Developmental and Comparative Immunology 33 (2009) 747–760

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Identification of a porcine DC-SIGN-related C-type lectin, porcine CLEC4G (LSECtin), and its order of intron removal during splicing: Comparative genomic analyses of the cluster of genes CD23/CLEC4G/DC-SIGN among mammalian species☆

Y.W. Huang, X.J. Meng *

Center for Molecular Medicine and Infectious Diseases, Department of Biomedical Sciences and Pathobiology, College of Veterinary Medicine, Virginia Polytechnic Institute and State University, 1410 Price's Fork Road, Blacksburg, VA 24061-0342, USA

ARTICLE INFO

Article history:
Received 20 October 2008
Received in revised form 16 December 2008
Accepted 23 December 2008
Available online 21 January 2009

Keywords:
C-type lectins
CLEC4G (LSECtin)
CD23
DC-SIGN (CD209)
L-SIGN (CD209L)
Pig
MicroRNA
Pre-mRNA splicing
Mammals

ABSTRACT

Human CLEC4G (previously named LSECtin), DC-SIGN, and L-SIGN are three important C-type lectins capable of mediating viral and bacterial pathogen recognitions. These three genes, together with CD23, form a lectin gene cluster at chromosome 19p13.3. In this study, we have experimentally identified the cDNA and the gene encoding porcine CLEC4G (pCLEC4G). Full-length pCLEC4G cDNA encodes a type II transmembrane protein of 290 amino acids. pCLEC4G gene has the same gene structure as the human and the predicted bovine, canis, mouse and rat CLEC4G genes with nine exons. A multi-species-conserved site at the extreme 3′-untranslated region of CLEC4G mRNAs was predicted to be targeted by microRNA miR-350 in domesticated animals and by miR-145 in primates, respectively. We detected pCLEC4G mRNA expression in liver, lymph node and spleen tissues. We also identified a series of sequential intermediate products of pCLEC4G pre-mRNA during splicing from pig liver. The previously unidentified porcine CD23 cDNA containing the complete coding region was subsequently cloned and found to express in spleen, thymus and lymph node. Furthermore, we compared the chromosomal regions syntenic to the human cluster of genes CD23/CLEC4G/DC-SIGN/L-SIGN in representative mammalian species including primates, domesticated animal, rodents and opossum. The L-SIGN homologues do not exist in non-primates mammals. The evolutionary processes of the gene cluster, from marsupials to primates, were proposed based upon their genomic structures and phylogenetic relationships.

1. Introduction

The C-type lectin receptor (CLR) family includes a large number of proteins that perform protein–carbohydrate interactions by binding to the polysaccharide chains present on the glycoprotein ligands in a calcium-dependent manner [1]. Numerous CLRs belong to pattern recognition receptors (PRRs) on the surface of antigen-presenting cells (APCs) that recognize foreign pathogens, and play key roles in host immune responses [1,2]. The type II CLRs are classified as their cytoplasmic tail (CT) located in the NH_2 terminus domain. Other type II CLRs domains include the transmembrane domain (TMD) following the CT, a single carbohydrate recognition domain (CRD) exposed extracellularly at the carboxyl terminus, and the neck domain between the TMD and CRD [1,2].

A human gene cluster of type II CLRs, CD23/CLEC4G/DC-SIGN/L-SIGN, which is localized in human chromosome 19p13.3, has received increasing interest during the past few years. Human CD23 (FCER2) is a low affinity IgE receptor that plays an important role in cell–cell adhesions, B cells survival and antigen presentation [3]. Human DC-SIGN (hDC-SIGN, CD209) was initially identified as an ICAM-3 binding protein mediating dendritic cells (DCs) and T cell interaction, an ICAM-2 binding protein regulating chemokine-induced trafficking of DCs across both resting and activated endothelium, and a HIV-1 gp120 receptor mediating transmission of HIV-1 to susceptible cells *in trans* [4–6]. The second hDC-SIGN homologue, hL-SIGN (CD209L), was subsequently shown to have

☆ The nucleotide sequences of the full-length cDNA and the gene of porcine CLEC4G and the cDNA of porcine CD23 reported in this paper have been deposited in the GenBank database with accession numbers EU814900, EU814899 and FJ545265.

