Characterization of human cd200 glycoprotein receptor gene located on chromosome 3q12-13

Jose María Vieitesa, Raul de la Torre a, María Angeles Ortega a, Trinidad Monteroa, Jose María Pecob, Antonio Sánchez-Pozo a, Angel Gil a, Antonio Suárez a, *

a Department of Biochemistry and Molecular Biology, University of Granada, 18071, Granada, Spain
b Department of Legal Medicine, Toxicology and Psychiatry, University of Granada, 18071, Granada, Spain

Received 7 February 2003; received in revised form 17 March 2003; accepted 25 March 2003
Received by G. Pesole

Abstract

An immunomodulatory membrane protein, CD200R displays an expression pattern restricted to myeloid cells in mice. It is the receptor for a ligand, CD200, expressed by a broad range of cell types. In this study, we describe the cloning and characterization of the human homologue of the CD200R gene. This gene maps closely to the CD200 gene on human chromosome 3q12-13. The human CD200R gene spans a region of 52 kb, consists of nine exons, and encodes a 348-amino-acid cell-surface protein consisting of two IgFF domains in a typical V/C2 arrangement. The 59-amino-acid cytoplasmic domain has two tyrosine residues, one of which is contained within a NPXY motif. In common with other IgSF genes, the CD200R gene can generate different protein isoforms through alternative splicing. An alternative spliceout form, which has not yet been described in mice, encodes a 188-amino-acid truncated soluble polypeptide containing only the V immunoglobulin domain. In contrast to murine CD200R protein, the human membrane-bound and soluble CD200R proteins have an insertion of 23 amino acids at position 23, encoded by exon 2, which generates a putative dihydroxyacid dehydratase domain. The splicing of exon 2 generates two new isoforms, encoding the membrane and soluble proteins but lacking the dihydroxyacid dehydratase domain. Northern-blot analysis shows that both membrane-bound and soluble isoforms are expressed in the thymus, liver, spleen and placenta. By RT-PCR, we have analyzed the expression of the four transcript variants in human placenta, spleen, liver, brain and kidney.

Keywords: Gene structure; Expression; Ig superfamily; Alternative splicing

1. Introduction

The immune system has evolved control mechanisms to differentiate between pathogens and self-antigens that depend on secreted factors and on the intimate communication of proteins on opposing cells. Many proteins that mediate this control in immune cells are proteins with IgSF domains. One member of the IgSF, CD200, is expressed on the surface of thymocytes, activated T cells, B cells, follicular neurons, kidney glomeruli, tonsil follicles, the syncytiotrophoblast and endothelial cells in rats (Wright et al., 2000). CD200 interacts with another protein (known as CD200 receptor or CD200R) expressed only in rat cells of the myeloid lineage (Wright et al., 2000).

The distribution and the molecular nature of the CD200/CD200R proteins suggest that they might be involved in the tissue-specific regulation of myeloid functions. Indeed, the phenotype of a CD200-deficient mouse has revealed defects in myeloid cellular biology and brain microglia appear to be more numerous and in a more activated state (Hoek et al., 2000). The CD200 interaction may be a potent target for immunomodulation since it has functional effects in limiting spontaneous abortion in an

Abbreviations: bp, base pair(s); cDNA, DNA complementary to RNA; DNase, deoxyribonuclease; dNTP, deoxyribonucleotide triphosphate; EST, expressed sequence tag; Ig, immunoglobulin(s); hCD200R, human CD200R protein; IgFF, immunoglobulin fold family; IgSF, immunoglobulin superfamily; kb, kilobase; kDa, kiloDalton(s); mCD200R, mouse CD200R protein; mRNA, messenger RNA; oligo, oligodeoxyribonucleotide; PCR, polymerase chain reaction; rCD200R, rat CD200R protein; RNase, ribonuclease; RT-PCR, reverse transcription-polymerase chain reaction; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na3 citrate pH 7.6; u, unit(s); Tm, annealing temperature.

