Two Variable Regions in Carcinoembryonic Antigen-related Cell Adhesion Molecule 1 N-terminal Domains Located in or Next to Monoclonal Antibody and Adhesion Epitopes Show Evidence of Recombination in Rat but Not in Human*

Received for publication, April 21, 2004, and in revised form, June 7, 2004 Published, JBC Papers in Press, June 7, 2004, DOI 10.1074/jbc.M404431200

Meghan M. Comegys‡‡, Sue-Hwa Lin‡, David Rand†, Deborah Brit††, Donna Flanagan††, Helen Callanan††*, Kate Brilliant††*, and Douglas C. Hixson††‡‡

From †Scottsdale, Arizona 85262, the *Department of Molecular Pathology, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, the ‡Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, and the ‡‡Department of Medicine, Rhode Island Hospital/Brown University, Providence, Rhode Island 02903

In this paper, we have characterized the structure, evolutionary origin, and function of rat and human carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) multifunctional Ig-like cell adhesion proteins that are expressed by many epithelial tissues. Restriction enzyme digestion reverse transcriptase-PCR analysis identified three cDNAs encoding novel CEACAM1 N-domains. Comparative sequence analysis showed that human and rat CEACAM1 N-domains segregated into two groups differing in similarity to rat CEACAM1-4L and human CEACAM1. Sequence variability analysis indicated that both human and rat N-domains possessed two variable regions, and one contained a major adhesive epitope. Recombination analysis showed that the group of rat but not human N-domains with high sequence similarity was derived at least in part by recombination. Binding assays revealed that three monoclonal antibodies with strong reactivity for the CEACAM1-4L N-domain showed no reactivity with CEACAM1-4S, an allele with a different N-domain sequence. CEACAM1-4S displayed adhesive activity efficiently blocked by a synthetic peptide corresponding to the adhesive epitope in CEACAM1-4L. Blocking analysis also showed that the adhesive epitope for rat CEACAM1 was located downstream from the equivalent human and mouse epitopes. Glycosylation analysis demonstrated O-linked sugars on rat CEACAM1-4S from COS-1 cells. However, this was not the alteration responsible for the lack of monoclonal antibody reactivity. When considered together with previous studies, our findings suggest an inverse relationship between functionality and amino acid sequence similarity to CEACAM1. Like IgG, the N-domain of CEACAM1 appears to tolerate 10–15% sequence diversification without loss of function but begins to show altered specificity or diminished functionality at higher levels.

Carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) is a member of a large family of multifunctional Ig-like cell adhesion molecules (CAMs) structurally related to carcinoembryonic antigen (CEA) (2, 3). CEACAM1 from both rodents and humans is composed of an ectodomain with an N-terminal Ig V-like domain (N-domain), three Ig-like C-domains, and a cytoplasmic (cyto) domain that through differential splicing varies in length from 6 to 71 amino acids (4–6). Multiple genes with unique N-domain sequences and a variety of splice variants have been reported in both rodents and humans (2, 3, 5–12). The major splice variants in rodents and humans have from 2 to 4 Ig-like domains and cyto domains with either 70–71 (L forms) or 9–10 amino acids (S forms) (1–3, 5, 6). In rodents, allelic variants (Ceacam1a and 1b) or separate genes differing in both the nucleotide and amino acid sequence of their N-terminal Ig domains (rats and mice) have also been described (1, 13).

Interest in the role of CEACAM1 in cancer has blossomed since early reports showed that this gene was lost or greatly down-regulated in rodent hepatocellular carcinomas (4–16) and colon carcinomas (13, 17, 18) and that restoration of CEACAM1–4L expression in human prostate (19, 20), mouse colon (21, 22), or human bladder carcinomas (25) produced a marked decrease or loss of tumorigenicity. Relatively recent studies have also shown that the ectodomain is not required for tumor suppression; the CEACAM1-4L cyto domain is necessary and sufficient for inhibiting tumorigenicity (24). Cell-cell adhesion, on the other hand, appears to be an activity that is mediated primarily by the N-terminal V-like domain (4, 25–27). Because all of the CEACAM1 alleles and genes have

*This work was supported by National Institutes of Health Grants 64856 (to S.-H. L.) and CA42714 and COBRE Center for Cancer Research Development Grant 1P20RR17695 (to D. C. H.). 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.

‡ Present address: 6009 E. Westland Dr., Scottsdale, AZ 85262.

§ Present address: 6009 E. Westland Dr., Scottsdale, AZ 85262.

To whom correspondence should be addressed: Dept. of Medicine, Division of Hematology and Oncology, Rhode Island Hospital, 593 Eddy St., Providence, RI 02903. Tel.: 401-444-8058; Fax: 401-444-8141; E-mail: dhixson@lifespan.org.

The abbreviations used are: CEACAM1, carcinoembryonic antigen-related cell adhesion molecule 1; mAb, monoclonal antibody; CEA, carcinoembryonic antigen; CAM, cell adhesion molecule; V, immunoglobulin-like N-terminal variable region; ER, restriction enzyme digestion; RT, reverse transcriptase; RT-PCR/RE, restriction enzyme digestion of reverse transcriptase-PCR products; PCR/RE, restriction enzyme digestion of genomic PCR products; N-domain, N-terminal IgV-like domain; cyto, the 71- or 10-amino acid cytoplasmic domain of CEACAM1; RER, restriction enzyme-resistant N-domain PCR product; pAb, polyclonal rabbit antiserum; UTR, untranslated region; Sf9, Spodoptera frugiperda insect cells; HBSS, Hanks’ balanced salt solution; MHV, mouse hepatitis virus; RSA, bovine serum albumin; PSG, pregnancy-specific glycoprotein; PBS, phosphate-buffered saline; IIF, indirect immunofluorescence; CEACAM1r, CEACAM1s, rat proteins encoded by the Ceacam1r and Ceacam1s genes.

