C-type Lectin CD209L/L-SIGN and CD209/DC-SIGN
Subjects: Molecular Biology
Submitted by: Nader Rahimi
(This entry belongs to Entry Collection "COVID-19")

Definition

COVID-19 pandemic continues to pose a serious threat to global public health with overwhelming worldwide socio-economic disruption. SARS-CoV-2, the viral agent of COVID-19, uses its surface glycoprotein Spike (S) for host cell attachment and entry. The emerging picture of pathogenesis of SARS-CoV-2 demonstrates that S protein, in addition, to ACE2, interacts with the carbohydrate recognition domain (CRD) of C-type lectin receptors, CD209L and CD209. Recognition of CD209L and CD209 which are widely expressed in SARS-CoV-2 target organs can facilitate entry and transmission leading to dysregulation of the host immune response and other major organs including, cardiovascular system. Establishing a comprehensive map of the SARS-CoV-2 interaction with CD209 family proteins, and their roles in transmission and pathogenesis can provide new insights into host-pathogen interaction with implications in therapies and vaccine development.

1. Introduction
Lectins are a diverse family of carbohydrate recognizing proteins that possess carbohydrate-recognition domain (CRD) or sulfated glycosaminoglycan (SGAG)-binding motif. Derived from the Latin word "lege"r", meaning "to select". Lectins were originally identified for their selective carbohydrate binding properties. However, now it is known that they can also mediate protein-protein, protein-lipid or protein-nucleic acid interactions. By virtue of their CRD, lectins have the ability to recognize specific carbohydrate structures on proteins, which in turn, mediate cell-cell and cell-pathogen interactions. There are currently fourteen structural families and three related subfamilies of lectins in human genome that span 76 different genes. The C-type (calcium-dependent) lectins with 66 gene members is one of the largest subgroups of the lectin superfamily that are further separated into multiple subgroups. One of these subgroups is the CD209/DC-SIGN (Dendritic cell-specific ICAM-3-grabbing non-integrin 1 also called CLEC4L, C-type lectin domain family 4 member L) subgroup, which includes CD209L/L-SIGN/CLEC4M (Liver/lymph node-specific ICAM-3-grabbing non-integrin, C-type lectin domain family 4 member M), CD23 and LSECtin/CLEC4G. Mouse genome encodes five homologues of human CD209 with a variable sequence homology to human CD209, but it is not clear whether their function is similar to human CD209L and CD209. Other major lectin subfamily proteins are the P-type lectins (mannose 6-phosphate (M6P) and the I-type lectins. SiglecS (Sialic acid-binding immunoglobulin-type lectins) are the best characterized I-type lectins. Given the role of CD209L and related proteins in diverse mechanisms of pathogen recognition and emerging evidence for the role of CD209 family proteins in SARS-CoV-2 entry and infection, this review article particularly has focused on the recent advances in the cellular and biochemical characterization of CD209 and CD209L and their roles in virial uptake, which could provide valuable insights pertinent to current pathobiological studies and therapeutic development of vaccines for SARS-Cov-2.

