CD97 and EMR2: receptors on the move

Kwakkenbos, M.J.

Citation for published version (APA):
Chapter 7

Recent evolution of the EGF-TM7 receptor EMR2 in primates: loss of ligand specificity for CD55 due to different mechanisms in humans and chimpanzees

Manuscript in preparation

Mark J. Kwakkenbos¹, Mourad Matmati¹, Walter Pouwels¹, Ole Madsen², Ronald Bontrop³, Peter J. Heidt³, and Jörg Hamann¹

¹Laboratory for Experimental Immunology, Academic Medical Center, University of Amsterdam, Amsterdam ²Department of Biochemistry, University of Nijmegen, Nijmegen, and ³Department of Immunobiology, Biomedical Primate Research Centre, Rijswijk, The Netherlands
Chapter 7

ABSTRACT

The human EGF-TM7 receptors CD97, EMR1, EMR2, EMR3, and EMR4 form a group of large heptahelical molecules predominantly expressed by cells of the immune system. Through their EGF domains, EGF-TM7 receptors bind cellular ligands. EMR2 possesses a chimeric structure with a seven-span transmembrane region closely related to EMR3 and an epidermal growth factor (EGF)-like domain region nearly identical to CD97. Human EMR2 and CD97 share ligand specificity for chondroitin sulfate but not for CD55, which is exclusively bound by CD97. We here show that this difference, which is caused by three amino-acid substitutions within the first two EGF domains, emerged after human speciation. Chimpanzee EMR2 can bind CD55, however, transcription of isoforms with high affinity for CD55 is diminished by alternative RNA splicing. Thus, different molecular mechanisms prevent CD55 binding by EMR2 in hominoids. Location of CD97 and EMR1-4 within two neighboring clusters on the short arm of human chromosome 19 and (on current evidence) absence of nonmammalian homologues suggest a young evolutionary history of these receptors. We suppose that the EGF-TM7 family evolved through a series of segmental intrachromosomal gene duplications. EMR2 likely originated only recently in a common ancestor of Old World monkeys and hominoids.

INTRODUCTION

Despite the striking similarity of their genomes, humans are more susceptible than other hominoids to diseases such as malaria, acquired immunodeficiency syndrome (AIDS), hepatitis B, malaria, Alzheimer’s disease, myocardial infarction, and epithelial cancer.1 The molecular basis of this difference is virtually unknown. We just begin to understand at what molecular levels the phenotypic divergence between us and our closest relatives might have developed.2 Gene inactivation might be one mechanism and has first been demonstrated for the human cytidine monophosphate-N-acetylneuraminic acid (CMP-Neu5Ac) hydroxylase gene (CAMAH), which synthesizes N-glycolyneuraminic acid (Neu5Gc).3 We recently reported inactivation of another nonpolymorphic gene after human speciation. Due to a one-nucleotide deletion in exon 8 that is not present in nonhuman primates, including chimpanzees, translation of EGF module-containing mucin-like receptor (EMR)4 terminates ahead of the transmembrane region.4 EMR4 is a member the EGF-TM7 family of adhesion class seven-span transmembrane (TM7) receptors.5,7 With the human genome unraveled, this family comprises five members, which are predominantly expressed on hematopoietic cells: CD97,8,9 EMR1,10 EMR2,11 EMR3,12 EMR4.4 EGF-TM7 receptors are characterized by an extended extracellular region with several tandemly arranged epidermal growth factor (EGF)-like domains at the N-terminus. As a result of alternative RNA splicing, individual family members express isoforms that possess variable numbers of EGF domains. A G protein-coupled receptor-proteolytic site (GPS) proximal to the first transmembrane domain gives rise to processing within the endoplasmic reticulum. Translated as single polypeptides, EGF-TM7 receptors are cleaved into an extracellular α subunit and a TM7/cytoplasmic β subunit, which noncovalently associate on the cell surface.

