Review

C-type lectin CD209L/L-SIGN and CD209/DC-SIGN: Cell adhesion molecules turned to pathogen recognition receptors

Nader Rahimi
Department of Pathology, School of Medicine, Boston University Medical Campus, Boston, MA 02118.
* Correspondence: Nader Rahimi; Departments of Pathology and Laboratory Medicine Boston University Medical Campus; 670 Albany St., Room 510, Boston, MA 02118;
E-mail: nrahimi@bu.edu

Abstract: C-type lectin CD209/DC-SIGN and CD209L/L-SIGN proteins are distinct cell adhesion and pathogen recognition receptors that mediate cellular interactions and recognize a wide range of pathogens including, viruses such as SARS, SARS-CoV-2, bacteria, fungi and parasites. Pathogens exploit CD209L family proteins to promote infection and evade the immune recognition system. CD209L and CD209 are widely expressed in SARS-CoV-2 target organs and can contribute to infection and pathogenesis. CD209L family receptors are highly susceptible to alternative splicing and genomic polymorphism, which may influence virus tropism and transmission in vivo. The carbohydrate-recognition domain (CRD) and the neck/repeat region represent the key features of CD209L family proteins that are also central to facilitating cellular ligand interactions and pathogen recognition. While, the neck/repeat region is involved in oligomeric dimerization, the CRD recognizes the mannose containing structures present on specific glycoproteins such as those found on the SARS-CoV-2 spike protein. Considering the role of CD209L and related proteins in diverse pathogen recognition, this review article discusses the recent advances in the cellular and biochemical characterization of CD209 and CD209L and their roles in viral uptake, which has important implications in understanding the host-pathogen interaction, the viral pathobiology and driving vaccine development of SARS-CoV-2.

Keywords: CD209; L-SIGN; CD209L; DC-SIGN; C-type lectin; Cell Adhesion Molecule; C-type lectin domain family 4 member M; CLEC4M; LSECtin; CLEC4G; SARS-CoV-2; COVID-19

Introduction

Lectins are a diverse family of carbohydrate recognizing proteins that possess carbohydrate-recognition domain (CRD) or sulfated glycosaminoglycan (SGAG)-binding motif [1,2]. Derived from the Latin word “legere”, 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 [3]. 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 [4,5]. There are currently fourteen structural families and three related subfamilies of lectins in human genome that span 76 different genes [6,7]. The C-type (calcium-dependent) lectins with 66 gene members is one of the largest subgroups of the lectin superfamily [6,8] that are further separated into multiple subgroups [7,9,10]. 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 CD209/DC-SIGN (and three other member genes, namely CD209/L-SIGN/CLEC4M (C-type lectin domain family 4 member M), CD23 and LSECtin/CLEC4G [11,12]. Mouse genome encodes five homologues of human CD209 with a variable sequence homology to human CD209 [13], 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 [14,15]. Given the role of CD209L and related proteins in diverse mechanisms of pathogen recognition and emerging evidence for the role of CD209L family proteins in SARS-CoV-2 entry and infection [16], this review article particularly has focused on the recent advances in the cellular and biochemical characterization of CD209 and CD209L and their roles in viral uptake, which could provide valuable insights pertinent to current pathobiological studies and therapeutic development of vaccines for SARS-Cov-2.

**CD209L/L-SIGN family proteins: Cell adhesion molecules turned to pathogen recognition receptors:**

The conserved physiological function of CD209L 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) [17-19]. CD23 acts as a low-affinity receptor for immunoglobulin E (IgE) and CR2/CD21 [20] and LSECtin interacts with CD44 on activated T cells [21]. A survey of current literature indicates that these receptors are also among the most common pathogen recognition receptors present in the human genome [22,23], CD209L and CD209 serve as receptors for Ebolavirus [24], Hepatitis C virus [25], human coronavirus 229E [26], human cytomegalovirus/HHV-5 [27], influenza virus [28], West-Nile virus [27], Dengue virus [29] and Japanese encephalitis virus [30]. Recently, we and others have shown that CD209 and CD209L is capable of recognizing SARS-CoV and SARS-CoV-2 [16,27,31,32]. In addition to its ability to recognize a plethora of viruses, CD209 is also known to recognize parasites such as leishmania amastigotes [33] and Yersinia pestis cocobacillus bacterium [34]. 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.

