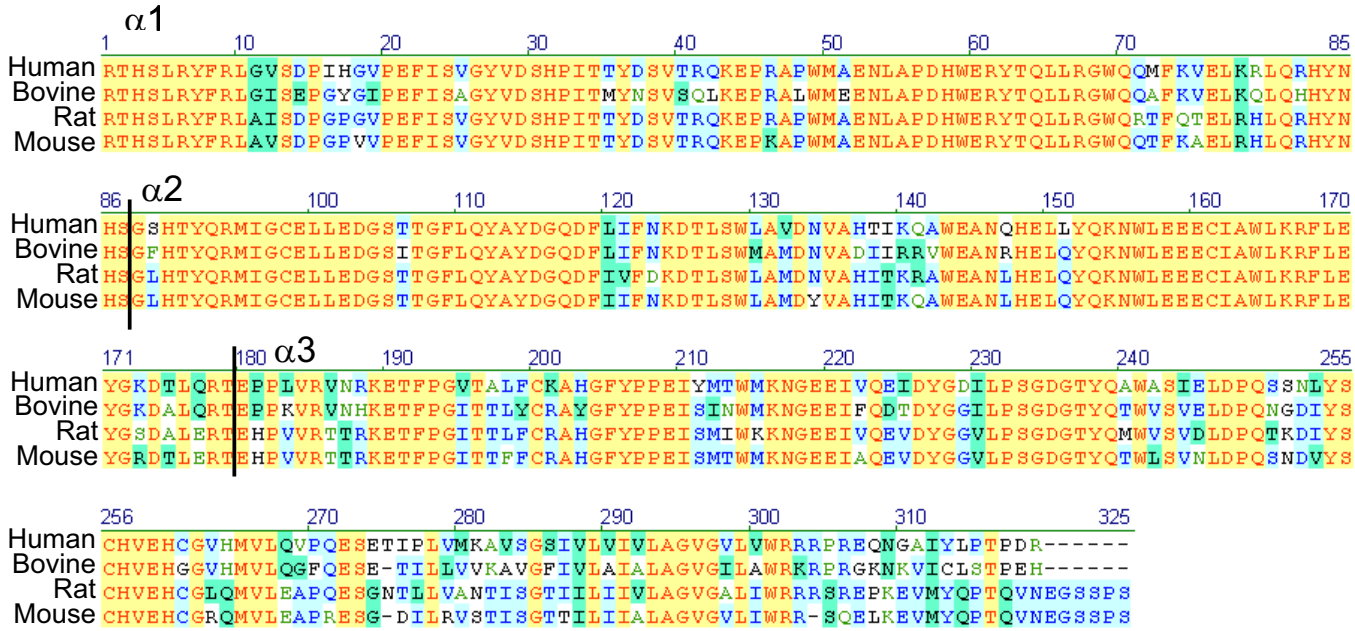


Fig. S1. Cloning of bovine *Mr1* cDNA and activation of the mouse MAIT cell hybridoma with coexpression of the bovine MR1 and bovine β2m. (*A*) cDNA of bovine *Mr1* was amplified with RT-PCR using putative primers identified from the bovine expression tagged sequence through the blastn search of the bovine ETS database in the NCBI server. The bovine β2m cDNA also was amplified from the same sample, and the amplification with mismatched 5′ end primer of *Mr1* and 3′ end primer of β2m was used as the negative control (Neg). (*B*) The exons of the bovine *Mr1* gene were predicted by matching to the Bos taurus contig sequences in GenBank with unknown chromosome location. (*C*) The bovine MR1 was expressed on HeLa cells with cotransduction of bovine β2m cloned in this experiment. The red line indicates FACS staining of the transduced HeLa cells using anti-MR1 antibody 26.5; the black line indicates the nontransduced HeLa cells. (*D*) The bovine MR1 was able to activate mouse MAIT cells partially blocked by anti-MR1 antibody 26.5. (*E*) The sequence alignment of bovine MR1 deduced from the cDNA with the complete coding regions (1; GenBank accession number EU792881) and partial coding sequence (2; EU841914). Our data match the other 2 bovine MR1 sequences from an independent report (FJ423041 and FJ423042).

A.



B.

	mouse	rat	bovine	Chimpanzee	Orangutan	monkey
human	89	87	84	100	98	94
mouse		93	81	89	89	88
rat			82	87	88	88
bovine				84	86	83
chimpanzee					98	94
orangutan						94

Fig. S2. Sequence homology of MR1 orthologous proteins. (*A*) Sequence alignment of human, bovine, rat, and mouse MR1 proteins using the Clustral X algorithm with Vector NTI. The α1, α2, and α3 domains were annotated. (*B*) Pairwise sequence identity in the α1 and α2 domains among different MR1 sequences.