84
Recent evolution of EMR2 in primates

Next to inactivation of EMR4 in the human genome, there is further evidence suggesting a rather recent evolution of the leukocyte-expressed EGF-TM7 receptors CD97 and EMR1-4. First, nonmammalian homologs have not been identified as yet. Second, the genes are located in close proximity on the short arm of human chromosome 19 within clusters in 19p13.1 (CD97, EMR2, and EMR3) and 19p13.3 (EMR1 and EMR4). Third, surveys of the mouse genome failed to identify orthologs for EMR2 and EMR3. Fourth, molecular cloning of EMR2 surprisingly unraveled a remarkable similarity with two other EGF-TM7 receptors. As shown in Figure 1, the C-terminal part of the stalk and the TM7 region are highly similar to EMR3. In contrast, the signal peptide, the EGF domains, and the most N-terminal part of the stalk region strongly resemble that of CD97. Only six out of 236 amino acids within the EGF domain region are different (97% amino acid identity).

EGF-TM7 receptors interact through their EGF domains with cellular ligands, an ability that is unique within the large superfamily of TM7 molecules. Due to their striking similarity, the largest isoforms of EMR2 and CD97 share specificity for the glycosaminoglycan side chain chondroitin sulphate (CS). This interaction is mediated by EGF domain 4, which is identical between EMR2 and CD97. In contrast, CD97, but not EMR2, binds CD55 (decay accelerating factor), a glycosylphosphatidyl inositol-linked molecule that prevents complement deposition on self cells. Three different amino acids within EGF domains 1 and 2, which together form the binding site for CD55, account for this divergence in ligand specificity.

In this study, we attempt to understand the parallel development of the molecular twins CD97 and EMR2 during primate evolution. We provide evidence that, first, EGF-TM7 receptors evolved through a series of segmental intrachromosomal gene duplications, second, EMR2 manifested only recently in Old World monkeys, and, third, EMR2 in hominoids lost its ability to bind CD55 due to different molecular mechanisms (mutations versus alternative splicing).
Chapter 7

MATERIALS AND METHODS

Animals
Primate species investigated within this study included chimpanzee (*Pan troglodytes*), orangutan (*Pongo pygmaeus*), rhesus macaque (*Macaca mulatta*), common baboon (*Papio hamadryas*), common marmoset (*Callithrix jacchus*), and cotton-top tamarin (*Saguinus oedipus*). Peripheral blood from healthy animals and genomic DNA samples were obtained from the Biomedical Primate Research Centre (Rijswijk, The Netherlands).

Non-primate genomic DNA analyzed were from three-toed sloth (*Bradypus tridactylus*), nine-banded armadillo (*Dasypus novemcintus*), common tenrec (*Tenrec ecaudatus*), Asian elephant (*Elephas maximus*), opossum (*Didelphis marsupialis*), and kangaroo (*Macropus rufus*).

Flow cytometry
Flow cytometry was performed by standard procedure on a FACSCalibur (Becton Dickinson, San Jose, CA). Whole blood from human or chimpanzee was incubated with biotinylated mAb CLB-CD97/3 (binds to stalk region of CD97), CLB-CD97/1 (binds to EGF domain 1 of CD97 and EMR2), CLB-CD97/2A1 (binds to stalk region of EMR2), CLB-CD97/15 (binds to EGF domain 4 of CD97 and EMR2), or control IgG. Streptavidin-APC (PharMingen, San Diego, CA) was used as second reagent. Prior to cytometry, erythrocytes were shocked using FACS lysing solution (Becton Dickinson).

RT-PCR, cDNA cloning, genomic PCR, and sequence analysis
Total RNA was isolated from human, chimpanzee, and rhesus macaque PBMC. For RT-PCR, first-strand cDNA was prepared using Superscript II reverse transcriptase (Invitrogen, Breda, The Netherlands). EGF domain-encoding sequences of human and chimpanzee CD97 and EMR2 isoforms were amplified by PCR (35 cycles, 30 s at 93°C, 30 s at 58°C, 60 s at 72°C) using specific primers (Table 1). PCR products were separated on a 1.5% agarose gel and photographed. The 5' ends of exon 3, encoding the first EGF domain, were separately amplified. Sequences of all PCR products were determined with the BigDye terminator cycle sequencing kit (Applied Biosystems, Warrington, UK).

<table>
<thead>
<tr>
<th>Table 1: Oligonucleotide primers used for RT-PCR and genomic PCR.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>CD97</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CD97</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>EMR2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>EMR2</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>EMR2</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Intronic sequences are in lower case.
Recent evolution of EMR2 in primates

The complete cDNA sequence of rhesus macaque EMR2 was derived by a combination of RT-PCR and 5’- and 3’-rapid amplification of cDNA ends (RACE) using the Smart RACE cDNA amplification kit (Clontech, Palo Alto, CA). Various primers, derived from the human sequence as well as specifically designed for rhesus macaque EMR2, were applied. Sequences were determined and the deduced amino acid sequence was compared with human EMR2 with ClustalW software.

