Extensive Repertoire of Membrane-bound and Soluble Dendritic Cell-specific ICAM-3-grabbing Nonintegrin 1 (DC-SIGN1) and DC-SIGN2 Isoforms

INTER-INDIVIDUAL VARIATION IN EXPRESSION OF DC-SIGN TRANSCRIPTS

Expression in dendritic cells (DCs) of DC-SIGN, a type II membrane protein with a C-type lectin ectodomain, is thought to play an important role in establishing the initial contact between DCs and resting T cells. DC-SIGN is also a unique type of human immunodeficiency virus-1 (HIV-1) attachment factor and promotes efficient infection in trans of cells that express CD4 and chemokine receptors. We have identified another gene, designated here as DC-SIGN2, that exhibits high sequence homology with DC-SIGN. Here we demonstrate that alternative splicing of DC-SIGN1 (original version) and DC-SIGN2 pre-mRNA generates a large repertoire of DC-SIGN-like transcripts that are predicted to encode membrane-associated and soluble isoforms. The range of DC-SIGN1 mRNA expression was significantly broader than previously reported and included THP-1 monocytic cells, placentas, and peripheral blood mononuclear cells (PBMCs), and there was cell maturation/activation-induced differences in mRNA expression levels. Immunostaining of term placenta with a DC-SIGN1-specific antiseraum showed that DC-SIGN1 is expressed on endothelial cells and CC chemokine receptor 5 (CCR5)-positive macrophage-like cells in the villi. DC-SIGN2 mRNA expression was high in the placenta and not detectable in PBMCs. In DCs, the expression of DC-SIGN2 transcripts was significantly lower than that of DC-SIGN1. Notably, there was significant inter-individual heterogeneity in the repertoire of DC-SIGN1 and DC-SIGN2 transcripts expressed. The genes for DC-SIGN1, DC-SIGN2, and CD23, another Type II lectin, co-localize to an ~85 kilobase pair region on chromosome 19p13.3, forming a cluster of related genes that undergo highly complex alternative splicing events. The molecular diversity of DC-SIGN-1 and -2 is reminiscent of that observed for certain other adhesive cell surface proteins involved in cell-cell connectivity. The generation of this large collection of polymorphic cell surface and soluble variants that exhibit inter-individual variation in expression levels has important implications for the pathogenesis of HIV-1 infection, as well as for the molecular code required to establish complex interactions between antigen-presenting cells and T cells, i.e. the immunological synapse.

The dissemination of human immunodeficiency virus-1 (HIV-1) and establishment of infection within an individual involve the transfer of virus from mucosal sites of infection to T cell zones in secondary lymphoid organs. How this happens is not precisely known. However, there is growing support for the notion that dendritic cells (DCs) present within the mucosal sites may play a central role in this process (1–15). The normal function of DCs is to survey mucosal surfaces for antigens, capture the antigens, process captured proteins into immunogenic peptides, emigrate from tissues to the paracortex of draining lymph nodes, and present peptides in the context of MHC (major histocompatibility complex) molecules to T cells (1). It is now generally believed that HIV-1 may subvert this normal trafficking process to gain entry into lymph nodes and access to CD4+ T cells. There is also evidence demonstrating that productive infection of DCs and the ability of DCs to capture virus with subsequent transmission to T cells is mediated through two separate pathways (Refs. 5 and 8; reviewed in Refs. 3 and 15). Thus, strategies designed to block mucosal transmission of HIV will require a clear understanding of the molecular determinants of not only virus infection but also of virus capture by DCs or other cell types that can subserve a similar function.

Two recent reports by Geijtenbeek et al. (16, 17) demonstrated that a mannose-binding, C-type lectin designated as

* This work was supported in part by National Institutes of Health Grants AI43279 and AI46326 (to S. K. A.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** Recipient of an Elizabeth Glaser Scientist award from the Elizabeth Glaser Pediatric AIDS Foundation and a Burroughs Wellcome Clinical Scientist Award in Translational Research. To whom correspondence should be addressed: Division of Infectious Diseases, Dept. of Medicine (Mail Code 7880), University of Texas Health Science Center at San Antonio, San Antonio, TX 78229-3900. Tel.: 210-567-6511; Fax: 210-567-4654; E-mail: ahujas@uthscsa.edu.

† The nucleotide sequence(s) reported in this paper has been submitted to the GenBank [¶] or EBI Data bank with accession number(s) AY042221 through AY042240.

¶ These authors contributed equally to this work.

The on-line version of this article (available at http://www.jbc.org) contains Supplementary Figs. 1–3.

¶¶ The on-line version of this article (available at http://www.jbc.org) contains Supplementary Figs. 1–3.
DC-SIGN (DC-specific, ICAM-3 grabbing, nonintegrin) may play a key role in DC-T cell interactions as well as in HIV pathogenesis. First, by binding to ICAM-3 expressed on T cells, DC-SIGN is thought to facilitate the initial interaction between DCs and naive T cells (17), setting the stage for subsequent critical events that lead to antigen recognition and the formation of a contact zone termed the immunological synapse (15, 16) confirmed an earlier observation that DC-SIGN is not a coreceptor for viral infection. As a first step in testing this hypothesis, we elucidated the gene and mRNA structure as well as the expression of DC-SIGN1 and DC-SIGN2 transcripts with variable expression levels may have important implications for the pathogenesis of HIV-1 infection and the generation of T cell immune responses.

**MATERIALS AND METHODS**

Cells, Cytokine Differentiation of DCs, and RNA—CD34⁺ peripheral hematopoietic progenitor cells (PBHP) and peripheral blood mononuclear cells (PBMCs) were isolated from healthy adult normal volunteers treated with granulocyte colony-stimulating factor (G-CSF, Amgen, CA) as described previously (27). The CD34⁺ PBHP cells were cultured in medium supplemented with 20 ng/ml of stem cell factor and 50 ng/ml granulocyte-macrophage colony-stimulating factor (R&D Systems, Minneapolis, MN). Tumor necrosis factor-α (10 ng/ml) was added on day 7, and on day 11 of culture IL-4 (10 ng/ml) was added to one-half of the cells. The cytokine-differentiated CD34⁺ PBHP cells were kept in culture for a total of 15 days. By day 14 of culture more than 99% of cells were CD33⁺ indicating that the predominant cell population was of the myeloid series (28). The proportion of cells that stained for T/B lymphocyte markers (CD3/CD19) was less than 1–3%. PBMCs were also isolated from 20 ml of blood obtained from normal donors who did not receive granulocyte colony-stimulating factor. An aliquot of these PBMCs were stimulated with PHA (5 µg/ml, Sigma) for 4 days. In some experiments IL-2 (50 units/ml, Life Technologies) was added to the culture medium after day 4. CD3 and CD28 monoclonal antibodies (PharMingen) were coated on tosyl-activated Dynal beads (Dynal, Lake Success, NY) and used to stimulate PBMCs (1:1 concentration). The placenta samples were from anonymous normal donors. mRNA from highly purified leukocyte subsets, including CD14⁺ monocytes, was also

| PCR primers used for DC-SIGN1 cDNA cloning (Fig. 1) | Southern blot analysis (Figs. 8 and 9) and PAGE analysis (Fig. 11) |
|-----------------------------------------------|---------------------------------------------------------------|
| 1–1 atg aag gac tcc aag gaa (S)               | 2–1 ggt acc atc tgg gga cag cgg gaa (S)                       |
| 1–2 aag tca cag tct ctc ctc ctc (AS)         | 2–2 gct cta gac tat tct cgt cag c (AS)                        |
| 1–3 ggt acc atc tgg gga cag cgg gaa (S)      | 2–3 atc tgg gga cag cgg gaa (S)                              |
| 1–4 ggc ctc agc ctc ccc ccc ccc (AS)         | 2–4 tgt gaa gaa gag cgg cag (AS)                              |
| 1–5 agg aac aec tga gag gcc gtt (Ex Ic)       | 1–6 tct gtg cct ggg ctt ctt cgc aag (Ex II)                   |
| 1–6 tct gtt cgg gcc tgt cgc aag (Ex Ic)       | 1–7 gag ctt cgg ggt gtc caa cgc ggc (Ex Ic-Ex III)            |
| 1–7 ggt acc aac atc tgg gga cag cgg gaa (S)   | 1–8 tgt ctt cgg cgg ctt cgc aag ggc gtt (Ex VI)               |

| DC-SIGN oligonucleotides used for Southern blot analysis (Figs. 8 and 9 and data not shown) | PCR primers used in generating DC-SIGN2 cDNAs (Fig. 5) |
|-------------------------------------------------------------------------------------------|---------------------------------------------------------|
| 2–3 ggt acc atc tgg gga cag cgg gaa (S)                                                   | 1–1 atg aag gac tcc aag gaa (S)                          |
| 2–4 tgt gaa gaa gag cgg cag (AS)                                                          | 2–2 gct cta gac tat tct cgt cag c (AS)                   |