* Corresponding author. Tel.: +1 540 231 6912; fax: +1 540 231 3426.
E-mail address: xjmeng@vt.edu (X.J. Meng).

0145-305X/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2008.12.007

similar but subtly distinct property of pathogen recognition to hDC-SIGN [7]. Both hDC-SIGN and hL-SIGN bind to asparagines-linked high-mannose glycans present on a broad spectrum of enveloped viruses [8]. Human CLEC4G (hCLEC4G, previously named LSECtin), co-expressed with hL-SIGN on liver and lymph node sinusoidal endothelial cell (LSECs), is the third hDC-SIGN-related CLR recently identified [9,10]. hCLEC4G was found to interact with the surface glycoproteins of Ebola virus and severe acute respiratory syndrome coronavirus (SARS-CoV) [11]. However, unlike hDC-SIGN and hL-SIGN, the hCLEC4G selectively binds to glycoproteins terminating in a disaccharide, GlcNAcβ1-2Man, resulting in its interacting with truncated glycans on glycoprotein of Ebola virus [12]. Since these four CLRs form a tight gene cluster and share overall protein domain structure, similar genomic organization and possible analogous function, it has been proposed that they are derived from a common ancestor [9].

DC-SIGN homologues have been experimentally identified from other mammalian species including non-human primates and mouse [13–15]. Interestingly, there exist seven mouse DC-SIGN paralogues, SIGNRs 1–5 as well as SIGNRs 7–8, on mouse chromosome 8A1.1, indicating widely divergent biochemical and probably physiological properties of DC-SIGN-related proteins in mouse [14]. However, none of these SIGNR molecules were shown to be the functional orthologue to hDC-SIGN. Recently, DC-SIGN homologues from domesticated animal species such as dog, cattle and horse have also been predicted from the completed genome projects. We recently reported the molecular cloning and characterization of the full-length cDNA and gene of porcine DC-SIGN (pDC-SIGN) in the absence of computer-based screening of DC-SIGN homologues in pig genome database [16]. Phylogenetic analysis revealed that pDC-SIGN, together with the putative bovine, canis and equine DC-SIGN, are more closely related to mouse SIGNR7 and SIGNR8 than to human DC-SIGN or other mouse SIGNR homologues and form a separate clade, indicating a distinct evolutionary pathway [16]. Since no other DC-SIGN sequences were detected in the bovine and canis genomes, the pDC-SIGN likely exists as a single gene analogously, although the relevant porcine genomic region has not been identified. Moreover, we observed that L-SIGN homologues only exist in human and non-human primates but not in other non-primate mammalian species [16]. Indeed, a previous study showed that the current L-SIGN gene, presented in apes such as chimpanzee and human but not in Old World monkeys (OWM) such as rhesus macaque, was newly duplicated from the ancestral DC-SIGN, whereas the older duplicator, CD209L2, was lost in human but still retained in OWM and apes [13]. The study indicated that the DC-SIGN/L-SIGN gene family in primates has undergone duplications and deletions during recent evolutionary processes [13].

Besides hCLEC4G, CLEC4G homologues in other mammalian species have not been experimentally identified although the gene information could be searched using computer programs from the genome databases. We report here the cloning and phylogenetic analysis of pCLEC4G cDNA and gene, its order of intron removal during pre-mRNA splicing, and the tissue distribution. In addition, the complete coding region of porcine CD23 (pCD23) cDNA was also determined. Furthermore, with the available information of CD23, DC-SIGN and CLEC4G genes obtained from the completed genome projects, we compare the chromosomal regions syntenic to the cluster of human genes CD23/CLEC4G/DC-SIGN/L-SIGN in the representative mammalian species including chimpanzee (*Pan troglodytes*), rhesus macaque (*Macaca mulatta*), cattle (*Bos taurus*), dog (*Canis lupus*), horse (*Equus caballus*), sheep (*Ovis aries*), pig (*Sus scrofa*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), and opossum (*Monodelphis domestica*). Finally, the evolutionary processes of the gene cluster, from marsupials to primates, were proposed based upon their genomic structures and phylogenetic relationships.