Genomic sequences of CD97 and EMR2 from several primates were amplified by PCR (35-40 cycles, 30 s at 93°C, 30 s at 58-60°C, 20-60 s at 72°C) using 100-500 ng genomic DNA per reaction. Degenerated primers were designed based on either human and other mammalian sequences (CD97) or on human and rhesus macaque sequences (EMR2) (Table 1). PCR products were separated on a 1.2% agarose gel, photographed, and sequenced.

Generation of Fc fusion proteins
To generate mouse Fc fusion constructs,22 the N-terminal part of the largest isoform of chimpanzee CD97 and EMR2 was amplified by PCR using the specific primers 5’-AAGCTTCCATGGAGGCCGCTTTCTTGC-3’ [an introduced HindIII site is depicted italic], and 5’-CTGGAGTCCACAGCCAGCTGCGCCGC-3’ [an introduced NotI site is depicted italic]. Gene-specific RT-PCR products derived from chimpanzee PBMC first-strand cDNA were used as template. PCR products were ligated into pGEM-T easy vector (Promega, Madison, WI) and after digestion with HindIII and NotI cloned in-frame immediately upstream to the CH3-CH2-hinge region sequence of mouse IgG2b, linked to a C-terminal biotinylation sequence, in pcDNA3.1/Neo(+) (kindly provided by M. Stacey, Sir William Dunn School of Pathology, Oxford). Generation of human CD97 and EMR2 Fc-constructs has been reported previously.14 In short, each four 225-cm² cell culture flasks with HEK 293 cells were transfected with the different constructs and after four days, conditioned OptiMEM 1 medium (Life Technologies Ltd, Paisly, Scotland), containing soluble recombinant protein was purified using a Protein A (Sigma, St. Louis, MO) column. After purification, recombinant protein was concentrated to ~0.5 ml using a 30-kDa molecular weight cut-off filter (Millipore, Bedford, MA), dialyzed against 10 mM Tris-HCl (pH 8) buffer, and incubated with 2 µl BirA enzyme and supplied substrates (Avidity, Denver, CO) overnight at room temperature. Excess biotin was subsequently removed by dialyses against 10 mM Tris-HCl (pH 7.3) buffer containing 10 mM CaCl² and 150 mM NaCl. The biotinylated proteins were then aliquoted and stored at -80°C after quantification by Bradford assay.

Ligand binding studies
Cell binding assays using biotinylated CD97- and EMR2-mouse Fc proteins coupled to fluorescent beads were performed as described previously.22 Briefly, 10 µl avidin-coated fluorescent beads (Spherotech Inc., Libertyville, IL) were washed with PBS/0.5% BSA and incubated with saturating amounts (> 1 µg) of biotinylated recombinant protein in a volume of 10 µl. After 1 h, nonbinding protein were removed by washing with PBS/0.5% BSA. The bead-protein complexes were sonicated immediately before addition to the cells (0.5 x 10⁶ cells/50 µl PBS/0.5% BSA). For blocking studies, 1 µg of mAb was added to the bead-protein complexes and incubated for 10 min at 4°C before adding the complexes to the cells. Cell-bead mixture, in a 96-well flat bottom plate, was spun at 1000 x g at 4°C for 10 min, incubated for another 50 min at 4°C, and finally resuspended in 300 µl of PBS for flow-cytometric analysis.
Human embryonic kidney (HEK) 293 cells and two mutants of Chinese hamster ovary (CHO) cells were used for ligand binding studies. In one mutant (PgsB-618, ATCC, Manassas, VA), glycosaminoglycan synthesis is mostly absent.²³ The other mutant (PgsD-677, ATCC) lacks heparan sulfate but not other glycosaminoglycans.²⁴ To generate transfectants, which express CD55 but not CS, PgsB-618 cells were transfected with full-length CD55 cDNA in pcDNA3 (Invitrogen). After selection with G418 (Invitrogen) at 500 µg/ml, stable clones were tested for CD55 expression by flow cytometry with the mAb CLB-CD97L/1, which binds to the first short consensus repeat of CD55.¹⁵ One positive clone was enriched for CD55-expressing cells by a single sort on a FACS Aria (Becton Dickinson) using the mAb CLB-CD97L/1.