**TABLE I**

| Oligonucleotides used in this study |
|------------------------------------|
| S, sense; AS, antisense; Ex, exon (orientation). |
FIG. 1. Molecular basis of the extensive repertoire of DC-SIGN1 mRNAs. a, schematic illustration of the molecular basis for generation of DC-SIGN1A mRNA transcripts. The top panel is a schematic illustration of the DC-SIGN1 gene. Horizontal lines indicate exons I–VI, and dashed lines illustrate the splicing events that lead to the formation of the prototypic exon II-containing DC-SIGN1A mRNA transcript that was originally described by Curtis et al. (20) and designated herein as mDC-SIGN1A Type I. H11001 indicates the translational start site in this prototypic DC-SIGN1 mRNA. Exon II is predicted to encode the TM domain (see Fig. 2) and the exon II-containing DC-SIGN1A mRNA transcripts are predicted to encode membrane-bound or mDC-SIGN1A isoforms, whereas mRNAs that lack this TM-encoding exon II are predicted to encode sDC-SIGN1A isoforms. Alternative splicing events that lead to the generation of mRNA transcripts that contain or lack the TM-encoding exon II can be deduced by joining the various exonic sequences indicated. The starting and ending nucleotide number of each exonic segment is separated by dots (e.g. join 1..46), and exonic segments are separated from each other by a comma (e.g. 1..46, 147..206, 981..1052). Note that we did not determine the length of the 5′ untranslated region of DC-SIGN1. sDC-SIGN1A Type I represents the prototypic exon II-lacking DC-SIGN1A mRNA. The translation initiation codon for all DC-SIGN1A mRNA transcripts resides in exon Ia. Positions shown in bold denote splicing sites that are distinct from those found in the prototypic mDC-SIGN1 (Type I) or sDC-SIGN1 (Type I) mRNA transcripts. An asterisk indicates the stop codon used by the DC-SIGN1A transcripts shown in this panel. The numbering system is based on the nucleotide sequence deposited under GenBank™ accession number AC008812, with the first nucleotide of the initiation Met codon of the prototypic mDC-SIGN1A (Type I) mRNA considered as +1. This nucleotide corresponds to residue 50622 in the nucleotide sequence in GenBank™ accession number AC008812. (The nucleotide numbers in AC008812 that correspond to each of the DC-SIGN1 nucleotide numbers shown in this figure can be found in the notes that accompany GenBank™/EBI Data Bank accession numbers AY042221–AY042233.) Note that the DC-SIGN1 is in the reverse orientation in this genomic contig. b, molecular basis for generation of DC-SIGN1B mRNA transcripts. The top panel is a schematic illustration of the DC-SIGN1 gene. Horizontal lines indicate exons I–VI, and dashed lines illustrate the splicing events that lead to the formation of the prototypic exon II-containing DC-SIGN1B mRNA transcript that was originally described by Curtis et al. (20) and designated herein as mDC-SIGN1B Type I. H11001 indicates the translational start site in this prototypic DC-SIGN1 mRNA. Exon II is predicted to encode the TM domain (see Fig. 2) and the exon II-containing DC-SIGN1B mRNA transcripts are predicted to encode membrane-bound or mDC-SIGN1B isoforms, whereas mRNAs that lack this TM-encoding exon II are predicted to encode sDC-SIGN1B isoforms. Alternative splicing events that lead to the generation of mRNA transcripts that contain or lack the TM-encoding exon II can be deduced by joining the various exonic sequences indicated. The starting and ending nucleotide number of each exonic segment is separated by dots (e.g. join 1..46), and exonic segments are separated from each other by a comma (e.g. 1..46, 147..206, 981..1052). Note that we did not determine the length of the 5′ untranslated region of DC-SIGN1. sDC-SIGN1B Type I represents the prototypic exon II-lacking DC-SIGN1B mRNA. The translation initiation codon for all DC-SIGN1B mRNA transcripts resides in exon Ia. Positions shown in bold denote splicing sites that are distinct from those found in the prototypic mDC-SIGN1B (Type I) or sDC-SIGN1B (Type I) mRNA transcripts. An asterisk indicates the stop codon used by the DC-SIGN1B transcripts shown in this panel. The numbering system is based on the nucleotide sequence deposited under GenBank™ accession number AC008812, with the first nucleotide of the initiation Met codon of the prototypic mDC-SIGN1B (Type I) mRNA considered as +1. This nucleotide corresponds to residue 50622 in the nucleotide sequence in GenBank™ accession number AC008812. (The nucleotide numbers in AC008812 that correspond to each of the DC-SIGN1 nucleotide numbers shown in this figure can be found in the notes that accompany GenBank™/EBI Data Bank accession numbers AY042221–AY042233.) Note that the DC-SIGN1 is in the reverse orientation in this genomic contig. c, DC-SIGN1 transcripts. The Splice donor and Splice acceptor sites for DC-SIGN1 transcripts that contain Exon II or lack Exon II are shown. The translation initiation codon for all DC-SIGN1 mRNA transcripts resides in exon Ia. Positions shown in bold denote splicing sites that are distinct from those found in the prototypic mDC-SIGN1B (Type I) or sDC-SIGN1B (Type I) mRNA transcripts. An asterisk indicates the stop codon used by the DC-SIGN1B transcripts shown in this panel. The numbering system is based on the nucleotide sequence deposited under GenBank™ accession number AC008812, with the first nucleotide of the initiation Met codon of the prototypic mDC-SIGN1B (Type I) mRNA considered as +1. This nucleotide corresponds to residue 50622 in the nucleotide sequence in GenBank™ accession number AC008812. (The nucleotide numbers in AC008812 that correspond to each of the DC-SIGN1 nucleotide numbers shown in this figure can be found in the notes that accompany GenBank™/EBI Data Bank accession numbers AY042221–AY042233.) Note that the DC-SIGN1 is in the reverse orientation in this genomic contig.
amplification was performed in a 100-

2 cDNAs were amplified using the PCR conditions described above in a Quanta III cassette for 15 h. The PCR products were cloned into TOPO vectors 2.1 or II (Invitrogen) and sequenced on both strands. To determine the genomic structure of DC-SIGN1, a series of sense and antisense orientation primers based on the cDNA sequence described by Curtis et al. (20) were designed (sequences not shown). Two expressed sequence tags (ESTs) that had homology to DC-SIGN1 were purchased from Research Genetics (Huntsville, AL, Image Clones 146996 and 240697) and sequenced on both strands.

Southern Blot Hybridization—One μg of total RNA was used for synthesizing cDNA by random primers (Superscript Preamplification System, Life Technologies, Inc.). One-tenth of the cDNA product was used for PCR amplification. The PCR amplification profile consisted of synthesizing cDNA by random primers (Superscript Preamplification System, Life Technologies, Inc.).

The DC-SIGN1B transcripts are also predicted to initiate translation at position 4335 – 4494, respectively, are indicated by daggers

The TNT®-coupled Reticulocyte Lysate System (Promega) was used to translate in vitro DC-SIGN1 cDNAs cloned into pcDNA4/Himax TOPO vector (Invitrogen). The 78-meric translated products were fractionated in a 9% acrylamide gel and were exposed to XAR-2 film (Kodak) in a Quanta III cassette.

Antibodies and Peptides—A synthetic peptide (NH2-CRDEEQFLS-PAPATPNNPPPA-COOH) derived from the C-terminal region of DC-SIGN1 was KLH-conjugated and used to immunize rabbits. The corresponding peptide sequence is absent in the DC-SIGN2. Rabbits were bled after 6 weeks to obtain polyclonal antiserum and were subsequently affinity-purified. Goat polyclonal antibodies (Ab) for CCR5 (sc-6128), PECA-1 (sc-1505), and the corresponding blocking peptides were purchased from Santa Cruz Biotechnology (Santa Cruz, CA).

The DC-SIGN1 blocking peptide was synthesized by Zymed Laboratories Inc. (San Francisco, CA).

polyacrylamide gels were dried, and autoradiography was performed as described above. For double staining, the sections were incubated in PBS for 5 min, and endogenous peroxidases were inhibited using a peroxidase blocking substrate kit (Vector Laboratories). Distilled water was used to block additional color development. The filters were exposed to Maxar (M) film (Kodak) at ~80°C in a Quanta III cassette for 15 h.

