CD97 and EMR2: receptors on the move

Kwakkenbos, M.J.

Citation for published version (APA):
Chapter 2

The EGF-TM7 family: a postgenomic view

*Immunogenetics, 55: 655-666, 2004*

Mark J. Kwakkenbos¹, Else N. Kop¹,², Martin Stacey³, Mourad Matmati¹, Siamon Gordon³, Hsi-Hsien Lin³ and Jörg Hamann¹

¹Laboratory for Experimental Immunology  
²Division of Clinical Immunology and Rheumatology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands  
³Sir William Dunn School of Pathology, University of Oxford, Oxford, U.K.
Chapter 2

ABSTRACT

With the human and mouse genome projects now completed, the receptor repertoire of mammalian cells has finally been elucidated. The EGF-TM7 receptors are a family of class B seven-span transmembrane (TM7) receptors predominantly expressed by cells of the immune system. Within the large TM7 superfamily, the molecular structure and ligand binding properties of EGF-TM7 receptors are unique. Derived from the processing of a single polypeptide, they are expressed at the cell surface as heterodimers consisting of a large extracellular region associated with a TM7 moiety. Through a variable number of N-terminal epidermal growth factor (EGF)-like domains, EGF-TM7 receptors interact with cellular ligands such as CD55 and chondroitin sulfate. Recent in vivo studies demonstrate a role of the EGF-TM7 receptor CD97 in leukocyte migration. The different number of EGF-TM7 genes in man compared to mice, the chimeric nature of EMR2 and the inactivation of human EMR4 point toward a still-evolving receptor family. Here we discuss the currently available information on this intriguing receptor family.

INTRODUCTION

Seven-span transmembrane (TM7) receptors represent the largest gene family in animal genomes. TM7 receptors respond to a diverse array of sensory and chemical stimuli such as light, taste, odor, pheromones, calcium ions, neurotransmitters, hormones and chemokines. Upon ligation, TM7 receptors regulate a variety of physiological processes via heterotrimeric G proteins which engage subsequent messenger molecules. Although G protein-coupled receptor (GPCR) is an abundantly used synonym for TM7 molecules there is growing evidence that some TM7 receptors signal through alternative, G protein-independent, mechanisms.

TM7 receptors are divided into several classes, which have most likely arisen independently through convergent evolution. Best established are the rhodopsin superfamily (class A), the secretin receptor superfamily (class B) and the metabotropic glutamate receptor superfamily (class C). Class B was originally defined as a family of insect and mammalian peptide hormone receptors. Subsequent identification of receptors with homologous TM7 moieties but different extracellular regions have led to the classification of class B TM7 receptors into three distinct families: B1, peptide hormone receptors, B2, TM7 receptors with a long N-terminal extracellular region (LNB-TM7 family) and B3, methuselah-like proteins.

The human LNB-TM7 family comprises approximately 30 members. They all possess large extracellular regions containing a number of well-defined protein domains. The structural diversity within the LNB-TM7 family is illustrated by molecules such as Cadherin EGF LAG seven-pass G-type receptor 1 (Celsr1), a mosaic protein containing over 15 repeats belonging to three different types of protein module. Another example is the very large G protein-coupled receptor 1 (VLG1), with ~6300 amino acids the largest cell surface protein yet identified. This review will focus on a subfamily of the LNB-TM7 receptors that solely contain extracellular epidermal growth factor (EGF)-like domains, referred to as EGF-TM7 receptors.

With the human genome now unravelled, the complete family comprises six members: CD97, EGF-like module-containing mucin-like receptor protein1
The EGF-TM7 family

Figure 1. Schematic structure of human (black) and mouse (gray) EGF-TM7 family members. EGF-TM7 receptors are expressed as noncovalently associated heterodimers consisting of an extracellular α subunit and a TM7/cytoplasmic β subunit. Due to alternative RNA splicing, they possess a variable number of N-terminal EGF domains, represented as triangles. Depicted here are the largest isoforms. Calcium-binding EGF-like domains are shown in gray. Small circles indicate the position of the processing site (GPS motif). Human EMR4 is shown with a dotted line because the human gene is inactivated by a one-nucleotide deletion in exon 8. See text for details.