RESULTS

Expression of CD97 and EMR2 on chimpanzee leukocytes
To confirm that CD97 and EMR2 are expressed in chimpanzees, we performed flow-cytometric analysis of peripheral blood leukocytes. As shown in Figure 2, both molecules are present on chimpanzee leukocytes. Like in humans, CD97 was found on lymphocytes, granulocytes, and monocytes whereas expression of EMR2 was restricted to monocytes and (at very low levels) granulocytes.

We also compared the presence of ligand-binding sites in CD97 and EMR2 with mAb to the respective EGF domains in both molecules. EGF domain 1, which is involved in the CD97-CD55 interaction and is possessed by all isoforms of both molecules, was detected on all types of leukocytes. In contrast, EGF domain 4, which binds CS and is found only in the largest isoform of CD97 and EMR2, was identified exclusively on monocytes.

![Figure 2. Expression of CD97 and EMR2 on peripheral blood leukocytes in human and chimpanzee. Cells were analyzed by flow cytometry with biotinylated mAb as indicated. Due to the high sequence similarity, CLB-CD97/1 and 1B5 recognize the respective EGF domains in both, CD97 and EMR2. Gray histograms represent control IgG stainings. Streptavidin-APC was used as a secondary reagent. Cell types were determined on the basis of forward- and side-scatter.](image-url)
Recent evolution of EMR2 in primates

Isoform transcription and sequence of EGF domains in chimpanzee CD97 and EMR2

Since binding specificity of CD97 and EMR2 isoforms depends on EGF domain composition\(^1\),\(^4\),\(^6\),\(^17\) and is affected by minor sequence differences between both molecules,\(^1\),\(^11\),\(^17\) we analyzed the EGF domain region of CD97 and EMR2 in chimpanzees. Using RT-PCR, differences in the isoform ratio between humans and chimpanzees for both CD97 and EMR2 were detected (Figure 3). Whereas in humans, the majority of CD97 transcripts encodes the smallest isoform possessing three EGF domains,\(^25\) a more equal distribution of isoforms was found in chimpanzees. A striking difference was detected for EMR2. Other than in humans, chimpanzee transcripts dominantly encode the largest isoform containing five EGF domains.

![Figure 3. Expression of CD97 and EMR2 on peripheral blood leukocytes in human and chimpanzee. Cells were analyzed by flow cytometry with biotinylated mAb as indicated.](image)

Due to the high sequence similarity, CLB-CD97/1 and 1B5 recognize the respective EGF domains in both, CD97 and EMR2. Gray histograms represent control IgG stainings. Streptavidin-APC was used as a secondary reagent. Cell types were determined on the basis of forward- and side-scatter.

We next sequenced the EGF domain regions of CD97 and EMR2 in humans and chimpanzees (Figure 4 and 5). Clearly, the high similarity found within the EGF domains between CD97 and EMR2 in humans also exists in chimpanzees. Chimpanzee CD97 and EMR2 differ at one amino acid in EGF domain 2, at four amino acids in EGF domain 3, and at two amino acids in EGF domain 4. Next, we compared the EGF domains of the homologs. Human and chimpanzee CD97 differ at two amino acid in EGF domain 4, the binding site for CS. Human and chimpanzee EMR2 differ at two amino acid in EGF domain 1, at two amino acids in EGF domain 2, and at one amino acid in EGF domain 3. Remarkably, the three amino-acid substitutions within the first two EGF domains that prevent CD55 binding by human EMR2 are not present in chimpanzees. On the other hand, chimpanzee EMR2 has acquired an independent mutation single alteration in EGF domain 2.

Ligand specificity of chimpanzee CD97 and EMR2 isoforms

Based on the amino acid sequence of the EGF domains, we wondered whether the largest isoform of chimpanzee CD97 can bind CS and whether chimpanzee EMR2 possess specificity for CD55. To answer these questions, we generated multivalent fluorescent probes loaded with recombinant Fc protein of the extracellular part of the largest isoform of chimpanzee CD97 and EMR2. We recently showed that human CD97 and EMR2 probes efficiently bind to HEK293 cells, which highly express both CD55 and CS (Kwakkenbos et al., submitted for publication). Preincubation of the probes with mAb CLB-CD97/1 (anti-EGF domain 1) and 1B5 (anti-EGF domain 4) allowed to discriminate between binding to CD55 or CS (Figure 6A).

