Reactivation by exon shuffling of a conserved *HLA-DR3*-like pseudogene segment in a New World primate species

Gaby G. M. Doxiadis^{†‡}, Marit K. H. van der Wiel[†], Herbert P. M. Brok[§], Natasja G. de Groot[†], Nel Otting[†], Bert A. 't Hart[†], Jon J. van Rood^{||†}, and Ronald E. Bontrop[†]

Departments of [†]Comparative Genetics and Refinement, [§]Animal Science, and ^{||}Immunobiology, Biomedical Primate Research Centre, P.O. Box 3306, 2280 GH, Rijswijk, The Netherlands; and ^{||}Department of Immunohematology and Blood Transfusion, Leiden University Medical Centre, E3-Q, P.O. Box 9600, 2300 RC, Leiden, The Netherlands

Contributed by Jon J. van Rood, February 1, 2006

The common marmoset (*Callithrix jacchus*), a New World monkey species with a limited MHC class II repertoire, is highly susceptible to certain bacterial infections. Genomic analysis of exon 2 sequences documented the existence of only one *DRB* region configuration harboring three loci. Two of these loci display moderate levels of allelic polymorphism, whereas the *-DRB*W12* gene appears to be monomorphic. This study shows that only the *Caja-DRB*W16* and *-DRB*W12* loci produce functional transcripts. The *Caja-DRB1*03* locus is occupied by a pseudogene, given that most of the transcripts, if detected at all, show imperfections and are present at low levels. Moreover, two hybrid transcripts were identified that feature the evolutionarily conserved peptide-binding motif characteristic for the *Caja-DRB1*03* gene. Thus, the severely reduced MHC class II repertoire in common marmosets has been expanded by reactivation of a pseudogene segment as a result of exon shuffling.

MHC | non-human primates | selection | immune response | evolution

The common marmoset (*Callithrix jacchus*) has its natural habitat in the swamps and Atlantic rainforest of northeastern Brazil. This species is used as an animal model to investigate several chronic and infectious diseases (1–5).

The classical *Mhc* class I and II genes play a key role in controlling various immune functions and generally exhibit a high degree of allelic polymorphism in different vertebrate species. Humans, apes, and Old World monkeys are known to share *Mhc* class I and II loci, but particular lineages have also been documented to predate speciation events that occurred during primate evolution (6).

Equivalents of HLA class II loci have been detected in common marmosets and display various levels of polymorphism (7–10). For instance, both of the apparently functional *Caja-DQA1* and *-DQB1* genes are oligomorphic, but the *-DP* locus seems to be inactivated (7). Despite extensive efforts, only three *DRB* loci have been identified thus far (7, 11, 12). The *Caja-DRB1*03* gene, related to the HLA-*DRB1*03* lineage that is characterized by the evolutionarily conserved EYSTS motif, is moderately polymorphic (7, 10, 13). Careful comparison of human and common marmoset nucleotide sequences provided evidence that these class II genes are not truly orthologous and arose as a result of convergent evolution (7, 13). The *Caja-DRB*W16* locus exhibits a degree of polymorphism comparable to *-DRB1*03*, whereas the *-DRB*W12* gene appears to be invariant. The monomorphic status of the latter gene, and its capacity to present particular myelin oligodendrocyte glycoprotein peptides, may explain why all common marmosets are susceptible to experimental autoimmune encephalomyelitis induction by immunization with particular autoantigens (14).

As a result of expansion and contraction processes, the primate *DRB* region shows extensive polymorphism with regard to the number and combination of loci present per haplotype, in

species as well as in individuals of a given species. For example, within rhesus macaques >30 different *DRB* region configurations have been defined, whereas in humans only 5 have been documented (15–17). Common marmosets, mostly born as non-identical twins, are natural bone-marrow chimeras because of a sharing of blood circulation through placental fusion during gestation. Although haplotype definitions are, therefore, complicated, phenotypic and segregation analyses suggested that every region configuration possesses one *Caja-DRB*W12* locus, together with at least one *-DRB1*03* and one *-DRB*W16* locus (7). Thus, common marmosets apparently lack *-DR* region configuration polymorphism. Moreover, because the few *Caja-DRB* loci themselves display only moderate or low degrees of allelic polymorphism, one might expect that these animals are susceptible to contracting particular diseases. Indeed, common marmosets in captivity are especially prone to die from certain viral and bacterial infections (3, 18).

Thus far, only genomic DNA analyses of exon 2 sequences have been performed, and nothing is known about the transcription and expression status of the various loci. To gain insight into the class II repertoire at the transcription level, a thorough cDNA analysis of the different *Caja-DRB* genes in a well defined population of animals was performed.