(EMR1),21 EMR2,22 EMR3,23 EMR424 and EGF-TM7-latrophilin-related protein (ETL).25 Four of these molecules have mouse orthologues: CD97,26,27 EMR1 (also known as F4/80),28,29 EMR4 (also called F4/80-like-receptor, FIRE)30,31 and ETL.32,33 Interestingly, surveys of the mouse genome4 have failed to identify EMR2 and EMR3 (Stacey and Lin, unpublished observation). More recently EGF-TM7 genes have also been identified in various other vertebrate species including rat, pig, cow and zebrafish (see below). Schematic structure and key characteristics of human and mouse EGF-TM7 receptors are depicted in Figure 1 and Table 1.

MOLECULAR STRUCTURE

The extracellular region possesses tandemly arranged EGF domains and a conserved cleavage motif

EGF-TM7 receptors possess distinct parts including a cleavable signal peptide, an N-terminal EGF domain region, a stalk containing a conserved cleavage site, a class B TM7 region and a short and poorly conserved cytoplasmic tail (Figures 1 and 2). Comprising several hundred amino acids, the extracellular regions of EGF-TM7 receptors represent one of the largest found in TM7 molecules. Located at the N-terminus of every EGF-TM7 receptor are several tandemly arranged EGF-like domains (Figures 1 and 2). EGF domains belong to the most broadly expressed protein modules in animals.24 Through the interaction with other protein modules, they are involved in various physiological processes including blood coagulation, fibrinolysis, neural development and cell adhesion.34 Within tandemly linked EGF domains, tertiary structures emerge through contacts between succeeding domains, which are stabilized by calcium ligation of the C-terminal module. Consequently, the EGF domains in EGF-TM7 receptors, except for the most proximal ones, possess a calcium-binding site. The calcium-binding EGF
### Table 1: Characteristics of human and mouse EGF-TM7 receptors. See text for details.

<table>
<thead>
<tr>
<th>Name</th>
<th>EGF domains</th>
<th>Expression pattern</th>
<th>Ligands</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD97</td>
<td>3-5:</td>
<td>Activated lymphocytes, monocytes, macrophages, dendritic cells, granulocytes, smooth muscle cells, malignant cells</td>
<td>CD55 (DAF), chondroitin sulfate</td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,5),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,5),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,4,5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR1</td>
<td>1-6:</td>
<td><em>Monocytes, dendritic cells</em></td>
<td>Unknown</td>
</tr>
<tr>
<td></td>
<td>(EGF4),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF4,5),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF4,5,6),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,4,5,6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR2</td>
<td>2-5:</td>
<td><em>Monocytes, macrophages, dendritic cells, granulocytes</em></td>
<td>Chondroitin sulfate</td>
</tr>
<tr>
<td></td>
<td>(EGF1,2),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,5),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,5),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,4,5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR3</td>
<td>2</td>
<td><em>Monocytes, macrophages, granulocytes</em></td>
<td>Expressed on monocyte-derived macrophages and activated neutrophils</td>
</tr>
<tr>
<td>EMR4</td>
<td>2</td>
<td><em>Monocytes, dendritic cells</em></td>
<td>Unknown</td>
</tr>
<tr>
<td>ETL</td>
<td>2</td>
<td><em>Smooth muscle cells</em></td>
<td>Unknown</td>
</tr>
<tr>
<td><strong>Mouse</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD97</td>
<td>3-4:</td>
<td>Lymphocytes, monocytes, macrophages, granulocytes</td>
<td>CD55 (DAF), chondroitin sulfate</td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,4),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,4),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,X3,4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR1</td>
<td>0-7:</td>
<td><em>Monocytes, macrophages, dendritic cells, eosinophils</em></td>
<td>Unknown</td>
</tr>
<tr>
<td>(F4/80)</td>
<td>(EGF4,5,6,7),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,6,7),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,4,5,6,7),</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EGF1,2,3,4,5,6,7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMR4</td>
<td>2</td>
<td><em>Monocytes, macrophages, dendritic cells</em></td>
<td>Expressed on B lymphoma cell line</td>
</tr>
<tr>
<td>(FIRE)</td>
<td></td>
<td></td>
<td>A20</td>
</tr>
<tr>
<td>ETL</td>
<td>3</td>
<td><em>Smooth muscle cells, hematopoietic stem cells</em></td>
<td>Unknown</td>
</tr>
</tbody>
</table>

