Table 4 Comparison of frequencies of non-synonymous and synonymous substitutions in proximal region (amino acid positions, 401–566) of extracellular domain of TLR2

	Non-synonymous	Synonymous	Total	*P* value (Fisher's exact test)
Comparison A: test for difference in frequencies of non-synonymous and synonymous changes in *Macaca mulatta TLR2*				
Changes SNPs)	5	0	5	0.358
No changes	382	111	493	
Total	387	111	498 (nucleotide positions, 1,201–1,698)	
Comparison B: comparison of number of SNPs in *M. mulatta* and *Macaca fuscata TLR2*				
M. mulatta	5	0	5	0.048
M. fuscata	0	2	2	
Total	5	2	7	

The change may contribute to certain functional properties, because it locates in the region contributing to ligand binding and interaction with dimerization partner of TLR2-TLR1 heterodimeric complex (Fig. 1a). Homology modeling was applied to determine the most stabilized structure for macaque TLR2-TLR1 dimers composed of different TLR2 variants. Because of the sequence conservation across primates including human, Tyr326 (*M. fuscata* type) is assumed to be ancestral and Asn326 (*M. mulatta* type) is derivative. The effect of single amino acid substitution at position 326 was evaluated by the comparison between Mamu-Hap1 and Mafu-Hap3 (Table 6, Supplementary figures 1 and 2). Relative binding affinity to Pam3CSK4 is attenuated by approximately 5% by amino acid substitution Tyr>Asn (100.0>94.5 Table 6), which is elucidated by structural change in ligand binding pocket composed of TLR2-TLR1 dimer in part (Supplementary Figures 1 and 2). The affinity to ligand is partially recovered by second substitutions in the membrane-proximal region of extracellular domain of TLR2 (Table 6), although these substitutions were not directly involved in ligand binding or dimerization with the TLR1 partner but were supposed to induce subtle change in intra-chain interaction of TLR2 (Suppremental Figures 3 and 4).

Discussion

The rhesus monkey, *M. mulatta* has been applied to many biomedical researches for a long time, because it provides a better relevance of model organism to humans in many pathological and physiological aspects. While the lack of inbred strains or closed colony with great animals is a drawback for the use of experimental model because of variations in biological response based on underlying genetic diversity, it turns to an advantage once the genetic variations are fully understood and are controlled adequately (Matano et al. 2004). Recently, complete genome of *M. mulatta* has

Table 5 Log-likelihood values and parameter estimates under different codon substitution models for ten *Macaca mulatta* TLR2 allelic variants

Model	Number of parameters	*l*: ln(likelihood)	Average ω (d_N/d_S)	Estimates of parameters
One ratio (M0)	1	−2,461.85	0.777	ω=0.777
Site-specific models				
M1a: neutral	2	−2,461.42	0.530	p_0=0.47 (p_1=1–p_0=0.53), ω_0=0.00
M2a: selection	4	−2,458.36	0.781	p_0=0.963, p_1=0 (p_2=1–p_0–p_1=0.037), ω_0=0.00, ω_2=21.29
M3: discrete	5	−2,458.36	0.782	p_0=0.002, p_1=0.961 (p_2=1–p_0–p_1=0.037), ω_0=0.00, ω_1=0.00, ω_2=21.30
M7: beta	2	−2,460.40	0.500	p=0.0074, q=0.0073
M8: beta and ω	4	−2,458.36	0.782	p_0=0.963, p=0.0050, q=4.116 (p_1=1–p_0=0.037), ω_s=21.30
Comparisons				
One ratio (M0) vs. discrete (M3)	χ^2=−2Δ*l*=6.98	*df*=4	*p*=0.14	
Neutral (M1a) vs. selection (M2a)	χ^2=−2Δ*l*=6.12	*df*=2	*p*=0.047	
Beta (M7) vs. beta and ω (M8)	χ^2=−2Δ*l*=4.08	*df*=2	*p*=0.13	

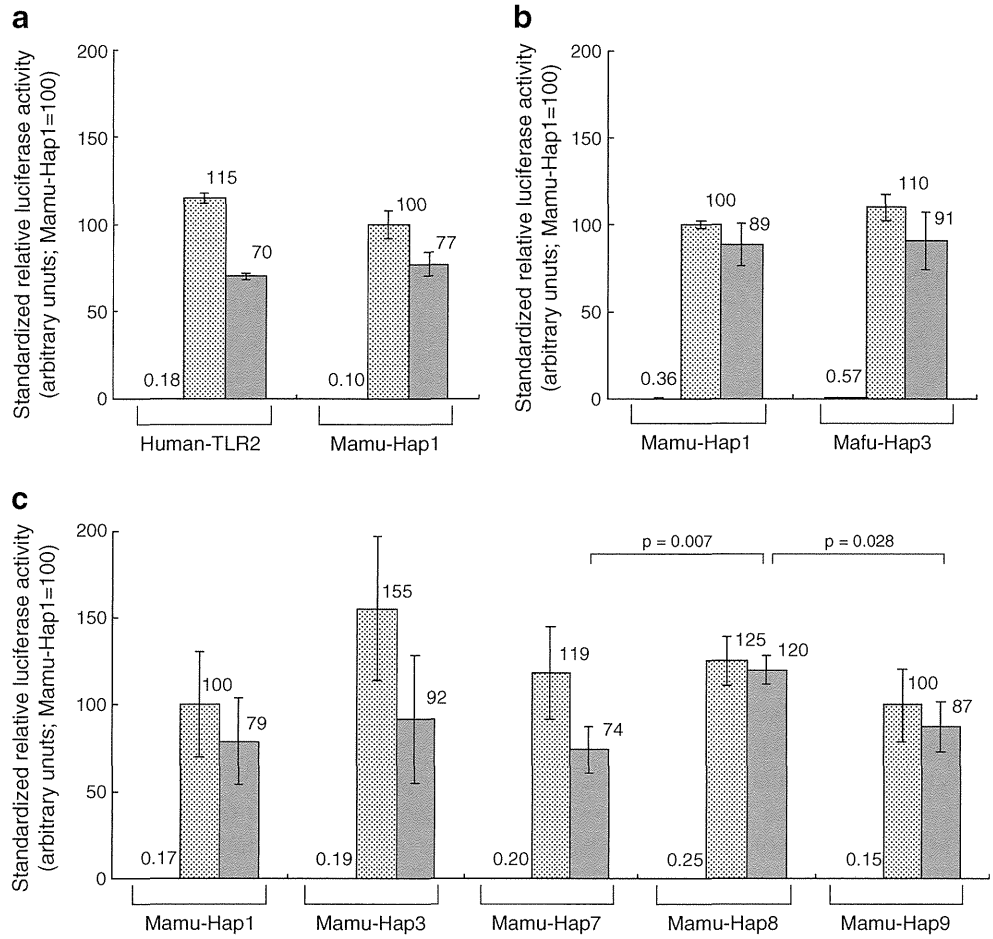


Fig. 6 Functional analysis of TLR2 alleles by luciferase reporter gene assay. Relative luciferase activity of each triplicate transfection is obtained as the ratio of relative light units (*RLUs*) of firefly luciferase assay (for NF-κB-dependent activity) to RLUs of *Renilla* luciferase assay (for constitutive basal transcriptional activity) and standardized by the mean value of those of triplicate Mamu-Hap1 transfections to be 100 for respective experiments. The results of representatives of repeated experiments at least two times are shown. Mean value of triplicate standardized relative luciferase activities without TLR2 ligand (*left*) that with 10 ng/ml of Pam2CSK4 (*middle*, *stippled bar*) and that with 100 ng/ml of Pam3CSK4 (*right*, *filled bar*) are depicted as *bar gragh with whiskers* of SEM. **a** A major allele of *M. mullata*, Mamu-Hap1, brought almost equivalent activities to the human TLR2. **b**, **c** *Macaca* TLR2 alleles carrying Tyr326 (Mafu-Hap3), Ala416 (Mamu-Hap3), Ile405-Val503-Val556 (Mamu-Hap7), Pro98-Val503 (Mamu-Hap8), and Val503-Ala545 (Mamu-Hap9) substitutions were compared

been uncovered (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), and efforts to reveal genetic polymorphism of this species have been carried out (Singh and Schmidtke 2005; Ferguson et al. 2007; Karl et al. 2008; Flynn et al. 2009; Blokhuis et al. 2009). On the other hand, the Japanese monkey, *M. fuscata*, an endemic species of Japan, also draws an attention of investigators as experimental models to human pathophysiology (Kawai et al. 1993; Isa et al. 2009). *M. mulatta* and *M. fuscata* are very close to each other in terms of species diversification (Smith et al. 2007), but geographic isolation and differential influence of dwelling environment cause species-specific characteristic through evolution. For example, *M. fuscata* is extremely labile to cerebral involvement of monkey malaria in comparison to tropical *Macaca* species (Kawai et al. 1993), presumably because of the absence of circulation of *Plasmodium* protozoa in wild *M. fuscata* population. It is suggested that the origin of current species belonging to genus *Macaca* had emerged in northern part of African continent, migrated through the Middle-East into India, and settled in Asian continent and islands (Hernandez et al. 2007). The population history of migration out of Africa and settlement in their destinations is very similar to that of divergent human ethnic groups. Infectious diseases such as malaria and arthropod-borne viral hemorrhagic fever would be a heavy burden to thrive in the tropical region in *Macaca* species as well as humans. Therefore, comparative studies of different *Macaca* species would provide valuable clues for the better understanding of host defense mechanisms of these species as well as those of humans against various infectious agents.

The previous studies on primate TLR genes revealed that the cytoplasmic TIR domains of *M. mulatta* TLRs shares high

Table 6 Molecular stability and interaction energy estimated for TLR ternary complex structure obtained by optimization through homology modeling

Ternary complex of TLR dimer and ligand	Total energy[a] (kcal/mol)	Ligand[b] (kcal/mol)	Interaction[c] (kcal/mol)	London dG score[d] (kcal/mol)	Relative affinity to ligand (macaque TLR2(Tyr326)=100)
Human TLR2 (27–506)-TLR1(25–475)-Pam3CSK4[e]	−11,680.30	−393.12	−110.59	−26.21	105.6
Macaque TLR2(Tyr326)-TLR1-Pam3CSK4 (*M. fuscata*, hap-3)	−9,101.94	−398.09	−110.41	−24.81	100.0 (reference)
Macaque TLR2(Asn326)-TLR1-Pam3CSK4 (*Macaca mulatta*, hap-1)	−9,114.51	−397.85	−109.13	−23.52	94.5
Macaque TLR2(Asn326-Ile405)-TLR1-Pam3CSK4	−9,808.11	−396.92	−110.80	−24.24	97.7
Macaque TLR2(Asn326-Ala416)-TLR1-Pam3CSK4	−9,524.80	−399.36	−109.04	−24.46	98.6

[a] Total potential energy between all the atoms of extracellular domains of TLR2-TLR1 heterodimeric receptor (position 27–506 of TLR2 and position 25–475 of TLR1) and a ligand Pam3CSK4 was calculated by Molecular Operating Environment (MOE) 2010.10(Chemical Computing Group Inc)

[b] Total potential energy between all the atoms of Pam3CSK4

[c] Total potential energy between the atoms of TLR2-TLR1 heterodimeric receptor and those of Pam3CSK4

[d] An index for affinity between receptor and ligand within the binding pocket obtained by Lig-X algorithm suite implemented in MOE 2010.10