Polyacrylamide Gel Electrophoresis—DC-SIGN1 and DC-SIGN2 cDNAs were amplified using the PCR conditions described above in a 100-μl reaction. The primers used for amplification were oligonucleotides 1-1 and 1-2 for DC-SIGN1 and oligonucleotides 2-3 and 2-4 for DC-SIGN2 (Table 1). An oligonucleotide that is DC-SIGN1 exon Ib-specific (oligonucleotide 1-3, Table 1) was used to amplify exons Ib-containing cDNAs. The amplified products were size-fractionated by electrophoresis on a 1.5% agarose gel. After denaturation in alkaline solution, the DNA was transferred to a nylon membrane (Amersham Pharmacia Biotech) by capillary action. Hybridization was performed with 32P-labeled oligonucleotide probes: (ii) an oligonucleotide derived from DC-SIGN1 sequences in exon Ib, exon Ic, exon II, and exon VI (oligonucleotides 1-4, 1-5, 1-6, and 1-8, respectively in Table 1); (iii) an oligonucleotide that had 11 nucleotides of the 3' end of exon 1c and 11 nucleotides of the 5' end of exon III of DC-SIGN1 (oligonucleotide 1-7, Table 1); (iii) an oligonucleotide that had identity with DC-SIGN2-specific exon II sequences (oligonucleotides 2-5 for DC-SIGN1 and oligonucleotides 1-1 and 2-2 for DC-SIGN2 (Table 1). An oligonucleotide that is DC-SIGN1 exon Ib-specific (oligonucleotide 1-3, Table 1) was used to amplify exons Ib-containing cDNAs. The amplified products were size-fractionated by electrophoresis on a 1.5% agarose gel. After denaturation in alkaline solution, the DNA was transferred to a nylon membrane (Amersham Pharmacia Biotech) by capillary action. Hybridization was performed with 32P-labeled oligonucleotide probes: (ii) an oligonucleotide derived from DC-SIGN1 sequences in exon Ib, exon Ic, exon II, and exon VI (oligonucleotides 1-4, 1-5, 1-6, and 1-8, respectively in Table 1); (iii) an oligonucleotide that had 11 nucleotides of the 3' end of exon 1c and 11 nucleotides of the 5' end of exon III of DC-SIGN1 (oligonucleotide 1-7, Table 1); (iii) an oligonucleotide that had identity with DC-SIGN2-specific exon II sequences (oligonucleotides 2-5 for DC-SIGN1 and oligonucleotides 1-1 and 2-2 for DC-SIGN2 (Table 1). One of the primers used for amplification was end-labeled with 32P to facilitate detection of the PCR products by autoradiography. Five μl of the PCR product was mixed with 15 μl of formamide dye (95% formamide, 10 mM EDTA, 0.2% bromphenol blue, 0.02% xylene cyanol) and boiled for 5 min. The mixture was then chilled and loaded on a 3 or 4% polyacrylamide gel containing 8 x urea and electrophoresed for 12 h at 200 V in a Protean II xi cell (Bio-Rad). The polyacrylamide gels were dried, and autoradiography was performed as described above.

In Vitro Translation—The TNT®-coupled Reticulocyte Lysate System (Promega) was used to translate in vitro DC-SIGN1 cDNAs cloned into pcDNA4/Himax TOPO vector (Invitrogen). The 78-meric translated products were fractionated in a 9% acrylamide gel and were contain exon Ib are designated as DC-SIGN1B mRNA transcripts. Sequence analysis of exon Ib-containing transcripts revealed two potential translation initiation sites (+1 or +101). Transcripts predicted to initiate translation at +101 in exon Ib may contain or lack the TM-encoding exon Ib. Dashed lines indicate the splicing events that lead to formation of the prototypic exon II-containing DC-SIGN1B mRNA transcript (mDC-SIGN1B Type I). Splicing out of the TM-encoding exon II generates transcripts designated as sDC-SIGN1B Types I–IV (Type I is the prototypic mDC-SIGN1B transcript). Splicing out of 20 m M Tris-HCl, 50 m M KCl, 1.5 m M MgCl2, 0.1 m M of each dNTP, 0.2 

ferred at 42 °C for 5 min and then stained for 30 min with the protein complex (Fig. 2). Color development was achieved using the glucose oxidase substrate kit (Vector Laboratories). Distilled water was used to block additional color development. For double staining, the sections were incubated in PBS for 5 min, and endogenous peroxidases were inhibited using a peroxidase block (Santa Cruz) for 5 min. Slides were then washed in PBS-Tween for 5 min, blocked with 5% bovine serum albumin for 30 min, and then incubated with one of the following: (i) PECA-1 Ab, (ii) PECA-1 Ab and its blocking peptide, (iii) CCR5 Ab, or (iv) CCR5 Ab and its blocking peptide. Subsequent steps for detection of goat primary antibodies was performed using the goat Immunocruz staining system according to the manufacturer's instructions (Vector Laboratories). Sections were incubated with diaminobenzidine for 10 min, and the reaction was stopped with distilled water. The sections were then dehydrated with graded alcohols and two washes in xylene and mounted with VectaMount™ (Vector Laboratories).

RESULTS

Genomic Organization of DC-SIGN1—In the course of identifying polymorphisms in DC-SIGN1, we identified several alternatively spliced DC-SIGN1 cDNAs (see below). To identify the genomic sequences homologous to these cDNAs, we determined the gene structure for human DC-SIGN1. Genomic DNA was subjected to PCR using primers corresponding to the known cDNA sequence (20); GenBank™ accession no. M98457), and the PCR products were cloned and sequenced. In addition, while this work was in progress, as part of the Human Genome Sequence Project, a ~143,619-bp contig of human chromosome 19p that contained DC-SIGN1 became available (GenBank™ accession no. AC008812). Other than a few polymorphisms, there was complete homology between the DC-SIGN1 genomic sequences that we had identified and those found in this contig (data not shown). Comparisons of the cDNA and genomic sequences revealed that the coding region of the previously described prototypic DC-SIGN1 cDNA (GenBank™ accession no. M98457) was encoded by six exons (Fig. 1a, top panel). The nomenclature for the exons was based on the alternate splicing events described below.
**FIG. 2.** Predicted structure and molecular diversity of membrane-bound and soluble DC-SIGN1 gene products with novel intra- and/or extracellular domains. 

- **a.** Gene organization of DC-SIGN1 and alternative splicing events that lead to the generation of the prototypic DC-SIGN1 protein product described by Curtis et al. (20). Boxes signify exons (I–VI) and dashed lines introns (I–V in black circles). The nucleotide length of the introns are shown in parentheses. The first nucleotide of the initiation Met codon of the prototypic DC-SIGN1A transcript is considered as +1. The stop codon used by the prototypic DC-SIGN1A isoform is denoted by an asterisk.

- **b.** Schematic illustrations of the molecular diversity and predicted structures of DC-SIGN1A and DC-SIGN1B isoforms generated by alternative splicing events. DC-SIGN1 variants that lack or contain exon Ib sequences (green box) are designated as DC-SIGN1A (b and c) or DC-SIGN1B (d and e) isoforms, respectively (Fig. 1). Predicted amino acid changes in the amino acid sequence of the prototypic DC-SIGN1A transcript.

- **c.** Schematic illustrations of the molecular diversity and predicted structures of DC-SIGN1A and DC-SIGN1B isoforms generated by alternative splicing events. DC-SIGN1 variants that lack or contain exon Ib sequences (green box) are designated as DC-SIGN1A (b and c) or DC-SIGN1B (d and e) isoforms, respectively (Fig. 1). Predicted amino acid changes in the amino acid sequence of the prototypic DC-SIGN1A transcript.

- **d.** Schematic illustrations of the molecular diversity and predicted structures of DC-SIGN1A and DC-SIGN1B isoforms generated by alternative splicing events. DC-SIGN1 variants that lack or contain exon Ib sequences (green box) are designated as DC-SIGN1A (b and c) or DC-SIGN1B (d and e) isoforms, respectively (Fig. 1). Predicted amino acid changes in the amino acid sequence of the prototypic DC-SIGN1A transcript.

- **e.** Schematic illustrations of the molecular diversity and predicted structures of DC-SIGN1A and DC-SIGN1B isoforms generated by alternative splicing events. DC-SIGN1 variants that lack or contain exon Ib sequences (green box) are designated as DC-SIGN1A (b and c) or DC-SIGN1B (d and e) isoforms, respectively (Fig. 1). Predicted amino acid changes in the amino acid sequence of the prototypic DC-SIGN1A transcript.
ternatively spliced exons identified in the DC-SIGN1 cDNAs (see below). Exons 1a and 1c encoded the majority of the cytoplasmic domain of the prototypic DC-SIGN1 cDNA (20). Exon II encoded 5 amino acids of the cytoplasmic domain and the entire transmembrane (TM) domain. Exon III encoded the repeats as well as a short stretch of amino acids that preceded the seven full repeats and the one-half repeat. Exons IV, V, and VI together encoded the predicted extracellular lectin-binding domain of DC-SIGN1.

**Extensive Structural Diversity of DC-SIGN1 Transcripts**—RT-PCR was used to amplify DC-SIGN1 cDNAs from PHA-activated PBMCs derived from normal human donors, human PBHP-derived mature DCs, and THP-1 monocytic cells. Sequence analyses of these PCR amplicons revealed several distinct cDNAs that shared homology to the previously reported prototypic DC-SIGN1 cDNA (Fig. 1, a and b; Ref. 20). These novel DC-SIGN1 transcripts differed from the originally reported cDNA sequence (GenBank™ accession no. M98457) by the presence or absence of stretches of sequences, indicating that they had arisen by a complex pattern of alternative splicing in the exons encoding the intra- or extracellular domains and/or by splicing out of exon II, the exon that encodes the predicted TM domain (Fig. 1, a and b). The predicted translation products of these transcripts are illustrated in Fig. 2 and shown in Supplementary Figs. 1 and 2.

Based on the structures predicted from their amino acid sequences, the DC-SIGN1 isoforms could be categorized into one of five major groups (Figs. 1, a and b, and 2; Supplementary Figs. 1 and 2), namely mDC-SIGN1A, sDC-SIGN1A, mDC-SIGN1B, sDC-SIGN1B, and truncated DC-SIGN1B (tDC-SIGN1B). The first group of transcripts designated as membrane-associated or mDC-SIGN1A isoforms had a Met (ATG) translation initiation codon within exon 1a and retained the exon predicted to encode the TM domain (exon II; Figs. 1a and 2a and b). These transcripts included the prototypic DC-SIGN1, designated here as mDC-SIGN1A Type I, as well as additional transcripts that are predicted to encode variable portions of the extracellular domain (Figs. 1a and 2a and b). For example, in mDC-SIGN1A Type II, the first 6 amino acids encoded by exon V are spliced out, whereas in mDC-SIGN1A Type III, some of the repeats encoded by exon III are spliced out (Figs. 1a and 2b).