In concurrence with the known monospecificity of human CD97-3EGF for CD55,\(^14\) binding was completely prevented when the probes were preincubated with CLB-CD97/1. Treatment with 1B5, in contrast, had no effect. Differently, binding of the human CD97-5EGF was efficiently blocked only by preincubation with both mAb, demonstrating its dual specificity for both CD55 and CS (Kwakkenbos et al., submitted)
Chapter 7

Figure 4. Alignment of the nucleotide sequences encoding the EGF domains of human and chimpanzee CD97 and EMR2. All EGF domains are encoded by separate exons with borders indicated by >< marks. Synonymous substitutions (amino acid-unchanging) are depicted in gray; nonsynonymous substitutions (amino acid-changing) are marked in black with the respective codons underlined. Chimpanzee sequences were derived from two individuals. A polymorphism found in one chimpanzee in EGF domain 4 (G/A → G/R) is indicated by a small circle.

Figure 5. Alignment of the amino acid sequences of the EGF domains of human and chimpanzee CD97 and EMR2. Amino acid variations are marked in black. Potential N-glycosylation sites are indicated by gray triangles. EGF domains 1 and 2 form the binding site for CD55, EGF domain 4 the binding site for CS.
Recent evolution of EMR2 in primates

for publication). Binding of chimpanzee CD97-5EGF could be blocked partially by CLB-CD97/1. Surprisingly, no additive inhibition was found when probes were coincubated with 1B5. To investigate whether the chimpanzee CD97-5EGF can bind CS, we compared probe binding to glycosaminoglycan-deficient (PgsB-618) and exclusively CS-expressing (PgsD-677) CHO cell mutants. Binding to CHO cells expressing solely CS indicates that chimpanzee CD97-5EGF indeed interacts with CS. Likely, the two different amino acids found in EGF domain 4 of chimpanzee CD97 (Figure 5) abolish the blocking effect of 1B5 on the interaction with CS.

Figure 6. (A) Ligand specificity of the largest isoforms of human and chimpanzee CD97 and EMR2. The smallest isoform of human CD97 is shown for comparison. Bars represent binding of fluorescent probes loaded with the extracellular region of the respective receptors to HEK293 cells, which express both CD55 and CS. Probes were preincubated with mAb to EGF domain 1 (CD55-binding site) and EGF domain 4 (CS-binding site) as indicated. Data shown are mean fluorescence intensity ± SD in percent from three separate experiments. (B) Binding of the largest isoform of human CD97 to glycosaminoglycan-deficient (PgsB-618) and exclusively CS-expressing (PgsD-677) CHO cell mutants. Due to phylogenetic restriction, hamster CD55 on CHO cell is not bound by primate CD97. Mean fluorescence intensities ± SD in percent from two experiments are shown. (C) Binding of human and chimpanzee CD97 and EMR2 isoforms to the glycosaminoglycan-deficient CHO cell mutant PgsB-618 stably expressing human CD55.

Human EMR2 does not bind CD55. Accordingly, binding of human EMR2-5EGF was efficiently blocked by 1B5 while CLB-CD97/1 had no effect. Binding of chimpanzee EMR2-5EGF, in contrast, could only be blocked completely when probes were preincubated with CLB-CD97/1 and 1B5 together. To confirm the ability of chimpanzee EMR2 to bind CD55, we transfected the glycosaminoglycan-deficient CHO cell mutant PgsB-618 with human CD55. As shown in Figure 6C, chimpanzee but not human EMR2-5EGF bound to the CD55-transfected cells.
Chapter 7

Detection of CD97 and EMR2 in primate and nonprimate genomes and cDNA cloning of rhesus macaque EMR2

The striking similarity with CD97 and EMR3 is indicative of a young evolutionary history of EMR2. To obtain phylogenetically older EMR2 sequences, we applied PCR to different primate genomes (data not shown). Using primer combinations that amplify sequences of low homology with other EGF-TM7 receptors (exon 11 (stalk region), exon 19 (cytoplasmic tail)), we could identify EMR2 sequences from Old World but not from New World monkeys. CD97 sequences, in contrast, were easily detected in all primates investigated.