[e] Molecular structure was optimized by MOE 2010.10 after retrieval of PDB data accession# 2Z7X (available at RCSB, http://www.rcsb.org/pdb/) and trimming of C-terminal 67 residues of fusion peptide derived from inshore hagfish Eptatretus VLRB.61

levels of homology with human counterparts (Sanghavi et al. 2004), and TIR domains of TLRs and other related genes had undergone purifying selection among seven primate species (Nakajima et al. 2008), by which the presence of strict functional constraint is suggested. On the other hand, it was also revealed that the extracellular domain of *Macaca* TLR4 have been uniquely evolved under the diversifying selection by comparison between primate species (Nakajima et al. 2008). While the coding polymorphisms in *M. mulatta TLR4* and *TLR5* (Ferguson et al. 2007), as well as microsatellite located in the second intron of *TLR2* (Yim et al. 2006), were reported, we investigated for the first time SNPs in the coding region of *TLR2* for *M. mulatta* and *M. fuscata* individuals and reach to the trace evidence for differential purifying and diversifying selections on certain parts of extracellular domain in the present study. It is of note that the shapes of sliding window plots of Tajima's D (Fig. 4) were quite different between *M. mulatta* and *M. fuscata* suggesting differential molecular evolution for these species after species diversification. Inter-specific differential molecular evolution is also prominent in the sliding window plot of K_a/K_s ratio for combined *M. mulatta* and *M. fuscata* allelic variants (Fig. 5). Further, we could obtain a trace of positive selection at one amino acid position 545 by application of maximal likelihood estimation for codon-wise evaluation of molecular evolution events. Certain polymorphisms located in extracellular domain of swine *TLR2* (Shinkai et al. 2006; Bergman et al. 2010) and those of bovine *TLR2* (Jann et al. 2008) were also suggested to have been under positive selection during the domestication and breeding of these mammalian species. It is probable that positive selection has been driven by the functional alteration adapting to the changes of dwelling environments. Further, one of several rare variants of human TLR2, Thr411Ile, which has been shown functionally deteriorated upon the Pam3CSK4-induced, NF-κB-driven reporter gene assay (Merx et al. 2007; Kormann et al. 2009), is located in the part C-terminal to the ligand binding and dimerization domain in the vicinity of membrane-proximal region where five of the non-synonymous substitutions were found in our *M. mulatta* population. Although we could not demonstrate apparent augmentation or deterioration accompanied with amino acid substitutions in *M. mulatta* alleles by luciferase reporter gene assay using HEK293 cells as recipients with saturating doses of lipopeptide ligands in the present study, the affinity to ligand appeared different and would be differentially tuned during the evolution of two *Macaca* species under distinct environmental selection pressure.

Acknowledgments This work was supported in part by Japan Initiative for Global Research Network on Infectious Diseases (J-GRID) and the Global Centers of Excellence Program of the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the Cooperation Research Program of Primate Research Institute, Kyoto University, Japan

References

Akira S, Uematsu S, Takeuchi O (2006) Pathogen recognition and innate immunity. Cell 124(4):783–801

Baena A, Mootnick AR, Falvo JV, Tsytskova AV, Ligeiro F, Diop OM, Brieva C, Gagneux P, O'Brien SJ, Ryder OA, Goldfeld AE

(2007) Primate TNF promoters reveal markers of phylogeny and evolution of innate immunity. PLoS One 2:e621

Bergman IM, Rosengren JK, Edman K, Edfors I (2010) European wild boars and domestic pigs display different polymorphic patterns in the Toll-like receptor (TLR) 1, TLR2, and TLR6 genes. Immunogenetics 62:49–58

Blancher A, Bonhomme M, Crouau-Roy B, Terao K, Kitano T, Saitou N (2008) Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). J Hered 99:254–264

Blokhuis JH, van der Wiel MK, Doxiadis GG, Bontrop RE (2009) Evidence for balancing selection acting on KIR2DL4 genotypes in rhesus macaques of Indian origin. Immunogenetics 61:503–512

Bochud PY, Hawn TR, Aderem A (2003) A Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling. J Immunol 170:3451–3454

Ferguson B, Street SL, Wright H, Pearson C, Jia Y, Thompson SL, Allibone P, Dubay CJ, Spindel E, Norgren RB Jr (2007) Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). BMC Genomics 8:43

Flynn S, Satkoski J, Lerche N, Kanthaswamy S, Smith DG (2009) Genetic variation at the TNF-alpha promoter and malaria susceptibility in rhesus (*Macaca mulatta*) and long-tailed (*Macaca fascicularis*) macaques. Infect Genet Evol 9:769–777

Hayasaka K, Fujii K, Horai S (1996) Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. Mol Biol Evol 13:1044–1053

Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, Muzny D, Gibbs R, Nielsen R, Bustamante CD (2007) Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science 316:240–243

Huang D, Qiu L, Wang R, Lai X, Du G, Seghal P, Shen Y, Shao L, Halliday L, Fortman J, Shen L, Letvin NL, Chen ZW (2007) Immune gene networks of mycobacterial vaccine-elicited cellular responses and immunity. J Infect Dis 195:55–69

Isa T, Yamane I, Hamai M, Inagaki H (2009) Japanese macaques as laboratory animals. Exp Anim 58:451–457

Jann OC, Werling D, Chang JS, Haig D, Glass EJ (2008) Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. BMC Evol Biol 8:288

Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee H, Lee JO (2007) Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell 130:1071–1082

Kang JY, Nan X, Jin MS, Youn SJ, Ryu YH, Mah S, Han SH, Lee H, Paik SG, Lee JO (2009) Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity 31:873–884

Karl JA, Wiseman RW, Campbell KJ, Blasky AJ, Hughes AL, Ferguson B, Read DS, O'Connor DH (2008) Identification of MHC class I sequences in Chinese-origin rhesus macaques. Immunogenetics 60:37–46

Kawai S, Aikawa M, Kano S, Suzuki M (1993) A primate model for severe human malaria with cerebral involvement: *Plasmodium coatneyi*-infected *Macaca fuscata*. Am J Trop Med Hyg 48:630–636

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kormann MS, Ferstl R, Depner M, Klopp N, Spiller S, Illig T, Vogelberg C, von Mutius E, Kirschning CJ, Kabesch M (2009) Rare TLR2 mutations reduce TLR2 receptor function and can increase atopy risk. Allergy 64:636–642

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299–306

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Ling B, Veazey RS, Luckay A, Penedo C, Xu K, Lifson JD, Marx PA (2002) SIVmac pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. AIDS 16:1489–1496

Lorenz E, Mira JP, Cornish KL, Arbour NC, Schwartz DA (2000) A novel polymorphism in the toll-like receptor 2 gene and its potential association with staphylococcal infection. Infect Immun 68:6398–6401

Matano T, Kobayashi M, Igarashi H, Takeda A, Nakamura H, Kano M, Sugimoto C, Mori K, Iida A, Hirata T, Hasegawa M, Yuasa T, Miyazawa M, Takahashi Y, Yasunami M, Kimura A, O'Connor DH, Watkins DI, Nagai Y (2004) Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial. J Exp Med 199:1709–1718

McMurray DN (2000) A nonhuman primate model for preclinical testing of new tuberculosis vaccines. Clin Infect Dis 30:S210–S212

Merx S, Neumaier M, Wagner H, Kirschning CJ, Ahmad-Nejad P (2007) Characterization and investigation of single nucleotide polymorphisms and a novel TLR2 mutation in the human TLR2 gene. Hum Mol Genet 16:1225–1232

Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T, Kimura A (2008) Natural selection in the TLR-related genes in the course of primate evolution. Immunogenetics 60:727–735

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Ogus AC, Yoldas B, Ozdemir T, Uguz A, Olcen S, Keser I, Coskun M, Cilli A, Yegin O (2004) The Arg753Gln polymorphism of the human toll-like receptor 2 gene in tuberculosis disease. Eur Respir J 23:219–223

Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, Inoue I, Hida M, Suzuki Y, Sugano S, Terao K, Kusuda J, Takahashi I (2008) Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. BMC Genomics 9:90

Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionaly and biomedical insights from the rhesus macaque genome. Science 316:222–234

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing Phylogenetic trees. Mol Biol Evol 4:406–425

Sanghavi SK, Shankarappa R, Reinhart TA (2004) Genetic analysis of Toll/Interleukin-1 receptor (TIR) domain sequences from rhesus macaque Toll-like receptors (TLRs) 1–10 reveals high homology to human TLR/TIR sequences. Immunogenetics 56:667–674

Schröder NW, Diterich I, Zinke A, Eckert J, Draing C, von Baehr V, Hassler D, Priem S, Hahn K, Michelsen KS, Hartung T, Burmester GR, Göbel UB, Hermann C, Schumann RR (2005) Heterozygous Arg753Gln polymorphism of human TLR-2 impairs immune activation by *Borrelia burgdorferi* and protects from late stage Lyme disease. J Immunol 175:2534–2540

Schwandner R, Dziarski R, Wesche H, Rothe M, Kirschning CJ (1999) Peptidoglycan- and lipoteichoic acid-induced cell activation is mediated by toll-like receptor 2. J Biol Chem 274:17406–17409

Shibata H, Yasunami M, Obuchi N, Takahashi M, Kobayashi Y, Numano F, Kimura A (2006) Direct determination of SNP haplotype of NFKBIL1 promoter polymorphism by DNA conformation analysis and its application to association study of chronic inflammatory diseases. Hum Immunol 67:363–373

Shinkai H, Tanaka M, Morozumi T, Eguchi-Ogawa T, Okumura N, Muneta Y, Awata T, Uenishi H (2006) Biased distribution of

single nucleotide polymorphisms (SNPs) in porcine Toll-like receptor 1 (TLR1), TLR2, TLR4, TLR5, and TLR6 genes. Immunogenetics 58:324–330

Singh KK, Schmidtke J (2005) Single nucleotide polymorphisms within the promoter region of the rhesus monkey tumor necrosis factor-alpha gene. Immunogenetics 57:289–292

Smith DG, McDonough JW, George DA (2007) Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the fascicularis group of macaques. Am J Primatol 69:182–198

Stevison LS, Kohn MH (2009) Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. Mol Ecol 18:2457–2475

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takahashi-Tanaka Y, Yasunami M, Naruse T, Hinohara K, Matano T, Mori K, Miyazawa M, Honda M, Yasutomi Y, Nagai Y, Kimura A (2007) Reference strand-mediated conformation analysis (RSCA)-based typing of multiple alleles in the rhesus macaque MHC class I Mamu-A and Mamu-B loci. Electrophoresis 28:918–924

Texereau J, Chiche JD, Taylor W, Choukroun G, Comba B, Mira JP (2005) The importance of Toll-like receptor 2 polymorphisms in severe infections. Clin Infect Dis 41:S408–S415

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yim JJ, Adams AA, Kim JH, Holland SM (2006) Evolution of an intronic microsatellite polymorphism in Toll-like receptor 2 among primates. Immunogenetics 58:740–745

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci U S A 95:3708–3713

Immunogenetics (2012) 64:131–141
DOI 10.1007/s00251-011-0568-y

ORIGINAL PAPER

Diversity of MHC class I haplotypes in cynomolgus macaques

Yusuke Saito · Taeko K. Naruse · Hirofumi Akari · Tetsuro Matano · Akinori Kimura

Received: 7 May 2011 / Accepted: 19 August 2011 / Published online: 1 September 2011

Abstract Cynomolgus macaques are widely used as a primate model for human diseases associated with an immunological process. Because there are individual differences in immune responsiveness, which are controlled by the polymorphic nature of the major histocompatibility (MHC) locus, it is important to reveal the diversity of MHC in the model animal. In this study, we analyzed 26 cynomolgus macaques from five families for MHC class I genes. We identified 32 *Mafa-A*, 46 *Mafa-B*, 6 *Mafa-I*, and 3 *Mafa-AG* alleles in which 14, 20, 3, and 3 alleles were novel. There were 23 MHC class I haplotypes and each haplotype was composed of one to three *Mafa-A* alleles and one to five *Mafa-B* alleles. Family studies revealed that there were two haplotypes which contained two *Mafa-A1* alleles. These observations demonstrated further the complexity of MHC class I locus in the Old World monkey.

Keywords Cynomolgus macaque · MHC · *Mafa* class I gene · Haplotype · Polymorphism

Yusuke Saito and Taeko K. Naruse contributed equally to this work.