*The EGF composition of isoforms from alternatively spliced receptors is depicted. For EMR2, also alternative splicing that could give rise to truncated, soluble receptors has been observed.*

*Data derived by detection of transcripts are given in italics.*
domains of the EGF-TM7 family belong to the fibrillin-like class I type and are arranged in a near-linear order.\textsuperscript{35}

A characteristic of CD97, EMR1, EMR2 and ETL is the existence of alternative RNA splicing. This results in the expression of various isoforms with differing numbers and arrangements of EGF domains. In contrast, EMR3 and EMR4 exist with two EGF domains. Closer examination of the EGF domain regions has led to some remarkable findings. Firstly, the EGF domains of CD97 and EMR2 are nearly identical,\textsuperscript{22} differing by only six amino acids. Secondly, unusual isoforms exist of mouse CD97 and mouse and rat ETL. In the largest isoform of mouse CD97, the second and third EGF domains are interrupted by a sequence of 45 amino acids that does not show homology to any known protein module.\textsuperscript{26,27} In rodent ETL, two calcium-binding EGF domains with an identical amino acid sequence are found.\textsuperscript{25,32}

A stalk region of 160-260 amino acids separates the EGF domain region from the GPCR-proteolytic site (GPS) (Figure 2). Low amino acid sequence conservation suggests that this region serves as a spacer between the ligand binding EGF domains and the TM7 part. Because of a serine/threonine content up to 21%, extensive O-linked glycosylation of the stalk region has been predicted as reflected by the designation EMR (where M refers to mucin). However, detailed biochemical analysis of mouse EMR1 provided no evidence for substantial O-linked glycosylation.\textsuperscript{36}

Proximal to the first transmembrane segment EGF-TM7 proteins possess a conserved cysteine motif (Figure 2). The first functional hint of this motif was revealed when Kelly and coworkers demonstrated intracellular processing of CD97.\textsuperscript{19} Translated as a single polypeptide, cleavage in the endoplasmic reticulum results in the formation of a heterodimer comprised of an extracellular \(\alpha\) subunit noncovalently associated with a TM7/cytoplasmic \(\beta\) subunit. Subsequent cloning of the calcium-independent receptor of \(\alpha\)-latrotoxin (CIRL) revealed a second heterodimeric LNB-TM7 receptor.\textsuperscript{37} Amino acid sequencing demonstrated the cleavage site to be within the cysteine box. Consequently, this motif was named GPCR-proteolytic site. Other non-TM7 membrane-associated proteins such as the sea urchin receptor for egg jelly protein (suREJ) and polycystin-1 also possess a GPS motif and are proteolytically processed at that site.\textsuperscript{38,39} A series of recent biochemical studies have confirmed the position of the cleavage and demonstrated that the GPS motif is absolutely necessary but not sufficient to facilitate the cleavage reaction.\textsuperscript{40-42} Additional N-terminal sequences, but not the TM7 region, are required. EMR2 cleavage was only observed when the complete stalk region was present.\textsuperscript{41}

Nearly all LNB-TM7 receptors have a GPS site and several have proven to be expressed at the cell surface as heterodimers.\textsuperscript{12,19,25,31,37,40,43,44} An exception might be EMR1, which has an imperfect GPS site, whose functionality still needs to be demonstrated. Although proteolytic processing of LNB-TM7 receptors is now generally acknowledged, the functional implications of a heterodimeric receptor structure are still elusive. Strikingly, ETL not only has a heterodimeric structure, but also associates pairwise at the cell surface.\textsuperscript{25}
Figure 2. Alignment of amino acid sequences of human (A) and mouse (B) EGF-TM7 receptors. EGF domains, the GPS motif and the seven hydrophobic transmembrane segments are indicated. An arrowhead shows the predicted processing site within the GPS motif. Inner and outer boarders of overlap between the transmembrane segments, found in the different sequences, are given. Positions of conserved amino acids are indicated by asterisks for identical residues and by colons for homologous residues. In human EMR4, a one-nucleotide deletion in exon 8 causes a shift to a different reading frame that results in a translation product terminating in the stalk region. The polypeptide sequence downstream from the termination site that, like in nonhuman primates, would result from a nonmutated gene is depicted in italics. In mouse CD97, the 45 amino acids between the second and third EGF domain that do not correspond to known protein modules are shaded. Amino acid sequences were aligned with the ClustalW software. The SMART program (http://smart.embl-heidelberg.de) was used to analyze protein architecture.
Chapter 2