1 The abbreviations used are: CEACAM1, carcinoembryonic antigen-related cell adhesion molecule 1; mAb, monoclonal antibody; CEA, carcinoembryonic antigen; CAM, cell adhesion molecule; V, immunoglobulin-like N-terminal variable region; ER, restriction enzyme digestion; RT, reverse transcriptase; RT-PCR/RE, restriction enzyme digestion of reverse transcriptase-PCR products; PCR/RE, restriction enzyme digestion of genomic PCR products; N-domain, N-terminal IgV-like domain; cyto, the 71- or 10-amino acid cytoplasmic domain of CEACAM1; RER, restriction enzyme-resistant N-domain PCR product; pAb, polyclonal rabbit antiserum; UTR, untranslated region; Sf9, Spodoptera frugiperda insect cells; HBSS, Hanks’ balanced salt solution; MHV, mouse hepatitis virus; RSA, bovine serum albumin; PSG, pregnancy-specific glycoprotein; PBS, phosphate-buffered saline; IIF, indirect immunofluorescence; CEACAM1r, CEACAM1s, rat proteins encoded by the Ceacam1r and Ceacam1s genes.

2 All abbreviations for CEACAM1 family members in mouse, rat, and human are according to the nomenclature reported in Beauchemin et al. (1).
unique N-domain sequences, it follows that these sequence variations may in turn produce changes in conformation that alter the adhesive properties of the N-terminal domain. This possibility is consistent with a recent report by Watt et al. (28) demonstrating that single amino acid mutations could alter monoclonal antibody (mAb) binding and cell-cell adhesion activity.

Over the past 10 years, considerable insight has been gained into the effects that changes in amino acid sequence have on the structure, tumor suppression, or cell-cell adhesion activity of N-domains from human and mouse CEACAM1. Surprisingly, however, there have only been a few studies of this nature for rat CEACAM1. Sippel et al. (25), for example, demonstrated that mutation of Ser-503 abrogated CEACAM1-mediated aggregation, whereas Estrera et al. (29) determined that phosphorylation of Ser-503 was essential for tumor suppression.

To further gain insight into the characteristics and origin of Ceacam1 N-domains, we have utilized genomic PCR to search the rat genome for additional N-domain encoding sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences. By using this approach, we were successful in identifying three novel N-domains by their resistance to REs that cut sequences.

In three different assays, mAb 362.50 reacted with peptide arrays spanning an epitope that was recognized most strongly by mAb 9.2. Because peptide arrays defined two N-terminal and one C-terminal epitope, differences in amino acid sequences of the six most similar rat N-domains, we examined the effects of naturally occurring differences in the amino acid sequences of the CEACAM1a and 1b N-domains, the two major functional properties of CEACAM1 N-domains as based on previous studies by Lin and co-workers (31). For flow cytometric analysis, cells in suspension were labeled by indirect immunofluorescence (IIF) as described previously (40) by using phyceroerythrin-conjugated anti-mouse IgG. CEACAM1 expression levels were analyzed according to the manufacturer’s protocol by using a Guava PC flow cytometer equipped with Guava Express software (Guava Technologies Inc., Hayward, CA). Data were presented as a graph of cell counts versus fluorescence intensity. Median and peak fluorescence values were determined using Guava Express software.

**Materials and Methods**

**Peptide Synthesis**—The numbering of the peptides corresponds to the published amino acid sequences of CEACAM1a-4L (5) with residue 1 corresponding to the initiation methionine. Peptides were synthesized and purified by high performance liquid chromatography, as described previously (34). Peptides corresponding to residues 75–72 (9.2A, NLQFQEQVFVYKGT) and 75–90 (9.2B, PDSEIARYHSDNMSK) were synthesized.

**Preparation of Antibodies**—The origin and characteristics of pAb 669 (35) and mAbs 362.50 and 5.4 (16, 36) have been described previously.

**Indirect Immunofluorescence Analysis**—At 24–48 h after transfection, COS cells were fixed with 4% formaldehyde and permeabilized with 0.1% Triton X-100 in PBS overnight at 4°C in the dark. Matrix was washed sequentially with PBS, 0.5% Nonidet P-40 detergent and 100 mM sodium acetate, pH 5.5.