Y. Saito · T. K. Naruse · A. Kimura (✉)
Department of Molecular Pathogenesis,
Medical Research Institute, Tokyo Medical and Dental University,
1-5-45 Yushima,
Bunkyo-ku, Tokyo 113-8510, Japan
e-mail: akitis@mri.tmd.ac.jp

Y. Saito · A. Kimura
Laboratory of Genome Diversity,
Graduate School of Biomedical Science,
Tokyo Medical and Dental University,
Tokyo, Japan

H. Akari
Primate Research Institute,
Kyoto University,
Inuyama, Japan

T. Matano
AIDS Research Center, National Institute of Infectious Diseases,
Tokyo, Japan

T. Matano
The Institute of Medical Science, The University of Tokyo,
Tokyo, Japan

Introduction

Non-human primates are widely used for immunological research because their immune system is similar to that of humans. In particular, the Old World monkeys such as cynomolgus macaques (crab-eating macaques, *Macaca fascicularis*) became a useful model for human infectious diseases including acquired immunodeficiency syndrome (AIDS) (Wiseman et al. 2007), severe acute respiratory syndrome (Lawler et al. 2006), and influenza (Kobasa et al. 2007) as well as in the transplantation field (Wiseman and O'Connor 2007). In the AIDS research, cynomolgus and rhesus macaques are important animal models for the development of vaccines against human immunodeficiency virus (HIV) or studies for susceptibility to HIV infection and/or development of AIDS (Matano et al. 2004; Loffredo et al. 2008; Tsukamoto et al. 2008; Burwitz et al. 2009; Mee et al. 2009; Aarnink et al. 2011a). To fully evaluate the results of immunological experiments in the macaque models, it is essential to characterize the genetic diversity of immune-related molecules which may control the individual differences in the immune response against foreign antigens and/or pathogens.

The major histocompatibility complex (MHC) is well known to control the immune-responsiveness to foreign

antigens. There are two classes of MHC molecules: one is the MHC class I molecule presenting peptides of intracellular origin to $CD8^+$ T cell and the other is the MHC class II molecule binding extracellular-derived antigenic peptides for presenting to $CD4^+$ T cell. It has been reported that the complexity of MHC genes in the rhesus and cynomolgus macaques is higher than that in humans (Kulski et al. 2004; Watanabe et al. 2006; Gibbs et al. 2007; Otting et al. 2007, 2008; Doxiadis et al. 2011). For example, *MHC class I* configurations in macaques are usually composed of one copy of highly transcribed major *MHC-A1*gene (*Mamu-A1*or *Mafa-A1*) and several other minor *MHC-A* genes (*Mamu-A2~A7* or *Mafa-A2~A6*) in addition to several *MHC-B* genes (*Mamu-B* or *Mafa-B*) (Watanabe et al. 2006; Otting et al. 2007, 2008, 2009; Naruse et al. 2010; Doxiadis et al. 2011), whereas each one copy of *MHC-A* and *-B* genes (*HLA-A* and *-B*) can be found in human *MHC class I* locus. In addition, other *MHC* loci showing lower expression levels, i.e., *HLA-B*-like gene (*Mamu-I* or *Mafa-I*) and *HLA-G*-like non-classical gene (*Mamu-AG* or *Mafa-AG*) have been identified (Slukvin et al. 2000; Urvater et al. 2000). The extent of genetic diversity is different, in part, depending on the geographic areas, as we have previously reported for *MHC class I* genes in rhesus macaque (Naruse et al. 2010). As for the cynomolgus macaques, *MHC class I* allelic diversity was reported for Indonesian (Pendley et al. 2008; Wu et al. 2008; Kita et al. 2009; Otting et al. 2009), Malaysian (Otting et al. 2009; Aarnink et al. 2011b), Mauritian (Budde et al. 2010), Vietnamese (Wu et al. 2008; Kita et al. 2009), and Philippino (Campbell et al. 2009; Kita et al. 2009) macaques, but information about the *MHC class I* haplotype remains insufficient.

In the present study, we have analyzed *MHC class I* loci in cynomolgus macaques originated from Indonesia, Malaysia, and the Philippines to obtain information on haplotype configuration. We report here further the complex nature of *MHC class I* loci in the Old World monkey, i.e., the presence of unique haplotypes carrying two *Mafa-A1* genes.

Materials and methods

Animals

A total of 26 cynomolgus macaques from five families were the subjects. Each family was composed of one or two males with one or two females and their offspring. They were maintained in the breeding colonies in Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Japan. The founders of the colonies were captured in Indonesia, Malaysia, and the Philippines. All care including blood sampling of animals were in accordance with the Guidelines for the Care and Use of Laboratory Animals published by the National Institute of Health (NIH Publication 85–23, revised 1985) and were subjected to prior approval by the local animal protection authority.

Sequencing analysis of cDNAs from *Mafa* class I genes

Total cellular RNA was extracted from whole blood by using RNAeasy (QIAGEN, Gmbh, Germany). Oligo(dT)-primed cDNA was synthesized using Transcriptor reverse transcriptase (Roche, Mannheim, Germany) according to the manufacturer's recommendations. Full-length cDNAs

Table 1 Primers used in PCR or sequencing of *Mafa* class I genes

Primer ID	Application	Direction	Sequence (5′–3′)	Position	Reference
5′ MHC_UTR	PCR	Sense	GGACTCAGAATCTCCCCAGACGCCGAG	5′ UTR	Karl et al. 2008
3′ MHC_UTR_A	PCR	Antisense	CAGGAACAYAGACACATTCAGG	3′ UTR	Karl et al. 2008
3′ MHC_UTR_B	PCR	Antisense	GTCTCTCCACCTCCTCAC	3′ UTR	Karl et al. 2008
5A long	PCR	Sense	ATGGCGCCCCGAACCCTCCTCCTG	Exon 1	Tanaka-Takahashi et al. 2007
3A	PCR	Antisense	TCACACTTTCAAGCCGTGAGAGA	Exon 7	Tanaka-Takahashi et al. 2007
5ASSP	PCR	Sense	ATGGCGCCCCGAACCCTCCTCCTGG	Exon 1	Tanaka-Takahashi et al. 2007
4R	PCR	Antisense	CCAGGTCAGTGTGATCTCCG	Exon 4	Tanaka-Takahashi et al. 2007
P000044	PCR	Sense	GATTCTCCGCAGACGCCCA	5′ UTR	Wu et al. 2008
P000023	PCR	Antisense	GGAGAACCAGGCCAGCAAT	Exon 5	Wu et al. 2008
P000076	Sequencing	Sense	GAGCAGCGACGGGACCGCA	Intron 1	Wu et al. 2008
P000060	Sequencing	Antisense	CCTGGGGCTCTCCCGGGTCA	Intron 2	Wu et al. 2008
P000096	Sequencing	Sense	TGTACTGAGTCTCCCTGATGG	Intron 2	Wu et al. 2008
P000098	Sequencing	Antisense	TTCATTCCCTCAGAGATTTT	Intron 3	Wu et al. 2008
P000055	Sequencing	Sense	CCCAGGTRCCTSTGTCCAGGA	Intron 3	Wu et al. 2008
P000281	Sequencing	Antisense	AGAGGGGAAAGTGAGGGGT	Intron 4	Wu et al. 2008

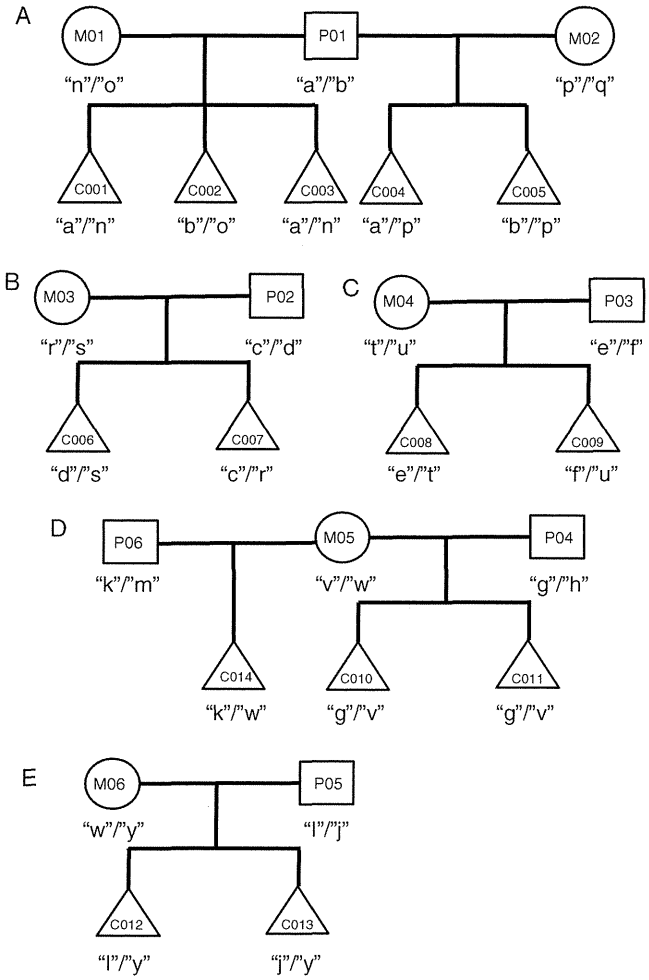


Fig. 1 Pedigree of cynomolgus macaques. The pedigrees of macaques analyzed in this study are shown. Founders were originated from Indonesia (**a**), Malaysia (**b**, **c**), and Philippines (**d**, **e**). *Open square*, *open circle*, and *open triangles* indicate father, mother, and offspring, respectively. The ID of each subject is noted in the symbol. *Mafa class I* haplotypes determined in this study are indicated under the subjects

for *Mafa* class I genes were amplified by polymerase chain reaction (PCR), as described previously (Tanaka-Takahashi et al. 2007; Naruse et al. 2010), by using locus-specific primer pairs as reported by Karl et al. (2008). Genomic gene and cDNA for *Mafa-A2* gene were analyzed according to the method described by Wu et al. (2008). The primers used in this study are listed in Table 1. To estimate the expression level of *Mafa-A* alleles, we also used an additional primer pair: MafaF (5′-TACGTGGACGA CACGCAGTT) and MafaR (5′-GGTGGGTCA CATGTGTCTTG). PCR was done under the condition of initial denaturation at 98°C for 10 s, 25 cycles of 98°C for 1 s, 64°C for 5 s, and 72°C for 20 s, followed by an additional extension at 72°C for 1 min, using Phusion Flash DNA polymerase (Finzymes, Espoo, Finland). The PCR products were cloned into pSTBlue-1 Perfectly Blunt vector (Novagen, WI, USA) according to the manufacturer's instructions and were transformed to NovaBlue Giga Singles™ competent cells (Merck Biosciences Japan, Tokyo, Japan). A total of 30 to 90 independent cDNA clones were obtained from each macaque for each locus and were sequenced on both strands by BigDye Terminator cycling system in an ABI 3730 automated sequence analyzer (Applied Biosystems, CA, USA).

Data analyses and nomenclature for *Mafa* class I allele

Nucleotide sequences of cDNA clones were aligned using the Genetyx software package (version 8.0, Genetyx Corp., Japan). When a cDNA sequence, which was represented by at least three clones, was independently obtained from at least two animals or repeatedly obtained from at least two independently prepared cDNAs from single animals, we considered it a real allele, not an artifact, and the sequences were submitted to the DNA Data Bank of Japan (DDBJ) database and to the Immuno Polymorphism Database for non-human primate MHC (http://www.ebi.ac.uk/ipd/mhc/sumit.html; Robinson et al. 2003) to obtain official nomenclature for the novel alleles of *Mafa-A* and *Mafa-B* genes. Neighbor-joining trees were constructed with Kimura's two-parameter method for a phylogenetic analysis of *Mafa-A* sequences spanning exons 2, 3, and a part of exon 4 obtained in this study by using the Genetyx software. Bootstrap values were based on 5,000 replications.