Figure 7. Alignment of the amino acid sequences of human and rhesus macaque EMR2. EGF domains, the GPS motif, and the seven hydrophobic transmembrane segments are indicated. An arrowhead shows the predicted processing site within the GPS motif. Identical amino acids in the rhesus macaque sequence are shown as dots.

We next determined the complete cDNA sequence of rhesus macaque EMR2 by a combination of RT-PCR and 5'- and 3'-RACE. Alignment with human EMR2 revealed an identity of 91% at the nucleotide level and 88% at the amino acid level (Figure 7). Regional similarity with CD97 and EMR3 very closely resembles that of its human homolog. Thus, the chimeric structure that characterizes EMR2 in humans (Figure 1) is found already in the rhesus macaque. Interestingly, the amino acid sequence of EGF domain 4 is identical between human and rhesus macaque EMR2 while five to ten differences are found in the other domains. Complete amino-acid conservation within the CS-binding site emphasizes the functional importance of this interaction.

CD97 has been identified previously in the genomes of humans, mice, rats (XM_341662), cows (NM_176661), and pigs. Using degenerated primers, designed to exon 18 (transmembrane segments 6 and 7), we screened genomes from more distantly related mammals (Figure 8). A sequence derived from the three-toed sloth represents the most ancient CD97 sequence found as yet.
Recent evolution of EMR2 in primates

Clade: Xenarthra
Order: Pilosa
Sloth
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT

Clade: Laurasiatheria
Order: Cetartiodactyla
Pig
ACCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT
Cow
ACCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT

Clade: Euarchontoglires
Order: Rodentia
Rat
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT
Mouse
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT

Order: Primates
Tamarin
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT
Marmoset
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT
Baboon
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT
Rhesus
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT
Human
AGCTTTCTCCGCTTTCGACAGTGGGTCTTTGGCCTGTTCCTCTTCAACCCGCACAGCTGGGTGCTGTCCTACACCTTCACCATCCTCAACT

Figure 8. Alignment of the nucleotide sequences encoding part of transmembrane domains 6 and 7 of CD97 in placental mammals. Orders and supraordinal clades for the different species are provided. Within one order, identical nucleotides are shown as dots. Positions conserved within all sequences are indicated by asterisks.

DISCUSSION

Different from related adhesion class receptors like ETL, which have homologs in fishes (AAH55171), the EGF-TM7 receptors CD97 and EMR1-4 have a young evolutionary history. As yet these molecules have exclusively been identified in two superordinal clades of placental mammals: Laurasiatheria (e.g. hoofed animals and carnivores) and Euarchontoglires (e.g. rodents and primates). We here report a more ancient trace of CD97 from a sloth, a representative of Xenarthra. This clade diverged from the common ancestor of Laurasiatheria and Euarchontoglires approximately 88 to 100 million years ago.29

Duplicative transposition of fragments ranging in size from a few to hundreds of kb has been an important mechanism in mammalian evolution.30 Genomic clustering on the short arm of chromosome 19 suggests that the five human EGF-TM7 receptor genes located there have evolved through a series of segmental intrachromosomal duplications. We propose three duplication events (Figure 9). In a first step, two genes with a tail-to-tail arrangement originated from a common ancestor of EGF-TM7 receptor genes. Different numbers of EGF domain-encoding exons (two versus ≥ five) manifested in these two genes. A second duplication could then have led to two pairs of tail-to-tail arranged genes with corresponding numbers of EGF domain-encoding exons (two in EMR3 and EMR4, five to seven in CD97 and EMRI). Finally, EMR2 arose when duplicated segments of EMR3 and CD97 merged in close centromeric proximity of their original loci.