2. Materials and methods

2.1. RNA extraction and reverse transcription PCR (RT-PCR)

Healthy crossbred conventional pigs of 7 weeks of age were used for the collection of tissue samples. Pigs were maintained in an isolated room under experimental conditions. Total RNA was isolated from homogenized pig liver using the RNeasy mini kit (Qiagen Inc.) followed by an RNase-free DNase I treatment. First-strand cDNA was synthesized from total RNA with SuperScript II reverse transcriptase (Invitrogen) using oligo-dT (Promega) as the reverse primer. A pair of gene-specific primers, PLST-F (5′-TATGCCCAGAGCAGGGCACC-3′) and PLST-R (5′-GGGCTAGGTCAG-CAGTTGTGC-3′), was designed for the amplification of the complete coding region of pCLEC4G cDNA according to a porcine cDNA sequence with the GenBank accession number AK232603. PCR was performed in 50 μL reaction with an Advantage 2 PCR kit (Clontech, Palo Alto, CA) using the following PCR parameters: 94 °C for 2 min, 30 cycles of 94 °C for 15 s, 60.0 °C for 30 s and 72 °C for 90 s, and a final incubation at 72 °C for 3 min. For the amplification of the complete coding region of pCD23 cDNA, pig spleen tissue as well as primers PCD23-F (5′-GCGCTCCCATGGAGGAAAGTTTA-TACTC-3′) and PCD23-R (5′-TGAACAGATGCTCAGCAAGTGGCCA-3′) was used. The primer sequences were based on the *S. scrofa* chromosome 2 clone CH242-334A8 (working draft sequence; GenBank accession number CU929919) that was released most recently on September 24, 2008. The obtained PCR products were individually excised, purified, and subsequently cloned into a pCR2.1 vector (Invitrogen) by TA cloning strategy followed by DNA sequencing.

2.2. Genomic PCR and gene sequencing

The same primers PLST-F and PLST-R were used for one-step genomic PCR, which was performed with a Platinum PCR HiFi Supermix kit (Invitrogen) using 150 ng of the pig genomic DNA (Novagen) in a total volume of 50 μL. The PCR condition was 35 cycles of 94 °C for 30 s, 68 °C for 4 min with an initial denaturation step at 94 °C for 2 min. The resulting fragment was cloned into a pCR2.1 vector by the TA cloning strategy. The M13 forward and reverse primers together with a gene-specific primer PLST-E3F (5′-CAGGATCTACTGAGGACAAACG-3′) were used for DNA sequencing.

2.3. Tissue distribution of porcine CLEC4G and CD23 gene expression detected by RT-PCR

Total RNA was isolated from ten homogenized pig tissues including spleen, duodenum, thymus, kidney, lung, lymph node, heart, bone marrow, liver and muscles using the RNeasy mini kit (Qiagen) followed by an RNase-free DNase I treatment, and cDNA was synthesized with SuperScript II reverse transcriptase (Invitrogen) using oligo-dT (Promega) as the reverse primer. To avoid the contamination of genomic DNA, PCR was performed in 50 μL reactions with Clontech's Advantage 2 PCR kit using primer PLST-E67F (5′-GAGAGTCCGGTTCCAGAACAGCTCCT-3′) spanning the boundary between exons 6 and 7, and primer PLST-E89R (5′-TCC-CCCAGATTCCAGTGGCTGAAG-3′) spanning the boundary of exons 8 and 9 of pCLEC4G gene sequence that had been determined by genomic sequencing. For the detection of pCD23 mRNA expression, primers PCD23-E89F (5′-CTACACGAGTCCAACGGCTCCGTG-3′) and PCD23-E1011R (5′-CGGGCTGCCAGTTGCTATAGTCCAG-3′) were used. The PCR parameters include 30 cycles of 95 °C for 20 s, 68 °C for 1 min with an initial denaturation step for 2 min. The house keeping gene, porcine glyceraldehyde 3-phosphate dehydrogenase (GAPDH), was also amplified using primers GAPDH5 (5′-GCTGAG-TATGTCGTGGAGTC-3′) and GAPDH3 (5′-CTTCTGGGTGGCAGTGAT-