* Corresponding author. Tel.: +34-95-824-3838; fax: +34-95-824-8960.
E-mail address: asuarez@ugr.es (A. Suárez).
animal model (Clark et al., 2001), prolonging graft survival (Gorczynski et al., 1999) and, in a model of tumor rejection, can reverse the protective effects of prior immunization with tumor cells expressing CD80 (Gorczynski et al., 2001).

Although the initial annotation of the Human Genome identified approximately 30,000 protein-coding genes, recent studies suggest that the number of human protein-coding genes may be higher (Katsanis et al., 2001; Das et al., 2001). While the human CD200 gene was mapped to chromosome 3 at position 3q12-13, the human CD200R gene was not annotated (McPherson et al., 2001). In this paper, we have used a bioinformatics approach combined with conventional molecular biology to characterize and map the human CD200R gene to chromosomal region 3q12-13.

2. Materials and methods

2.1. Bioinformatics analysis

Protein and nucleotide queries for CD200 glycoprotein receptor of rat and mouse were used for searches for similarities and translations of all reading frames (BLASTN and tBLASTN searches) in Human Genome draft sequence (http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&f&ORG=Hs). Exon predictions from genomic sequences were obtained with GeneMachine software (Makalowska et al., 2001). Sequence data were analyzed using MacMolly Tetra (version 3.8, Soft Gene GmbH). Nucleotide-sequence similarity searches were performed in EST database (http://www.ncbi.nlm.nih.gov/BLAST). Only ESTs with scores higher than 200 (E values lower than 6e-65) were considered ESTs for putative cd200r isoforms. Protein and nucleotide queries for CD200 glycoprotein receptor of rat and mouse were used for searches for similarities and translations of all reading frames (BLASTN and tBLASTN searches) in Human Genome draft sequence (http://www.ncbi.nlm.nih.gov/genome/seq/page.cgi?F=HsBlast.html&f&ORG=Hs). Exon predictions from genomic sequences were obtained with GeneMachine software (Makalowska et al., 2001). Sequence data were analyzed using MacMolly Tetra (version 3.8, Soft Gene GmbH). Nucleotide-sequence similarity searches were performed in EST database (http://www.ncbi.nlm.nih.gov/BLAST). Only ESTs with scores higher than 200 (E values lower than 6e-65) were considered ESTs for putative cd200r isoforms. Multiple sequence alignments were calculated by the ClustalW software package (Thompson et al., 1994). Protein sequences were analyzed and conserved domains detected with PFAM, PROSITE, SMART, BLOCKS, PRODOM, PREDICT, SIGFIND, NetOGlyc and TMHMM (available at http://www.motif.genome.ad.jp, http://www.expasy.ch/prosite, http://www.cbs.dtu.dk/services and http://www.smart.embl-heidelberg.de/).

2.2. Northern blotting

A human multiple-tissue Northern blot (Beckton–Dickinson) containing purified poly(A) mRNA from various tissues was hybridized with a cd200r-specific cDNA probe. The probe was labeled with (α-32P)dCTP using the Rediprime™ II random primer labeling system (Amersham Pharmacia Biotech) and purified on Sephadex G-50 micronuclear columns. The membrane was hybridized with the 32P-labelled probe in NorthernMax prehyb/hyb buffer solution (Ambion) at 42 °C overnight. Afterwards, the membrane was washed with 2 × SSC and 0.1% SDS at 65 °C for 15 min and once again with 0.2 × SSC and 0.1% SDS at 65 °C for 15 min. The membrane was exposed overnight to a Electronic Autoradiography Instant Imager (Perkin–Elmer). After cd200r gene-expression analysis, the membrane was stripped following the manufacturer’s instructions and hybridized with a human Glyceraldehyde-3-phosphate dehydrogenase fragment used as the control probe. Hybridization, washing and exposure conditions were performed as above.

