CD97 and EMR2: receptors on the move

Kwakkenbos, M.J.

Citation for published version (APA):
Chapter 8

Inactivation of the EGF-TM7 receptor EMR4 after the Pan-Homo divergence

*European Journal of Immunology*, 33, 1365-1371, 2003

Jörg Hamann¹, Mark J. Kwakkenbos¹, Esther C. de Jong², Henk Heus³, Anne S. Olsen⁴, and René A.W. van Lier¹

¹Laboratory for Experimental Immunology
²Department of Cell Biology and Histology, Academic Medical Center, University of Amsterdam, Amsterdam
³Target Discovery Unit, NV Organon, Oss, The Netherlands
⁴DOE Joint Genome Institute, Walnut Creek, CA, USA.
Chapter 8

ABSTRACT

We here report on the identification of a novel human EGF-TM7 receptor, designated EMR4. Like most EGF-TM7 receptor genes, EMR4 is localized on the short arm of chromosome 19, in close proximity to EMR1. Remarkably, due to a one-nucleotide deletion in exon 8, translation of human EMR4 would result in a truncated 232-amino acid protein lacking the entire seven-span transmembrane region. This deletion is not present in nonhuman primates, including chimpanzees, suggesting that EMR4 became nonfunctional only after human speciation, about five million years ago. Thus, EMR4 surprisingly accounts for a genetic difference between humans and primates related to immunity.

INTRODUCTION

The EGF-TM7 family is a group of class B seven-span transmembrane (TM7) receptors predominantly expressed by cells of the immune system.1,2 Family members CD97, EGF module-containing mucin-like receptor (EMR) 1-4 and EGF-TM7-latrophilin-related protein (ETL) are characterized by an extended extracellular region with a variable number of N-terminal epidermal growth factor (EGF)-like domains.3-15 Proximal to the first transmembrane segment, EGF-TM7 receptors possess a G protein-coupled receptor-proteolytic site (GPS).16 This site gives rise to proteolytic 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.4,13,14,17

EGF-TM7 receptors interact through their EGF domains with cellular ligands, an ability that is unique within the large superfamily of TM7 molecules.16 Two molecular interactions have been identified thus far. CD97 binds CD55/decay accelerating factor,5-6,19,20 a glycosylphosphatidylinositol-linked molecule that prevents complement deposition on self cells by inhibiting C3/C5 convertases. More recently, EMR2 and CD97 were demonstrated to interact with chondroitin sulfate B (dermatan sulfate) (Stacey, et al., submitted). Based on the unique hybrid structure and the ligand-binding properties, a role of EGF-TM7 receptors in leukocyte adhesion and migration has been suggested.1,2,19

In this article, we describe the cloning and initial characterization of a novel human member of the EGF-TM7 family, designated EMR4, and report its inactivation after the Pan-Homo divergence.

MATERIALS AND METHODS

cDNA cloning and sequence analysis

A survey of the DDBJ/EMBL/GenBank databases for novel EGF-TM7 receptor genes identified several exons of the thus far unknown family member EMR4 in a working draft sequence of human chromosome 19 (accession no. AC020895). Based on this sequence, primers were designed and used to derive the complete EMR4 cDNA sequence by a combination of RT-PCR from human PBMC first-strand cDNA and 5'- and 3'-rapid amplification of cDNA ends (RACE) from human spleen Marathon-Ready
Characterization of the EGF-TM7 family member EMR4
cDNA (Clontech, Palo Alto, CA). Sequences were determined with the BigDye terminator cycle sequencing kit (Applied Biosystems, Warrington, GB). Amino acid sequences were compared with the ClustalW software. Protein architecture was analyzed using the SMART program.21