Table 2 *Mafa* class I alleles found in the cynomolgus macaques

Locus	Number of observed alleles	Number of novel alleles (%)	Number of observed alleles in macaques from different regions[a]		
			Indonesian	Malaysian	Philippino
Mafa-A	32	14 (43.7%)	9 (3), 33.3%	12 (8), 66.7%	11 (3), 27.3%
Mafa-B	46	20 (43.5%)	13 (5), 38.5%	20 (15), 75.0%	18 (1), 5.6%
Mafa-I	6	3 (50.0%)	2 (1), 50.0%	4 (3), 75.0%	2 (0), 0%
Mafa-AG	3	3 (100%)	0 (0), 0%	2 (2), 100%	1 (1), 100%
Total	87	40 (45.5%)	24 (9), 37.5%	38 (28), 73.7%	32 (5), 15.6%

[a] The number and frequency of novel alleles are indicated in parentheses

Results

Identification of *Mafa* class I alleles in cynomolgus macaques

We determined the nucleotide sequences of cDNA clones for *Mafa-A* and *-B* loci in 26 cynomolgus macaques from one family of Indonesian origin (six haplotypes), two families of Malaysian origin (eight haplotypes), and two families of Philippino origin (nine haplotypes) (Fig. 1). When the observed alleles were segregated in the family or when at least three clones with identical sequences were observed from two independent PCR for an individual, the nucleotide sequences were considered to be real and not artifacts. As shown in Table 2, 32 *Mafa-A*, 46 *Mafa-B*, 6 *Mafa-I*, and 3 *Mafa-AG* sequences were obtained in this study. Among them, 14 (43.7%), 20 (43.5%), 3 (50.0%), and 3 (100%) were novel alleles of *Mafa-A*, *Mafa-B*, *Mafa-I*, and *Mafa-AG* loci, respectively (Table 2).

Table 3 Alleles of *Mafa-A* locus identified in the cynomolgus macaques

Locus	Allele name	Novelty[a]	Accession number[b]	Origin[c]	Identical *Mamu* and/or *Mane* alleles[d]	Origin and reference of known alleles[e]
A1	*A1*001:01*		AM295828	Malaysian		Utrecht, Otting et al. 2007
A1	*A1*002:01:02*	Novel	AB569214	Indonesian		
A1	*A1*008:02*		EU392108	Philippino		Philippino, Campbell et al. 2009
A1	*A1*008:03-like*	Novel	AB647187	Philippino		
A1	*A1*018:06*		FM246489	Indonesian		Utrecht, Otting et al. 2007
A1	*A1*019:05*		AB447616	Indonesian		Indonesian, Kita et al. 2009
A1	*A1*023:01*	Novel	AB569216	Malaysian		
A1	*A1*032:05*	Novel	AB569215	Malaysian		
A1	*A1*052:02*		EU392105	Philippino	*Mamu-A1*052:01/03/06*	Philippino, Campbell et al. 2009
A1	*A1*054:01*		AB154771	Malaysian		Tsukuba, Uda et al. 2004
A1	*A1*056:02*	Novel	AB569218	Malaysian		
A1	*A1*062:05*	Novel	AB569219	Malaysian		
A1	*A1*068:02*	Novel	AB569217	Malaysian		
A1	*A1*074:02*		AB447606	Philippino		Philippino, Kita et al. 2009
A1	*A1*079:01*		AB154773	Malaysian		Tsukuba, Uda et al. 2004
A1	*A1*089:02*		EU392104	Philippino		Philippino, Campbell et al. 2009
A1	*A1*093:01*		EU392103	Philippino		Philippino, Campbell et al. 2009
A1	*A1*094:01*		EU392111	Philippino		Philippino, Campbell et al. 2009
A1	*A1*097:01*		AB447576	Indonesian	*Mamu-A1*109:01*	Indonesian, Kita et al. 2009
A1	*A1*103:01*	Novel	AB583236	Indonesian		
A1	*A1*124:01*	Novel	AB583237	Malaysian		
A2	*A2*05:13-like*	Novel	AB647189	Philippino		
A2	*A2*05:16*		AM295878	Indonesian		Utrecht, Otting et al. 2007
A2	*A2*05:34-like*	Novel	AB647190	Philippino		
A3	*A3*13:03*		EU392112	Philippino		Philippino, Campbell et al. 2009
A3	*A3*13:15*	Novel	AB583238	Malaysian		
A3	*A3*13:16*	Novel	AB583240	Indonesian		
A4	*A4*14:01*		AM295880	Indonesian		Utrecht, Otting et al. 2007
A4	*A4*14:02*		AM295881	Malaysian		Utrecht, Otting et al. 2007
A6	*A6*01:05*	Novel	AB583239	Malaysian		

[a] New alleles are indicated as novel

[b] Nucleotide sequences were submitted to a public database and given accession numbers

[c] Origin of cynomolgus macaques

[d] Identical sequences were found in *Mamu* or *Mane* alleles

[e] Origin and references in which each known allele was first reported. Utrecht and Tsukuba indicate that the alleles were found in colonies maintained in the University of Utrecht, The Netherlands, and Tsukuba primate center, Japan, respectively

Table 4 Alleles of *Mafa-B* locus identified in the cynomolgus macaques

Allele name[a]	Novelty[b]	Accession number	Origin[c]	Identical to *Mamu* and/or *Mane* alleles[d]	Origin and reference of known alleles[c]
*B*002:03*	Novel	AB569224	Indonesian, Malaysian		
*B*004:01*		EU203722	Indonesian		Indonesian, Pendley et al. 2008
*B*007:01:01*		AY958137	Philippino	*Mamu-B*007:02/03*	Mauritian, Krebs et al. 2005
*B*007:01:02*		EU392135	Philippino		Philippino, Campbell et al. 2009
*B*007:01:03*	Novel	AB569223	Indonesian		
*B*007:03*		FM212802	Philippino		Indonesian or Malaysian, Otting et al. 2009
*B*011:02*	Novel	AB569229	Malaysian		
*B*013:08*		EU392114	Indonesian, Philippino		Philippino, Campbell et al. 2009
*B*017:01*		EU392119	Philippino		Philippino, Campbell et al. 2009
*B*018:01*		AY958138	Indonesian	*Mamu-B*018:01*	Mauritian, Krebs et al. 2005
*B*030:02*		AY958134	Malaysian	*Mamu-B*030:03:01*	Mauritian, Krebs et al. 2005
*B*032:01*	Novel	AB569237	Malaysian		
*B*033:02*		EU392118	Philippino		Philippino, Campbell et al. 2009
*B*043:01*	Novel	AB569230	Malaysian	*Mamu-B*043:01*	
*B*056:01*		AY958131	Indonesian	*Mamu-B*056:01*	Mauritian, Krebs et al. 2005
*B*056:02*		EU392128	Philippino		Philippino, Campbell et al. 2009
*B*057:03*	Novel	AB569231	Malaysian	*Mamu-B*057:06*	
*B*060:04*	Novel	AB569226	Indonesian		
*B*061:01*		AB195445	Malaysian	*Mamu-B*061:04:01, Mane-B*061:01*	Tsukuba, Uda et al. 2005
*B*061:02*	Novel	AB569233	Malaysian		
*B*064:02*		FM212804	Philippino		Indonesian or Malaysian, Otting et al. 2009
*B*068:04*	Novel	AB569236	Malaysian	*Mamu-B*068:04 , Mane-B*nov078*	
*B*069:02*		FM212842	Malaysian		Indonesian or Malaysian, Otting et al. 2009
*B*074:01:02-like*	Novel	AB647188	Philippino		
*B*074:02*	Novel	AB569228	Malaysian	*Mamu-B*074:01 /02*	
*B*076:04*	Novel	AB569232	Malaysian		
*B*081:01*	Novel	AB569225	Indonesian		
*B*089:01:01*		EU392131/ FJ178820	Philippino		Philippino, Campbell et al. 2009
*B*089:01:02*		EU392125	Indonesian, Malaysian, Philippino	*Mamu-B*089:01, Mane-B*089:02*	Philippino, Campbell et al. 2009
*B*090:01*		AB195436	Malaysian		Tsukuba, Uda et al. 2005
*B*091:01*	Novel	AB569240	Malaysian	*Mamu-B*091:02*	
*B*092:01:01*	Novel	AB569227	Malaysian	*Mamu-B*092:02*	
*B*095:01*		EU392113/ AY958148	Philippino		Mauritian, Krebs et al. 2005 Philippino, Campbell et al. 2009
*B*104:03*		EU392126	Philippino		Indonesian, Pendley et al. 2008 Philippino, Campbell et al. 2009
*B*116:01*		EU392123	Philippino		Philippino, Campbell et al. 2009
*B*121:01*		AB195455	Indonesian		Tsukuba, Uda et al. 2005
*B*124:01:02*	Novel	AB569235	Malaysian		
*B*136:02*		EU203720	Indonesian,		Indonesian, Pendley et al. 2008
*B*137:03*		EU392117/ EU203723	Indonesian, Philippino		Indonesian, Pendley et al. 2008
*B*137:04*	Novel	AB569239	Malaysian		

Table 4 (continued)

Allele name[a]	Novelty[b]	Accession number	Origin[c]	Identical to *Mamu* and/or *Mane* alleles[d]	Origin and reference of known alleles[e]
*B*138:02*	Novel	AB569234	Malaysian		
*B*151:02:02*	Novel	AB569222	Indonesian		
*B*155:02*	Novel	AB569238	Malaysian		
*B*157:01*		EU392121	Philippino		Philippino, Campbell et al. 2009
*B*158:01*		EU392122	Philippino		Philippino, Campbell et al. 2009
*B*160:01*		EU606042	Philippino		-

[a] New alleles are indicated as novel

[b] Nucleotide sequences were submitted to a public database and given accession numbers

[c] Origin of cynomolgus macaques

[d] Identical sequences were found in *Mamu* or *Mane* alleles

[e] Origin and references in which each known allele was first reported. Utrecht and Tsukuba indicate that the alleles were found in colonies maintained in the University of Utrecht, The Netherlands, and Tsukuba primate center, Japan, respectively

The *Mafa-A* alleles found in this study are listed in Table 3, where 21 alleles were from the major *Mafa-A1* locus, while the remaining 11 alleles were from the minor *Mafa-A* loci, 3 from *Mafa-A2*, 3 from *Mafa-A3*, 2 from *Mafa-A4*, and 1 from *Mafa-A6* alleles (Table 3). The major *Mafa-A1*alleles were defined by the sequence similarity to the known *Mafa-A1* alleles to be given official nomenclatures by IPD, except for *Mafa-A1*008:03*-like allele, and we confirmed that the frequencies of cDNA clones for *Mafa-A1* alleles were over 10% in each macaque. Similarly, alleles of minor *Mafa-A* genes, *Mafa-A2*, *-A3*, *-A4*, and *-A6* were defined by sequence similarity to the known alleles. They, except for two novel *Mafa-A2* alleles, were also given official names by IPD. On the other hand, a total of 46 *Mafa-B* alleles (Table 4) as well as 6 *Mafa-I* and 3 *Mafa-AG* alleles (Table 5) were identified. It was found that 2 out of

Table 5 Alleles of *Mafa-AG* and *Mafa-I* locus identified in the cynomolgus macaques

Locus	Allele name[a]	Novelty[b]	Accession number[c]	Origin[d]	Identical to *Mamu* and/or *Mane* alleles[e]	Origin and reference of known alleles[e]
AG	*AG*04:03*	Novel	AB569221	Malaysian		
AG	*AG1 like-1*	Novel	AB569220	Malaysian		
AG	*AG1 like-3*	Novel	AB583241	Philippino		
I	*I*01:01:01*		EU392139	Philippino		Philippino, Campbell et al. 2009
I	*I*01:09/01:08*		AB195465/AB195464	Indonesian, Malaysian		Tsukuba, Uda et al. 2005
I	*I*01:15*		FM246493	Philippino	*Mamu-I*01:06* , *Mamu-I*01:08:01*	Indonesian or Malaysian, Otting et al. 2009
I	*I*01:15 like-1*	Novel	AB569241	Indonesian, Malaysian		
I	*I*01:15 like-2*	Novel	AB569242	Malaysian	*Mamu-I*03:01:01*, *Mamu-I*01:07:01* , *Mamu-I*01:06:05*	
I	*I*01:18 like*	Novel	AB569243	Malaysian		

[a] Official allele names were not obtained for *AG1 like-1*, *AG1 like-3*, *I*01:15 like-1*, *I*01:15 like-2*, and *I*01:18 like* due to the limited sequence information

[b] New alleles are indicated as novel

[c] Nucleotide sequences were submitted to a public database and given accession numbers

[d] Origin of cynomolgus macaques

[e] Identical sequences were found in *Mamu* or *Mane* alleles

[f] Origin and references in which each known allele was reported. Tsukuba indicates that the alleles were found in colonies maintained in the Tsukuba primate center, Japan

21 (9.5%) *Mafa-A1a* alleles and 12 out of 46 (26.1%) *Mafa-B* alleles had identical sequences to *Mamu-A1* and *Mamu-B* alleles, respectively, implying a genetic admixture of cynomolgus macaques with rhesus macaques during the evolution (Otting et al. 2007; Bonhomme et al. 2009; Otting et al. 2009). Because we determined the nucleotide sequences only for exons 2, 3, and 4, two novel *Mafa-AG* alleles and three novel *Mafa-I* alleles were not given official names. As for the geographic distribution of *Mafa* class I alleles, there was no overlapping of *Mafa-A* alleles originated from different regions (Table 3), while there were a few *Mafa-B* and *Mafa-I* alleles commonly observed in macaques from different regions (Tables 4 and 5, respectively). When we looked into the presence of novel alleles in the geographic distribution, most of the novel alleles were obtained from Malaysian macaques, while almost all of the alleles found in Philippino macaques were not novel (Table 2).