The second group of transcripts was designated as sDC-SIGN1A. sDC-SIGN1A transcripts also had a Met (ATG) translation initiation codon within exon 1a, but the exon predicted to encode the TM domain (exon II) was spliced out, suggesting the synthesis of soluble forms of DC-SIGN1A (Figs. 1a, 2c, and 3a). The prototypic version of this class of transcripts, designated as sDC-SIGN1A Type I, lacked only the TM-containing exon II, whereas additional splicing events resulted in sDC-SIGN1A Types II–IV (Figs. 1a and 2c).

The exon Ib-containing DC-SIGN1 cDNAs were collectively designated as DC-SIGN1B transcripts and are predicted to encode the third (mDC-SIGN1B), fourth (sDC-SIGN1B), and fifth (truncated DC-SIGN1B) category of DC-SIGN1 isoforms (Figs. 1b and 2d and e). Notably, exons 1a, Ib, and 1c are sequences that are not interrupted by an intron and that collectively comprise exon I. There is a Met (ATG) translation initiation codon within exon Ib, and thus DC-SIGN1B transcripts can potentially initiate translation at two sites: +1 or +101 (Figs. 1b, 2d, and 3b). The sequence flanking the +101 position has a strong Kozak consensus sequence for initiation of translation (GCCATGG). The deduced amino acid sequence of transcripts that commence translation at the downstream Met codon (i.e. +101) in exon Ib differed from mDC-SIGN1A or sDC-SIGN1A isoforms only in the predicted cytoplasmic domain. These transcripts could be further categorized into those that had (DC-SIGN1B) or lacked (sDC-SIGN1B) the TM-encoding exon II (Figs. 1b, 2d, and 3c). Notably, prototypic m- or sDC-SIGN1 differed from m- or sDC-SIGN1B (Type I) by only 14 amino acids in the predicted N terminus encoded by exon Ib (Figs. 2 and 3c; Supplementary Figs. 1 and 2). Finally, usage of the Met codon in exon 1a in DC-SIGN1B transcripts predicted the production of a truncated protein of 41 amino acids (nucleotides +1 to +123; Figs. 1b, 2d, and 3b), and these isoforms were designated as truncated DC-SIGN1B isoforms (tDC-SIGN1B). To minimize the possibility that the exon Ib-containing DC-SIGN1 transcripts (i.e. DC-SIGN1B mRNAs) reflected PCR amplification of pre-mRNA contaminating the mRNA preparations, we confirmed the presence of these transcripts in poly(A⁺) RNA (see below, and data not shown).

Splicing events generated sDC-SIGN1-A or -B transcripts that are predicted to encode novel C termini (Figs. 1, a and b, 2, c and e, and 3, d–f). In some instances, the splice junctions for the DC-SIGN1 mRNA did not obey the consensus rules for 5′-intron/exon boundaries (Fig. 3c). Based on the splicing events in exons III–VI, these exons could be further subdivided (e.g. exon IIIa, IIIb, etc.). However, we refrained from doing so, recognizing that based on mRNA expression analyses there are probably additional splice variants that have not been discovered as of yet (see below). The model shown in Fig. 4 summarizes the predicted DC-SIGN1 gene structure, the primary transcript, mature mDC-SIGN1 (A or B) and sDC-SIGN1 (A or B) mRNAs, and a schema of the potential processing events underlying the formation of the mature messages. Collectively, the findings illustrated in Figs. 1–4 demonstrate that the DC-SIGN1 gene is subject to highly complex alternative splicing events, generating a wide array of transcripts that are predicted to encode for an extensive repertoire of membrane-associated as well as soluble DC-SIGN1 isoforms with variable intra- and/or extra-cellular regions.

**DC-SIGN2, a Gene with Structural Homology to DC-SIGN**—By searching the GenBank™ data bases, we found a cDNA (28) and two ESTs (Image Clones 146996 and 240697) that had high overall sequence homology with the DC-SIGN1 transcripts that we had identified. The cDNA and ESTs differed from each other by the presence or absence of additional stretches of sequences. To determine whether the cDNA and ESTs represented allelic versions of the DC-SIGN1 gene or products of a novel gene, RT-PCR was performed on human placenta mRNA using primers specific to those found in the cDNA and ESTs. Sequence analyses of the PCR products revealed additional novel cDNAs.

**differences among the isoforms, the source(s) from which their transcripts were cloned, and the length of the message (in nucleotides (nt)) and the predicted translated product (amino acids (aa)) are indicated to the right of the schema depicting the structural domains present in a given variant.** Panels e and e depict the transcripts that encode the isoforms predicted to lack the TM domain (i.e. isoforms lacking exon II). An in-frame initiation codon present at +101 in the exon Ib is predicted to commence translation at an intact open reading frame but with a novel cytoplasmic tail (see Figs. 1b and 3b). The splicing out of exon V (panel e, sDC-SIGN1B Type III) is predicted to generate a soluble variant with a novel C-terminal sequence (blue box). Similarly, splicing events in sDC-SIGN1 Type IV are predicted to result in a novel C terminus (Fig. 3e). † and †† denote the amino acid lengths of the DC-SIGN1B translated sequences that are predicted to initiate translation at either +1 in exon 1a (41 amino acids) or +101 in exon Ib (varying lengths). ∆ denotes skipping of the indicated exons/sequences. Because of splicing events in exons III–VI, these exons can be further subdivided (e.g. exon IIIA, IIIB, etc.); however, these demarcations are not indicated.
FIG. 3. Generation of sDC-SIGN and DC-SIGN1B isoforms and alignment of deduced amino acid sequences of prototypic mDC-SIGN1A, sDC-SIGN1A, mDC-SIGN1B, and DC-SIGN2. a, deduced amino acid sequences at the junctions of exon Ic and exon III generated by the splicing out of the TM-encoding exon II. b, deduced amino acid sequence of the N-terminal end of transcripts that contain exons Ia and Ib, i.e. DC-SIGN1B isoforms. The nucleotide sequence of exon Ia (red), Ib (black), and the initial portion of exon Ic (blue) are shown. The open reading frame initiated at the Met (ATG) codon in exon Ia is predicted to give rise to a truncated protein of 41 amino acids (MSD⁰⁹⁹⁴⁰...PRLstop), terminating with a stop codon in exon 1b (†/H11001–124). The open reading frame initiated from a start codon at †/H11001–103 in Exon Ib is predicted to encode DC-SIGN1B products that except for the N terminus (MASACPGSDFTSIHS, amino acid sequence in green) are identical to the prototypic DC-SIGN1A isoforms (see panel c). c, alignment of the deduced amino acid sequences of prototypic mDC-SIGN1A, sDC-SIGN1A, mDC-SIGN1B, and mDC-SIGN2 isoforms. Dots and dashes represent sequence identities and gaps, respectively. The cytoplasmic (CYT) domain, TM domain, the extracellular (EC) domain, which includes the repeats and the lectin binding domain, are indicated along with the deduced amino acid sequences encoded by the six exons of the prototypic mDC-SIGN1A (20) and mDC-SIGN1B. The first amino acid (I) of each of the repeats is in red. Note that the predicted amino acid sequences of mDC-SIGN1A Type I and mDC-SIGN1B Type I are identical beyond the first 14 amino acids of the cytoplasmic domain. The underlined sequence denotes the peptide sequence that was used for raising antiserum and is not found in DC-SIGN2. d-f, splicing patterns of exon II-lacking transcripts that are predicted to encode sDC-SIGN1 isoforms with novel C termini. Stop codons are boxed. The antisense orientation primer used for PCR amplification is underlined in panel f.
with sequences identical to the previously described cDNA/ESTs but distinct from DC-SIGN1 (A or B) transcripts, suggesting that they were alternatively spliced products of a distinct gene and not allelic variants of DC-SIGN1 (Fig. 5). The predicted translation products of these transcripts are illustrated in Fig. 6 and are also shown in Supplementary Fig. 3.

Genomic sequences identical to the novel DC-SIGN-like mRNAs that we had discovered as well as the previously identified cDNA (28) and ESTs were found 15.8 kb centromeric to DC-SIGN1, DC-SIGN2 transcripts in which the exon predicted to encode the TM domain (exon III) was spliced in or out were not found and were designated mDC-SIGN2 or sDC-SIGN2 isoforms, respectively (Figs. 5a and 6a–c). Additional alternative splicing events generated mDC-SIGN2 or sDC-SIGN2 transcripts, which are predicted to encode isoforms with varied extracellular domains (Figs. 5c, d and 6b, b and c). Notably, of the >30 DC-SIGN2 transcripts that we cloned and sequenced from the placenta of a normal donor, 21 cDNAs were found that contained sequences corresponding to intron IV, and in this particular placenta sample, we were unable to identify a prototypic mDC-SIGN2 transcript (Fig. 5). These findings provided the first clue that there might be significant interindividual variability in the repertoire of DC-SIGN2 transcripts expressed in term placenta. The discovery of DC-SIGN2 transcripts with distinct splicing patterns that contained intron IV and/or lacked exon VI from multiple sources (Fig. 6; e.g. ESTs and this study) indicated that the splicing patterns that we found were not aberrant or random events but rather may represent fairly common processing events.