2. CD209/DC-SIGN and CD209L/L-SIGN Family Proteins: Cell Adhesion Molecules Turned to Pathogen Recognition Receptors:
The conserved physiological function of CD209 family proteins is to mediate cell-cell adhesion by functioning as high affinity receptors for intercellular adhesion molecules 2 and 3 (ICAM2 and ICAM3/CD50). CD23 acts as a low-affinity receptor for immunoglobulin E (IgE) and CR2/CD21 and LSECtin interacts with CD44 on activated T cells. A survey of current literature indicates that these receptors are also among the most common pathogen recognition receptors present in the human
genome\textsuperscript{[21]}\textsuperscript{[22]}. CD209L and CD209 serve as receptors for Ebolavirus\textsuperscript{[23]}, Hepatitis C virus\textsuperscript{[24]}, human coronavirus 229E\textsuperscript{[25]}, human cytomegalovirus/HHV-5 \textsuperscript{[26]}, influenza virus\textsuperscript{[27]}, West-Nile virus\textsuperscript{[26]}, Dengue virus\textsuperscript{[28]} and Japanese encephalitis virus\textsuperscript{[29]}. Recently, we and others have shown that CD209 and CD209L is capable of recognizing SARS-CoV and SARS-CoV-2 \textsuperscript{[21]}\textsuperscript{[22]}\textsuperscript{[23]}\textsuperscript{[24]}\textsuperscript{[25]}\textsuperscript{[26]}\textsuperscript{[27]}\textsuperscript{[28]}\textsuperscript{[29]}\textsuperscript{[30]}\textsuperscript{[31]}. In addition to its ability to recognize a plethora of viruses, CD209 is also known to recognize parasites such as leishmania amastigotes\textsuperscript{[32]} and Yersinia pestis coccobacillus bacterium\textsuperscript{[33]}. The complete list of viruses that are recognized by CD209, CD209L and LSECtin are shown (Table 1). To date, it is not known whether CD23 is involved in any pathogen recognition.

\textbf{Table 1.} List of known pathogens recognized by CD209, CD209L and LSECtin lectin family proteins. The data is extracted from the publications available through PubMed.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Pathogen Name</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD209</td>
<td>HIV-1 and HIV-2</td>
<td>[34] [35]</td>
</tr>
<tr>
<td></td>
<td>Ebolavirus</td>
<td>[23] [24]</td>
</tr>
<tr>
<td></td>
<td>Cytomegalovirus</td>
<td>[25] [26]</td>
</tr>
<tr>
<td></td>
<td>Hepatitis C virus</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Dengue virus</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>Measles virus</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>Herpes simplex virus</td>
<td>[30]</td>
</tr>
<tr>
<td></td>
<td>Influenza virus A</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td>SARS-CoV-2</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td>SARS-CoV</td>
<td>[33]</td>
</tr>
<tr>
<td></td>
<td>MERS</td>
<td>[34]</td>
</tr>
<tr>
<td></td>
<td>Japanese encephalitis virus</td>
<td>[35]</td>
</tr>
<tr>
<td></td>
<td>Lassa virus</td>
<td>[36]</td>
</tr>
<tr>
<td></td>
<td>Respiratory syncytial virus</td>
<td>[37]</td>
</tr>
<tr>
<td></td>
<td>Rift valley fever virus</td>
<td>[38]</td>
</tr>
<tr>
<td></td>
<td>Uukuniemi virus</td>
<td>[39]</td>
</tr>
<tr>
<td></td>
<td>West-Nile virus</td>
<td>[40]</td>
</tr>
<tr>
<td>CD209L</td>
<td>Ebolavirus</td>
<td>[23]</td>
</tr>
<tr>
<td></td>
<td>Hepatitis C virus</td>
<td>[24]</td>
</tr>
<tr>
<td></td>
<td>HIV-1</td>
<td>[25] [26]</td>
</tr>
<tr>
<td></td>
<td>Human coronavirus 229E</td>
<td>[27]</td>
</tr>
<tr>
<td></td>
<td>Human cytomegalovirus/HHV-5</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>Influenza virus</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>SARS-CoV</td>
<td>[30]</td>
</tr>
<tr>
<td></td>
<td>SARS-CoV-2</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td>West-Nile virus</td>
<td>[32]</td>
</tr>
</tbody>
</table>
t is increasingly evident that viruses exploit host lectin receptors like the CD209L family proteins and others for two major reasons; to promote infection of target cells and evade the immune recognition system. In many cases lectin receptors such as CD209 and CD209L are employed as functional portals for viral recognition and infection. However, in some other cases, they may also enable infection of target cells via trans-infection (i.e., cell captures the pathogen without entry and then passes it to another cell, which is also a replication-independent mechanism [50]. For example, CD209 expressed in DCs can bind to HIV envelope glycoprotein, gp120, without triggering cell-virus fusion[31]. The interaction of CD209 with gp120 appears to be complex as it can lead to both positive and negative outcomes for virus, perhaps depending to cell type in which CD209 is expressed. In some cases, CD209-captured virions are internalized and targeted to the lysosome for degradation[52][53]. However, in cases which HIV-1 receptor and co-receptors (CD4 and CCR5/CXCR4) are present on the host cells, CD209 can facilitate infection by transferring the virus to immune cells[54][55]. Additionally, it was found that CD209-dependent capture of HIV-1 virions could transiently protect virions from degradation, which ultimately leads to viral infectivity[56][57][58], suggesting that CD209 positive DCs capture and internalize HIV-1 virions and homest them to lymph nodes[59]. However, the Trojan horse model of HIV transmission by CD209 was challenged by various studies[60]. Studies on B lymphocytes and platelets indicate that CD209 expressed in these cells successfully mediate the entry of and infection by HIV-1[61][62][63]. Similarly, CD209 and CD209L interact with Ebola virus glycoprotein and mediate infection of endothelial cells via both cis- and trans-infection[23][30]. Likewise, the recent findings also support for CD209L-mediated cis- and trans-infection of SARS-CoV-2[13][64]. Aside from the role of CD209 and CD209L in cis- and trans-infection and transmission, the recognition of these receptors by pathogens also can impact the host defense mechanism against these pathogens. For instance, CD209-dependent viral entry and infection can initiate signaling events in host cells that compromise immune responses and promote infection of DCs[65][66]. Interestingly, although CD209L and CD209 are devoid of any enzymatic activity, however upon interaction with pathogens, they can stimulate activation of multiple protein kinases, GTPases and phosphatases[67].