Generation, expression and detection of mFc fusion proteins
To generate mFc fusion constructs,13 the N-terminal part of EMR4 was amplified by RT-PCR from human PBMC first-strand cDNA using the specific primers 5'-CTGACCACGTCAAGCTT/GAGAG-3' [nt 2–25 of the (+) strand, an introduced HindIII site is depicted italic], 5'-GAAATTGGATCCTTTCTGGAAGAAG-3' [nt 551–575 of the (-) strand, an introduced BamHI site is depicted italic], 5'-GGATGCATTCAGAGGCAGAATCCCAAGAG-3' [nt 730–755 of the (-) strand, an introduced BamHI site is depicted italic], and 5'-CAGCACAGGATCCTCCTTGGGGGC-3' [nt 1038–1061 of the (-) strand, an introduced BamHI site is depicted italic]. PCR products were ligated 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). COS cells were transfected with the expression constructs and after 1 day, medium was replaced by serum-free Opti-MEM. Five days later, culture supernatant was collected and 5 μl of each transfection were analyzed by immunoblot for soluble EMR4-mFc protein using HRP-labeled anti-mouse Ig (DAKO, Glostrup, Denmark).

Chromosomal localization
EMR4 was identified in genomic sequence (accession number AC020895) as described. Coding sequences for previously identified EGF-TM7 receptor genes were localized in chromosome 19 genomic sequence by blast analysis with the corresponding cDNA sequences. The cosmids and BAC from which genomic sequence was derived had been previously localized on the chromosome 19 map by a combination of restriction mapping and high-resolution pronuclear FISH22,23; for updated map see http://bbrp.llnl.gov/bbrp/genome/html/chrom-map.html). Mouse chromosomal locations for regions corresponding to the human EGF-TM7 receptor gene locations on chromosome 19 were derived from the mouse-human comparative map constructed by Kim et al.24 and the location of homologous reference genes in mouse genomic sequence.25

RT-PCR on human DC subsets
Blood DC of the myeloid and plasmacytoid lineage were isolated from human PBMC according to the protocol of Dzionek et al.26 In short, after depletion of CD19+ B cells using anti-CD19 microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany), cells were labeled with PE-conjugated mAb directed against BDCA-1, BDCA-3 or BDCA-4 for 15 min at 4°C in FcR-blocking reagents (Miltenyi Biotec). Next, anti-PE microbeads were added and incubated for another 15 min at 4°C. BDCA-1+, BDCA-3+ and BDCA-4+ cells were subsequently isolated using immunomagnetic cell sorting (MACS, Miltenyi Biotec). To obtain >98% pure cell populations, the enriched fractions were further purified by FACS sorting using a FACSVantage (Becton Dickinson, San Jose, CA). From the sorted cells, first-strand cDNA was prepared and amplified by PCR (35–40 cycles, 30
s at 93°C, 30 s at 58°C, 40 s at 72°C) using primer combinations depicted in Table 1. PCR products were separated on a 1% agarose gel.

Table 1. Oligonucleotide primers used for RT-PCR on human DC.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Orientation</th>
<th>Sequence (5' to 3')</th>
<th>Fragment size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD97</td>
<td>Sense</td>
<td>CCGAGGTACCATCCAGAATG</td>
<td>301</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>GGATCTGCGGCCTCAGCTGC</td>
<td></td>
</tr>
<tr>
<td>EMR1</td>
<td>Sense</td>
<td>CCAGTGTCGACACAGGGGCATG</td>
<td>473</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>GCCGTCTTGGAAGCGGATGGC</td>
<td></td>
</tr>
<tr>
<td>EMR2</td>
<td>Sense</td>
<td>CTAGGCTGTGCTCACTAACT</td>
<td>582</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>CTAGTTACCGGCTGGGTTTGGAGGTG</td>
<td></td>
</tr>
<tr>
<td>EMR3</td>
<td>Sense</td>
<td>GAACCTCTAGGTTGATGCTGC</td>
<td>373</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>GGTGCTGTGACTGAGCTGGC</td>
<td></td>
</tr>
<tr>
<td>EMR4</td>
<td>Sense</td>
<td>GCTGTGTCGAAATGCTGGCATCG</td>
<td>318</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>GTATGCAATGTTGATCGCAATCG</td>
<td></td>
</tr>
<tr>
<td>ETL</td>
<td>Sense</td>
<td>GGGTCTTTCGTTCACTCTTGGAC</td>
<td>734</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>GCTGCGCAAATAACCTACCTACCAC</td>
<td></td>
</tr>
<tr>
<td>GADPH</td>
<td>Sense</td>
<td>GTGAAAGTCCAGGATCGTAC</td>
<td>298</td>
</tr>
<tr>
<td></td>
<td>Antisense</td>
<td>TGAAGACGCCAGTGAGCTC</td>
<td></td>
</tr>
</tbody>
</table>