2.3. RT-PCR and cDNA cloning

Human tissue samples were obtained from the School of Medicine, University of Granada. Total RNA was purified according to Chomczynski and Sacchi (Chomczynski and Sacchi, 1987). To exclude possible amplification of contaminating genomic DNA, the DNA was digested with 10 μl of RNase-free DNase for 30 min at 37 °C. cDNA was synthesized with 1 μg of total DNase-treated RNA and 0.5 μg of oligo-dT16 or 0.1 μg of a specific oligo in 20 μl of a solution containing 1× first-strand buffer, 10 mmol/l dithiothreitol, 500 mmol/l each dNTP, 40 μl of RNase inhibitor and 200 μl of Superscript II RNase H Reverse Transcriptase (Invitrogen). All reverse-transcription reactions were performed in duplicate with equal amounts of total RNA. PCR reactions were performed with an Applied Biosystems DNA thermal cycler model 9700 as follows: an initial denaturing for 5 min at 94 °C; then cycles of 94 °C for 15 s, the specific annealing temperature for 15 s and 72 °C for a time depending on fragment length (1 kb/1 min); final extension at 72 °C for 5 min. The primer sequences and annealing temperatures are listed in Table 1. The PCR products were fractionated on 2% agarose gels and visualized after ethidium bromide staining. The PCR fragments were purified from agarose gels using the Qiaex II gel-extraction kit (Qiagen) and cloned into pGEM-T vector (Promega). Sequences were analyzed in a 377 ABI autosequencer (Applied Biosystems) After assembly, the sequences were deposited in GenBank under accession numbers AF495380, AF497548, AF497549, and AF497550.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Primer namea</th>
<th>Tm (°C)b</th>
<th>Sequence (5’-3’)c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F1</td>
<td>58</td>
<td>TGGACCTCAAGTGGAAAAG</td>
</tr>
<tr>
<td>1</td>
<td>F1a</td>
<td>58</td>
<td>TGGACCTCAAGTGGAAAAG</td>
</tr>
<tr>
<td>4</td>
<td>F4</td>
<td>60</td>
<td>GTTACAGCTGAGTTGATGTTG</td>
</tr>
<tr>
<td>4–6</td>
<td>R4–6</td>
<td>58</td>
<td>TCCAGGCTGAGTACTCACCTTG</td>
</tr>
<tr>
<td>5</td>
<td>R5</td>
<td>58</td>
<td>TGGAGACCACACATGAGTTG</td>
</tr>
<tr>
<td>9</td>
<td>R9</td>
<td>60</td>
<td>TTACAATAATTAGTCTCCTG</td>
</tr>
<tr>
<td>9a</td>
<td>R9a</td>
<td>58</td>
<td>TTACAATAATTAGTCTCCTG</td>
</tr>
</tbody>
</table>

a F refers to sense orientation; R indicates antisense orientation.
b Tm refers to the annealing temperature used in RT-PCR analysis.
c Indicates the specific splice-junction between exons 4 and 6, skipping exon 5.
3. Results and discussion

3.1. Identification and structure of CD200R

Although the human CD200 gene was mapped to chromosome 3 at position 3q12-13, the human CD200R gene was not initially located (McPherson et al., 2001). We used a bioinformatics approach combined with conventional molecular biology to characterize and map the human CD200R gene to chromosomal region 3q13. With the use of BLAST algorithms, the nucleotide and amino acid sequence of rat and mouse CD200R (GenBank accession no. NM_023953 and NM_021325) were compared with the Human Genome draft sequence. The results showed that two different regions within human chromosome 3 contained sequences with different degrees of homology to rat

---

**Fig. 1.** Genomic organization, splicing pattern and domain structure of CD200R. (A) Genomic structure of the human CD200R gene. Exons are depicted on scale as black boxes and introns as thin horizontal lines. The names and relative positions of primers used for RT-PCR experiments are indicated by arrows. (B) The mRNA variants produced by the human CD200R gene. The transcribed genomic 125 nt sequence between exons 4 and 5 is indicated as a gray box. (C) Domain structure of CD200R1 and CD200R2 and the soluble proteins, sCD200R1 and sCD200R2.
and mouse CD200R genes at both nucleotide and amino acid levels. In addition, both genomic regions were independently scanned for identification of predicted exons and genes with GeneMachine software (Makalowska et al., 2001) which indicated that exon 4 (Fig. 1A) was a highly probable exon.