**Radiolabeling Procedures and Immunoprecipitation Analysis**—Retrovirally transduced COS cells stably expressing either the gene for Ceacam1-4L or Ceacam1-4S expression plasmids, COS-1 cultures growing in chamber slides were fixed in acetone and labeled by an indirect immunofluorescence protocol as described previously (40). After blocking with 1% fetal bovine serum (FBS) and 5% normal goat serum (NGS), matrix was washed sequentially with 0.2 M triethanolamine and 0.5% Triton X-100 in 0.5 M NaCl. Matrix was pelleted at 10,000 × g and washed sequentially with PBS, 0.5% Triton X-100 in PBS and 0.2 M triethanolamine, pH 8.0. The IgG was cross-linked to the matrix using 50 mM dimethyl pimelidimide-2HCl (Pierce) in 0.2 M triethanolamine, pH 8.0, for 45 min with constant end-over-end mixing. Matrix was washed three times with 0.2 M triethanolamine and 0.5% Triton X-100 in 0.5 M NaCl. Prior to deglycosylation and immunoblot analysis, 5 × 10⁶ transfected COS cells or 1 × 10⁶ rat hepatocytes isolated by a collagenase perfusion protocol were harvested with 100 µl of Affi-Gel protein-A agarose (Bio-Rad) suspended in PBS overnight at 4°C with constant end-over-end mixing.
Western blotting detection system from Amersham Biosciences accord-
mal Fischer rat hepatocytes were developed with a chemiluminescent
pH 7.4.
mAbs 9.2 and 362.50 purified on protein-G-Sepharose were labeled with
remaining digested samples were incubated an additional 18 h at 37
min. Aliquots (50
were terminated by adding 0.25 volume of 5
with 0.5 units of
O
CEACAM1a-4L, CEACAM1b-4S, and CEACAM1a-4L and -4S from nor-
from Cappel/Organon Teknika. Blots prepared with deglycosylated
cellulose filters were labeled with mAb 9.2, mAb 362.50, or pAb 669 by
stained by using a silver staining kit (Bio-Rad). After transfer, nitro-
onto nitrocellulose membrane filters (Schleicher & Schuell) was carried
stant voltage of 200 V. Transfer of proteins separated by SDS-PAGE
coating. Liver
samples were snap-frozen in liquid nitrogen according to the Easy DNA kit (Invitrogen) specifications. Liver
cDNAs. No template controls were also run. The 485-bp products of the
GD1AS, 5
in liquid nitrogen according to the Easy DNA kit (Invitrogen) specifi-
cations for small amounts of tissue (100–200 mg) were dissolved in liquid
agarose gels containing ethidium bromide.
no template controls were also run. The 485-bp products of the
\[5\text{-UTR/3'}\text{UTR} -\text{.arrow by guest on July 25, 2018http://www.jbc.org/Downloaded from]
Cloning of Ceacam1 N-terminal Ig Domains—Ceacam1 N-domain cDNAs were amplified with the GD1S and GD1AS primers using genomic DNA isolated from Fischer 344 rat liver. PCR products were ligated into the pCR2 TA cloning vector (Invitrogen) according to the manufacturer’s specifications. The ligation mix was then used to transform competent INViTRO (Invitrogen) cells. Blue-white screening of recombinants was subsequently performed according to the protocol provided by the manufacturer. Plasmid DNA obtained from transformed white colonies using a Qiagen Spin Miniprep Kit (Qiagen, Valencia, CA) was screened for the presence of Ceacam1 N-domain insert by both restriction digest and RT-PCR protocol and A2 constructs described above. A2 constructs were digested with HincII and RE-resistant (RER) N-domains were then digested with HincII to eliminate any plasmids containing the Ceacam1 N-domain cDNA. To obtain a population of PCR products enriched in RER N-domains, the initial PCR products were sequentially digested to completion with HincII and XbaI, enzymes specific for the Ceacam1 and Ceacam10 N-domains. The digested PCR products were successively digested with HincII and XbaI, plasmids containing the Ceacam1 N-domain cDNA were amplified with the GD1S and GD1AS primers as described above. For the cloning of Ceacam10 N-domain, genomic PCR products were directly digested with the pCR2 TA cloning vector. The ligation mix was then digested with HincII to eliminate any contaminating Ceacam1 N-domain containing plasmids. The Ceacam10 and RER N-domain-enriched ligation mixtures used to transform competent INViTRO cells and plasmid DNA from recombinant white colonies were screened by RE digest and PCR/RE assays. The authenticity of the Ceacam1 and Ceacam10 N-domain inserts and the sequence of the unique RER N-domain sequences were confirmed by automated sequencing as described previously (19), using the following vector-specific primers: forward SP6 primer (5′-CTATTTAGGTGACACTATAG-3′) and reverse T7 promoter primer (5′-TAATACGACTCACTATAGGG-3′). Automated sequence readouts were transferred into the MacVector program for Macintosh and were aligned with the known sequences of the Ceacam1, 1, 1, and 10 N-domains.

Cloning and Sequence Analysis of N-domain PCR Product Resistant to RE Digestion—The undigested 314-bp PCR products remaining from the sequential HincII/XbaI digestion were cloned into the pCR2 TA cloning vector (Invitrogen), and forward (T7) and reverse (SP6) primers were used to obtain the sequence both strands of DNA from each clone. N-domain accuracy of the sequenced clones was confirmed by automatic sequencing as described previously (19), using the following vector-specific primers: forward SP6 primer (5′-CTATTTAGGTGACACTATAG-3′) and reverse T7 promoter primer (5′-TAATACGACTCACTATAGGG-3′). Automated sequence readouts were transferred into the MacVector program for Macintosh and were aligned with the known sequences of the Ceacam1, 1, 1, and 10 N-domains.

Cloning and Sequence Analysis of N-domain PCR Product Resistant to RE Digestion—The undigested 314-bp PCR products remaining from the sequential HincII/XbaI digestion were cloned into the pCR2 TA cloning vector (Invitrogen), and forward (T7) and reverse (SP6) primers were used to obtain the sequence both strands of DNA from each clone. N-domain accuracy of the sequenced clones was confirmed by automatic sequencing as described previously (19), using the following vector-specific primers: forward SP6 primer (5′-CTATTTAGGTGACACTATAG-3′) and reverse T7 promoter primer (5′-TAATACGACTCACTATAGGG-3′). Automated sequence readouts were transferred into the MacVector program for Macintosh and were aligned with the known sequences of the Ceacam1, 1, 1, and 10 N-domains.

COS-1 cells (American Type Culture Collection, Manassas, VA) were grown in 2-well plastic chamber slides in Dulbecco’s modified Eagle’s medium high glucose (Invitrogen) containing 10% fetal bovine serum (Hyclone, Logan, UT). At 60–70% confluence, cultures were transfected for 3 h at 37°C with 15 μg of one of the two expression plasmids mixed with 70 μl of N-[1,2,3-dioleoyl]glycerol]N,N,N-trimethylammonium salts (Roche Applied Science) or Lipofectamine (Molecular Probes) according to the manufacturer’s instructions. Cultured COS-1 cells, AML12, a well differentiated, nontumorigenic cell line generated from the liver of a TFG-a transgenic mouse (48) were transfected by essentially the same protocol. The AML-12 cell line was kindly provided by Dr. Nelson Fausto, University of Washington, Seattle, WA. The construction of the retroviral expression plasmid LNCL containing the Ceacam1-4L cDNA has been described previously (19). Another retroviral expression plasmid, LNCS, containing the Ceacam1-4S cDNA was also constructed for these experiments (19). Retroviral plasmids were transfected into the ecotropic BOSC23 (49) and the amphotropic PA317 (50) retroviral packaging cell lines as described previously (19). The human colon carcinoma cell lines (M-HuCC) stably transfected with M-HuCC cells carrying either the Ceacam1-4L or Ceacam1-4S cDNAs were kindly provided by Dr. Sue-Hwa Lin. COS-1, PA290, and PA317 cells were examined 48 h post-transfection by IIF with polyclonal antibody (pAb) 669 to determine the efficiency of transfection before further manipulation of the cells.