A unique differential splicing event was observed that distinguished DC-SIGN2 mRNAs that contained (mDC-SIGN2) or lacked (sDC-SIGN2) the TM-encoding exon III. Among the DC-SIGN2 cDNAs that we cloned and sequenced, all sDC-SIGN2 transcripts contained sequences corresponding to exon IVa, but none of the transcripts that had the TM-encoding exon III, i.e. mDC-SIGN2 transcripts contained exon IVa sequence (Figs. 5, a and e and 6c). Exon IVa is predicted to encode a short hydrophobic stretch of amino acids (Fig. 5c and Supplementary Fig. 3).

It should be noted that intron I of the DC-SIGN2 gene corresponds to exon Ib of the DC-SIGN1 gene. Similar to the scenario observed in DC-SIGN1B, the use of an alternative translational start site at position 111 of intron I in DC-SIGN2 is predicted to encode isoforms with a novel intracellular domain. However, DC-SIGN2 transcripts that contained intron I sequences were not found in the cDNA clones that we have sequenced thus far.

Although these findings provided evidence for extensive alternative splicing events within the region preceding the lectin-binding domain (i.e. region that encodes the repeats) of DC-SIGN2, it was conceivable that in some individuals the variation in the number of repeats could be because of allelic variation. For example, it was conceivable that one allele could encode for eight repeats whereas the other allele could encode for seven repeats. To determine this, we amplified the genomic DNA that spanned the region between exon III and intron IV from normal donors. We found that in some instances, one allele encoded seven repeats whereas the other allele encoded
Gene and mRNA Structure of DC-SIGN1 and DC-SIGN2

**FIG. 5.** Molecular basis of the generation of DC-SIGN2 transcripts that are predicted to encode membrane-bound and soluble isoforms. **a,** the splicing patterns were inferred by comparing the cDNAs cloned with the genomic sequences of DC-SIGN2. The numbering system is based on the nucleotide sequence deposited under GenBank™ accession number AC008812, with the first nucleotide of the initiation Met codon of the prototypic mDC-SIGN2 (Type I) mRNA transcript is considered as **H11001**. This nucleotide corresponds to residue 66378 in the nucleotide sequence in GenBank™ accession number AC008812. (The nucleotide numbers in AC008812 that correspond to each of the DC-SIGN2 nucleotide numbers shown in this figure can be found in the notes that accompany GenBank™/EBI Data Bank accession numbers AY042234–AY042240.) The top panel is a schematic illustration of the DC-SIGN2 gene. Horizontal lines are exons (I–VIII), and dashed lines illustrate the splicing events that lead to the formation of the prototypic exon III-containing DC-SIGN2 mRNA transcript (mDC-SIGN2 Type I). Exon III is predicted to encode the TM domain (Fig. 6), and the exon III-retaining DC-SIGN2 mRNAs are predicted to encode membrane-bound or mDC-SIGN2 isoforms, whereas mRNAs that lack this TM-encoding exon III are predicted to encode soluble or sDC-SIGN2 isoforms. Alternative splicing events that lead to the generation of DC-SIGN2 mRNAs that contain or lack the TM-encoding exon III can be deduced by joining the various exonic sequences indicated. The starting and ending nucleotide number of each exonic segment is separated by dots (e.g. join 1..46), and exonic segments are separated from each other by a comma (e.g. join 1..46, 127..210). Asterisks indicate the stop codon utilized by most m- or sDC-SIGN2 transcripts, whereas the daggers indicate the stop codon used by most m- or sDC-SIGN2 transcripts, whereas the daggers indicate the stop codon utilized by most mDC-SIGN2 mRNAs Type V and VI at positions 5608–5610. Sequences corresponding to exon IVa were found only in the sDC-SIGN2 transcripts. Note that repeats 3–5 cannot be distinguished from each other; hence the splice junctions for mDC-SIGN2 type VI and sDC-SIGN2 Type I transcripts cannot be inferred. **b,** DC-SIGN2 mRNA transcripts that use non-canonical splice donor and/or acceptor sites. **c** and **d,** alternative splicing events that lead to the generation of sDC-SIGN2 isoforms with novel C termini. **e** denotes skipping of the indicated exons/sequences. Stop codons are boxed, deduced amino acid sequences encoded by exon IVa and alignment of the region bridging exon II and exon IVb in sDC-SIGN2 and mDC-SIGN2 isoforms.
FIG. 6. Predicted structure and molecular diversity of membrane-bound and soluble DC-SIGN2 gene products. a, gene organization of DC-SIGN2. Boxes are exons (I–VIII), and dashed lines are introns (I–VII in black circles). The nucleotide lengths of the introns are shown in parentheses. The first nucleotide of the initiation Met codon of the prototypic mDC-SIGN2 (Type I) mRNA transcript is considered as +1 (Fig. 5a). An asterisk denotes the stop codon found in the prototypic mDC-SIGN2 transcript. The box with vertical hatch lines represents the 3′-untranslated region. The predicted structure of the prototypic mDC-SIGN2 protein product is shown in panel b. Exons I, II, and a portion of exon III encode a short cytoplasmic domain (Cyt, open boxes); the TM domain (boxes with forward slash) is encoded by sequences in exon III. Exons IVb–VII encode the predicted extracellular domain of the prototypic mDC-SIGN2, and this includes a short stretch of sequence just proximal to the repeats (box with horizontal lines), the seven full repeats and one half-repeat (numbered black boxes), and the lectin-binding domain (box with backward slash). The green box represents the alternatively spliced exon IVa that is found only in those isoforms that lack the TM-encoding exon III. Image Clone 240607 was a partial cDNA clone that contained exons V, VI, and VII (data not shown). The alignment of the deduced amino acid sequences of the DC-SIGN2 isoforms depicted in this figure is shown in supplementary Fig. 3. b and c, schematic illustration of the molecular diversity and predicted structures of DC-SIGN2 isoforms generated by alternative splicing events. Panels b and c depict the transcripts that encode the isoforms that are predicted to contain the TM domain (mDC-SIGN2 isoforms) and isoforms that lack the TM domain (sDC-SIGN2 isoforms), respectively. Predicted amino acid (aa) differences among the isoforms and the source(s) from which their transcripts were cloned are indicated to the right of the schema depicting the structural domains present in a given variant. Retention of intron IV leads to formation of a novel C terminus in mDC-SIGN2 types II and IV and sDC-SIGN2 type II (blue box). Because of splicing out of exon VI, a novel C terminus is predicted to form in mDC-SIGN2 types V and VI (red box). The sDC-SIGN2 isoforms exclusively contain a short hydrophobic stretch of amino acids because of the presence of exon IVa (green box). The yellow box represents intron VII. Δ, denotes skipping of the indicated exons/sequences.
eight repeats. These findings suggested that in addition to alternative splicing, a variation in the number of repeats encoded in the \textit{DC-SIGN2} gene could be another source for variability in generating the DC-SIGN2 mRNA repertoire. Additional studies are under way to characterize the nature and frequency of this genetic polymorphism (\textit{i.e.} variability in number of repeats) in different ethnic populations. Studies are also underway to determine whether there is variability in the number of repeats in the \textit{DC-SIGN1} gene.

Additional inspection of the genomic contig from chromosome 19p13.3, demonstrated that the gene for the low affinity immunoglobulin \textit{Fc} receptor (CD23), another Type II lectin (23, 24), was situated \(43.3\) kb telomeric to \textit{DC-SIGN1} (Fig. 7). Thus, \textit{DC-SIGN1} (CD209), \textit{DC-SIGN2} (CD209L), and \textit{CD23} form a cluster of highly related genes, suggesting that they may have arisen by gene duplication of an ancestral gene, and notably alternative splicing events in all three genes lead to the generation of multiple transcripts (Figs. 1–7) (28, 30, 31).

Expression of DC-SIGN1 Is Not Restricted to DCs—Given the aforementioned findings, we asked whether the DC-SIGN1 transcripts were expressed in a complementary manner. That is, does a given cell type express only one DC-SIGN1 transcript, similar to the exclusive expression of odorant receptors in olfactory neurons (32), or are different DC-SIGN1 variants expressed in a combinatorial manner? In the first scenario, a given cell type could potentially be classified into one of five groups depending on which DC-SIGN1 transcript it expressed. In the second scenario, distinct transcripts could be coex-

\textbf{FIG. 7.} Colocalization of DC-SIGN1 (CD209), DC-SIGN2 (CD209L), and CD23 to within \(\sim 85\) kbp of chromosome 19p13.3. All three genes are subject to highly complex splicing events (23, 30, 31).