Sequencing of exon 8 in human and nonhuman primates

A 182-bp fragment from exon 8 was amplified by PCR (35 cycles, 30 s at 94°C, 30 s at 58°C, 30 s at 72°C) with the specific primers 5'-TTTCTTCAGAGAGCACTGCAG-3' [nt 698–709 of the (+) strand, nt 1–9 belong to the preceding splice acceptor] and 5'-GGCGAAAAGTCAGGAACAGG-3' [nt 852–873 of the (-) strand]. As template, genomic DNA was used from Caucasians (n=8, donors recruited from our laboratory, kindly provided by E. Remmerswaal), sub-Saharan Africans (n=9, Coriell Institute for Medical Research, Camden, NJ) and different nonhuman primates including chimpanzee (Pan troglodytes, n=3), orangutan (Pongo pygmaeus, n=2), rhesus macaque (Macaca mulatta, n=3) and crab-eating macaque (Macaca fascicularis, n=3) (Biomedical Primate Research Center, Rijswijk, The Netherlands). PCR products were sequenced as described.

RESULTS AND DISCUSSION

cDNA cloning, sequence analysis and chromosomal localization of EMR4

Evidence for the existence of a novel EGF-TM7 receptor gene on human chromosome 19 was received during a survey of the public databases. In line with the designation of other EGF-TM7 family members, we called this novel molecule EMR4. The complete cDNA sequence of the low-copy gene EMR4, comprising 2,732 nucleotides (nt), was determined (Figure 1). Alignment of the cDNA sequence with the sequence of human chromosome 19 indicated that EMR4 consists of 16 exons. Unexpectedly, the single open-reading frame was found to shift in exon 8 to a different reading frame that terminates ten codons downstream. Comparison of the cDNA sequence with the recently identified murine homologue of EMR4, also designated F4/80-like receptor
Characterization of the EGF-TM7 family member EMR4

Figure 1. Nucleotide and deduced amino acid sequence of human EMR4. A bar indicates the codon (nt 715–716) with a one-nt deletion. This deletion results in a shift to a different reading frame that terminates ten codons downstream. Without the frameshift mutation, a polypeptide with homology to other EGF-TM7 receptors would be generated, the sequence of which is depicted in italic. The signal sequence and the seven hydrophobic transmembrane segments are underlined. EGF domains are shaded and the GPS motif is underlined by a broken line. Potential N-glycosylation sites are boxed. Exon borders as determined by structural analysis of the EMR4 gene are indicated.

(FIRE),12,13 showed that this frameshift is caused by a one-nt deletion in codon 223. Due to this deletion, translation would result in a truncated 232-amino acid polypeptide, whereas a 663-amino acid polypeptide would be synthesized without a frame shift (Figure 1).

Analysis of the amino acid sequence identified a hydrophobic signal sequence (residues 1–19). A full-length mature polypeptide would consist of 644 amino acids, with a predicted molecular mass of 72 kDa. Close to the N-terminus, EMR4 possesses two EGF domains (residues 31–69 and 70–120), the second one containing the consensus sequence for a class I calcium-binding site. The EGF domain region is followed, subsequently, by a stalk region (residues 121–339), a class B TM7 region (residues 340–586) and a cytoplasmic region (residues 587–663). The presence of a GPS immediately upstream to the TM7 region indicates that EMR4, like other EGF-TM7 receptors, is likely to traffic to the plasma membrane by a default process.27
receptors, would be proteolytically processed into an extracellular α chain and a TM7/cytoplasmic β chain, which reassociate prior to expression at the cell surface. The predicted overall structure of EMR4 (Figure 2A) is similar to other members of the EGF-TM7 family. Pairwise comparison revealed an amino acid sequence identity of 45% with EMR3, 38% with EMR2, 34% with EMR1, 29% with CD97 and 26% with ETL. The amino acid sequence identity between human EMR4 and its murine homologue is 59%.