Expression of Ceacam1-4L and Ceacam1-4S Isoforms of CEACAM in Retrovirally Transduced Cells—Culture medium containing the LNCL (Ceacam1-4L) or LNCS (Ceacam1-4S) ecotropic retrovirus was produced in BOSC23 packaging cell lines and used to infect PA317 cells. Culture supernatants containing amphotropic retroviruses produced by PA317 cells as described by Comegys et al. (19) were used to infect the COS-1 cells, the AC1 rat-derived transplantable hepatocellular carcinoma cell lines, 25T and 25T, and the PA317 cells, AML12, a well differentiated, nontumorigenic cell line (48). Infection of COS-1, 25T, 1682A, and PA317 cells were cultured in the presence of 750 μg/ml G418, respectively, for 7–14 days, and drug-resistant colonies were selected by drug cloning. Drug-resistant colonies were maintained in Dulbecco’s modified Eagle’s medium (GIBCO, Paisley, Scotland), or in Dulbecco’s modified Eagle’s medium (Invitrogen). Drug-resistant colonies were identified by 10% Rehataulin fetal bovine serum (Intergen, Purchase, NY). Clonal lines derived by limited dilution cloning were screened by IIF with pAb to identify clones expressing the highest levels of CEACAM1 protein. To eliminate contamination with negative cells, CEACAM1-4L-expressing cells were selected from COS-1, 25T, and 1682A clones that were selected by IIF (Dynabeads M-450, Dynal, Inc., Vista, CA) according to the manufacturer’s instructions. To ensure a high percentage of product with the A overhangs needed for TA cloning, the gel-purified product enriched for RER N-domains was reamplified using the GD1S and GD1AS primers. Following ligation of enriched ligation mixtures used to transform competent INV (Invitrogen) cells, blue-white screening of recombinant white colonies was performed by 10% Rehataulin fetal bovine serum (Intergen, Purchase, NY). Clonal lines derived by limited dilution cloning were screened by IIF with pAb to identify clones expressing the highest levels of CEACAM1 protein. To eliminate contamination with negative cells, CEACAM1-4L-expressing cells were selected from COS-1, 25T, and 1682A clones that were selected by IIF (Dynabeads M-450, Dynal, Inc., Vista, CA) according to the manufacturer’s instructions.

Expression of Ceacam1-4L and Ceacam1-4S Isoforms of CEACAM in Retrovirally Transduced Cells—Culture medium containing the LNCL (Ceacam1-4L) or LNCS (Ceacam1-4S) ecotropic retrovirus was produced in BOSC23 packaging cell lines and used to infect PA317 cells. Culture supernatants containing amphotropic retroviruses produced by PA317 cells as described by Comegys et al. (19) were used to infect the COS-1 cells, the AC1 rat-derived transplantable hepatocellular carcinoma cell lines, 25T and 25T, and the PA317 cells, AML12, a well differentiated, nontumorigenic cell line (48). Infection of COS-1, 25T, 1682A, and PA317 cells were cultured in the presence of 750μg/ml G418, respectively, for 7–14 days, and drug-resistant colonies were selected by drug cloning. Drug-resistant colonies were maintained in Dulbecco’s modified Eagle’s medium (GIBCO, Paisley, Scotland), or in Dulbecco’s modified Eagle’s medium (Invitrogen). Drug-resistant colonies were identified by 10% Rehataulin fetal bovine serum (Intergen, Purchase, NY). Clonal lines derived by limited dilution cloning were screened by IIF with pAb to identify clones expressing the highest levels of CEACAM1 protein. To eliminate contamination with negative cells, CEACAM1-4L-expressing cells were selected from COS-1, 25T, and 1682A clones that were selected by IIF (Dynabeads M-450, Dynal, Inc., Vista, CA) according to the manufacturer’s instructions. To ensure a high percentage of product with the A overhangs needed for TA cloning, the gel-purified product enriched for RER N-domains was reamplified using the GD1S and GD1AS primers. Following ligation of the resulting PCR product into the pCR2 TA cloning vector, the ligation mix was digested with HindIII to eliminate any contaminating Ceacam1 N-domain containing plasmids. The Ceacam10 and RER N-domain-enriched ligation mixtures used to transform competent INViTRO cells and plasmid DNA from recombinant white colonies were screened by RE digest and PCR/RE assays. The authenticity of the Ceacam1 and Ceacam10 N-domain inserts and the sequence of the unique RER N-domain sequences were confirmed by automated sequencing as described previously (19), using the following vector-specific primers: forward SP6 primer (5′-CTATTTAGGTGACACTATAG-3′) and reverse T7 promoter primer (5′-TAATACGACTCACTATAGGG-3′). Automated sequence readouts were transferred into the MacVector program for Macintosh and were aligned with the known sequences of the Ceacam1, 1, 1, and 10 N-domains.
FIG. 1. RT-PCR and PCR analysis of CEACAM isoforms expressed in rat liver. A, schematic diagram of the two major CEACAM isoform cDNAs, adapted from Culic et al. (5) and Earley et al. (8). Nucleotides are numbered according to Culic et al. (5). Location of the primers used to amplify N-domain sequences in both RT-PCR, (S-UTR/Sig and D2AS) and genomic PCR (GD1S and GD1AS) analyses are shown. Sig, signal sequence; N1a or N1b, A1, B, and A2, Ig-like domains; TM, transmembrane domain; cyto, cyto domain. Narrow solid boxes represent the untranslated regions; different patterns in the N-domain boxes denote sequence differences between the CEACAM1α and CEACAM1β N-domains.