Mafa class I haplotypes identified in the family study

We could identify the *Mafa-A* and *Mafa-B* alleles composing 23 different haplotypes from the segregation studies (Table 6). It was found that one to three expressing *Mafa-A*

Table 6 *Mafa class I* haplotypes identified in the cynomolgus macaques

ID[a]	Origin[b]	Haplotype[c]	*Mafa-A1* (major)	*Mafa-A* (minor)	*Mafa-AG*	*Mafa-B* (major)	*Mafa-B* (minor)	*Mafa-I*
P01	Indonesian	"a"	A1*002:01:02	A3*13:16		B*136:02		I*01:09/01:08
		"b"	A1*103:01			B*007:01:03, B*121:01	B*151:02:02	
P02	Malaysian	"c"	A1*023:01			B*090:01	B*011:02, B*074:02	
		"d"	A1*068:02			B*043:01	B*030:02, B*057:03	I*01:15 like-2
P03	Malaysian	"e"	A1*001:01, A1*032:05		AG1 like-1	B*068:04, B*124:01:02	B*032:01, B*061:01, B*089:01:02	
		"f"	A1*079:01		AG*04:03	B*061:02, B*138:02	B*155:02	
P04	Philippino	"g"	A1*089:02	A2*05:13-like, A3*13:03		B*137:03		
		"h"	A1*008:02			B*104:03		
P05	Philippino	"i"	A1*094:01			B*007:01:02	B*160:01	
		"j"	A1*008:02		AG1 like-3	B*157:01	B*017:01, B*089:01:02, B*116:01	I*01:01:01, I*01:15
P06	Philippino	"k"	A1*08:03-like	A2*05:34-like		B*074:01:02-like		
		"m"	A1*089:02	A3*13:03		B*007:03, B*064:02	B*089:01:01	
M01	Indonesian	"n"	A1*018:06	A2*05:16, A4*14:01		B*002:03		I*01:15 like-1
		"o"	A1*097:01			B*056:01	B*089:01:02	
M02	Indonesian	"p"	A1*097:01			B*137:03	B*013:08	
		"q"	A1*019:05			B*018:01	B*004:01, B*060:04, B*081:01	
M03	Malaysian	"r"	A1*054:01			B*002:03		I*01:15 like-1
		"s"	A1*056:02	A4*14:02		B*076:04		I*01:18 like
M04	Malaysian	"t"	A1*062:05			B*069:02	B*137:04	
		"u"	A1*124:01	A3*13:15		B*091:01		
M05	Philippino	"v"	A1*074:02, A1*093:01			B*007:01:01, B*158:01		
		"w"	A1*093:01			B*007:01:02	B*160:01	
M06	Philippino	"w"	A1*093:01			B*007:01:02	B*160:01	
		"y"	A1*052:02			B*033:02, B*095:01		

[a] ID of founder animals as indicated in Fig. 1

[b] Origin of cynomolgus macaques

[c] Haplotypes were determined from studies of family as shown in Fig. 1

alleles and one to five expressing *Mafa-B* alleles consisted of *Mafa class I* haplotype, similar to the *Mamu class I* haplotypes in rhesus macaques (Naruse et al. 2010). Of particular interest was that there were two haplotypes, "e" (Malaysian founder P03) and "v" (Philippino founder M05), carrying two different *Mafa-A1* genes (Fig. 1; Table 6). Because previous studies have demonstrated that there is usually only one *Mafa-A1* allele on a chromosome (Otting et al. 2007), while the presence of two *Mamu-A1* alleles on the same haplotype was suggested in rhesus macaques (Naruse et al. 2010; Doxiadis et al. 2011), we performed further analyses.

The family studies showed that the *Mafa-A1*alleles consisting of haplotype "e", *Mafa-A1*001:01* and*Mafa-A1*032:05*, or haplotype "v", *Mafa-A1*074:02* and *Mafa-A1*093:01*, did not carry accompanying minor *Mafa-A* genes (Table 6). When we constructed a phylogenetic tree of *Mafa-A* alleles identified in this study (Fig. 2), it was found that *Mafa-A1*001:01* was mapped in the neighbor of *Mafa-A3*gene, raising a possibility that one of the two alleles on the same chromosome might be a minor *Mafa-A* allele and not the major *Mafa-A1* allele. To test the possibility, we investigate the expression level of *Mafa-A* alleles composing of haplotypes "e" and "v". For this purpose, other primer pairs were designed within the sequences completely shared by these alleles to amplify the *Mafa-A* cDNAs to avoid a possibility of affecting the efficacy of PCR by mismatches with the primer sequences. The cloning and sequencing analysis revealed that both *Mafa-A1*001:01* and *Mafa-A1*032:05* on the haplotype "e" were observed at similar frequencies among the cDNA clones of *Mafa-A* alleles in P03and C008 (Fig. 1): 29.7% and 33.3% in P03 and 22.5% and 17.5% in C008, respectively. Similarly, frequencies of haplotype "v" alleles, *Mafa-A1*074:02* and *Mafa-A1*093:01*, in cDNA clones were 59.5% and 40.5%, respectively, in M05, while those in C010 were 23.3% and 26.7% and 31.4% and 17.1% in C011, respectively. The frequencies of cDNA clones varied in different individuals presumably due to the allelic competition with the alleles of another haplotype in each individual (Fig. 1), but they were much higher than the frequencies of the minor *Mafa-A* allele (*Mafa-A3*13:03*) clones: 3.3% and 2.9% in C010 and C011, respectively. These observations indicated that two *Mafa-A* alleles were considered to be major *Mafa-A1* alleles in both haplotypes "e" and "v".

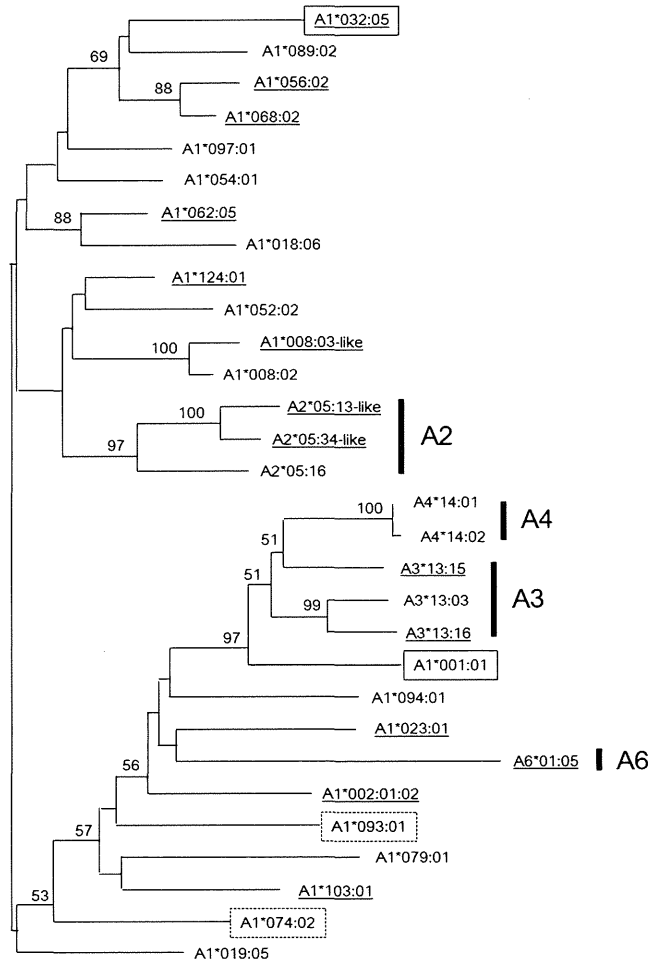


Fig. 2 Phylogenetic tree of *Mafa-A* alleles. A phylogenetic tree of the *Mafa-A* alleles detected in this study was constructed by using the neighbor-joining method with a bootstrap value of 5,000 replications. Values more than 50% are indicated as percentages. Novel alleles were *underlined*. *Mafa-A*1 alleles consisting of haplotype "e" are *boxed*, while the *stippled boxes* represent the alleles on haplotype "v". Alleles of minor *Mafa-A* genes, *Mafa-A2*, *A3*, *A4*, and *A6*, are also indicated

Discussion

Native cynomolgus macaques are widespread throughout the islands of Southeast Asia into mainland Asia. They are mainly found in Indonesia, Malaysia, and the Philippines, then Burma, India, Vietnam, Cambodia, Laos, and Thailand (Lang 2006). It was suggested that the founding population of Mauritian macaques was introduced from Indonesia (Pendley et al. 2008; Campbell et al. 2009). More than 40% of *Mafa* class I alleles observed in this study were novel, even though there have been many reports on the analysis of *Mafa class I* genes, demonstrating that the diversity of MHC in the cynomolgus macaques still needs to be investigated. When we considered the origin of founders, 73.7% (28/38) were novel in alleles found in Malaysian macaques, while only 15.6% (5/32) were novel alleles in Philippino macaques (Table 2). The geographic distribution of novel alleles may be due to the fact that the Malaysian macaques had not been extensively analyzed before (Otting et al. 2007;

Pendley et al. 2008; Kita et al. 2009). In the present study, *B*089:01:02* was found in individuals among Indonesian, Malaysian, and Philippino macaques in different *Mafa-B* haplotypes (Table 6). Likewise, *B*137:03* was found in Indonesian and Malaysian macaques (Table 4). In addition, shared alleles among the cynomolgus macaques, rhesus macaques, and pig-tailed macaques (*Macaca nemestrina*) were noted (Tables 3, 4, and 5). These observations indicated that the diversity of *MHC class I* genes is similar not only in the cynomolgus macaque population but also among the Old World monkeys, suggesting that the *MHC class I* polymorphisms might be generated before the divergence of Old World monkeys and/or there were admixtures of the Old World monkeys.