### 3. CD209/DC-SIGN and CD209L/L-SIGN Family Proteins and Coronaviruses:

The evolutionarily conserved mechanism by which human coronaviruses including, CoV-229E, NL63, OC43, HKU1, MERS-CoV, SARS-CoV, SARS-CoV-2 recognize the host cells rely on the viral glycoprotein spike (S) that interacts with specific receptors on the host target cells. Intriguingly, the S protein appears to be highly adept and can interact with different types of host receptors. For examples, the S protein of CoV-229E and transmissible gastroenteritis virus (TGEV) employ CD13 (aminopeptidase N) as a receptor for entry and infection of target cells[68][69], whereas S protein of CoV-NL63 and HKU1 interact with glycan-based receptors carrying 9-O-acetylated sialic acid (9-O-Ac-Sia)[70][71]. The S protein of MERS-CoV uses Dipeptidyl peptidase 4 (DPP4/CD26) and carinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) as attachment or entry receptors for infection[72][73]. CEACAM5 appears to facilitate MERS-CoV infection by enhancing the attachment of the virus to the host cell surface[73]. The S protein of Filoviridae Marburg virus, SARS-CoV [26][31] and SARS-CoV-2[16][74] can employ the carbohydrate-recognition domain (CRD) containing CD209L and CD209 lectins as attachment or entry receptor.
CD209L is broadly expressed in human lung, kidney epithelium and endothelium\[16\]. Furthermore, human endothelial cells are permissive to SARS-CoV-2 infection and interference with CD209L activity via shRNA or soluble CD209L inhibited SARS-CoV-2 entry and replication\[16\]. Remarkably, the S protein of human coronaviruses including, NL63\[75\], SARS-CoV\[76\] and SARS-CoV-2\[16\] can also employ angiotensin-converting enzyme 2 (ACE2) as an entry receptor for infection, suggesting that both ACE2 and the lectin family proteins, CD209L and CD209, contribute to the spread of these pathogens in vivo. Previous studies on SARS-CoV demonstrated a direct role for CD209L and CD209 in infection by acting as entry receptors for SARS independent of ACE2\[27\][78][79][80]. Curiously, CD209L can physically interact with ACE2\[16\], suggesting both ACE2-dependent and independent mechanisms for CD209L-mediated viral entry. However, the underlying mechanism of CD209L and CD209 mediated SARS-CoV-2 infection is not fully understood and requires further investigation.