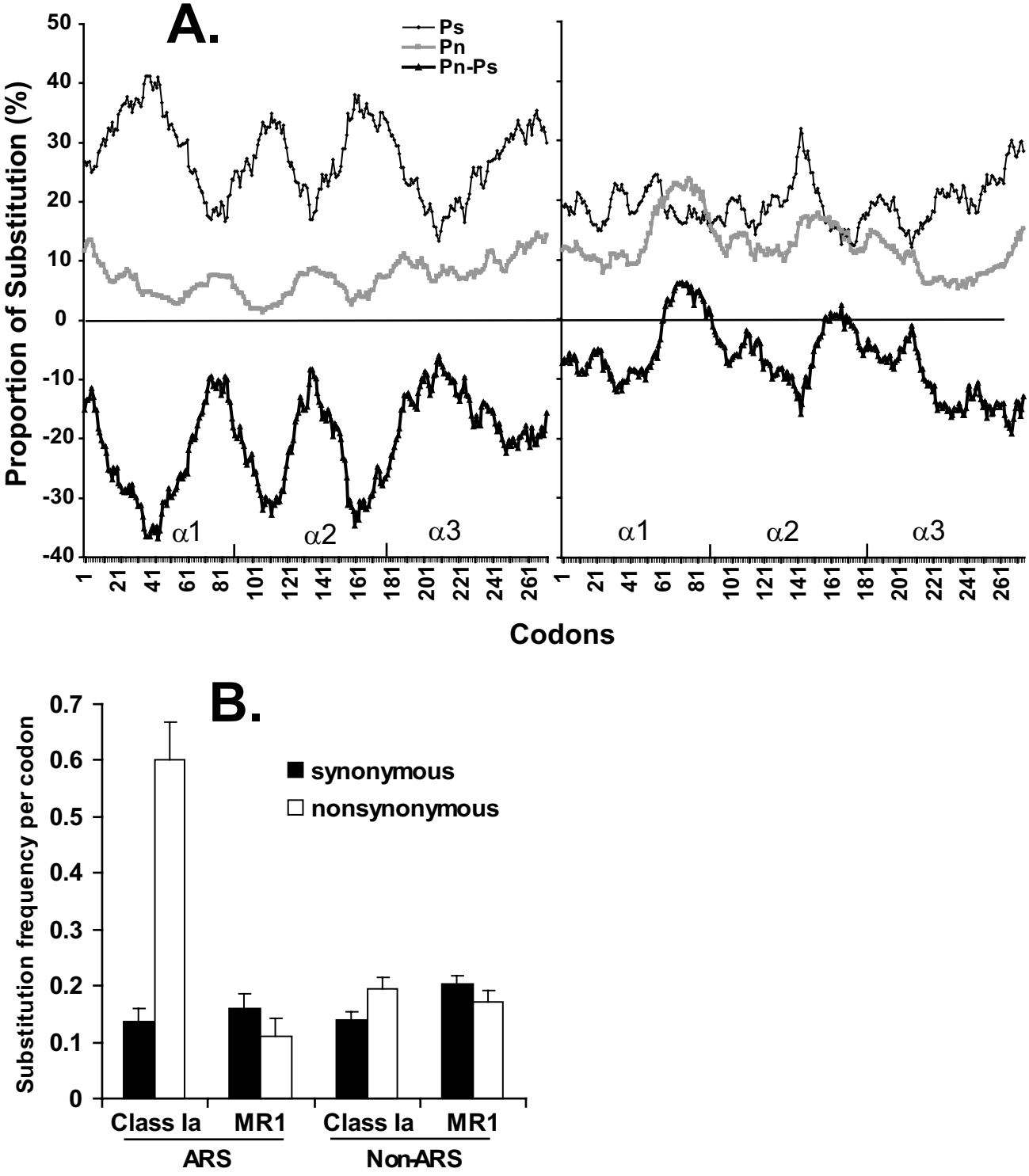


Fig. S3. Synonymous substitution is predominant in *Mr1*. (*A*) The *Mr1* (*Left*) and *Mhc* class Ia (*Right*) molecules from 7 mammalian species with determined *Mr1* cDNA sequences were analyzed for the proportion of synonymous (*Ps*), nonsymonymous (*Pn*) substitution, and their discrepancy (*Pn-Ps*), which were plotted to corresponding codons. (*B*) Averaged pairwise comparisons for synonymouse and nonsynonymous substitution per codon were performed and plotted for the ARS and non-ARS regions.