**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>[35,36]</td>
</tr>
<tr>
<td></td>
<td>Ebolavirus</td>
<td>[37,38]</td>
</tr>
<tr>
<td></td>
<td>Cytomegalovirus</td>
<td>[39,40]</td>
</tr>
<tr>
<td></td>
<td>Hepatitis C virus</td>
<td>[39]</td>
</tr>
<tr>
<td></td>
<td>Dengue virus</td>
<td>[41]</td>
</tr>
<tr>
<td></td>
<td>Measles virus</td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>Herpes simplex virus</td>
<td>[43]</td>
</tr>
<tr>
<td></td>
<td>Influenza virus A</td>
<td>[44]</td>
</tr>
<tr>
<td></td>
<td>SARS-CoV-2</td>
<td>[45]</td>
</tr>
<tr>
<td></td>
<td>SARS-CoV</td>
<td>[31]</td>
</tr>
<tr>
<td></td>
<td>MERS</td>
<td>[46]</td>
</tr>
<tr>
<td></td>
<td>Japanese encephalitis virus</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>Lassa virus</td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td>Respiratory syncytial virus</td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td>Rift valley fever virus</td>
<td>[50]</td>
</tr>
<tr>
<td></td>
<td>Uukuniemi virus</td>
<td>[50]</td>
</tr>
</tbody>
</table>
It 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 [58]. For example, CD209 expressed in DCs can bind to HIV envelope glycoprotein, gp120, without triggering cell-virus fusion [59]. 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 [60,61]. 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 [62,63]. Additionally, it was found that CD209-dependent capture of HIV-1 virions could transiently protect virions from degradation, which ultimately leads to viral infectivity [17,64]. These observations is consistent with the ‘Trojan horse hypothesis’ of HIV transmission, where CD209 expressed in DCs capture and internalize HIV-1 virions and homes them to the target cells rich lymph nodes [65,66]. However, the Trojan horse model of HIV transmission by CD209 was recently challenged [67]. Studies on B lymphocytes and platelets indicate that CD209 expressed in these cells successfully mediate the entry of and infection by HIV-1 [68-70]. Similarly, CD209 and CD209L interact with Ebola virus glycoprotein and mediate infection of endothelial cells via both cis- and trans-infection [24,37]. Likewise, the recent findings also support for CD209L-mediated cis- and trans-infection of SARS-CoV-2 [16,71]. 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 [72,73]. 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 [74].
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 (CoV229E-S) employs CD13 (aminopeptidase N) as a receptor for entry and infection of target cells [75,76], whereas S protein of CoV-NL63 and HKU1 interact with glycan-based receptors carrying 9-O-acetylated sialic acid (9-O-Ac-Sia) [77,78]. The S protein of MERS-CoV uses Dipeptidyl peptidase 4 (DPP4/CD26) and carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5) as attachment or entry receptors for infection [79,80]. CEACAM5 appears to facilitate MERS-CoV infection by enhancing the attachment of the virus to the host cell surface [80]. The S protein of Filoviridae Marburg virus [81], SARS-CoV [81,82] and SARS-CoV-2 [16] 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 [83], SARS-CoV [84] 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 [82,85]. 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.

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

CD209L 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). The neck/repeat region on LSECtin/CLEC4G is replaced with a coil-coil motif (Figure 1A). Central to recognition of cellular and pathogen glycoproteins, is the presence of the CRD on the C-terminus of CD209L family proteins, which is paramount to recognition of mannose containing structures present on specific glycoproteins.