Results and Discussion

***Caja-DRB* Genes and the Presence/Absence of Corresponding Transcripts.** In our panel of 15 selected marmosets, a total of 22 full-length cDNAs could be defined (Fig. 1). Nearly all of the *Caja-DRB*W16* cDNAs defined earlier by exon 2 analyses are present and characterized by abundant transcription levels. The *Caja-DRB*W16* transcripts have all of the characteristics necessary to produce bona fide gene products (Fig. 2). In some animals, three distinct transcripts were detected. This finding is probably due to the chimeric status of the animals, although we cannot exclude the possibility that some region configurations harbor duplicated loci.

Thus far, the *Caja-DRB*W12* locus has been considered to be monomorphic. During this study, a second allele differing by only two codons producing amino acid replacements was observed and designated *Caja-DRB*W1202* (Fig. 2). A specially designed set of primers allowed us to detect at least one transcript in all animals tested (Fig. 1). These results, and

Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AM114038–AM114059 and AM180875).

[†]To whom correspondence may be addressed at: Biomedical Primate Research Centre, Lange Kleiweg 139, 2288 GJ, Rijswijk, The Netherlands. E-mail: doxiadis@bprc.nl.

^{††}To whom correspondence may be addressed. E-mail: vanrood@europdonor.nl.

© 2006 by The National Academy of Sciences of the USA

Fig. 2. Alignment of deduced Caja-DRB amino acid sequences, given in conventional one-letter code. A dash or dot indicates identity with the consensus or deletion of an amino acid, respectively, whereas * indicates that this part of the sequence is unknown. Amino acids or deletions, which are typical for the -DRB1*03 and -DRB*W16 lineages, are shown in yellow or blue, respectively. Consequently, the names of the hybrid genes, -DRB1*0307 and -DRB1*0312, are shown in both blue and yellow. L, leader peptide and first four amino acids of the coding sequence; β_1 , β_1 domain; β_2 , β_2 domain; TM, transmembrane region; CYT, cytoplasmic tail; 3'UT, 3' untranslated region.

Functional Implications. MHC class II molecules are heterodimers composed of an α chain and a β chain. Most of the polymorphic codons of the *Mhc-DRB* gene map to exon 2 encoding the β_1 domain, which plays a crucial role in the

formation of the peptide-binding site (26). The β_1 domain of the *Caja-DRB1*0307* and *-DRB1*0312* gene products possesses the EYSTS motif (Fig. 2), which is characteristic for the *Mhc-DRB1*03* lineage and is present in many non-human

primate species (6, 27). Peptide-binding and T cell recognition studies demonstrated that this motif might have been conserved throughout primate evolution as a result of functional constraints and selective advantages (28). The importance of this motif is further illustrated by the fact that it has been generated several times *de novo* during evolution in New World primate and other species through convergent evolution (7, 29, 30). Consequently, the limited peptide-binding reper-

toire of the common marmoset has been expanded significantly by the reactivation of a segment from a pseudogene, as the result of an exon-shuffling event. The MHC region is scattered with pseudogenes and pseudogene segments (31–35). This study clearly documents that an intact, but old, polymorphic gene segment may be recruited to again play an active role in immune response.

Materials and Methods

Animals. For this study, common marmosets were selected from two breeding colonies: 4 animals from the Institute of Technology, Zürich-Schwerzenbach, and 11 animals from the Biomedical Primate Research Centre's pedigreed breeding colony. Most of these animals were analyzed for *Caja-DRB* polymorphism by exon 2 sequence analyses, as well as full-length cDNA sequencing of *Caja-DRB* transcripts.

DNA/RNA Isolation and Exon 2 Amplification of Genomic DNA. Genomic DNA was extracted from immortalized B lymphocytes by a standard salting-out procedure, and RNA was isolated by using the RNeasy Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. Amplification of exon 2 was performed as described recently for rhesus macaques by using generic *DRB* primers (36), or with primers specific for common marmosets (7, 10).

cDNA Synthesis, Amplification, and RT-PCR. cDNA was synthesized from freshly extracted mRNA by using the Universal RiboClone cDNA synthesis system (Promega) according to the manufacturer's recommendations. Full-length sequences were amplified by PCR from 10 μ l of cDNA by using primers developed for human *DRB5'* and 3' untranslated sequences (37). PCR mixtures and conditions were as described in ref. 38, employing a commercially available *Taq* polymerase (Invitrogen). RT-PCR was performed directly on RNA by using the Qiagen OneStep RT-PCR kit according to the manufacturer's guidelines. The primers used for RT-PCR are the same as those indicated above (37). In addition, the following internal primers specific for *Caja-cDRB* exon 1 and exon 4, respectively, were used: 5' *Caja-cDRB*ex1, 5'-TGC RTG GCA GYG CTG ACA GTG-3'; and 3' *Caja-cDRB*ex4, 5'-GCC CTG CCC CAA GGA AGA GC-3'. Twenty-five amplification cycles then followed, with 45 s at 94°C, 45 s at 60°C, and 1 min at 72°C. A final extension step for 30 min at 72°C was added.