CHROMOSOMAL LOCALIZATION, GENE ORGANIZATION AND EVOLUTION

The EGF-TM7 family arose by gene duplication and exon shuffling

Table 2 summarizes the organization of human EGF-TM7 receptor genes. The structure of CD97, EMR1, EMR2, EMR3 and EMR4, and to a lesser extent ETL, is similar. Variability in exon numbers is mainly due to differences in the number of EGF domains. Except for ETL, all human EGF-TM7 receptor genes are located on the short arm of human chromosome 19 within clusters in 19p13.1 (CD97, EMR2 and EMR3) and 19p13.3 (EMR1 and EMR4) (Figure 3). These clusters, separated by approximately 7.4 Mb, diverged by gene duplication in a common vertebrate ancestor as long as 250 million years ago.45 Close proximity of EMR1 and EMR4 (genes about 12 kb apart) as well as CD97, EMR3 and EMR2 (genes about 190 and 57 kb apart, respectively) suggest that additional rounds of gene duplication and diversification have contributed to the current extent of the EGF-TM7 family in man. The presence of Cd97, Emrl and Emr4 in the mouse indicates that these genes evolved prior to divergence of rodents and primates about 70 million years ago. The completion of further mammalian genomes should show whether EMR2 and EMR3 have evolved later on in primates or have been lost in rodents. The sequence of EMR2 (see below) favours a recent origin of this EGF-TM7 receptor.

Table 2: Genomic organization of human EGF-TM7 receptors genes.

<table>
<thead>
<tr>
<th></th>
<th>CD97</th>
<th>EMR1</th>
<th>EMR2</th>
<th>EMR3</th>
<th>EMR4</th>
<th>ETL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'/SP</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>EGF</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Stalk</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>TM7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>C/3'</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>16</td>
<td>16</td>
<td>15</td>
</tr>
</tbody>
</table>

Depicted are the number of exons in relation to the 5' noncoding region and the signal peptide (5'/SP), the EGF domain region (EGF), the stalk region (Stalk), the TM7 region (TM7) and the cytoplasmic and 3' noncoding region (C/3'). Given is also the total number of exons. Exon numbers have been derived from the human genome and references.30,68,28,24,28,22,25

The broad distribution of modules like EGF domains in vertebrate proteins is due to a mechanism known as exon shuffling.46 As a prerequisite, these modules need to be encoded by individual exons with compatible ends. Indeed, all EGF domains of EGF-TM7 receptors are encoded by exons that start and stop after the first nucleotide of a codon. The presence of mobile modules clearly identifies EGF-TM7 receptors as modern animal proteins.
The EGF-TM7 family

Thus far, except for one gene in zebrafish (Genbank accession number AAH55171), EGF-TM7 receptors seem to be restricted to mammals. The gene found in zebrafish is most closely related to ETL emphasizing the (also for structure, chromosomal location and expression) secluded position of ETL within the EGF-TM7 family. Sequencing of other vertebrate genomes is required to define whether ETL indeed represents a more ancienly evolved family member.

![Diagram of EGF-TM7 receptors]

**Figure 3.** Location of human EGF-TM7 receptor genes on the short arm of Chr 19. Two sections of the map in the 19p13.3 and 19p13.1 region are shown. Numbers along scale bar indicate approximate distances (in Mb) from p-telomere. Several reference genes are given. EGF-TM7 family members are depicted in bold, with transcriptional orientation indicated by an arrow to the right of the name. Extents of homology to mouse chromosomes are shown by bars at far right. The ETL gene is localized on human chromosome 1p32-p33 and mouse chromosome 3, respectively. This figure has been adapted with kind permission from reference. See also references.

**EMR2 represents an intermediate, closely related to both CD97 and EMR3**

Molecular cloning of EMR2 surprisingly revealed a remarkable similarity with two other members of the EGF-TM7 family (Figures 2 and 4, Table 3). The amino acid sequence of the \( \beta \) part is 79% identical with that of EMR3. Due to this similarity, EMR3 was discovered by chance when primers designed to the TM7 region of EMR2 coamplified a related but not identical sequence. In contrast, the \( \alpha \) subunit of EMR2 is highly similar to CD97 (57% amino acid identity) and nearly identical for the signal peptide, the EGF domains (see above) and the most upstream part of the stalk region. The striking homology of EMR2 with CD97 and EMR3 suggests that this EGF-TM7 receptor evolved through a combination of gene duplication and gene recombination.