B, analysis of PCR products from control plasmid DNAs using 5'-UTR/Sig and D2AS primer pair followed by RE analysis with DraI (D) specific for C1β, HincII (H) specific for C1α, or both enzymes (D + H). Undigested (U) samples appear adjacent to (left side) digested samples, and product sizes are indicated on the right. Lane 1, 1-kbp DNA ladder; 500-bp marker indicated at the left; lanes 2–5, pCDM8 C1β-4S DNA; and lanes 6–9, pCDM8 C1α-4L DNA. C, analysis of RT-PCR products from liver RNAs of different rat strains using 5'-UTR/Sig/D2AS primer pair followed by RE analysis described in B. Lane 1, 1-kbp DNA ladder; 500-bp marker indicated at the left; L1 (lanes 2–5): German Fischer; L2 (lanes 6–9): American Fischer; L3 (lanes 10–13): ACI; L4 (lanes 14–17): Japanese Fischer and L5 (lanes 18–21): Sprague-Dawley. D, analysis of PCR products from control plasmid DNAs followed by RE digestion. N-domains present in genomic DNA from American Fischer rat liver were amplified by primer pair GD1S/GD1AS and then subjected to RE analysis with DraI (D), specific for CEACAM1β; HincII (H), specific for CEACAM1α; XbaI (X), specific for CEACAM10; a combination of DraI and HincII (D + H); or a combination of XbaI and HincII (X + H). Undigested (U) samples appear adjacent to (left side) digested samples, and product sizes are indicated on the right. Lane 1, 50–2000-bp ladder, sizes marked at left; lanes 2–5, pCDM8 CEACAM1β-4S DNA, and lanes 6–9, pCDM8 CEACAM1α-4L DNA. E, analysis of PCR products from American Fischer liver genomic DNA followed by RE analysis described above. Lane 1, 50–2000-bp ladder, sizes marked at left; Lanes 2–7, genomic DNA from American Fischer rat liver. Note: remaining 314-bp product after XbaI and HincII digestion in lane 7 contains the newly identified N-domains.
Briefly, sterile 100-mm Petri dishes coated with 10 \( \mu g/ml \) unconjugated goat anti-rabbit IgG (Pierce) in carbonate buffer (15 mM sodium carbonate, 35 mM sodium bicarbonate, pH 9.5) for 90 min at room temperature were washed sequentially with Hank's-buffered saline solution (HBSS) (Invitrogen) and HBSS containing 0.1% bovine serum albumin (BSA) and 0.2 mM EGTA (HBSS-BE). Infected cells (1–2 \( \times 10^7 \)) trypsinized from culture plates were collected in HBSS-BE containing 4.2 \( \mu g/ml \) aprotinin (Sigma), washed in HBSS-BE, resuspended in a 1:200 dilution of pAb 669 in HBSS-BE, and incubated on a rotator at 4°C for 20–30 min. Following two washes in HBSS-BE, antibody-coated cells were plated into goat anti-mouse IgG-coated plates, incubated for 40 min at 4°C, swirled to redistribute unbound cells, and incubated an additional 30 min at 4°C. Dishes were then washed six times with HBSS-BE, decanting and discarding the supernatants from each wash. After the final wash, culture medium with serum was added to each plate, and bound cells were released in a stream of medium generated with a pipette.

RESULTS

The CEACAM1\(^b\) Allele Is Not Present in the Fischer Rat Genome—PCR of control plasmid DNA pCDM8 carrying the Ceacam1\(^a\)-4L or Ceacam1\(^b\)-4S cDNAs defined in this report is indicated by the symbols shown on the figure. Regions of high variability depicted in Fig. 3 are designated by light background shading (A).
sulting in fragments of 352 and 133 bp, whereas Ceacam1b-derived products were cut only by DraI resulting in fragments of 245 and 240 bp (Fig. 1B). RT-PCR analysis of total RNA from rat liver of several different strains, including American Fischer 344, German Fischer 344, Japanese Fischer 344, ACI, and Sprague-Dawley, using the 5′-UTR/Sig and D2AS primers also resulted in the amplification of 485-bp products (Fig. 1C). RE analysis of these 485-bp products revealed that they could be cut with HincII, resulting in fragments of 352 and 133 bp identical in size to those produced by digestion of control Ceacam1a plasmid DNA (Fig. 1C). However, none of the 485-bp products amplified from rat liver contained the DraI enzyme site unique to the Ceacam1b N-domain (Fig. 1C), indicating that both of the two major cyto domain splice variants expressed in the liver of these rat strains and substrains had the Ceacam1a N-domain.

To determine whether the lack of Ceacam1b-4S transcripts was due to the absence of this allele in the genome, PCR/RE analysis using primers confined to the N-domain of Ceacam1, designated GD1S and GD1AS (Fig. 1A), was performed on genomic DNA from rat liver. PCR of pCDM8 plasmids carrying Ceacam1a-4L or Ceacam1b-4S cDNAs and genomic DNA from American Fischer rat liver resulted in products of the predicted 314-bp size (Fig. 1D). Digestion of the CEACAM1a-derived and CEACAM1b-derived control PCR products with HincII and DraI produced fragments of the expected sizes of 234/80 and 192/122 bp, respectively (Fig. 1D). In addition to DraI and HincII, PCR products derived from genomic DNA were cut with XbaI, the RE unique to the Ceacom1b N-domain, which, although not expressed by rat liver, would be present in the genomic DNA. DraI and/or HincII digestion of N-domains amplified from American Fischer liver genomic DNA only produced 234- and 80-bp fragments consistent with the presence of a Ceacam1a but not a Ceacam1b N-domain (Fig. 1E). This demonstrated that the Ceacam1 allele is not present in the genome of this inbred strain of rats, thereby eliminating the need to digest with DraI when searching for RE-resistant N-domains. Subsequent XbaI digestions produced predicted fragment sizes of 190 and 124 bp (Fig. 1E), and the results were consistent with the presence of the Ceacam1b N-domain.