In this study, we determined the haplotype structure of *Mafa* class I locus by family studies and a total of 23 haplotypes were identified. Among them, haplotypes "i" and "w" carried identical *Mafa-B* alleles but different *Mafa-A* alleles (Table 6), suggesting that there were haplotypes originated by a recombination between the *Mafa-A* and *Mafa-B* loci. We showed that the *Mafa class I* haplotypes were usually composed of one to three *Mafa-A* alleles and one to five *Mafa-B* alleles, similar to the *Mamu class I* haplotypes, of which usually one *MHC-A1* gene and a few (one to three) *MHC-B* genes were highly transcribed (Otting et al. 2007, 2008; Naruse et al. 2010; Doxiadis et al. 2011). As for the *MHC-A* locus in the cynomolgus macaques, highly transcribed *Mafa-A1*gene and other minor *Mafa-A* genes, such as *Mafa-A2*, *-A3*, *-A4*, and *-A6* could be detected. It was reported that 87% of cynomolgus macaques had at least one *Mafa-A2* alleles (Wu et al. 2008). However, only 3 out of 23 (13.0%) haplotypes carried a *Mafa- A2* allele in this study (Table 6). We could not exclude a possibility that the strategy of our study might not be sufficient to detect the *Mafa-A* genes with low expression and/or the alleles with mismatches at the primer site, based on the number of clones within a PCR sample. Such a possibility is unlikely because we used the primer pairs which could cover the known *Mafa-A2* alleles, although there might be novel *Mafa-A2* alleles having different sequences at the primer binding sites. Therefore, we might underestimate the complexity of *Mafa* class I alleles in this study. High-throughput pyrosequencing methods may be a useful strategy to avoid the possibility of missing alleles, as described by several investigators (Wiseman et al. 2009; Budde et al. 2010; Aarnink et al. 2011b). In addition, because it was reported that the cell surface expression of Mamu class I molecule was varied depending on the locus and allelic structure (Rosner et al. 2010), locus- and allele-dependent expression of Mafa class I molecule at the cell surface will be required.

The most important finding in this study was that we demonstrated evidence for the presence of haplotypes carrying two major *MHC-A1* genes on the same chromosome from the family studies and additional cloning studies. Interestingly, we and others have reported similar phenomena in rhesus macaques (Naruse et al. 2010; Doxiadis et al. 2011). In addition, several haplotypes carried multiple major *Mafa-B1* alleles (Table 6), similar to the *Mamu-B1* locus (Otting et al. 2008; Doxiadis et al. 2011). The *raison d'etre* of multiple major *MHC class I* genes/alleles on the same chromosome may be that they play an immunological role as the "double lock strategy" (Doxiadis et al. 2011) in which the double *MHC-A1* alleles of high transcription level might be favorable to present peptide to CD8+ T cells. However, there is another unique haplotype which carries no *MHC-A1*allele in cynomolgus macaques (Otting et al. 2007) and maybe in rhesus macaques (Doxiadis et al. 2011). These observations suggested that the diversity of MHC in the Old World monkey is far more complicated than in humans.

In summary, we investigated 26 cynomolgus macaques from five families for the diversity of *MHC class I* alleles and haplotypes. A total of 87 alleles were identified, of which 40 were novel. There were 23 different haplotypes, and two of them carried two *MHC-A1* genes, demonstrating further the complexity of *MHC class I* locus in the Old World monkey.

Acknowledgments We thank Ms. Yukiko Ueda for her technical assistance. We acknowledge Dr. Natasja de Groot and Dr. Nel Otting for assigning nomenclature of *Mafa class I* alleles. This work was supported in part by research grants from the Ministry of Health, Labor and Welfare, Japan; the Japan Health Science Foundation; the program of Founding Research Centers for Emerging and Reemerging Infection Disease; the program of Research on Publicly Essential Drugs and Medical Devices; and Grant-in-Aids for scientific research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. This work was also supported by a program of support for women researchers from the Tokyo Medical and Dental University.

References

Aarnink A, Dereuddre-Bosquet N, Vaslin B, Le Grand R, Winterton P, Apoil PA, Blancher A (2011a) Influence of the MHC genotype on the progression of experimental SIV infection in the Mauritian cynomolgus macaque. Immunogenetics 63:267–274

Aarnink A, Apoil PA, Takahashi I, Osada N, Blancher (2011b) Characterization of MHC class I transcripts of a Malaysian cynomolgus macaque by high-throughput pyrosequencing and EST libraries. Immunogenetics. doi:10.1007/s00251-011-0550-8

Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B (2009) Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). J Hered 100:158–169

Budde ML, Wiseman RW, Karl JA, Hanczaruk B, Simen BB, O'connor DH (2010) Characterization of Mauritian cynomolgus macaque major histocompatibility complex class I haplotypes by high-resolution pyrosequencing. Immunogenetics 62:773–780

Burwitz BJ, Pendley CJ, Greene JM, Detmer AM, Lhost JJ, Karl JA, Piaskowski SM, Rudersdorf RA, Wallace LT, Bimber BN, Loffredo JT, Cox DG, Bardet W, Hildebrand W, Wiseman RW, O'Connor SL, O'Connor DH (2009) Mauritian cynomolgus macaques share two exceptionally common major histocompatibility complex class I alleles that restrict simian immunodeficiency virus-specific CD8+ T cells. J Virol 83:6011–6019

Campbell KJ, Detmer AM, Karl JA, Wiseman RW, Blasky AJ, Hughes AL, Bimber BN, O'Connor SL, O'Connor DH (2009) Characterization of 47 MHC class I sequences in Filipino cynomolgus macaques. Immunogenetics 61:177–187

Doxiadis GG, de Groot N, Otting N, Blokhuis JH, Bontrop RE (2011) Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. Immunogenetics 63:73–83

Gibbs RA, Rogers J, Katze MG et al (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234

Karl JA, Wiseman RW, Campbell KJ, Blasky AJ, Hughes AL, Ferguson B, Read DS, O'Connor DH (2008) Identification of MHC class I sequences in Chinese-origin rhesus macaques. Immunogenetics 60:37–46

Kita YF, Hosomichi K, Kohara S, Itoh Y, Ogasawara K, Tsuchiya H, Torii R, Inoko H, Blancher A, Kulski K, Shiina T (2009) MHC class I A loci polymorphism and diversity in three Southeast Asian populations of cynomolgus macaques. Immunogenetics 61:635–648

Kobasa D, Jones SM, Shinya K, Kash JC, Copps J, Ebihara H, Hatta Y, Kim JH, Halfmann P, Hatta M, Feldmann F, Alimonti JB, Fernando L, Li Y, Katze MG, Feldmann H, Kawaoka Y (2007) Aberrant innate immune response in lethal infection of macaques with the 1918 influenza virus. Nature 445:319–323

Krebs KC, Jin Z, Rudersdorf R, Hughes AL, O'Connor DH (2005) Unusually high frequency MHC class I alleles in Mauritian origin cynomolgus macaques. J Immunol 175:5230–5239

Kulski JK, Anzai T, Shiina T, Inoko H (2004) Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol Biol Evol 21:2079–2091

Lang KC (2006) http://pin.primate.wisc.edu/factsheets/entry/long-tailed_macaque

Lawler JV, Endy TP, Hensley LE, Garrison A, Frinz EA, Lesar M, Baric RS, Kulesh DA, Norwood DA, Wasieloski LP, Ulrich MP, Slezak TR, Vitalis E, Hugging JW, Jahrling PB, Paragas J (2006) Cynomolgus macaque as an animal model for severe acute respiratory syndrome. PLoS Med 3:677–686

Loffredo JT, Bean AT, Beal DR, León EJ, May GE, Piaskowski SM, Furlott JR, Reed J, Musani SK, Rakasz EG, Friedrich TC, Wilson NA, Allison DB, Watkins DI (2008) Patterns of CD8+ immunodominance may influence the ability of Mamu-B*08-positive macaques to naturally control simian immunodeficiency virus SIVmac239 replication. J Virol 82:1723–1738

Matano T, Kobayashi M, Igarashi H, Takeda A, Nakamura H, Kano M, Sugimoto C, Mori K, Iida A, Hirata T, Hasegawa M, Yuasa T, Miyazawa M, Takahashi Y, Yasunami M, Kimura A, O'Connor DH, Watkins DI, Nagai Y (2004) Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial. J Exp Med 199:1709–1718

Mee ET, Berry N, Ham C, Sauermann U, Maggiorella MT, Martinon F, Verschoor EJ, Heeney JL, Le Grand R, Titti F, Almond N, Rose NJ (2009) Mhc haplotype H6 is associated with sustained control of SIVmac251 infection in Mauritian cynomolgus macaques. Immunogenetics 61:327–339

Naruse TK, Chen Z, Yanagida R, Yamashita T, Saito Y, Mori K, Akari H, Yasutomi Y, Miyazawa M, Matano T, Kimura A (2010) Diversity of MHC class I genes in Burmese-origin rhesus macaques. Immunogenetics 62:601–611

Otting N, deVos-Rouweler AJM, Heijmans CMC, de Groot NG, Doxiadis GG, Bontrop RE (2007) MHC class I A region diversity and polymorphism in macaque species. Immunogenetics 59:367–375

Otting N, Heijmans CM, van der Wiel M, de Groot NG, Doxiadis GG, Bontrop RE (2008) A snapshot of the Mamu-B genes and their allelic repertoire in rhesus macaques of Chinese origin. Immunogenetics 60:507–514

Otting N, Doxiadis GG, Bontrop RE (2009) Definition of Mafa-A and -B haplotypes in pedigreed cynomolgus macaques (*Macaca fascicularis*). Immunogenetics 61:745–753

Pendley CJ, Becker EA, Karl JA, Blasky AJ, Wiseman RW, Hughes AL, O'Connor SL, O'connor DH (2008) MHC class I characterization of Indonesian cynomolgus macaques. Immunogenetics 60:339–351

Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P, Marsh SGE (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucl Acid Res 31:311–314

Rosner C, Kruse PH, Lübke T, Walter L (2010) Rhesus macaque MHC class I molecules show differential subcellular localization. Immunogenetics 62:149–158

Slukvin II, Lunn DD, Watkins DI, Golos TG (2000) Placental expression of the nonclassical MHC class I molecule Mamu-AG at implantation in the rhesus monkey. Proc Natl Acad Sci USA 97:9104–9109

Tanaka-Takahashi Y, Yasunami M, Naruse T, Hinohara K, Matano T, Mori K, Miyazawa M, Honda M, Yasutomi Y, Nagai Y, Kimura A (2007) Reference strand-mediated conformation analysis-based typing of multiple alleles in the rhesus macaque MHC class I Mamu-A and Mamu-B loci. Electrophoresis 28:918–924

Tsukamoto T, Dohki S, Ueno T, Kawada M, Takeda A, Yasunami M, Naruse T, Kimura A, Takiguchi M, Matano T (2008) Determination of a major histocompatibility complex class I restricting simian immunodeficiency virus Gag241-249 epitope. AIDS 22:993–998

Uda A, Tanabayashi K, Yamada YK, Akari H, Lee YJ, Mukai R, Terao K, Yamada A (2004) Detection of 14 alleles derived from the MHC class I A locus in cynomolgus monkeys. Immunogenetics 56:155–163

Uda A, Tanabayashi K, Fujita O, Hotta A, Terao K, Yamada A (2005) Identification of the MHC class I B locus in cynomolgus monkeys. Immunogenetics 57:189–197

Urvater JA, Otting N, Loehrke JH, Rudersdorf R, Slukvin II, Piekarcyk MS, Golos TG, Hughes AL, Bontrop RE, Watkins DI (2000) Mamu-I: a novel primate MHC class I B-related locus with unusually low variability. J Immunol 164:1386–1398

Watanabe A, Shiina T, Shimizu S, Hosomichi K, Yanagiya K, Kita YF, Kimura T, Soeda E, Torii R, Ogasawara K, Kulski JK, Inoko H (2006) A BAC-based contig map of the cynomolgus macaque (*Macaca fascicularis*) major histocompatibility complex genomic region. Genomics 89:402–412

Wiseman RW, O'Connor DH (2007) Major histocompatibility complex-defined macaques in transplantation research. Transplant Rev 21:17–25

Wiseman RW, Wojcechowskyj JA, Greene JM, Blasly AJ, GoponT SomaT, FriedrichT C, O'Connor SL, O'Connor DH (2007)

Simian immunodeficiency virus SIV mac239 infection of major histocompatibility complex-identical cynomolgus macaques from Mauritius. J Virol 81:349–361

Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres E Jr, Wright C, Harkins T, O'Connor DH (2009) Major histocompatibility complex genotyping with massively parallel pyrosequencing. Nat Med 15:1322–1326

Wu J, Bassinger S, Montoya GD, Chavez L, Jones CE, Holder-Lockyer B, Masten B, Williams TM, Prilliman KR (2008) Allelic diversity within the high frequency Mamu-A2*05/Mane-A2*05(Mane-A*06)/Mafa-A2*05 family of macaque MHC-A loci. Tissue Antigens 72:29–38