**Inactivation of EMR4 in humans**

EMR4 has changed during most recent evolution. Human EMR4 contains a one-nucleotide deletion in exon 8 that results in termination of the translation product ahead of the TM7 region (Figure 2). The frame shift mutation is not present in nonhuman primates, including chimpanzees, suggesting that EMR4 became nonfunctional only after human speciation. Molecular differences between humans and primates related to immunity are largely restricted to multi-gene families like the MHC or the NK cell receptors. EMR4 is the first single copy gene that, on current evidence, is nonfunctional in humans but still potentially active in chimpanzees.
**Chapter 2**

**Table 3:** Amino acid sequence identity (%) between EGF-TM7 family members. Identities higher than 50% are in bold. Pairwise alignments were performed with the ClustalW software.

<table>
<thead>
<tr>
<th></th>
<th>hCD97</th>
<th>mCD97</th>
<th>hEMR1</th>
<th>mEMR1</th>
<th>hEMR2</th>
<th>hEMR3</th>
<th>hEMR4</th>
<th>mEMR4</th>
<th>NETL</th>
</tr>
</thead>
<tbody>
<tr>
<td>mCD97</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hEMR1</td>
<td>28</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mEMR1</td>
<td>30</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hEMR2</td>
<td>61</td>
<td>43</td>
<td>34</td>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hEMR3</td>
<td>34</td>
<td>33</td>
<td>37</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>hEMR4</td>
<td>29</td>
<td>27</td>
<td>34</td>
<td>34</td>
<td>38</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mEMR4</td>
<td>27</td>
<td>27</td>
<td>33</td>
<td>32</td>
<td>34</td>
<td>41</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NETL</td>
<td>25</td>
<td>23</td>
<td>25</td>
<td>28</td>
<td>26</td>
<td>30</td>
<td>26</td>
<td>26</td>
<td>78</td>
</tr>
</tbody>
</table>

*Human and mouse orthologues are indicated by the prefix h and m, respectively.

**EXPRESSION**

**EGF-TM7 receptors are predominantly expressed by hematopoietic cells**

Receptors of the EGF-TM7 family are expressed by cells of the immune system and by smooth muscle cells. The expression of human and mouse CD97, mouse EMR1, human EMR2 and mouse EMR4 has been studied with monoclonal antibodies (mAbs), whereas data on the other family members are restricted to the transcription level. Three main patterns of cellular distribution can be distinguished between CD97, the EMR group and ETL (Table 1).

CD97 is broadly expressed, found on most hematopoietic cells including activated lymphocytes, monocytes, macrophages, dendritic cells and, granulocytes. As yet, CD97 is the only EGF-TM7 receptor detected on lymphocytes. Rapid upregulation during lymphocyte activation initially led to the definition of CD97 as an activation marker. Expression on myeloid cells is to a lesser extent upregulated by cellular activation. Next to hematopoietic cells, smooth muscle cells abundantly express CD97.

On current evidence, the expression of the four EMR molecules is restricted to myeloid cells. The EMR1 mAb (F4/80) has been used extensively by Gordon and coworkers in order to describe the mononuclear phagocyte system in the mouse. Due to its characteristic presence on many tissue macrophages, EMR1 is broadly used as a defining marker of this cell type in the mouse. Monocytes (low levels), CD8+ myeloid dendritic cells, Langerhans cells and eosinophilic granulocytes also express mouse EMR1. Other hematopoietic cells as well as macrophages localized to T cell areas, e.g. in the white pulp of the spleen, are F4/80+. Macrophage activation results in downregulation of mouse EMR1. Transcription data from human EMR1 are too limited for comparison.

EMR2 is expressed on monocytes, tissue macrophages, dendritic cells and neutrophilic granulocytes (low levels). In contrast to the rather equally expressed CD97, presence of EMR2 on peripheral blood CD14+CD16+ cells is lower than on the
The EGF-TM7 family

more mature CD14-CD16- cells. Interestingly, BDCA4+ lymphoid dendritic cells also seem to express EMR2. Distribution of EMR3 is probably similar to EMR2 as revealed by the detection of transcripts in monocytes, macrophages and neutrophils.