The Fischer Rat Genome Contains Three Novel N-domains Resistant to RE Digestion—After sequential and complete digestion of the 314-bp amplified N-domain products from genomic DNA with HincII and XbaI, a residual product resistant to digestion (Fig. 1E) was consistently observed, suggesting the existence of unidentified N-domains. Ceacam1b could not have been present in the resistant product because it contains the Ceacam1b HincII site. The possibility that these RER N-domains were derived from Ceacam9 and Ceacam11 also seemed unlikely because the primers chosen for amplification did not match the corresponding sequences in these two genes. The undigested 314-bp PCR products remaining from the

![Graph](image-url)
sequential HincII/XbaI digestion were cloned into the pCR2 TA cloning vector (Invitrogen), and forward (T7) and reverse (SP6) primers were used to obtain insert sequences off both strands of DNA from several clones. To verify the accuracy of the PCR-derived inserts, sequences obtained from clones determined by RE digestion to contain the Ceacam1 and the Ceacam10 N-domains were shown to be identical to published sequences for these two N-domain isoforms (4, 6, 8). From the HincII/XbaI-resistant clones, three (designated Nx, Ny and Nz) were identified that contained 4–24 nucleotide changes relative to the Ceacam1a N-domain and had nucleotide changes in the cleavage sites for HincII and XbaI. RE-RT-PCR analysis using total RNA from small intestine, colon, heart, lung, liver, kidney, brain, and prostate as template and the same primer set used for genomic PCR and total RNA from liver (Fig. 1A) did not produce any product that was resistant to RE digestion (data not shown), suggesting these new N-domain variants were not expressed at mRNA levels detectable by RT-PCR/RE analysis.

Sequence Alignment Delineates Groups of N-domains in Rat and Human Displaying High and Low Similarity to CEACAM1—Comparative amino acid sequence analysis of rat CEACAM1 N-domains defined two groups with high (CEACAM1b, CEACAM10, Nx, Ny, and Nz) and low (CEACAM9 and CEACAM11) sequence similarity (Fig. 2A). When sequence variability in five amino acid increments was determined for the high similarity group N-domains (Fig. 3), two regions of sequence variability (V-regions) were detected. Moreover, these V-regions were located adjacent to or within the regions corresponding to binding sites identified previously for opa proteins (55, 56) and MHV (57–59) and adhesive domains mediating intercellular adhesion (25, 28, 33) (Fig. 3). N- and C-terminal V-regions were also found in the low similarity group, but the second V-region appeared to extend into the A1 domain. To determine whether human CEACAM1 genes exhibited similar features, the same analysis was performed on human (h) CEACAM1 N-domains. As was found in the rat, human N-domains could be divided into groups with low and high amino acid similarity to hCEACAM1 (Fig. 2B). Scanning for sequence differences in five amino acid increments, also identified two V-regions in the low similarity group that were in approximately the same location as their rat counterparts (Fig. 3) and a third N-terminal V-region that corresponded to a small peak on the shoulder of the first V-domain in the rat. In contrast, well defined V-regions could not be discerned in the low similarity group (Fig. 3).

Variable Sites in the Rat High Similarity N-domain Group Are at Amino Acid Altering Positions—Casual inspection of the rat sequence data for the high similarity group (Fig. 2A) indicated that most of the variable sites between the six rat N-domain sequences were at amino acid altering (nonsynony-
mous) positions. If simple purifying selection were operating in the N-domain, one would expect more variation at the synonymous (no change in amino acid) sites. To quantify this, patterns of synonymous differences per synonymous site \( (dS) \) and nonsynonymous differences per nonsynonymous site \( (dN) \) were analyzed using MEGA 1.02 (46). This confirmed that amino acid variation per amino acid-altering site is greater than the synonymous variation per synonymous site. This difference was significant at the 5% level (Table I). The patterns of synonymous and nonsynonymous differences were also examined for the larger data set of the six rat sequences plus two homologous mouse sequences and a single human sequence. As before, the variation at amino acid-altering sites was absolutely greater than that for synonymous sites, but the difference was not statistically significant (Table I).

Recombination Is Involved in the Generation of Rat but Not Human N-domains—By using all six rat N-domains in the high similarity group as the sample, DNAsp2.2 software identified five recombination events at the sites shown in Fig. 4A and B. By sequentially removing and adding sequences from the analysis, it was apparent that pairs of sequences that include

![Fig. 5. Reactivity of mAb 362.50 and pAb 669 with COS-1 stably expressing CEACAM cDNAs. A, immunoprecipitation analysis of radiolabeled proteins from COS-1 cells expressing CEACAM isoforms. Components immunoprecipitated with normal rabbit serum, pAb 669, mAb 324.50 (as a control), or mAb 362.50 from detergent lysates of radioiodinated COS-1 cells stably transduced with CEACAM1^4L and CEACAM1^4S retroviruses were subjected to SDS-PAGE. Molecular weights of immunoprecipitated proteins are indicated at the left. I, CEACAM1^4L immunoprecipitated with normal rabbit serum (lane 1) and pAb 669 (lane 2). II, CEACAM1^4S immunoprecipitated with normal rabbit serum (lane 3) and pAb 669 (lane 4). III, CEACAM1^4L immunoprecipitated with mAb 324.50 (lane 5) and mAb 362.50 (lane 6). IV, CEACAM1^4S immunoprecipitated with mAb 324.50 (lane 7) and mAb 362.50 (lane 8). Note mAb 362.50 did not immunoprecipitate the CEACAM1^4S isoform. B–E, indirect immunofluorescence analysis of CEACAM1 expressing COS-1 cells. Clones of COS-1 cells transduced with CEACAM1^4L or CEACAM1^4S retroviruses were grown in chamber slides, fixed in acetone, and labeled with pAb 669 (B and D) and mAb 362.50 (C and E). Length of bars = 60 μm. Consistent with the results from immunoprecipitation analysis, COS-1 cells transduced with CEACAM1^4L retrovirus showed strong reactivity with pAb 669 (B) and mAb 362.50 (C). In contrast, cells expressing the CEACAM1^4S isoform that were strongly labeled by pAb 669 (D) showed no significant reactivity with mAb 362.50 (E).