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

Butyl Pocket Formation in the Vitamin D Receptor Strongly Affects the Agonistic or Antagonistic Behavior of Ligands

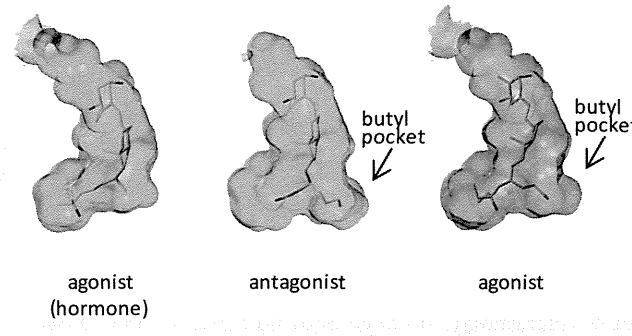
Nobuko Yoshimoto,[†,‡,⊥] Yuta Sakamaki,[†,§,⊥] Minoru Haeta,[†] Akira Kato,[†] Yuka Inaba,[†] Toshimasa Itoh,[†] Makoto Nakabayashi,[∥] Nobutoshi Ito,[∥] and Keiko Yamamoto*[,†,‡]

[†]Laboratory of Drug Design and Medicinal Chemistry and [‡]High Technology Research Center, Showa Pharmaceutical University, 3-3165 Higashi-Tamagawagakuen, Machida, Tokyo 194-8543, Japan

[§]Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University, 2-3-10 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062, Japan

[∥]Graduate School of Biomedical Science, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo 113-8510, Japan

ABSTRACT: Previously, we reported that 22*S*-butyl-25,26,27-trinor-1α,24-dihydroxyvitamin D_3 **2** represents a new class of antagonist for the vitamin D receptor (VDR). The crystal structure of the ligand-binding domain (LBD) of VDR complexed with **2** showed the formation of a butyl pocket to accommodate the 22-butyl group and insufficient interactions between ligand **2** and the C-terminus of VDR. Here, we designed and synthesized new analogues **5a–c** and evaluated their biological activities to probe whether agonistic activity is recovered when the analogue restores interactions with the C-terminus of VDR. Analogues **5a–c** exhibited full agonistic activity in transactivation. Interestingly, **5c**, which bears a 24-diethyl group, completely recovered agonistic activity, although **3c** and **4c** act as an antagonist and a weak agonist, respectively. The crystal structures of VDR-LBD complexed with **3a**, **4a**, **5a**, and **5c** were solved, and the results confirmed that butyl pocket formation in VDR strongly affects the agonistic or antagonistic behaviors of ligands.



INTRODUCTION

Active vitamin D_3, 1α,25-dihydroxyvitamin D_3 **1a** (Chart 1), is a hormone that plays a major role in calcium homeostasis. This hormone is also involved in cell differentiation and proliferation and immunomodulation.[1] The actions of **1a** and its active analogues are mediated by the vitamin D receptor (VDR), which is a member of the nuclear receptor superfamily that includes the receptors for steroid, retinoid, and thyroid hormones.[2] These nuclear receptors function through a common mechanism by which the receptors regulate the transcription of their target genes. VDR forms a heterodimer with retinoid X receptor (RXR), and upon ligand binding, VDR changes to an active conformation that provides the activation function 2 (AF2) surface to allow binding of a coactivator.[3]

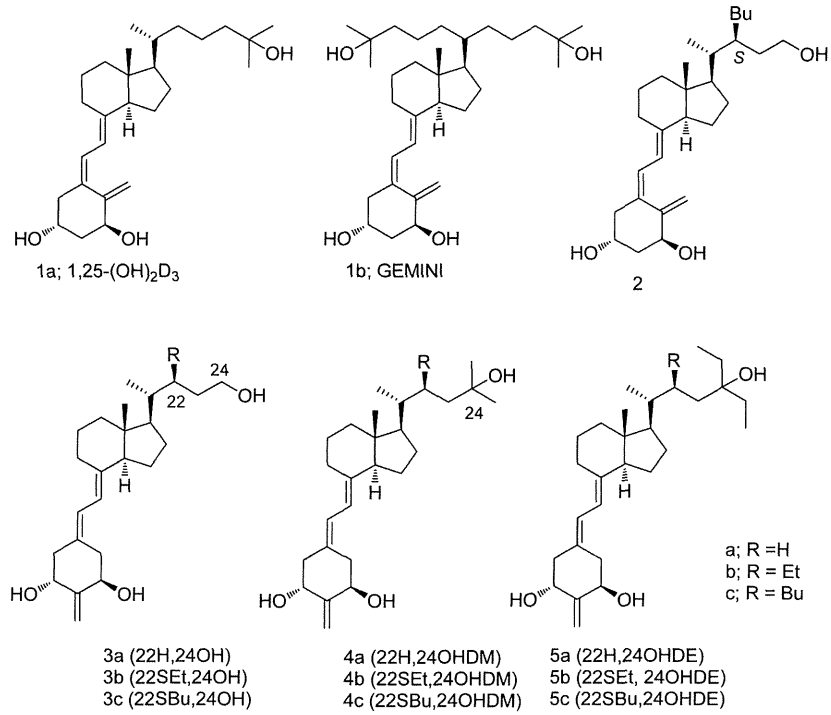
Since the discovery of **1a**, thousands of vitamin D analogues have been synthesized,[3] of which more than 10 have been clinically used to treat metabolic bone diseases and skin diseases such as psoriasis, immune disorders, and malignant tumors.[4,5] All these clinically used analogues are VDR agonists. VDR antagonists have been synthesized by several groups and are classified structurally into two types. The first group is composed of analogues containing a bulky side chain, such as carboxylic ester (ZK series) compounds[6] and adamantane compounds,[7] while the second group is composed of analogues lacking a bulky side chain, such as (23*S*)-25-dehydro-1α-hydroxyvitamin D_3-26,23-lactone (TEI9647)[8] and its derivatives.[8−11] VDR antagonism is believed to be based on an unstable conformation of VDR generated upon ligand binding. A VDR that accommodates an antagonist would prevent the heterodimerization with RXR and/or the recruitment of coactivators. However, the molecular basis for VDR antagonism is not clearly understood.

Recently, we reported that 22*S*-butyl-25,26,27-trinor-1α,24-dihydroxyvitamin D_3 **2** acts as a third type of VDR antagonist.[12] The crystal structure of the ligand binding domain (LBD) of VDR complexed with **2** shows the formation of an extra cavity to accommodate the butyl group.[12] We have termed this extra cavity the "butyl pocket". Several research groups have solved crystal structures of VDR-LBD complexed with a variety of ligands.[13−25] Most of the crystal structures show the canonical Moras' active conformation[13] of VDR-LBD and similar architectures of the ligand binding pocket (LBP). Exceptions are the complex of zebrafish VDR-LBD with "GEMINI",[16,20] which has two identical side chains,[26,27] and the complex of VDR-LBD with 22-butyl analogue **2**.[12] The crystal structure of the zebrafish VDR-LBD/GEMINI complex revealed the formation of a new cavity, which is an extension of the original LBP, in order to accommodate the second side chain. The crystal structure of the VDR-LBD/**2** complex also revealed the

Received: February 21, 2012
Published: April 18, 2012

dx.doi.org/10.1021/jm300230a | *J. Med. Chem.* 2012, 55, 4373–4381

Chart 1. Structures of 1a and Vitamin D Analogues



formation of a cavity similar to that of the VDR-LBD/GEMINI complex. These observations suggest that the region around helix 6, loop 6–7, and the N-terminus of helix 7 is somewhat flexible. Interestingly, although both GEMINI **1b** and 22-butyl analogue **2** induce the formation of an extra cavity, **1b** acts as a VDR agonist whereas 22-butyl analogue **2** acts as a VDR antagonist. We hypothesized that the contrasting behaviors of **1b** and **2** against VDR were derived from differences in hydrophobic interactions between the side chain terminus of each ligand and the C-terminus of VDR.

To investigate the effects on antagonism of the induction of the butyl pocket and of insufficient interactions of the ligand with the C-terminus of VDR, we synthesized a series of vitamin D_3 analogues with and without a 22-alkyl substituent and evaluated their biological activities.[28] We found that these synthetic ligands show strong antagonist, partial agonist, or full agonist activities, depending on the primary or tertiary alcohol at the side chain terminus and on the length of the 22-alkyl group. The results suggested that both the induction of a butyl pocket and insufficient hydrophobic interactions with the VDR C-terminus are necessary for VDR antagonism.

The present study was conducted to further confirm the molecular basis of agonistic and antagonistic activity of a new class of ligands that induce the formation of a butyl pocket. Here, we report the design, synthesis, and biological evaluation of new analogues in order to probe whether agonistic activity increases when the analogue exhibits increased interactions with the C-terminus of VDR. Furthermore, X-ray crystallographic analyses of VDR-LBD complexed with those analogues, which contain or lack a 22-butyl substituent, are also reported.

DESIGN AND SYNTHESIS

Design. Previously we showed that a 22-butyl analogue with a 24-primary alcohol, **3c**, is a VDR ligand but does not activate VDR whereas the corresponding 24-tertiary alcohol **4c** activates VDR, albeit weakly.[28] The results implied that the introduction of a more hydrophobic substituent at the 24-position would enhance VDR activation. Therefore, we designed 22-butyl analogue **5c** with a diethyl substituent instead of a dimethyl at C(24). 22-Ethyl analogue **5b** and 22-H analogue **5a**, both of which have a diethyl substituent at C(24), were also designed as counterparts of **5c**. As described in earlier paper, we selected the 2-methylene-19-nor structure because of its superior chemical stability, convenient synthesis, and its increased biological activity compared to the original A-ring structure.[28,29]

Synthesis of the 1,24-Dihydroxyvitamin D_3 Analogues. As shown in Scheme 1, 1,24-dihydroxyvitamin D_3 derivative **5a** was synthesized from ester **6**. Ester **6**, prepared by the procedure reported previously,[28] was treated with ethyllithium in THF to afford 24,24-diethylated compound **7**. Diethyl compound **7** was deprotected with CSA in methanol to provide desired compound **5a**. 22-Ethyl analogue **5b** was derived from 22-ethyl ester **8**.[28,30] Ester **8** was treated with EtMgBr to afford 24,24-diethylated compound **9**. Compound **9** was deprotected with CSA in methanol to provide desired compound **5b**. 22-Butyl analogue **5c** was derived from 22-butyl ester **10**.[28,30] Ester **10** was treated with EtMgI to afford 24,24-diethylated compound **11**. Compound **11** was deprotected with CSA in methanol to provide desired compound **5c**.