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 [86]. Typically, CRD has two conserved disulfide bonds and up to four Ca²⁺ binding sites, depending on the specific family of lectin. Amino acid residues with the carbonyl side chains are involved in coordinating Ca²⁺ in the CRD, and these residues also directly bind to carbohydrates leading to a ternary complex formation between a carbohydrate in a glycan, the Ca²⁺ 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 [87].
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 [88,89], 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 CD209L family proteins [90,91]. 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 [92,93]. 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 [94]. Interestingly, a recent analysis revealed that N-glycosylation of SARS-CoV-2 spike protein is predominantly oligomannose-type glycans [95], which may account for the strong binding of SARS-CoV-2 spike protein with CD209L and CD209 [16]. 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[16], suggesting that N-glycosylation of CD209L may generate a hindrance for the CRD-mediated glycoprotein interaction [16] 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 [96].
A corollary to the function of CD209L and CD209 in pathogen and cellular ligand recognition is that these receptors are highly susceptible to alternative splicing and genomic polymorphism, which may significantly influence their core functions. Analysis of CD209L via uniprot
(https://www.uniprot.org/), a freely accessible resource of protein sequence, revealed that CD209L mRNA can generate at least 9 alternatively spliced variants (Figure 3A). In many cases, the CRD is either completely or partially deleted (Figure 3A), which generates a carbohydrate binding decoy of CD209L. In some other cases, the transmembrane domain is deleted, which results in the soluble form of CD209L. Yet in other cases, multiple deletions occur simultaneously leading to generation of soluble CD209L proteins without CRD (Figure 3A). Some of the alternatively spliced variants of CD20L also carry a deletion on the neck region (Figure 3A). Similar to CD209L, CD209 also due to alternative splice mechanism can also yield 11 different variants (Figure 3B). However, the tissue expression profiles and the potential function of these alternatively spliced variants in normal cellular ligand recognition, and more importantly in pathogen interaction remains largely unknown. Moreover, various recent studies have shown a distinct genomic polymorphism, specifically, in the tandem-neck-repeat region of CD209L [100,101] and on the promoter CD209 [102], which is linked to the pathogenesis of tuberculosis [103] and resistance to HIV-1 [104]. These polymorphic changes most likely influence the physiological function of CD209L family proteins as well as the role that CD209L family proteins play in binding to other pathogens, like SARS and SARS-CoV-2.

Expression profile of CD209L family proteins in human tissues and cells:

Survey of the published data indicates that CD209/DC-SIGN is predominantly expressed on the monocyte-derived dendritic cells (DCs), and on DCs of immature and mature in lymphoid tissue, lymph nodes and spleen [19,105], whereas CD209L is predominantly expressed in human type II alveolar cells of lung, liver, kidney and lymph nodes [18,32,45,106]. Our analysis of expression of CD209L, CD209 and LSECtin mRNAs through the Human Expression Atlas (https://www.proteinatlas.org/), a publically available dataset, revealed that CD209L expression
is relatively restricted to a few human organs. The highest levels of CD209L was observed in liver and lymph node followed by placenta, lung and ovary (Figure 4A). CD209, on the other hand, is broadly expressed in human tissues and organs at the various levels. The highest levels of CD209 expression was observed in the lymph node followed by the adipose tissue, small intestine and rectum (Figure 4B). Similar to CD209, LSECtin/CLEC4G was also broadly expressed in human tissues and organs. The highest levels of LSECtin was observed in liver and lymph node, followed with DCs, monocytes, adipose tissue, heart muscle and cerebellum (Figure 4C). Considering the current SARS-CoV-2 pandemic, it is worthwhile to compare the expression profiles CD209L family proteins with ACE2. Human ACE2 is considered an important entry receptor for SARS-CoV-2 [84,107,108] and was previously reported to be widely expressed in the lung, vascular system and other organs [109]. However, a recent study demonstrated that ACE2 is expressed at very low levels and only in a small subset of lung epithelial cells [110] and low-to-undetectable levels in endothelial cells[111]. In agreement with the recent observations, our analysis of ACE2 mRNA through the Human Expression Atlas revealed that ACE2 is highly expressed only in intestinal tissues (small intestine, colon and duodenum) and at the low levels in the hearth muscles, kidney, gallbladder and testis, but undetectable in lung (Figure 4D). The observed limited/low ACE2 expression pattern in human tissues suggests that SARS-CoV-2 could use alternative receptors for cell entry in a cell-type dependent manner as CD209L and CD209 appear to be more broadly expressed in human organs and tissues than the ACE2. Given the wider tissue expression patterns of CD209 and CD209L in SARS-CoV-2 target tissues, they could play important roles in the pathogenesis of SARS-CoV-2.

SARS-CoV-2, CD209L family proteins and endothelial cells:

CD209L is expressed on endothelial cells of various organs including, the lymph nodes sinuses [112], liver sinus endothelial cells [112,113], capillary endothelial cells of the placenta [113], the endothelial cells of the gastrointestinal tract [114], kidney endothelial cells [16] and pulmonary endothelial cells [16,32,115]. In addition, other CD209L family lectin, LSECtin/CLEC4G is also expressed in liver sinusoidal endothelial cells and in the lymph node [116]. CD209L expressed in endothelial cells can interact with ICAM3 which is expressed in DCs, T-cells [117,118] and endothelial cells [18,119] generating multiple immunoregulatory interactions between T-cells, DCs and endothelial cells (Figure 5). Moreover, endothelial cells express both ICAM3 [18,119] and ICAM2 [120], capable of generating homotypic interactions between endothelial cells (Figure 5) that regulate endothelial function and angiogenesis. ICAM2 and ICAM3 belong to immunoglobulin (Ig) superfamily cell adhesion molecules. ICAM2 and ICAM3 are highly glycosylated at the ectodomain
and contain two and five Ig domains, respectively. Consistent with this idea, loss of ICAM2 in mouse results in impaired endothelial cell migration and angiogenesis [120]. However, expression of ICAM2, ICAM3 and CD209L in endothelial cells, the role of this CD209L/ICAM2/ICAM3 axis in endothelial biology is largely unknown. A recent study demonstrated that knocked down of CD209L by shRNA in immortalized endothelial cells, HUVEC-TERT, increases cell migration, but inhibits capillary tube formation/\textit{in vitro} angiogenesis [16], underscoring the importance of CD209L pathway in the regulation of normal endothelial function.

Recent studies demonstrate that human vascular system is a direct target of SARS-CoV2 [16,121,122]. Although primarily infects the lungs, the SARS-CoV-2 virus also targets multiple other organs including, the cardiovascular system, gastrointestinal tract, and the kidneys [123-126]. Severe endothelial cell injury, vascular thrombosis with micro-angiopathy, occlusion of alveolar capillaries, and angiogenesis are commonly observed in COVID-19 patients [127]. COVID-19 patients suffer from distinct endothelial cell injury (\textit{i.e.}, endothelitis), and altered angiogenesis [127] with widespread microvascular thrombosis [128,129]. These observations coupled with the fact that vascular endothelial dysfunction also plays crucial roles in the pathogenesis of COVID-19 [23,130], underscores the pivotal role of endothelial cells in the pathobiology of SARS-CoV-2 infection. Collectively, the data points to potentially important role for CD209L family proteins in the pathogenesis of SARS-CoV-2. Therefore, CD209L not only can act as a SARS-COV-2 entry receptor, but also performs critical functions in the angiogenic responses of endothelial cells, suggesting that SARS-COV-2, by exploiting CD209L, could undermine CD209L function in endothelial cells, leading to endothelial cell injury and altered angiogenesis.

![Figure 5. CD209L/L-SIGN mediates multiple cell-cell interactions between endothelial cells, DCs and T-cells](image)

**Figure 5. CD209L/L-SIGN mediates multiple cell-cell interactions between endothelial cells, DCs and T-cells.** CD209L cellular ligands (ICAM2 and ICAM3) are expressed in DCs, T-cells and endothelial cells leading multiple immunoregulatory and other interactions between T-cells, DCs and endothelial cells. CD209L/ICAM2/ICAM3 interactions in endothelial cells can regulate endothelial adhesion and angiogenic properties of endothelial cells.
Conclusion and perspective:

Infectious diseases have been foremost among the threats posed to human health and survival through history. The novel coronavirus disease-2019 (COVID-19) pandemic continues to pose a serious threat to global public health with overwhelming worldwide socio-economic disruption. The emerging picture of pathogenesis of SARS-CoV-2 is that in addition to dysregulation of the host immune response in lung, also targets other major organs including, cardiovascular system and others, which may account for COVID-19 induced mortality. The vascular system, particularly, the pulmonary endothelium may play a pivotal role in the pathogenesis of COVID-19 via engagement of CD209L and other related receptors. To date, many aspects of SARS-CoV-2 transmission, infection, and treatment remain unclear. Not only can CD209L function as an entry receptor it can also contribute to the pathogenesis of COVID-19. Establishing a comprehensive map of the SARS-CoV-2 interaction with CD209L family proteins, and their roles in in endothelial functions and injury can provide new insights into pathogenesis of COVID-19 and offers a bona fide treatment modality.

Author Contributions: Nader Rahimi conceived and wrote the manuscript.

Funding: This work was supported in part through grants from CTSI grant UL1 TR001430, Malory Fund, Department of Pathology, Boston University and a grant from The Evans Center for Interdisciplinary Biomedical Research ARC on COVID-19.

Acknowledgments: Author thanks Razie Amraei and Rachel Ho for reading and commenting on the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

References:

11 of 17


Londrigan, S.L.; Turville, S.G.; Tate, M.D.; Deng, Y.-M.; Brooks, A.G.; Reading, P.C. N-linked glycosylation facilitates sialic acid-independent attachment and entry of influenza A viruses into cells expressing DC-SIGN or L-SIGN. Journal of virology 2011, 85, 2990-3000.