RACE. 5' RACE was performed directly on RNA by using the 5' RACE system for rapid amplification of cDNA ends, Version 2.0 (Invitrogen) according to the manufacturer's guidelines. The 3' primer used for first-strand cDNA synthesis was a gene-specific primer for *Caja-DRB1*03* and *-DRB*W16*: Caja GSP1 5'RACE W16-DRB1 5'-GGA AGG TCC AGT CTC CAT TC-3'. For the PCR of dC-tailed cDNA, a nested gene-specific primer was used:

Fig. 4. Schematic representation of the exon-shuffling event producing a novel *Caja-DRB* gene.

Caja GSP2 5'RACE all, 5'-ACC TGA CTT CAA TGC TGC CTG G-3'. The cycling parameters were 2 min at 94°C, followed by 32 cycles of 45 s at 94°C, 45 s at 60°C, and 90 s at 72°C. A final extension step was added for 30 min at 72°C.

We performed a 3' RACE directly on the RNA by using the 3' RACE system for rapid amplification of cDNA ends, Version D (Invitrogen) according to the manufacturer's guidelines. Use of 3' RACE takes advantage of the natural poly(A) tail found in mRNA as a generic priming site for PCR. The primers used for amplification of the target cDNA were Caja-DRB*W12, -DRB*W16, and -DRB1*03 exon-2-specific primers, as published (10). The cycling parameters were 3 min at 94°C, followed by 25 cycles of 45 s at 94°C, 45 s at 60°C, and 1 min at 72°C. A final extension step was performed for 30 min at 72°C.

Cloning and Sequencing. Genomic exon 2 amplicons and/or full-length *Caja-DRB* PCR products were cloned and sequenced by using the InsT/Aclone PCR product cloning kit (Fermentas, St. Leon-Roth, Germany) as described (36). Sequencing reactions were performed with the BigDye Terminator cycle sequencing kit (Applied Biosystems), and the samples were run on the ABI 3100 genetic analyzer (Applied Biosystems). Analysis of the sequences was performed with SEQUENCE NAVIGATOR (Applied Biosystems).

We thank Donna Devine for editing the manuscript and Henk van Westbroek for help in creating the artwork. This study was supported in part by National Institutes of Health Grant 1R24 RR16038-01 and European Union Project EUPEAH QLRI-CT-2002-02758.