To characterize this new human CD200R gene, we conducted PCR-homology cloning using a set of primers (F4 and R9, Table 1) designed on the basis of the gene sequence predicted by the BLAST and GeneMachine analyses. The amplification was performed as a touchdown PCR, reducing the annealing temperature by 1 °C until reaching the touchdown temperature of 60 °C. A set of negative controls for all the components of the PCR reaction except for cDNA, replaced by placenta RNA, or containing just one primer were used to test for the presence of nonspecific amplified products and to exclude the amplification of CD200R pseudogene. Screening of several human placenta samples resulted in the isolation of a single 760-bp fragment. Its sequence shared total identity with five genomic DNA sequences from positions 117355 to 117430 kb of human chromosome 3 (McPherson et al., 2001).

A BLAST search of the non-redundant (nr) and expressed sequence tag (EST) databases using the 760-bp sequence predicted by the BLAST and GeneMachine software and confirmed by RT-PCR (primers F4 and R9a, Table 1).

3.2. Analysis of putative gene products

The full-length open-reading frame (accession no. AF495380) contains 1046 nucleotides (Fig. 1B, isoform 1) and encodes 348 amino acid residues (Fig. 1C, CD200R1).

The human CD200R gene spans a region of 52 kb and consists of nine exons. All splice sites agree with the GT/AG consensus sequence (Mount, 1982).

The alignment of the deduced amino acid sequence with those of rCD200R and mCD200R shows a sequence identity of 53 and 52%, respectively. This result suggests that this protein is related to rCD200R and mCD200R; therefore, we named it hCD200R. This hCD200R contains a signal peptide cleavable between positions 24–25 (SIGFIND v.2.1). There are 9 potential N-linked and one potential O-linked glycosylation sites in hCD200R (NetOGlyc v.2.0). Hydrophobicity studies predict that hCD200R is a plasma membrane protein with a high probability of having one transmembrane domain between amino acids 267–289 (TMMHM v.2.0). These predicted molecular features are very similar to those of rCD200R (Wright et al., 2000). A long N- (amino acids 25–266) and a short C-terminal tail (amino acids 290–348) are assumed to be oriented to the outer and inner faces of cell membrane, respectively. The 59-amino-acid cytoplasmic domain of hCD200R contains two tyrosine residues, one of which is contained within a NPXY motif as in rCD200R. The rCD200R intracellular domain is phosphorylated when rat macrophages are treated with pervanadate, a tyrosine phosphatase inhibitor (Wright et al., 2000). This feature suggests that engagement of CD200 may lead to intracellular signals affecting macrophage function via tyrosine phosphorylation in human CD200R.

Structurally, hCD200R contains two immunoglobulin fold-family (IgFF) domains in a typical V/C2 set arrangement with the conserved cysteine, tyrosine and threonine residues which define the canonical B to F strand IgFF organization (Halaby et al., 1999). Many IgSF proteins have both membrane-bound and soluble forms, created either by cleavage from the membrane structure at some stage, as seen in CD48 molecule (Smith et al., 1997), or by deletion of domains during alternative splicing in mRNA processing as with CD150w (Cocks et al., 1995). A set of primers was designed to screen for the presence of splice variants (Table 1).

It is noteworthy that a cDNA fragment contained a 125-bp insertion (accession no. AF497549) that corresponds to the genomic sequence between exons 4 and 6, and results in a frameshift leading to a premature stop codon and the production of a truncated soluble form of hCD200R that comprizes the V domain of the molecule (Fig. 1B, isoforms 3 and 4). This is a unique feature of human CD200R gene expression since the soluble variant has not yet been described in mice, and may have important consequences for the biological activity of this immunomodulatory molecule.