Figure 4. Phylogenetic trees of human and mouse EGF-TM7 family members. Amino acid sequence analysis has been performed separately for the extracellular α subunit and the TM7/cytoplasmic β subunit. To avoid effects from domain number, only the first two EGF domains were included for alignment of the α subunits. Analysis was performed with the MEGA program (http://megasoftware.net) using the neighbor-joining method.

Using a panel of mAbs, expression of mouse EMR4 has been demonstrated on monocytes, macrophages and CD8+ myeloid dendritic cells. Expression is restricted to a subpopulation of these cell types and declines after cellular activation. Expression of EMR4 shows similarities to EMR1, which might be explained by coordinated regulation of the very closely located genes (see Figure 3). However, unlike mouse EMR4, mouse EMR1 is found at much higher levels on macrophages compared to dendritic cells. Transcripts of human EMR4 have been detected, albeit at low levels, in monocytes and myeloid dendritic cells (Matmati, unpublished observation).

In contrast to the other EGF-TM7 receptors, ETL is mainly expressed in smooth muscle. Rat ETL transcripts have been identified in cardiac myocytes and bronchiolar and vascular smooth muscle cells. Expression of ETL is not restricted to smooth muscle as indicated by the identification of mouse ETL in a cDNA library of hematopoietic stem cells. We have found human ETL mRNA in the B lymphoblastic cell line Ramos but have failed to detect it in any type of peripheral blood mononuclear cells (PBMC) (Kwakkenbos and Matmati, unpublished observation).
Chapter 2

Aberrant expression of EGF-TM7 receptors at sites of inflammation

Increased CD97 expression has been implicated in several diseases linked with autoimmunity. In a descriptive study, expression of CD97 on brain tissue of patients with multiple sclerosis (MS) was compared to tissue of healthy controls. MS is a chronic inflammatory demyelinating disease of the central nervous system. While no CD97 staining was found in the white matter of healthy controls, infiltrating T cells, macrophages and microglia observed in tissue of MS patients expressed CD97.

Another study described the expression of CD97 in synovial tissue of patients with rheumatoid arthritis (RA). RA is a chronic inflammatory disease affecting synovial tissue in multiple joints. It is believed that the chronic inflammation seen in this disease, is sustained by the interaction of fibroblast-like synoviocytes and macrophages. Fibroblast-like synoviocytes are characterised by high expression of CD55, which is a ligand of CD97 (see next section). A close association between the CD97+ intimal macrophages and CD55+ fibroblast-like synoviocytes in RA synovium was observed. The interaction between CD97 and CD55 might therefore maintain and amplify synovial inflammation and be accountable for the specific architecture of the intimal lining layer. Next to CD97, intimal macrophages also abundantly express EMR2 (Kop et al., manuscript in preparation).

Increased expression of CD97 in the synovium is accompanied by detectable levels of soluble CD97 in the synovial fluid. The mechanism of CD97 release is not known but might include augmented matrix metalloproteinase activity in the synovium. Attempts to detect soluble CD97 in cerebrospinal fluid of MS patients or to correlate serum levels with CD97 expression on tumors failed.

Expression of CD97 in malignancies

Initiated by the detection of CD97 in tumor cell lines of different origin, Aust and coworkers investigated expression of CD97 and EMR2 in various epithelial tumors. CD97+ malignant cells were found in thyroid, colorectal, gastric, esophageal and pancreatic carcinomas. In the same tumor entities, little if any expression of EMR2 was detectable.

Investigation of CD97 in malignancies led to two interesting observations. Firstly, expression levels of CD97 are related to dedifferentiation and tumor stage as demonstrated in thyroid carcinomas. Secondly, different lines of evidence suggest a role of CD97 in tumor migration and invasiveness: (1) Expression levels of CD97 correlate with the in vitro migration and invasion capacity of colorectal tumor cell lines. (2) Migration and invasiveness of the fibrosarcoma cell line HT-1080 can be increased 2-3 fold by inducing expression of CD97. (3) Scattered tumor cells at the invasion front of colorectal and gastric carcinomas are stronger CD97+ than tumor cells in solid formations of the same tumor.