![Fig. 6. Localization of the binding sites for mAbs 9.2 and 362.50. A, schematic diagram of full-length CEACAM1^4L (CEACAM1) and deletion mutant lacking one or more Ig domains (adapted from Cheung et al. (4)). Ig-like domains (N, A1, A2; transmembrane domain (TM); cyto domain. B and C, immunoblot analysis of deletion mutant proteins expressed in Sf9 insect cells. Lysates from cells infected with recombinant viruses were loaded on the gel as follows: lane 1, wild-type baculovirus-infected cells (negative control); lane 2, full-length CEACAM1^4L; lane 3, ΔN, lane 4, ΔA1; lane 5, ΔB; lane 6, ΔA1, B, A2; lane 7, A1, B, 2, A2. Molecular weight markers are indicated at the left. B, sizes of the various mutants and the levels of expression in lysates as detected by immunoblot analysis with pAb 669 (4). All of the deletion mutant proteins are recognized by this antibody. C, immunoblot analysis of the same cell lysates with mAb 9.2. Mutants were not recognized by this antibody, and the antibody displayed reduced reactivity with the ΔA1 mutant.

Ceacam1^p contributes a number of sites identified as having experienced recombination (Fig. 4A). For example, Ceacam1^p, -10, Ns, Nv, and N, showed one recombination event between
positions 140 and 370, but if the N-domains for Ceacam1a were exchanged for Ceacam1b, four recombination events were identified. At the very least, these data indicate that recombination has played a significant role in the evolutionary history of this locus. Surprisingly, when the same analysis was performed on the group of human CEACAM1 sequences with low sequence variability, no evidence of recombination was detected with any combination of N-domains from CEACAM1, -3, -5, and -6.

mAb Raised against CEACAM1a Show Differential Reactivity with CEACAM1a and -1b—To determine whether the naturally occurring amino acid sequence differences in the high similarity group of rat N-domains resulted in structural or functional changes, we focused on the CEACAM1a and -1b N-domains. We first determined whether pAb 669 and mAb 362.50 would show differential reactivity with CEACAM1b. Radioimmunoprecipitation analysis was carried out on cell lysates of 125I surface-labeled COS cells stably transduced with retroviruses carrying either the Ceacam1a-4L or Ceacam1b-4S cDNAs. mAb 324.5 failed to immunoprecipitate CEACAM1b-4S displayed detectable levels of reactivity in IIF assays with mAb 362.50 (Fig. 5E) or mAb 9.2 (not shown). Because the amino acid sequences of CEACAM1a and -1b only differ in the N-terminal Ig and cyto domains, these results suggested that variations in primary sequence were directly or indirectly responsible for differences in the conformation and/or accessibility of mAb epitopes.

The Binding Sites for mAb 9.2 and mAb 362.50 Are Located in the N-terminal Ig Domain—To define further the recognition domain for the two mAbs, immunoblot analysis using CEACAM1a-4L proteins with various domains deleted (Fig. 6A) was performed. Western blots from SDS-PAGE gels loaded with Sf9 insect cell lysates and labeled with pAb 669 expressed similar amounts of CEACAM1a-4L proteins (Fig. 6B). Immunoblot analysis also showed that mAb 9.2 reacted with full-length CEACAM1a-4L expressed in Sf9 cells (Fig. 6C). When tested with CEACAM1a-4L deletion mutants, mAb 9.2 did not

![Fig. 7. Mapping of epitopes recognized by mAb 9.2 using peptide arrays. Membranes containing an overlapping series of 10 amino acid peptides that spanned the first Ig domain were used to define linear epitopes recognized by mAb 9.2. A and C, autoradiograms of representative arrays that had been incubated with radiiodinated mAb 9.2 or mAb 362.50, respectively. Consensus binding patterns for each mAb developed from four different arrays are shown in B and D. The different shadings indicate the relative intensity of the binding to each peptide spot. E, autoradiogram of an array corresponding to epitope B following incubation with radiiodinated mAb 362.50. The primary sequence of the six overlapping peptides spanning the B epitope region in the N-domain from CEACAM1a-4L (sequences 1A–1F) and CEACAM1b-4S (sequences 2A–2F) are indicated above each spot in the array. mAb 362.50 reacted strongly with peptides 1B/1C and 2B/2C but showed different patterns of reactivity with the remaining peptides.](https://www.jbc.org/content/138/1/35072)
show detectable reactivity with ΔN, suggesting that the epitopes for mAb 9.2 were located in this domain. Similar results were obtained with mAb 362.50 (data not shown). Further analysis with mAb 9.2 showed that this antibody was unable to recognize the ΔA1, B, A2 deletion mutant and displayed decreased reactivity with the ΔA1 mutant (Fig. 6C), suggesting that mAb 9.2 recognized an epitope near the C terminus of the N-domain.

Sequences Recognized by mAb 9.2 and mAb 362.50 Are Located in the C and G β-Strands—High resolution mapping of the epitopes recognized by mAb 9.2 and mAb 362.50 was carried out by using overlapping arrays of 10 amino acid peptides that spanned the CEACAM1α-4L N-domain (Table II). When binding assays were performed on four separate arrays from three different suppliers using 125I-labeled mAbs and a one-step labeling protocol, the strongest and most consistent reactivity for mAb 9.2 mapped to peptides 18–20, 26–28, and 52–56 (Fig. 7, A and B). The defined epitope sequences of YWYKGT (epitope A), YIRSDN (epitope B), and VQFRVYPA (epitope C) (Table II) corresponded, respectively, to the rat C β-strand, C' and part of the C'C loops, and G β-strand (Fig. 2A). The strongest binding for mAb 362.50 mapped to single peptides with strong to moderate reactivity (peptides 40, 52, and 56) and two overlapping series of peptides (peptides 18–20 and 25–29). mAb 362.50 showed the most consistent reactivity with peptides 19–20 and 26–28 (Fig. 7, C and D). These two series defined overlapping sequences, respectively, of YWYKGTTL (epitope A') and YIRSDN (epitope B) (Table II). To determine whether variations in the primary sequence of the mAb 362.50 epitope in CEACAM1α and CEACAM1β were responsible for the differential reactivity seen in IIF and IP assays, a peptide array composed of 13-mer peptides spanning epitope B (peptide 25–31) from CEACAM1α-4L and CEACAM1β-4S was synthesized and probed with radiiodinated mAb 362.50. As shown in Fig. 7E, mAb 362.50 strongly reacted with two N-domain peptides in CEACAM1α (1B and 1C) and their corresponding peptides in CEACAM1β (2B and 2C) but showed opposing reactivities with the 1A/2A and 1D/2D peptides. These findings...
indicated that the differences in primary sequence had produced subtle differences in the location of epitope B.