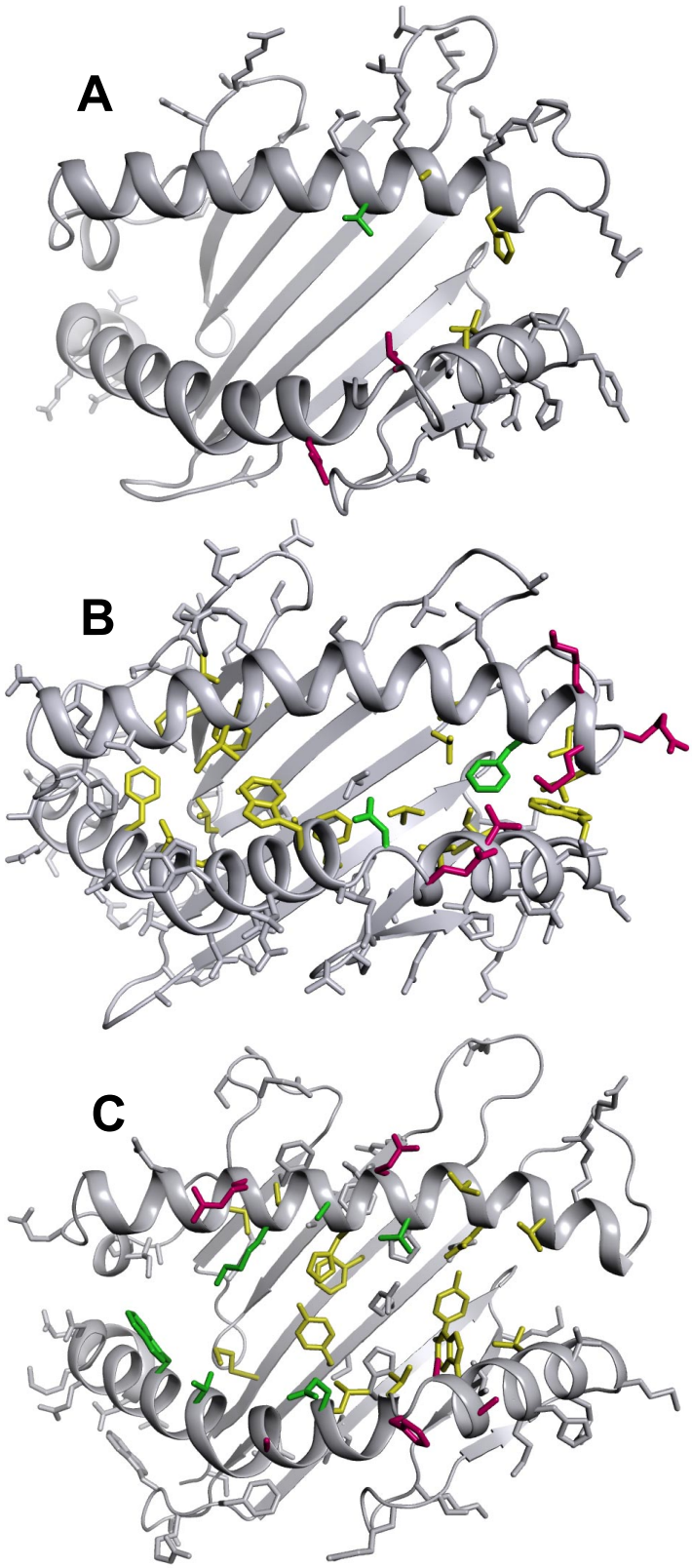


Fig. S4. Variable residues of MR1, CD1d, and class Ia molecules from human, bovine, rat, and mouse. (*A*) The MR1 variable residues are shown with side chains using the homology model. The interacting sites with TCR (magenta), putative ligand (yellow), or both (green) are identified through alignment with the HLA-A2 structure shown in (*C*). The variable residues from CD1d (*B*) and class Ia molecules (*C*) are shown using the crystal structures of CD1d/α-GalCer/i$V\alpha$24TCR (2PO6) (12) and HLA-A2/Tax/TCR (1AO7) (13) molecules. Ligand or TCR interacting sites were calculated with the HBplus program using a contact distance of $\leq$4.5 Å (14).