LIGAND BINDING AND FUNCTION

EGF-TM7 receptors bind cellular ligands via their EGF domains

The ability of EGF domains to interact with other protein modules suggested that EGF-TM7 receptors might be involved in cell-cell interactions. Indeed, cell adhesion studies have demonstrated that both lymphocytes and erythrocytes specifically bind to human CD97 transfectants. Generating a ligand-specific mAb, the ligand was identified
The EGF-TM7 family

as CD55/decay accelerating factor (DAF), a molecule that protects host cells from complement-mediated damage by accelerating the decay of C3/C5 convertases. Erythrocytes deficient in CD55, due to either paroxysmal nocturnal hemoglobinuria (PNH) or the Inab phenotype, are not capable of adhering to CD97-transfected cells. Characterization of mouse CD97 demonstrated the interaction with CD55 to be phylogenetically restricted.

A series of studies has verified the critical role of the EGF domains in CD55 binding (Leemans et al., submitted for publication). Antibody blocking, domain deletion and domain swapping experiments showed that the first two EGF domains of CD97 mediate the interaction. A third, tandemly arranged EGF domain is necessary for the structural integrity of the binding region. Interestingly, in human CD97 the presence of additional EGF domains reduces the binding affinity. Another remarkable observation is that EMR2 does not bind CD55. Although the first two EGF domains differ by only three amino acids from that of CD97, this small difference dramatically alters ligand specificity.

Many of the cell–cell interactions within the immune system are of low affinity and transient in nature. Surface plasmon resonance experiments confirmed this observation for the CD97-CD55 interaction. Contacts with human CD97 (EGF1,2,5) are of low affinity (86 μM) and have a rapid off-rate (at least 0.6 s⁻¹). Because of the expected low affinity, multivalent probes have been used in the search for ligands for the other EGF-TM7 receptors. Recombinant soluble protein of the extracellular part of EGF-TM7 receptors containing a biotinylation sequence was biotinylated in vitro and coupled to avidin-coated fluorescent beads. Multivalency increases the avidity of the molecular association allowing detection by flow cytometry and immunohistology. Use of this approach has resulted in localization of an EMR3 ligand at the surface of monocyte-derived macrophages and activated human neutrophils. A ligand for mouse EMR4 was detected on the B lymphoma cell line A20. Recently, we identified chondroitin sulfate as ligand for the largest isoform of EMR2 and CD97 (Kwakkenbos et al., manuscript in preparation). Chondroitin sulfate is a glycosaminoglycan side chain abundantly found as component of cell surfaces proteoglycans and in extracellular matrixes. Binding to chondroitin sulfate is mediated by the fourth EGF domain, whose amino acid sequence is identical in CD97 and EMR2 and which is present only in the largest isoform of both molecules.

Differences in ligand specificity of CD97 and EMR2 isoforms raised the question, whether alternative RNA splicing is regulated under physiological or pathological conditions. A first study addressing this problem failed to detect substantial variation in the ratio of human CD97 isoforms expressed by several cell types.

A role for EGF-TM7 receptors in cell migration

The almost exclusive expression of most EGF-TM7 receptors on leukocytes strongly points to a role in immunity. The unusual structure of the α subunit and the ability of this region to bind cellular ligands suggest that this function might relate to cell adhesion and migration. Evidence for this assumption was provided when migration and invasiveness of tumor cells was found to correlate with CD97 expression (see above). In a recent investigation, we used a panel of newly generated mAbs to the EGF domain region of mouse CD97 to evaluate the function of CD97 in leukocyte migration in vivo (Leemans et al., submitted for publication). Because of their fast migration and
Chapter 2

their important role in the eradication of pathogens we choose neutrophils as a model system. In dextran sulphate sodium (DSS)-induced experimental colitis we showed that homing of adoptively transferred neutrophils to the colon was significantly delayed when cells were preincubated with CD97 mAb. The consequences of this defect in neutrophil migration for host defense became evident in a murine model of *Streptococcus pneumoniae*-induced pneumonia. Mice treated with CD97 mAb displayed a reduced granulocytic inflammatory infiltrate that was associated with a significantly enhanced outgrowth of bacteria in the lungs and a strongly diminished survival.

Different from infections, leukocyte recruitment is detrimental in autoimmune diseases like rheumatoid arthritis. We recently found that the severity of collagen-induced arthritis (CIA), a model of RA, was diminished in mice treated with CD97 mAb (Kop et al., manuscript in preparation). Together, these findings indicate an essential role for CD97 in leukocyte migration.