Figure 2. (A) Schematic structure of human EMR4. Due to a frameshift mutation translation would result in a truncated soluble protein. The part of the molecule lacking in the truncated translation product is depicted dotted. EMR4 has two EGF domains, the second one possessing a calcium binding site. (B) Generation of EMR4-mFc fusion constructs. The cDNA sequence of EMR4 up to codon 176, 236 or 338 was ligated in-frame immediately upstream to the CH3-CH2-hinge region sequence of mouse IgG2b. The resulting recombinants contain the extracellular part of EMR4 upstream from the one-nt deletion (1), including the one-nt deletion (2) or upstream from the TM7 region (3), respectively. (C) Expression and detection of EMR4-mFc fusion proteins. Constructs were transfected into COS cells and culture supernatants were analyzed by immunoblot for secreted EMR4-Fc protein. In constructs 2 and 3, the frame shift induced by the one-nt deletion prevented generation of a fusion protein.

To verify the effect of the frameshift mutation on the translation of EMR4, several fusion proteins were generated in which the sequence encoding the extracellular region of EMR4 was ligated in-frame upstream to the CH3-CH2-hinge region sequence of mouse IgG2b. Analysis of the resulting translation products demonstrated that synthesis of EMR4-mouse Fc (mFc) fusion proteins was abrogated in those constructs that possessed the one-nt deletion, but not in a construct that only contained the amino acid sequence upstream from the deletion (Figure 2B, C).

By integration of genomic sequence with the physical map of chromosome 19, we assigned EMR4 to 19p13.3 just proximal to EMR1 (Figure 3). The two genes are transcribed from opposite strands in a tail-to-tail arrangement, with the 3'-termini about 12 kb apart. The organization of the EMR4 gene, consisting of 16 exons (Figure 1), is highly similar to that of other EGF-TM7 receptor genes.

Transcription of EMR4 in dendritic cells
Characterization of the murine homologue of EMR4 revealed expression on monocytes, macrophages and dendritic cells. On DC, mouse EMR4, like the murine homologue of EMR1, F4/80, is mainly expressed by the CD8+ cells, which are considered to represent myeloid DC. To investigate whether transcriptional regulation has been conserved in human EMR4, we examined transcription of EGF-TM7 receptors in human blood DC subsets, defined by the expression of blood DC antigen (BDCA) markers. Whereas BDCA-1 and BDCA-3 are expressed on myeloid DC, BDCA-4 is
Characterization of the EGF-TM7 family member EMR4

Figure 3. Location of EGF-TM7 receptor genes on human chromosome 19p13. Two sections of the map in the 19p13.3 and 19p13.1 region are shown. Numbers along scale bar indicate approximate distance (in Mb) from p-telomere, as determined by a combination of restriction mapping and high-resolution pronuclear FISH. Several reference genes are indicated to the right of the scale. Five EGF-TM7 genes (ETL is on chromosome 11) are indicated in bold, with transcriptional orientation indicated by an arrow to right of the gene name. Accession numbers of genomic sequence from BAC or cosmid clones containing these EGF-TM7 genes are shown to the left of the scale. The chromosomal locations of the homologous mouse genes are shown to the far right.