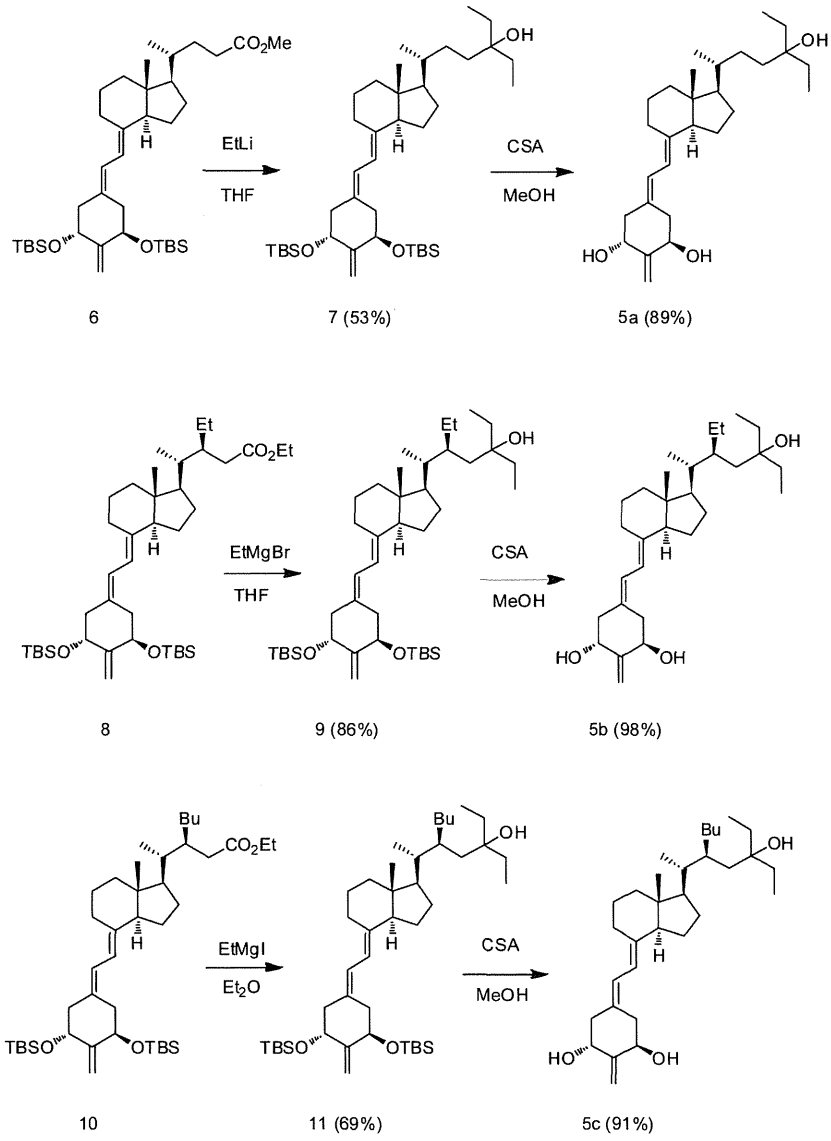
BIOLOGICAL ACTIVITIES

VDR Binding. Binding affinity to VDR was evaluated by a competitive binding assay using recombinant human VDR-LBD expressed as a C-terminus GST-tagged protein using the pGEX-VDR vector[28,31] in *Escherichia coli* BL21. The results are summarized in Table 1 including previously reported results.[28] Three synthetic compounds **5a–c** showed specific binding to VDR, indicating they are ligands for VDR. In particular, **5a** showed stronger affinity than **1a**.

Transactivation. The ability of synthetic compounds **5a–c** to induce transcription of a vitamin D-responsive gene was tested using the mouse osteopontin luciferase reporter gene

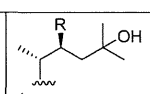
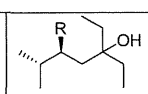
dx.doi.org/10.1021/jm300230a | J. Med. Chem. 2012, 55, 4373–4381

Scheme 1. Synthesis of Compounds 5a–c



assay system in Cos7 cells.[12] The results are shown in Figure 1, together with the results for **3a–c** and **4a–c**.[28] New analogues **5a–c** showed concentration-dependent transcriptional activity and acted as potent agonists. Of the 22-butyl analogues, new analogue **5c** completely restored full agonistic activity. In the 22-H analogues, the EC_{50} of transcriptional activity decreased in the order 24OH **3a** > 24OHDM **4a** > 24OHDE **5a**. In the 22-Et analogues, it is interesting that **5b** acted as a full agonist while **3b** and **4b** are partial agonists.

Table 1. VDR Binding Affinity of Synthetic Analogues 3–6, IC_{50} (nM)[a]

Compd.	1a	 3	 4	 5
a: R = H	0.08	2.7	0.10	0.029
b: R = Et		0.88	0.72	0.38
c: R = Bu		0.13	0.21	0.41

[a]Competitive binding of **1a** and synthetic compounds **5a–c** to the human vitamin D receptor. Affinity of related compounds (**3a–c** and **4a–c**) was cited from previous report[28] to compare with **5a–c**. The experiments were carried out in duplicate. The IC_{50} values are derived from dose–response curves and represent the compound concentration required for 50% displacement of radiolabeled 1α,25-dihydroxyvitamin D_3 from the receptor protein.

X-RAY CRYSTAL STRUCTURE

We attempted to crystallize the VDR-LBD complex with the 22-H analogues (**3a**, **4a**, and **5a**) and 22-butyl analogues (**3c**, **4c**, and **5c**). We were able to obtain crystals in the presence of, but not in the absence of, a coactivator peptide, which is the same result as in our previous report.[12] Good quality crystals of VDR complexed with **3a**, **4a**, **5a**, or **5c** were obtained, but good crystals accommodating antagonists **3c** or **4c** were not obtained. As a result, we could solve the crystal structure of the complex with **3a**, **4a**, **5a** (Figure 2), or **5c** (Figure 3) but not the complex with **3c** or **4c**. The crystallographic analysis data are summarized in Table 2. The overall protein structure of VDR-LBD complexed with **3a**, **4a**, **5a**, or **5c** adopted the canonical Moras' active conformation observed in the complex with **1a**.[13] Coactivator peptide is also closely bound to the AF2

dx.doi.org/10.1021/jm300230a | J. Med. Chem. 2012, 55, 4373–4381

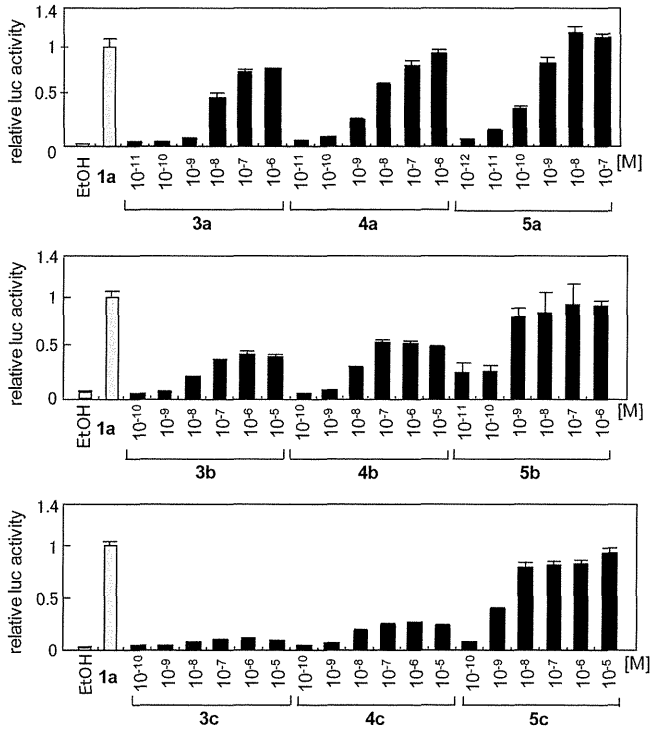


Figure 1. Transactivation of compounds **5a–c** in Cos7 cells, together with the results of **3a–c** and **4a–c** previously reported.[28] Transcriptional activity was evaluated by the dual luciferase assay using a full-length human VDR expression plasmid (pCMX-hVDR), a reporter plasmid containing three copies of mouse osteopontin VDRE (SPPx3-TK-Luc), and an internal control plasmid containing sea pansy luciferase expression constructs (pRL-CMV) in Cos7 cells as described previously.[12] Luciferase activity of 10^{-8} M **1a** was defined as 1.

surface. These results indicate that the complexes with **3c** or **4c**, for which good quality crystals were not obtained, are difficult to adopt the canonical Moras' active conformation.

The entire main-chain structure of the VDR-LBD/**5c** complex is similar to the complex with **1a** or antagonist **2**. The 1α-, 3β-, and 25-hydroxyl groups of **5c** form pincer-type hydrogen bonds, as is the case with most vitamin D analogues (Figure 3a). Formation of a butyl pocket to accommodate the 22-butyl group was clearly observed (Figure 3b). The residues within 4.5 Å from each ligand are summarized in Table 3; these residues in **5c** are identical with those, except Leu400, in **1a**.

In the complexes with **3a**, **4a**, or **5a**, the side chain conformations of the amino acid residues lining the LBP are superimposed onto those of the VDR-LBD/**1a** complex. In the crystallographic analysis, ligands **4a** and **5a** docked in the VDR are well-defined, whereas the electron density map of the side chain of ligand **3a** in the VDR is unclear, probably because of its flexibility. Therefore, the side chain conformation of **3a** was tentatively determined based on the electron density map and the conformation that allowed hydrogen bonding with His393 and/or His301. Ligands **3a**, **4a**, and **5a** were accommodated into the VDR-LBP in a manner similar to **1a**. The hydroxyl groups at the 1α and 3β positions form pincer-type hydrogen bonds, evident in the VDR-LBD/**1a** complex (Figure 2a). Furthermore, the 24-hydroxyl group of **4a** and **5a** forms hydrogen bonds with both His393 and His301, and that of **3a** forms a hydrogen bond with His393 (Figure 2a). As shown in Table 3, compared with the complex with **1a**, complexes with **3a** and **4a** showed reduced interactions.

DISCUSSION

All synthetic analogues with a 24-hydroxyl group at the side chain showed specific binding for hVDR (Table 1). Therefore, compounds **5a–c** synthesized here are true ligands for hVDR. Interestingly, compound **5a** showed stronger affinity for hVDR

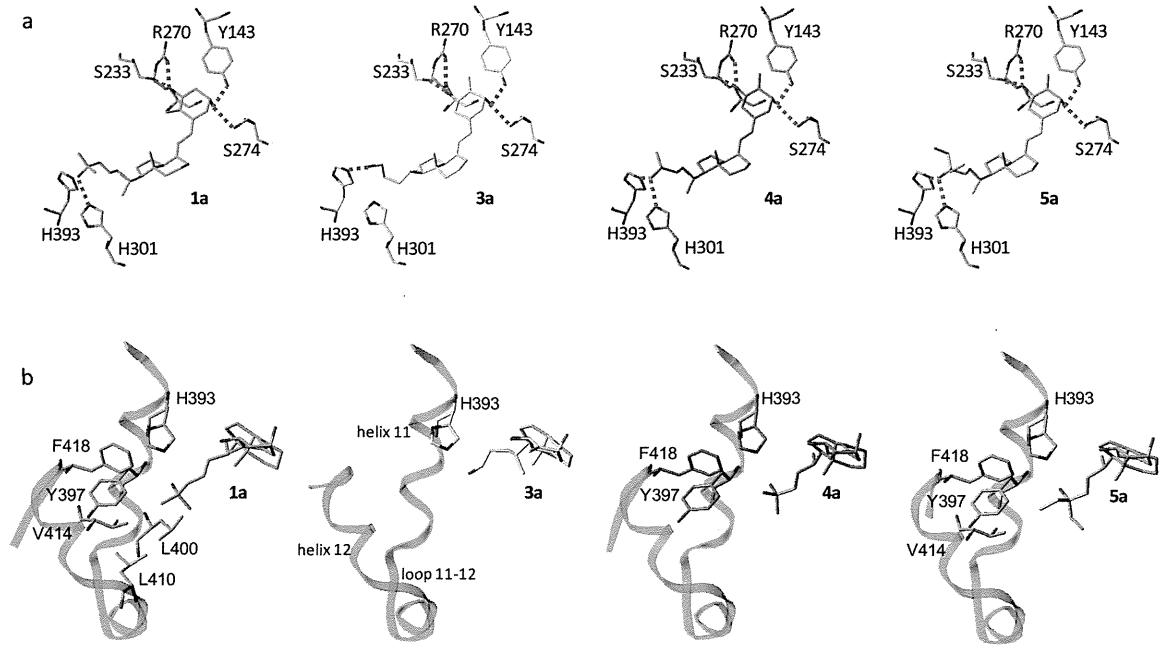


Figure 2. Crystal structures of VDR-LBD complexed with **1a**, **3a**, **4a**, and **5a**: (a) hydrogen bonds between VDR-LBD and **1a** (green), **3a** (gray), **4a** (purple), or **5a** (yellow); (b) interactions between the VDR C-terminus (helix 11, loop 11–12, and helix 12) and the ligands. The C-terminus (helix 11, loop 11–12, and helix 12) is presented as a green ribbon, and the ligands are presented in the same colors as in (a). Residues within 4.5 Å from each ligand are presented.

dx.doi.org/10.1021/jm300230a | J. Med. Chem. 2012, 55, 4373–4381