\textbf{FIG. 8.} Expression of DC-SIGN1 transcripts that lack or contain the TM-encoding exon in DCs and THP-1 cells. \textit{a}, the overall experimental strategy for the findings shown in panels \textit{b}–\textit{h} and in Fig. 9 are shown. Total RNA (1 \(\mu\)g) isolated from DCs derived from cytokine-differentiated CD34 \textsuperscript{+} PBHPs, PBMCs, placenta, THP-1 cell line, or other cell lines (data not shown) was reverse-transcribed with oligo(dT) primers. The resulting cDNA was PCR-amplified using DC-SIGN1A (primers 1-1 and 1-2) or DC-SIGN1B (primers 1-3 and 1-2)-specific primers. The PCR amplicons were fractionated by agarose gel (1.5\%-\%) electrophoresis, transferred to nylon membrane, and hybridized with the indicated radiolabeled probes. The blots were washed and then exposed for 15 h. \textit{b}–\textit{e}, specificity of the radiolabeled oligomers used. Nylon membranes spotted with the indicated DNA listed on the top of each blot were hybridized with the radiolabeled probe: \textit{b}, Exon VI (Ex VI); \textit{c}, Ex II; \textit{d}, Ex Ic-Ex III; \textit{e}, Ex Ib. Ex VI probe hybridizes all DC-SIGN1 (A or B) transcripts; Ex II probe hybridizes all DC-SIGN1 (A or B) transcripts that contain the TM-encoding exon II; Ex Ic-Ex III probe hybridizes DC-SIGN1 (A or B) transcripts that lack the TM-encoding exon II. (\textit{f}) DC-SIGN1 (A or B) expression in CD34 \textsuperscript{+} PBHPs-differentiating DCs cultured in the presence or absence of IL-4. Note, activation-induced differences in the levels of DC-SIGN1 expression (compare hybridizing signal in DCs \(\pm\) IL-4). The probes used are indicated to the right of each blot. \textit{g}, cDNAs amplified using DC-SIGN1B-specific primers from DCs derived from cytokine-differentiated CD34 \textsuperscript{+} PBHPs or THP-1 cells were fractionated by gel electrophoresis and Southern blot-hybridized with the radiolabeled Ex VI oligomer or a radiolabeled oligomer that is specific to DC-SIGN1B. Note, the additional hybridizing signal observed in the \textit{THP-1} lane. \textit{h}, DC-SIGN1 (A or B) expression in THP-1 cells obtained from ATCC. The molecular weight makers (in bp) are indicated to the left of the autoradiographs. Shorter exposures of panels \textit{f} and \textit{g} revealed a ladder of hybridizing signals, but they could not be completely fractionated by agarose gel electrophoresis (data not shown).
pressed in variable patterns to confer specific properties onto the expressing cells, with the variability being dependent on the ratio of expression of the different DC-SIGN1 mRNAs. An additional level of complexity could be that the expression patterns varied depending on the activation state and/or maturation stage of the cell.

To address the aforementioned question, a RT-PCR-based strategy that included Southern blot hybridization was used to determine the expression of DC-SIGN1 mRNAs in primary human cells and human cell lines (Fig. 8a). To perform semi-quantitative RT-PCR, in initial experiments we determined the number of PCR cycles wherein the hybridizing signal for DC-SIGN1 cDNAs were in the linear range (30 cycles), and PCR was performed using equal (1 μg) amounts of mRNA from each cell/tissue type.

To increase the specificity and to estimate the relative amounts of DC-SIGN1 mRNAs that had or lacked the TM-encoding exon II, five procedures were adopted. First, PCR was performed using unlabeled oligonucleotides specific for DC-SIGN1 or DC-SIGN1B (Table I), and the PCR products containing the DC-SIGN1 cDNAs were transferred to a membrane, and hybridized using DC-SIGN1 (A or B)-specific internal primers. This strategy assured that the hybridizing signal contained the DC-SIGN1-specific sequence and not nonspecific amplification.

Second, because DC-SIGN1 and DC-SIGN2 transcripts shared high sequence homology, the specificity of the nested radiolabeled DC-SIGN1 probes and washing conditions were optimized in control experiments using cloned DC-SIGN1 and DC-SIGN2 cDNAs (Fig. 8, b–c). Four nested radiolabeled oligomers were used in these hybridization studies (Fig. 8a and Table I). (i) The exon VI oligomer was designed to hybridize DC-SIGN1A and DC-SIGN1B transcripts regardless of whether they contained or lacked the TM-encoding exon II. This oligomer hybridized specifically to mDC-SIGN1, and a very faint cross-hybridizing signal was detected in mDC-SIGN2 cDNAs (Fig. 8b). (ii) The exon II oligomer was designed to hybridize transcripts that contained the TM-encoding exon II, i.e. mDC-SIGN1 (A or B) mRNAs. This probe specifically hybridized mDC-SIGN1 but not sDC-SIGN1, mDC-SIGN2, or sDC-SIGN2 cDNAs (Fig. 8c and data not shown). (iii) The exon Ic-exon III oligomer is specific for sDC-SIGN1 (A or B) DNA, i.e. transcripts that lacked exon II. Notably, this probe did not hybridize to DC-SIGN-1 or -2 transcripts that contained the exon II-encoding TM domain or to sDC-SIGN2 DNA (Fig. 8d and data not shown). (iv) The exon Ib oligomer was designed from a region that is not found in DC-SIGN1A transcripts, and in hybridization studies it was specific to m- or sDC-SIGN1B cDNAs (Fig. 8e and data not shown).

Third, to confirm that the DC-SIGN1 PCR primers used to generate the cDNAs were specific, the Southern blots shown in Figs. 8, f–h and 9 were stripped of radioactivity and reprobed with primers specific to DC-SIGN2. On rehybridization, DC-SIGN2 cross-hybridizing signals were not detected.

Fourth, because of the very faint cross-hybridization signals observed with the exon VI probe (Fig. 8b), we designed oligomers specific to DC-SIGN1 exon Ic (oligomer 1–5; Table I) and DC-SIGN2 exon II (oligomer 2–6; Table I). A set of Southern blots identical to those shown in Figs. 8 and 9 were hybridized with either a radiolabeled DC-SIGN1 exon Ic or DC-SIGN2 exon II probe. Hybridizing signals obtained with the DC-SIGN1 exon Ic probe were identical to those observed previously with the DC-SIGN1 exon VI probe. In contrast, a hybridizing signal was not detected with the DC-SIGN2 exon II probe, indicating that the mRNA expression patterns observed using the strategy outlined was specific for DC-SIGN1. As a final step to increase specificity and validate the expression pattern of DC-SIGN1 and DC-SIGN2 transcripts, cDNAs were synthesized from multiple different normal donors and cell lines.

An example from four separate experiments demonstrating
the cell and tissue expression of DC-SIGN1 transcripts is shown in Figs. 8 and 9. We first focused on the expression of DC-SIGN1 mRNA in CD34+/PBHP cells cytokine-differentiated toward the DC lineage (Fig. 8f). m- and sDC-SIGN1 (A or B) cDNAs were abundantly expressed in mature DCs, i.e. CD34+/PBHPs cytokine-differentiated for 15 days but not at earlier time points (Fig. 8, f and g). In addition to the prominent hybridizing signals of 1–1.3 kb in length, several hybridizing bands that were <1 kb in length were also detected (see below and data not shown). Notably, the hybridizing signal in CD34+/PBHPs differentiated with IL-4 was stronger than that observed in DCs cultured without IL-4 (day 15; Fig. 8, f and g), suggesting that the expression level of DC-SIGN1 mRNA may be dependent on the maturational/activation state of DCs. On longer exposures, faint hybridizing signals were evident at day 8 and 12 cytokine-differentiated CD34+/PBHPs, suggesting that the expression of DC-SIGN1 in immature DCs was significantly lower than that in mature DCs derived from CD34+/PBHPs.

In addition to DCs, m- and sDC-SIGN1 (A or B) transcripts were expressed in other antigen-presenting cells such as highly purified resting CD14+ monocytes (data not shown) as well as THP-1 and U937 cells, two monocytic cell lines (Fig. 8, g and h, and data not shown). Expression of DC-SIGN1 transcripts was confirmed in two independent sources of THP-1 cells (ATCC and National Institutes of Health AIDS repository; data not shown). Because it was difficult to control for differences in the labeling and hybridizing efficiencies of the different probes required to differentiate between the exon II-containing or -lacking DC-SIGN1 transcripts, it was not possible to assess in a quantitative manner their relative abundance in DCs or THP-1 cells. Nevertheless, the findings shown in Fig. 8 indicated that both m- and sDC-SIGN1 (A or B) transcripts are abundantly expressed in DCs and THP-1 cells.

Weak expression of DC-SIGN1 mRNA was detected in resting PBMCs obtained from eight normal donors (Fig. 9a and data not shown). In contrast, abundant expression for m- and sDC-SIGN1 (A or B) transcripts was detected in all eight PBMC samples after stimulation with PHA (Fig. 9, b–d and data not shown) as well as in PBMCs activated with CD3/CD28 (Fig. 9e). DC-SIGN1-specific hybridizing signals were evident in PBMCs activated with PHA for 4 days but not in PBMCs cultured in PHA (days 1–4) plus IL-2 (days 5–12; Fig. 9e). Notably, there was inter-individual variation in the expression of DC-SIGN1 transcripts in PHA-activated PBMCs (Fig. 9, b and d; e.g. compare hybridizing signals in donors 2 and 4 versus donors 1, 3, and 5 in panel d).

Because of our interest in the potential role of HIV attachment factors such as DC-SIGN1 in mother-to-child transmission of the virus, we also determined whether DC-SIGN1 is expressed in the placenta. Notably, we detected both inter-individual variation in the levels of DC-SIGN1 expression as well as heterogeneity in the repertoire of transcripts expressed (Fig. 9f; compare pattern of hybridizing signals in donor 2 versus 1 and 3). The expression of DC-SIGN1 in placenta was confirmed by immunohistochemical staining of term placentae. DC-SIGN1 expression colocalized with that of PECAM, an endothelial cell marker, as well with CCR5 (Fig. 10). The double immunostaining in Fig. 10C indicates that DC-SIGN1 is coexpressed along with CCR5 in placental villi, and the distribution pattern of CCR5+/DC−SIGN1+ cells is consistent with their expression in villous macrophages.