Figure 4. Transcription of EGF-TM7 receptors in human DC. Blood DC were sorted into highly pure populations expressing either BDCA-1 (myeloid DC), BDCA-3 (myeloid DC) or BDCA-4 (plasmacytoid DC). From 40,000, 20,000 and 30,000 cells, respectively, first-strand cDNA was synthesized and analyzed by RT-PCR for transcripts of CD97, EMR1, EMR2, EMR3, EMR4, ETL and GADPH (control). The number of PCR cycles was 35, except for EMR4 (40 cycles). As control, first-strand cDNA from PBMC or from the lymphoblastic B cell line Ramos (ETL) was used.

expressed on plasmacytoid DC that are likely of lymphoid origin. As delineated in Figure 4, three transcription patterns were found. First, EMR3 and ETL are not detectable in DC. Second, CD97 and EMR2 are present in all three subsets. Third, EMR1 and EMR4 are transcribed in BDCA-1+ cells and at very low levels in BDCA-3+ cells, but not in BDCA-4+ cells. Although comparison of human and murine DC is hampered by the different expression of phenotypic markers, like CD8 and CD4, these data suggest that expression of EMR4 and EMR1, in both species, is restricted to myeloid DC. Possibly, the remarkable similarity in the presence of EMR4 and EMR1 in DC is due to coordinated transcriptional regulation of the very closely located genes.
Analysis of the one-nucleotide deletion in human and nonhuman primates

The completely preserved exon-intron organization and regulated transcription in DC suggest that the frameshift mutation in exon 8 of *EMR4* manifested during recent evolution. To investigate this, we examined the DNA sequence of this region from various human and nonhuman primates. No polymorphism at the one-nt deletion site was detected in either, Caucasians or sub-Saharan Africans (data not shown, for details see Sect. 3). However, as shown in Figure 5, an additional nt was found in all primates studied, including chimpanzees. Although the presence of inactivating mutations elsewhere in the primate *EMR4* genes cannot be excluded, this indicates that EMR4 lost its function only after human speciation that occurred approximately 5 million years ago. It makes *EMR4* one of the very few genes and, to our knowledge, the only one encoding an individual molecule involved in immunity that, on current evidence, are nonfunctional in humans but still might be active in chimpanzees. The only biochemically characterized gene that became inactive after divergence from chimpanzee is the cytidine monophosphate-N-acetylneuraminic acid (CMP-Neu5Ac) hydroxylase gene (*CMAH*), which contains a 92-nt deletion in humans. Consequently, the great apes, but not humans, express high levels of the sialic acid derivative N-glycolyNeuraminic acid (Neu5Gc). Another example, the C-type lectin gene *Ly49L*, became nonfunctional due to a splice donor site mutation after divergence of the common ancestor of chimpanzee and humans from gorilla, approximately 8 million years ago. The function of *Ly49L* is still elusive.

Concluding remarks

Taken together, we here provide evidence that the EGFTM7 receptor gene *EMR4* became inactive in humans due to a frameshift mutation. Whether a truncated, soluble form of EMR4 is secreted by human leukocytes remains to be demonstrated with specific antibodies. The most interesting questions, actually, deal with the identification of the ligand and the function of EMR4. Binding of multivalent, soluble mouse EMR4 probes to the B lymphoma cell line A20 suggests that EMR4, like other EGF-TM7 receptors, mediates cellular interactions. Unravelling the physiological role of EMR4 might provide the unique opportunity to study a genetic difference between our species and our evolutionary closest relatives, a challenge that is even more intriguing since humans and chimpanzees differ substantially in their susceptibility to diseases like malaria, epithelial cancers, Alzheimer's disease and AIDS.
ACKNOWLEDGEMENTS

We thank Ronald Bontrop, Erik Eldering and Peter de Knijff for their comments and suggestions. We are indebted to Hsi-Hsien Lin and Martin Stacey for sharing data on mouse EMR4 prior to publication. This work was supported by grant no. 901-07-208 from the Netherlands Organization for Scientific Research (NWO). J.H. is a fellow of the Royal Netherlands Academy of Arts and Sciences. Work at LLNL/JGI was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, under contract no. W-7405-Eng-48. Nucleotide sequence data reported are available in the DDBJ/EMBL/GenBank databases under accession number AF489700.

REFERENCE LIST

(14) Nechiporuk T, Urness LD, Keating MT. ETL, a novel seven-transmembrane receptor that is developmentally regulated in the heart. ETL is a member of the secretin family and belongs to the epidermal growth factor-seven-transmembrane subfamily. J Biol Chem. 2001;276:4150-4157.
Chapter 8