4. Topology of CD209L/L-SIGN and CD209/D-SIGN

CD209 family proteins are type II transmembrane glycoprotein receptors (i.e., C-terminus is exposed outside the lipid bilayer and N-terminus resides in the cytosol). The ectodomain of CD209 and CD209L is composed of the neck region followed by the CRD. These domains represent the most distinct and functional features of these two receptors (Figure 1A). The neck/repeat region is composed of 23 amino acids which is repeated seven times in CD209L and CD209 and three times in CD23/FCER (Figure 1A). However, the neck/repeat region on LSECtin/CLEC4G is replaced with a coil-coil motif, which is also involved in protein-protein interaction (Figure 1A). Central to recognition of cellular and pathogen glycoproteins, is the presence of the CRD on the C-terminus of CD209 family proteins, which is paramount to recognition of mannose containing structures present on specific glycoproteins.

Figure 1. CD209 family proteins. (A) Graphic presentation of CD209 family proteins and the key domain information. The schematic of domains do not directly correlate to the number of amino acids in each domain. (B) Crystal structure of a typical CRD complexed with carbohydrate and the position of the Ca\(^{2+}\) ion, which makes a tertiary complex between lectin and carbohydrate structure.

CRD is a 110–130 amino acid long with a double-looped, two-stranded anti-parallel β-sheet connected by two α-helices and a three-stranded anti-parallel β-sheet\[81\]. Typically, CRD has two conserved disulfide bonds and up to four Ca\(^{2+}\) binding sites, depending on the specific family of lectin. Amino acid residues with the carbonyl side chains are involved in coordinating Ca\(^{2+}\) in the CRD, and these residues also directly bind to carbohydrates leading to a ternary complex formation between a carbohydrate in a glycan, the Ca\(^{2+}\) ion, and amino acids within the CRD. A typical CRD-carbohydrate interaction is shown (Figure 1B). Amino acid sequence alignment of CD209L with CD209 illustrates that these proteins are highly conserved, suggesting that they likely evolved through gene duplications. There are at least two putative internalization motifs at the cytosolic N-terminus tail of CD209 and CD209L, indicating that both CD209 and CD209L upon interaction with pathogens are capable of undergoing internalization and delivering the pathogen inside the target cells. The internalization motifs are di-leucine (LL) and tyrosine(Y)-based (Figure 2B), but, the key tyrosine residue in the tyrosine-based internalization motif on the CD209L is replaced with histidine (H) (Figure 2B), indicating that CD209L undergoes internalization solely via di-leucine motif\[82\].
Figure 2. Amino acid sequence homology of CD209 and CD209L: (A) The schematic of CD209L is shown. (B) Alignment of the amino acids of human CD209 and CLEC4M (gene encoding for CD209L called C-type lectin domain family 4 member M, CLEC4M). The key common features of CD209L and CD209L, including potential PTMs and ion bindings are highlighted.