- Kyuregyan, K. K., Poleschuk, V. F., Zamyatina, N. A., Isaeva, O. V., Michailov, M. I., Ross, S., Bukh, J., Roggendorf, M. & Viazov, S. (2005) *Virus Res.* **114**, 154–157.
- Means, R. E. (2004) *Virology* **326**, 67–78.
- Lanford, R. E., Chavez, D., Notvall, L. & Brasky, K. M. (2003) *Virology* **311**, 72–80.
- Jacob, J. R., Lin, K. C., Tennant, B. C. & Mansfield, K. G. (2004) *J. Gen. Virol.* **85**, 2525–2533.
- Brok, H. P., Bauer, J., Jonker, M., Blezer, E., Amor, S., Bontrop, R. E., Laman, J. D. & 't Hart, B. A. (2001) *Immunol. Rev.* **183**, 173–185.
- Bontrop, R. E., Otting, N., de Groot, N. G. & Doxiadis, G. G. (1999) *Immunol. Rev.* **167**, 339–350.
- Antunes, S. G., de Groot, N. G., Brok, H., Doxiadis, G., Menezes, A. A., Otting, N. & Bontrop, R. E. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 11745–11750.
- Bidwell, J. L., Lu, P., Wang, Y., Zhou, K., Clay, T. M. & Bontrop, R. E. (1994) *Eur. J. Immunogenet.* **21**, 67–77.
- Gyllenstein, U., Bergstrom, T., Josefsson, A., Sundvall, M., Savage, A., Blumer, E. S., Giraldo, L. H., Soto, L. H. & Watkins, D. I. (1994) *Immunogenetics* **40**, 167–176.
- Wu, M. S., Tani, K., Sugiyama, H., Hibino, H., Izawa, K., Tanabe, T., Nakazaki, Y., Ishii, H., Ohashi, J., Hohjoh, H., et al. (2000) *J. Mol. Evol.* **51**, 214–222.
- Trtkova, K., Kupfermann, H., Grahovac, B., Mayer, W. E., O'hUigin, C., Tichy, H., Bontrop, R. & Klein, J. (1993) *Immunogenetics* **38**, 210–222.
- Middleton, S. A., Anzenberger, G. & Knapp, L. A. (2004) *Immunogenetics* **55**, 785–790.
- Kriener, K., O'hUigin, C. & Klein, J. (2001) *Hum. Immunol.* **62**, 1–14.
- Brok, H. P., Uccelli, A., Kerlero de Rosbo, N., Bontrop, R. E., Roccatagliata, L., de Groot, N. G., Capello, E., Laman, J. D., Nicolay, K., Mancardi, G. L., et al. (2000) *J. Immunol.* **165**, 1093–1101.
- Doxiadis, G. G., Otting, N., de Groot, N. G., de Groot, N., Rouweler, A. J., Noort, R., Verschoor, E. J., Bontjer, I. & Bontrop, R. E. (2003) *Immunogenetics* **55**, 540–551.
- Doxiadis, G. G., Otting, N., de Groot, N. G., Noort, R. & Bontrop, R. E. (2000) *J. Immunol.* **164**, 3193–3199.
- Slierendregt, B. L., Otting, N., van Besouw, N., Jonker, M. & Bontrop, R. E. (1994) *J. Immunol.* **152**, 2298–2307.
- Potkay, S. (1992) *J. Med. Primatol.* **21**, 189–236.
- Harton, J. A., Van Hagen, A. E. & Bishop, G. A. (1995) *Immunity* **3**, 349–358.
- Khalil, H., Brunet, A., Saba, I., Terra, R., Sekaly, R. P. & Thibodeau, J. (2003) *Int. Immunol.* **15**, 1249–1263.
- Smiley, S. T., Rudensky, A. Y., Glimcher, L. H. & Grusby, M. J. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 241–244.
- Pinet, V., Vergelli, M., Martin, R., Bakke, O. & Long, E. O. (1995) *Nature* **375**, 603–606.
- Robinson, J., Waller, M. J., Parham, P., de Groot, N., Bontrop, R., Kennedy, L. J., Stoeck, P. & Marsh, S. G. (2003) *Nucleic Acids Res.* **31**, 311–314.
- Long, M. (2001) *Curr. Opin. Genet. Dev.* **11**, 673–680.
- Gilbert, W. (1978) *Nature* **271**, 501 (lett.).
- Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. (1993) *Nature* **364**, 33–39.
- Bontrop, R. E., Otting, N., Slierendregt, B. L. & Lanchbury, J. S. (1995) *Immunol. Rev.* **143**, 33–62.
- Geluk, A., Elferink, D. G., Slierendregt, B. L., van Meijgaarden, K. E., de Vries, R. R., Ottenhoff, T. H. & Bontrop, R. E. (1993) *J. Exp. Med.* **177**, 979–987.
- Andersson, L., Gustafsson, K., Jonsson, A. K. & Rask, L. (1991) *Immunogenetics* **33**, 235–242.
- Lundberg, A. S. & McDevitt, H. O. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6545–6549.
- Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., Lush, M. J., Povey, S., Talbot, C. C., Jr., Wright, M. W., et al. (2004) *Nat. Rev. Genet.* **5**, 889–899.
- Kelley, J., Walter, L. & Trowsdale, J. (2005) *Immunogenetics* **56**, 683–695.
- Kulski, J. K., Anzai, T., Shiina, T. & Inoko, H. (2004) *Mol. Biol. Evol.* **21**, 2079–2091.
- Kulski, J. K., Shiina, T., Anzai, T., Kohara, S. & Inoko, H. (2002) *Immunol. Rev.* **190**, 95–122.
- Mayer, W. E., O'hUigin, C. & Klein, J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10720–10724.
- Penedo, M. C., Bontrop, R. E., Heijmans, C. M., Otting, N., Noort, R., Rouweler, A. J., de Groot, N., de Groot, N. G., Ward, T. & Doxiadis, G. G. (2005) *Immunogenetics* **57**, 198–209.
- Lekutis, C. & Letvin, N. L. (1995) *Hum. Immunol.* **43**, 72–80.
- de Groot, N., Doxiadis, G. G., de Groot, N. G., Otting, N., Heijmans, C., Rouweler, A. J. & Bontrop, R. E. (2004) *J. Immunol.* **172**, 6152–6157.