*In vivo* data on other EGF-TM7 receptors are scarce and limited to mouse EMR1. It has been reported that the release of inflammatory cytokines from spleen cell cultures exposed to heat-killed *Listeria monocytogenes* was inhibited by the mAb F4/80. More recent investigation of EMR1-deficient mice however failed to detect defects in the defense of *L. monocytogenes* infection. Development of resident macrophage populations in EMR1-deficient mice was found to be normal.

Do EGF-TM7 receptors signal through G proteins?
Certainly the most important limitation in our understanding of EGF-TM7 receptors is the absence of data elucidating signal transduction. Several approaches to demonstrate signalling of EGF-TM7 receptors by either overexpression in reporter cell lines, stimulation with purified ligand molecules like CD55 or site-directed introduction of mutations that could give rise to constitutive activation in class B TM7 peptide hormone receptors have failed as yet (Garritsen, van Puijenbroek, van Elsal, Covineau and Laburthe, personal communication). This failure might be explained in different ways. One is EGF-TM7 receptors might engage G-protein-independent signalling pathways. Through recent years, a growing number of nonclassical signalling cascades engaged by TM7 receptors have been reported. Secondly, EGF-TM7 receptors might facilitate cell-cell or cell-extracellular matrix interactions without engaging signalling cascades. Although possible, this option is unlikely as the TM7 regions resemble that of class B TM7 peptide hormone receptors, which are known to engage Gα.

**CONCLUDING REMARKS**

With completion of the human and mouse genome project, the search for new human EGF-TM7 receptors is finished. Draft sequences of other vertebrate genomes are forthcoming, which will hopefully shed more light on the evolution of the EGF-TM7 family.

Structurally, EGF-TM7 receptors belong to the most complex TM7 molecules. Intracellular processing and transcriptional regulation give rise to heterodimeric surface expression and variable numbers of EGF domains. We are just beginning to understand the implications of these characteristics.
EGF-TM7 receptors are the only TM7 molecules hitherto known to bind cellular ligands. Several lines of evidence suggest a role in leukocyte migration. Better understanding of the physiological function and the underlying molecular mechanisms will depend on the generation and in-depth analysis of genetically modified mice. The relatively smaller size of the EGF-TM7 family in mice might thereby be an advantage as less molecular redundancy could be expected.

EGF-TM7 receptors are chimeric molecules consisting of an extracellular α subunit with the ability to bind cellular ligands and a TM7/cytoplasmic β subunit with the potential to induce intracellular signal cascades. Linking the action of both subunits at a molecular level will be the final challenge in EGF-TM7 biology.

ACKNOWLEDGEMENTS

We thank René A.W. van Lier for his comments and suggestions, Lisa A. Gilhuijs-Pederson for help with the phylogenetic analysis and Laurie Gordon for the information about the locus position of the EGF-TM7 genes. This work was supported by grants from the Netherlands Organization for Scientific Research (NWO, 901-07-208), the Landsteiner Foundation for Bloodtransfusion Research (LSBR, 0109), the Dutch Arthritis Association (RF, 99-2-14), the Wellcome Trust and the British Heart Foundation (PG/02/144). J.H. is a fellow of the Royal Netherlands Academy of Arts and Sciences.

REFERENCE LIST

(9) Segre GV, Goldring SR. Receptors for secrerin, calcitonin, parathyroid hormone (PTH)/PTH-related peptide, vasoactive intestinal peptide, glucagonlike peptide 1, growth hormone-releasing hormone, and glucagon belong to a newly discovered G-protein-linked receptor family. Trends Endocrinol Metab. 1993;4:309-314.
Chapter 2


(14) Hadjantonakis AK, Formstone CJ, Little PF. mCelsrl is an evolutionarily conserved seven-pass transmembrane receptor and is expressed during mouse embryonic development. Mech Dev. 1998;78:91-95.


(25) Nechiporuk T, Uarness 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.


The EGF-TM7 family

(38) Mengerink KJ, Moy GW, Vacquier VD. suREJ3, a polycystin-1 protein, is cleaved at the GPS domain and localizes to the acrosomal region of sea urchin sperm. J Biol Chem. 2002;277:943-948.
Chapter 2