The ectodomain of CD209 and CD209L is composed of the neck region followed by the CRD. These domains also represent the most distinct and functional features of these two receptors. The neck region which is a repeat of 23 amino acids (Figure 2B), is involved in protein dimerization/oligomerization, and may also contribute to increased pathogen recognition and concentration of pathogens at the cell surface. The neck region forms an α-helical coiled-coil fold that is thought to stabilize the oligomerization of CD20 family proteins. The presence of CRD on the ectodomain is paramount to recognition of mannose, fucose- or galactose-containing structures on the pathogens and the cellular ligands by CD209 and CD209L. It is thought that within the CRD a highly conserved EPN motif (Glu-Pro-Asn) is responsible for recognition of mannose, fucose- or galactose-containing structures. Yet, despite a high degree of homology of the amino acid residues in the CRD of CD209L and CD209, there is evidence for differential recognition of oligosaccharide structures by these receptors. For example, CD209L appears to prefer mannose oligosaccharides but not fucose-containing carbohydrates such as LewisX (LeX) glycans. Interestingly, a recent analysis revealed that N-glycosylation of SARS-CoV-2 spike protein is largely oligomannose-type glycans, which may account for the strong binding of SARS-CoV-2 spike protein with CD209L and CD209. Furthermore, while the ectodomain of CD209L contains two N-glycosylation sequons, at sites N92 and N361 (Figure 2B), only N92 is occupied. Curiously, removal of N-glycosylation on the CD209L increases the binding of CD209L with the SARS-CoV-2 spike protein, suggesting that N-glycosylation of CD209L may generate a hindrance for the CRD-mediated glycoprotein interaction and may have impact in virus tropism and transmissibility in vivo. A similar hindering mechanism for ligand-receptor interaction by N-glycosylation was reported for an unrelated receptor tyrosine kinase, vascular endothelial receptor-2 (VEGFR-2) interaction with its ligand. Another important, yet poorly understood aspect of CD209 family proteins is their cytoplasmic N-terminus domain, which is vital for their signal transduction relays. To date, there is no evidence for potential posttranslational modifications (PTMs) or a direct protein interaction between the cytoplasmic N-terminus domains of CD209L and CD209 with the signaling proteins. Unlike many of their counterpart receptors, the cytoplasmic N-terminus domains of CD209L and CD209 contain no conserved immunoreceptor tyrosine-based inhibitory (ITIM, V/IXXXXL/I/V) motif, which interacts with the Src-homology 2 (SH2) domain containing proteins. While, the cytoplasmic N-terminus domains of CD209L...
contains no tyrosine residue, the cytoplasmic N-terminus domain of CD209 contains one tyrosine residue with a weak sequence homology to ITIM motif (Figure 2B). But, there is no experimental evidence whether the key tyrosine (Y31) residue is phosphorylated and recruits any SH2 domain signaling proteins to CD209. Furthermore, there are multiple serine/threonine residues (four on the CD209 and 8 on the CD209L) on the cytoplasmic N-terminus domains of CD209L and CD209 (Figure 2B), which potentially could be phosphorylated. Similarly, there are multiple lysine (K) residues on the cytoplasmic N-terminus domain of CD209L and CD209 with potential to undergo ubiquitination. In particular, K5, which is could be phosphorylated. Similarly, there are multiple lysine (K) residues on the cytoplasmic N-terminus domains of CD209L and CD209 (Figur...
the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that

Mou, H.; Raj, V.S.; van Kuppeveld, F.J.M.; Rottier, P.J.M.; Haagmans, B.L.; Bosch, B.J. The receptor binding domain of
deh Jong, M.A.W.P.; de Witte, L.; Bolmstedt, A.; van Kooyk, Y.; Geijtenbeek, T.B.H. Dendritic cells mediate herpes

Soilleux, E.J.; Riley, J.L., et al. DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages

Cormier, E.G.; Durso, R.J.; Tsamis, F.; Boussemart, L.; Manix, C.; Olson, W.C.; Gardner, J.P.; Dragic, T. L-SIGN (CD209L)
and DC-SIGN (CD209) mediate transinfection of liver cells by hepatitis C virus. Proc Natl Acad Sci U S A 2004, 101,


Mou, H.; Raj, V.S.; van Kuppeveld, F.J.M.; Rottier, P.J.M.; Haagmans, B.L.; Bosch, B.J. The receptor binding domain of the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that


Keywords

C-type Lectin CD209L/L-SIGN and CD209/DC-SIGN

Retrieved from https://encyclopedia.pub/7056