Weak DC-SIGN1-specific hybridizing signals of ~1.2 kb in length were also observed in MG63 (osteoblast) cells, HSB-2 (T cells), and MC116 cells, a B-cell line (data not shown). DC-SIGN1 expression was observed in the T cell line, HUT78; however, only an ~300 and ~600 bp hybridizing signal was detected in this cell type (data not shown). The presence or absence of hybridizing signals of varying sizes in T cells might reflect differences in the activation states of these cell lines. A ladder of hybridizing bands was also observed in HL-60 cells, a granulocytic cell line (data not shown).

The strongest hybridizing signals for DC-SIGN1 in mature DCs, PBMCs, placenta, and THP-1 cells were in the 1,000–1,300-bp range (Figs. 8 and 9), which was concordant with the large number of transcripts identified in this size range by direct cDNA sequencing (Figs. 1 and 2). However, the strong intensity of the hybridizing signals at ~1–1.3 kb masked the ladder of hybridizing bands that was evident on shorter exposures (data not shown) or as seen in Fig. 8g (Ex VI probe) as well as in Fig. 9d. Furthermore, we found it difficult to resolve this ladder of hybridizing bands using horizontal gel electrophoresis. To circumvent this limitation, the DC-SIGN1 sense orientation oligomer used in the aforementioned experiments was radiolabeled and used in PCR, and the amplicons were resolved on a polyacrylamide gel (Fig. 11a). The findings of these experiments revealed a ladder of PCR amplicons in all cell types having size ranges concordant with the lengths of the DC-SIGN1A or DC-SIGN1B cDNAs that we had identified by direct sequencing (Fig. 11a and data not shown). This ladder of PCR amplicons is consistent with the notion that DC-SIGN1 undergoes extensive splicing events to generate a large repertoire of transcripts of varying lengths. Fig. 11b illustrates that on gel electrophoresis, the lengths of the transcripts in the 1–1.3 kb size range may appear deceptively similar, and direct sequencing may be necessary to distinguish their unique sequence characteristics.

We next determined whether the transcripts predicted to encode membrane-associated and soluble DC-SIGN1 isoforms are translated in vitro (Fig. 11c). The in vitro translated products of the predicted sizes (epitope tag plus coding region) for both DC-SIGN1A and DC-SIGN1B products confirmed the integrity of the
coding regions of the transcripts shown in Figs. 1 and 2.

Expression Pattern of DC-SIGN2 Transcripts—To determine the expression of DC-SIGN2 transcripts, a strategy similar to that used to examine the expression of DC-SIGN1 was adopted (Fig. 12 and data not shown). In initial experiments, we observed that akin to DC-SIGN1, DC-SIGN2 transcripts were expressed in the placenta, and concordant with our isolation of cDNAs of varied lengths from this tissue, a ladder of amplicons was observed in some placental samples (Fig. 12a). However, in these initial experiments, we found that there was extensive inter-individual heterogeneity in not only the expression levels but also the repertoire of transcripts expressed. For example, we found that placenta from donor 3 lacked a transcript in the size range for the prototypic mDC-SIGN2 mRNA (Fig. 12a). This finding was notable because it may explain, in part, why we were unable to directly clone mDC-SIGN2 Type I transcripts from mRNA derived from this placenta sample. In agreement with our cDNA cloning studies, all four placenta samples had transcripts in the size ranges that were consistent with a repertoire of mDC-SIGN2 mRNA. As indicated earlier, variability in the length of the DC-SIGN2 transcripts may be accounted for, in part, by the variation in the number of repeats present on a given allele.

To extend and confirm these findings, we determined the expression of DC-SIGN2 transcripts in 10 additional term placentae from normal donors. Consistent with our initial studies (Fig. 12a), we found that there was striking heterogeneity in both the levels of expression of DC-SIGN2 transcripts as well as in the repertoire of transcripts expressed. Notably, despite equal expression for actin in all placental samples, we were unable to detect transcripts for DC-SIGN2 in 4 of the 10 placental samples, and only 2 of 10 placenta mRNA samples (samples 11 and 12) had transcripts with lengths corresponding to the prototypic mDC-SIGN2 mRNA.

DC-SIGN2 amplicons were also found in mature DCs (day 15 cytokine differentiated CD34⁺ PBHPs; Fig. 12c). However, using a RT-PCR Southern blot hybridization strategy similar to that shown in Fig. 8a, the expression of DC-SIGN2 in CD34⁺ PBHP-derived mature DCs was found to be significantly lower than that of DC-SIGN1 (data not shown). Using the Southern blot hybridization strategy, weak expression for DC-SIGN2 was also detected in THP-1 monocytic cells, whereas expression was not detected in CaCo2 (colorectal adenocarcinoma), RD (rhabdomyosarcoma), HUT 78 (T cell), MC116 (B cell) cells, or resting or activated PBMCs (data not shown).

DISCUSSION

DCs are thought to act as “Trojan horses,” capturing HIV in the mucosal surfaces for transport to the T cell areas of draining lymphoid tissues. The proficiency of DCs in interacting with T cells makes them prime candidates for enhancing viral infection. Recent reports indicate that DC-SIGN, a surface receptor with high expression in DCs, may play an important role in DC-T cell as well as DC-HIV interactions (16, 17). We have significantly extended these initial reports by (i) discovering that complex alternative splicing events in DC-SIGN (designated here as DC-SIGN1) pre-mRNA generates a wide repertoire of DC-SIGN1 transcripts. These DC-SIGN1 transcripts are predicted to encode both membrane-associated (mDC-SIGN1-A or -B) as well as soluble (sDC-SIGN1-A or -B) isoforms with varied intracellular and/or extracellular ligand (gp120/ICAM-3) binding domains. (ii) We have identified another highly homologous gene designated here as DC-SIGN2. Similar to DC-SIGN1, alternative splicing of DC-SIGN2 pre-mRNA also generates a wide repertoire of DC-SIGN2 transcripts that are predicted to encode membrane-associated and soluble isoforms. (iii) Interestingly, in addition to DC-SIGN1 (original version) and DC-SIGN2, we found that the low affinity immunoglobulin e Fe receptor (CD23) also maps to chromosome 19p13.3, forming a cluster of highly related genes that all undergo highly complex alternative splicing events (23, 24).

In contrast to previous reports (16, 17), we found that DC-SIGN1 mRNA expression is not restricted to DCs but is significantly broader and includes THP-1 monocytic cells, resting CD14⁺ monocytes, PBMCs, and placenta. Immunostaining with a DC-SIGN1-specific antibody indicated that DC-SIGN1 is expressed on placental endothelium as well as on CCR5⁺ cells. The distribution of these CCR5⁺ DC-SIGN1⁺ cells is con-
Alternative splicing is often tightly regulated in a cell type- or developmental stage-specific manner. Coordinated changes in alternative splicing patterns of multiple pre-mRNAs are an integral component of gene expression programs, like those involved in nervous system differentiation (46) and apoptotic cell death (47). Similar programs are also likely to exist during T cell and DC differentiation (48–50). In addition to cellular differentiation, the pattern of splicing can be influenced by the activation of particular signaling pathways (51–57). Notably, in our studies we found that the expression pattern of DC-SIGN1 transcripts may depend, in part, on the cell maturation/activation state (Figs. 8 and 9).

It is known that alternative splicing can generate mRNA structures that can take many different forms (33–35). Exons can be spliced into mRNA or skipped. Introns that are normally excised can be retained in the mRNA. The positions of either 5′ or 3′ splice sites can shift to make exons longer or shorter. In addition to these changes in splicing, alterations in transcriptional start site or polyadenylation site also allow the production of multiple mRNAs from a single gene. It is remarkable that nearly all of these variations in mRNA structure were observed in DC-SIGN1 and DC-SIGN2 transcripts (Figs. 1–6).

An emerging paradigm is the observation that proteins involved in cell-cell contact or recognition often exhibit a high degree of molecular diversity. Examples include genes for cadherins, cadherin-related neuronal receptors, olfactory receptors, and neurexins in the nervous system (36, 38–41, 58) and for immunoglobulin and T cell receptor genes in the immune system (59–61). In this context, it is notable that DC-SIGN1-mediated binding of DCs to ICAM-3 on resting T cells is thought to be a key initial adhesion step in the multistep process that leads to the formation of the immunological synapse and the activation of resting T cells (17). Thus, DC-SIGN1 (and potentially DC-SIGN2) demonstrates the generality of the features found in certain other genes involved in cell-cell adhesion/recognition. These common features include extensive alternative splicing events, cell type- and activation-specific expression, and a similar domain structure with distinct patterns of shared and divergent sequences.

In this report, we have demonstrated the genomic basis for the generation of not only several membrane-associated but also potentially soluble forms of DC-SIGN1 and DC-SIGN2. Furthermore, our studies suggest that the expression levels of DC-SIGN1 transcripts that lack the TM-coding exon are not minor variants of the overall pool of DC-SIGN1 mRNAs. Remarkably, the skipping of the TM-coding exon is observed in several type II membrane proteins that belong to the C-type animal lectin family (30, 31, 62–64), suggesting that this is an evolutionarily conserved property.