The alignment of CD200R primary sequences reveals an insertion of 23 amino acids at position 23, encoded by exon 2, which are not present in mice CD200R (Fig. 2). This exon is alternatively spliced in the cDNAs (Fig. 1B, isoforms 2 and 4) that encode the membrane-bound and soluble proteins (accession no. AF497548 and AF497550). The presence of these amino acids generates a putative dihydroxyacid dehydratase domain comprised between positions 21 to 98 in hCD200R (PROSITE) (Fig. 1C, CD200R1 and sCD200R1). Dihydroxyacid dehydratase (DH) belongs to the hydroxylase class of enzymes, contains a catalytically-active prosthetic [4Fe-4S] group and participates in the mitochondrial biosynthesis of valine and isoleucine (Flint, 1996). Several lines of structural evidence imply that hCD200R is not a DH enzyme. Firstly, DH proteins are soluble enzymes consisting of a single polypeptide chain of 400–600 amino acids. Secondly, the amino acid identity of this putative hCD200R domain to the consensus DH sequence is partial and very low (18.75%), and the critical cysteine residues necessary to form [4Fe-4S] cluster are not present. Further investigation will help to explore the implication of this domain on the binding activity of CD200 proteins.
3.3. mRNA-expression patterns

High stringency Northern blot analysis of CD200R transcripts reveals that two major transcripts of approximately 1.4 and 2.4 kb are expressed in the thymus, spleen, liver and placenta (Fig. 3A). The size of the transcripts confirmed that the CD200R gene is transcribed in two isoforms of different sizes. There were no detectable CD200R transcripts in the brain, heart, muscle, colon, kidney, small intestine, lung or leukocytes.

The same set of primers generated to study splice variants (Table 1) were used in RT-PCR reactions with several human-tissue cDNA samples in order to confirm the expression of mRNA transcripts. These primers generate DNA products that extend from exon 1 to either exon 4–6 (Fig. 4A) or to exon 5 (Fig. 4B). In each variant, we could identify two fragments corresponding to the inclusion or skipping of exon 2. The mRNA transcript resulting from alternative splicing of exon 2 in CD200R would be 69 bp shorter than the longer transcript. Using this approach, we were able to detect the expression of the four transcript variants in human placenta, spleen, liver, brain and kidney.

As RT-PCR is a far more sensitive assay to detect mRNA transcripts than is Northern blot hybridization, these results could indicate that CD200R may be expressed at very low levels in some tissues, below the detection limit of RNA dot-blot analysis. Alternatively, it may be highly expressed only in tissues during physiological or pathological states that should be studied in order to understand fully how this gene functions. Further assays are needed to provide a
A complete explanation of the possible role of those variants or their ratio in different tissues during physiological or pathological states.

3.4. Conclusions

1. We have characterized the genomic structure of the novel human gene \textit{CD200R} on chromosome 3q12-13.
2. The gene spans a region of 52 kb, comprises nine exons and can generate four different alternatively spliced transcripts.
3. mRNA transcripts encode for a cell surface receptor and for a soluble protein (due to the insertion of exon 5 which creates a frameshift leading to a premature stop codon).
4. The alternative splicing of exon 2 in both transcripts produces the deletion of 23-amino-acids in the receptor and soluble proteins.
5. Northern blot and RT-PCR analyzes confirm the expression of the \textit{CD200R} gene in human placenta, spleen, liver, brain and kidney.

Acknowledgements

This work was supported by research projects (ref. CICYT PM97-0171 and 1FD97-0410) funded by Spanish Science and Technology Ministry. We thank Dr. Enrique García-Olivares for providing human samples.

References

Flint, D.H., 1996. Escherichia coli contains a protein that is homologous in function and N-terminal sequence to the protein encoded by the \textit{niS} gene of Azotobacter vinelandii and that can participate in the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase. J. Biol. Chem. 271, 16068–16074.