Transcriptional orientation, exon number, and clustering on the short arm of human chromosome 19 provide evidence in support of the first two duplication events. Presence of Cd97, Emr1, and Emr4 on syntenic regions on mouse chromosome 8 and 17, respectively, suggest that these duplications occurred prior to the divergence of primates and rodents about 64-74 million years ago.31 It is tempting to speculate that Emr3 has existed in ancestral rodent genomes but later disappeared during mouse evolution.
Tracing back the third duplication event that gave rise to EMR2 is more difficult. We detected EMR2 in Old World monkeys but not in phylogenetically older primates. Whether EMR2 indeed evolved only after branching of Old World and New World monkeys over 35 million years ago is uncertain. It is reasonable, however, to assume that in a common ancestor of Old World monkeys and hominoids EMR2 has acquired an independent function and thus protection from inactivating mutations. Analysis of the human genome revealed that approximately 5% of the euchromatic portion consists of interspersed duplications that arose over the past 35 million years. The emergence of genes with new functions from such segmental duplications is a very rare but evolutionary highly important event. Notably, in the case of EMR2, a novel functional gene likely arose from segments from two related genes.

Since its “birth”, ligand specificity of EMR2 has changed. We here demonstrate that mutations within EGF domains 1 and 2 that prevent CD55 binding occurred only after human speciation. This makes EMR2 one of the very few known proteins, which possess a functional difference in humans and chimpanzees. Previously reported examples are Siglec-L1, a lectin recognizing sialic acids, and FOXP2, a transcription factor involved in human speech and language. In all cases, subtle amino-amino substitutions in the human lineage have modified protein function.

Chimpanzee EMR2 can bind CD55. Notably, we found that the largest, CS-binding isoform of EMR2 is predominantly transcribed in chimpanzee peripheral blood leukocytes. Human CD97 isoforms substantially differ in their affinity for CD55, which is about ten times lower for the largest isoform compared to the smallest isoform. Accordingly, unlike the smaller isoforms, CD97-5EGF has little capacity to bind CD55 when expressed at physiological levels like on erythrocytes, leukocytes, or fibroblast-like synoviocytes and Kwakkenbos et al. and Kop et al., submitted for publication). Likewise, we found only very weak binding of chimpanzee EMR2-5EGF to CD55 expressed on leukocytes (data not shown). Little if any transcription of the smaller isoforms of chimpanzee EMR2 (Figure 3) suggests that binding to CD55 is diminished through
Recent evolution of EMR2 in primates

alternative RNA splicing. Amazingly, different mechanisms (mutations versus alternative splicing) seem to have developed in humans and chimpanzees in order to reduce the affinity of EMR2 for CD55.

CS expressed on fibroblasts and on B cells (but not on other leukocytes) has recently been identified as major ligand of the largest isoform of human EMR2 and CD97 (14 and Kwakkenbos et al., submitted for publication). The molecular consequences of the different ligand interactions engaged by CD97 and EMR2 are still poorly understood. Recent in vivo evidence that CD97 has a function in cell migration18 is not conclusive about the role of the two ligands. Neutrophil migration was similarly impaired by blocking of the CD55- and the presumed CS-binding site of mouse CD97 with mAb. Functional studies on EMR2 have not been reported as yet and are hampered by the absence of a mouse homolog. Our findings indicate that CS binding is essential for the function of EMR2. Complete amino-acid conservation of the fourth EGF domain, which provides the CS-binding site, in humans, chimpanzees, and rhesus macaques, emphasize the importance of this interaction.

We previously reported a gene-inactivating frame shift mutation in the human EMR4 gene that is not present in nonhuman primates. The novel differences, identified between human and chimpanzee EMR2, raise the question what drives the rapid evolution of EGF-TM7 receptors in primates. Fast evolution of the polymorphic MHC and natural killer cell receptor families has been attributed to coevolution with human pathogens.35-38 A better knowledge on the function of EGF-TM7 receptors is necessary to identify selective constraints imposed on their progenitors. Genome sequencing of the chimpanzee1 and other primates, in combination with analysis of transcription and molecular function, will tell us whether EGF-TM7 receptors indeed have evolved more rapidly than other families of immunity-related molecules and whether they account for differences, related to immunity, between us and the great apes.

ACKNOWLEDGEMENTS

We thank Rene van Lier for his comments and suggestions. This work was supported by grants from the Netherlands Organization for Scientific Research (NWO, 901-07-208) and the Landsteiner Foundation for Bloodtransfusion Research (LSBR, 0109). J.H. is a fellow of the Royal Netherlands Academy of Arts and Sciences.

REFERENCE LIST

(7) Bjarnadóttir TK, Fredriksson R, Höglund PJ et al. The human and mouse repertoire of the adhesion family
Chapter 7


Recent evolution of EMR2 in primates