Because DC-SIGN-1 and -2 lack a leader sequence, it is not clear whether loss of the hydrophobic TM-encoding exon would limit the ability of these molecules to traverse across the endoplasmic reticulum membrane, resulting in their retention in the cytoplasm. However, there are examples among the lectin family wherein molecules lacking the secretory signal are externalized by mechanisms other than the classical secretory pathway (65). Notably, certain other cytoplasmic proteins lacking a signal sequence are externalized by mechanisms other than the classical secretory pathway. These include IL-1 (66), fibroblast growth factor (67, 68), and others (69, 70). Alternatively, these TM-lacking DC-SIGN isoforms may function as intracellular molecules. For example, the invariant or γ chain, another type II membrane protein, is responsible for targeting the Class II αβ dimers to the endocytic pathway that influences the delivery of antigens (71).

We found that the mRNA expression pattern for DC-SIGN1 was broader than reported previously (17). For example, we
cloned the transcripts of DC-SIGN1 from THP-1 cells and PBMCs. Expression of DC-SIGN1 mRNA, albeit low was detected in resting PBMCs. In contrast, in PBMCs stimulated with PHA or CD3/CD28 (stimulation of the T cell receptor) there was an increase in DC-SIGN1 mRNA expression. In studies not shown, DC-SIGN1 mRNA expression in PBMCs also increased significantly after stimulation with PMA and ionomycin, a calcium ionophore; this form of stimulation is known to activate the PKC pathway in T cells by bypassing the T-cell receptor. In ongoing studies we are investigating the precise cell types in resting as well as in PHA-, CD3/CD28-, and PMA/ionomycin-activated PBMC cultures that express DC-SIGN1 mRNA. It is difficult at the present moment to reconcile the differences between our findings and those of Geijtenbeek et al. (17) whose studies indicated that the expression of DC-SIGN1 is DC-specific. By using a PCR-based strategy they found no mRNA expression for DC-SIGN1 on THP-1 cells, granulocytes, PBMCs activated for 2 days with PHA and IL-2, or peripheral blood leukocytes (17). The reasons for this discrepancy remain unclear but could be related to differences in PCR conditions or primer design. We are currently in the process of generating monoclonal antibodies to determine whether there is a discordance between the levels of DC-SIGN1 mRNA and protein expression. Notably, there are several examples of tissue- or cell type-specific regulation of translation, including that for IL-2 (72–80).

We found that the genes for DC-SIGN1 (CD209), DC-SIGN2 (CD209L), and CD23 colocalize to an ∼55-kb region of chromosome 19p13.3. Alternative splicing events in CD23 generates several transcripts including two isoforms (FcεRIa/CD23a and FcεRIIb/CD23b) that differ only at the N-terminal cytoplasmic region (23, 30, 31). Interestingly, FcεRIa (CD23a) and FcεRIIb (CD23b) exhibit differences in their tissue expression, and IL-4 differentially regulates their expression (30, 81). These two CD23 isoforms also have differential functions in allergic reactions, immunity to parasitic infections, and B cell development (30, 81). As a corollary, we found that alternative splicing of DC-SIGN1 pre-mRNA also leads to the generation of transcripts that are predicted to encode distinct N-terminal regions (DC-SIGN1-A and -B) and that IL-4 differentially regulates the expression of DC-SIGN1 in DCs. There is growing evidence that lectins, including CD23, can serve as cell surface transducers of signals from the outside to the inside of the cell (82, 83); in this context, we are currently investigating whether DC-SIGN1-A and -B isoforms activate distinct intracellular signaling pathways.

The biological properties of this large repertoire of DC-SIGN1 and -2 isoforms with respect to their roles in HIV pathogenesis and DC-T cell interactions remains unknown. Changes in splicing have been shown to determine the ligand binding of growth factor receptors and cell adhesion molecules (33, 35). The mDC-SIGN1 and mDC-SIGN2 isoforms with varied extracellular domains may bind ligands, including gp120, with varied avidity. Furthermore, in addition to ICAM-3, this extensive array of membrane-associated DC-SIGN1s (and potentially mDC-SIGN2s) may mediate cell-cell contact via interactions with a larger number of specific ligands or adhesion molecules of different protein families. Studies are currently underway to determine whether, similar to the findings in other gene systems, an alternative splice variant of DC-SIGN-1 or -2 cross-regulates or antagonizes the biological activities of the other isoforms (47, 84–91). For example, an alternatively spliced isoform of CD40 influences the function of the prototypic full-length CD40 isoform (91). An intriguing possibility is that the DC-SIGN-1 and -2 isoforms lacking the transmembrane domain if secreted may act as natural competitive inhibitors of DC-SIGN/ICAM-3/HIV binding interactions in vivo, or alternatively, they may function in regulating the expression of the membrane forms of DC-SIGN. Furthermore, lectin-binding domains can oligomerize (92–96), and potentially this oligomerization among the varied membrane forms of DC-SIGN1 or between DC-SIGN1 and DC-SIGN2 isoforms in cell types in which they are coexpressed may further increase the repertoire and specificity of DC-SIGN-like surface proteins available for mediating cell-cell contact.

The prototypic membrane-associated DC-SIGN1 (mDC-SIGN1 Type I) and DC-SIGN2 (mDC-SIGN2 Type I) isoforms have been shown to mediate gp120 adhesivity and potentiate in trans the infection of T lymphocytes by HIV (16, 26). By mRNA expression studies and immunostaining, expression for DC-SIGN1 was detected in both placental endothelial cells and CCR5-expressing cells in which distribution was consistent with placental macrophages (HoBauer cells), a cell type that can support HIV infection (97). We also detected DC-SIGN2 transcripts in the placenta; and while this manuscript was in review, using a DC-SIGN2-specific antiserum, Pohlmann et al. (26) documented expression for DC-SIGN2 in the placental endothelium but not macrophages. The expression of both DC-SIGN1 and DC-SIGN2 in the placenta has important implications for vertical transmission of HIV-1. However, pertinent to our search for genetic determinants that account for the significant inter-individual variability in susceptibility to HIV infection, our studies indicate that DC-SIGN1 and DC-SIGN2 gene expression in the placenta and other cell types may be highly variable. We examined a large panel of placenta samples, and found inter-individual variation with respect to both the levels of expression as well as the repertoire of transcripts expressed. Notably, in some instances, we were unable to detect expression for the prototypic mDC-SIGN2 transcripts in placenta, and transcripts that contained intron IV appeared to be more abundant than the prototypic isoform. Conceivably, inter-individual variation in the generation of DC-SIGN isoforms could account, in part, for host differences in susceptibility to HIV-1 infection, especially vertical transmission.

In summary, while searching for polymorphisms in the gene for DC-SIGN1, we identified another homologous gene designated here as DC-SIGN2 that recently has been shown to serve also as an HIV attachment factor. Notably, we found that alternative splicing of DC-SIGN1 and DC-SIGN2 generates a wide array of transcripts that are predicted to encode both membrane-associated and soluble isoforms. Determining the functional properties of this extensive repertoire of DC-SIGN1 and DC-SIGN2 isoforms in vivo is likely to pose a daunting task, and in this respect it will be important to develop reagents that can discriminate between the different isoforms. In addition, the inter-individual heterogeneity in DC-SIGN expression, especially DC-SIGN2 in placenta, introduces an unanticipated degree of complexity with regard to dissecting the determinants of HIV susceptibility. Nevertheless, this plethora of DC-SIGN-like molecules will serve as a powerful tool to probe HIV-host cell interactions as well as DC-T cell interactions and as a potential target for a novel means to block these interactions. Based on the striking parallels between DC-SIGN-1 and -2 and other alternatively spliced type II membrane proteins such as CD23, we hypothesize that the diverse DC-SIGN isoforms have pleiotropic activities and that they may interact with additional, as yet undiscovered molecules.

Acknowledgments—We thank Dr. R. A. Clark for unwavering support and critical reading of the manuscript. We thank P. Melby, A. Valente, E. Gonzalez, N. Sato, M. Dolan, M. Quiniones, and J. Allan for insightful advice and M. D. Gamez for superb technical assistance. We thank Drs. R. Reddick and P. Valente for assistance in reading the immunohistochemistry slides. S. K. A. and S. S. A. thank A. S. Ahuja for guest on July 25, 2018
Extensive Repertoire of Membrane-bound and Soluble Dendritic Cell-specific ICAM-3-grabbing Nonintegrin 1 (DC-SIGN1) and DC-SIGN2 Isoforms: INTER-INDIVIDUAL VARIATION IN EXPRESSION OF DC-SIGN TRANSCRIPTS

Srinivas Mummidi, Gabriel Catano, LeeAnn Lam, Angelina Hoefle, Vanessa Telles, Kazi Begum, Fabio Jimenez, Seema S. Ahuja and Sunil K. Ahuja

J. Biol. Chem. 2001, 276:33196-33212. doi: 10.1074/jbc.M009807200 originally published online May 3, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M009807200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2001/08/28/276.35.33196.DC1

This article cites 97 references, 29 of which can be accessed free at http://www.jbc.org/content/276/35/33196.full.html#ref-list-1