Epitope B Is Located within the Major Adhesive Site in the CEACAM1a-4L-N-domain—In previous reports (25, 28, 33, 55–59), the adhesive sites for CEAl and the binding domains for MHV and opa proteins were localized to the C β-strand and CC’ loop domains. Because one of the mAb epitopes was within this region, it was of interest to determine whether the two mAbs could block cell adhesion. When we tested the ability of the mAb 362.50 and mAb 9.2 to block CEACAM1a-4L-mediated cell aggregation, both mAbs were also able to inhibit, to some extent, the adhesion of SF9 cells expressing full-length CEACAM1a-4L, whereas the control mAb 324.50, which was against TuAg1 (60), had no effect (Fig. 8A). Significant net inhibition (72%) of cell aggregation was observed with 2 μg of mAb 9.2, whereas only 24% of net inhibition was obtained with a similar amount of mAb 362.50. These results suggested that mAb 9.2 was recognizing an epitope close to or within the adhesion sequence.

To determine which mAb 9.2 epitope was involved in cell-cell adhesion, peptides corresponding to mAb 9.2 and mAb 362.50 peptides with extensions toward the N and C termini of the Adhesive site were synthesized and used for blocking of CEACAM1a-4L-mediated aggregation of SF9 cells (Table III). Of the eight peptides tested, only peptide 9.2B, containing epitope B, could completely inhibit adhesion at 30 μg/ml. This result suggested that sequences around epitope B are critical for adhesion. To define further the adhesion epitope, three peptides with deletions from the C terminus (peptide 9.2B-1, 8.0, and L1) and two with deletions from the N terminus (9.2B-3 and 1.0) of peptide 9.2B were synthesized and tested for their effect on CEACAM1a-4L-mediated adhesion. As shown, peptides 8.0 (Fig. 8B) and 9.2B-1 (Table III) maintained adhesion blocking activity. In contrast, peptide 1.0 (Fig. 8B) and 9.2B-3 (Table III) were no longer able to block adhesion indicating that the three N-terminal amino acids but not the five C-terminal amino acids of 9.2B are essential for adhesion blocking.

When these experiments were repeated using COS-1 cells expressing either CEACAM1a-4L or CEACAM1b-4S, peptide 9.2B with the B epitope again showed the highest blocking activity for both isoforms (Fig. 8C) despite the differences in primary sequence of the B epitope in CEACAM1a-4L and CEACAM1b-4S. The levels of expression of the two isoforms (dark line) relative to cells labeled by IIF without a primary anti-CEACAM1 antibody (light dashed line) are shown in Fig. 8D. Taken together, these results define the adhesion epitope for CEACAM1 to be PDSEIARYIRS, a sequence encompassing the entire C β-strand and part of the CC’ and C/C’ loops (Fig. 2A). In addition, the results show that, despite the differences in amino acid sequence, the 9.2B peptide also blocked aggregation mediated by CEACAM1b-4S, a finding in keeping with the ability of mAb 362.50 to recognize the B epitope in both the 1a and 1b allele.

Removal of O-Linked Glycans Does Not Restore mAb Reactivity with CEACAM1a—Previous studies have shown that glycosylation can alter the conformation and binding properties of membrane receptors and Ig-like proteins including mouse Ceacam1 (61). To determine whether the amino acid sequence differences between CEACAM1a and CEACAM1b N-domains altered patterns of glycosylation, we searched for sites of O-glycosylation by using the rules proposed by Pisano et al. (43). By this method, a total of nine possible sites for addition of O-linked sugars were identified in the CEACAM1b N-domain (Fig. 9A), only two of which were also present in CEACAM1a. Analysis using the NetOGlyc 3.0 Prediction Server, a collection of artificial neural networks that recognize O-glycosylation sites based on sequence context and surface accessibility (62), identified a single site unique to the CEACAM1b N-domain at Thr-91 (Fig. 9A). In addition, three N-glycosylation sites shared by CEACAM1a and CEACAM1b N-domains were identified (Fig. 9A) by using the consensus sequence NX(S/T), where X is anything but Pro (63). These findings raised the possibility that the differential use of common glycosylation sites in or near mAb epitopes or occupation of unique glycosylation sites created by differences in primary sequence might contribute to altered mAb reactivity. Analysis of the three epitope sites showed that none of the seven potential O-glycosylation sites unique to CEACAM1b-N-domain resided within epitope C for mAb 9.2. However, sequence analysis identified an additional O-glycosylation site (not shown) and a single N-glycosylation site at the N terminus of the A1-domain (Fig. 9A) that were located in close proximity to epitope C. Epitope B for mAb 362.50, on the other hand, overlapped a region that not only differed in primary sequence in the two N-domains but was also flanked by the O-glycosylation site at Thr-91 unique to CEACAM1b. In addition, mAb 362.50 epitope B contained a single N-glycosylation site common to both isoforms.

To determine whether the loss of mAb reactivity involved differential O-glycosylation, CEACAM1a-4L and CEACAM1b-4S isolated on immobilized pAb 669 from Nonidet P-40 extracts of transfected COS cells or Fischer 344 hepatocytes were subjected to deglycosylation by sequential neuraminidase and O-glycanase digestion. Removal of sialic acid with neuraminidase (64). Immunoblot analysis revealed that CEACAM1a-4L from transfected COS cells reacted strongly with pAb 669 both before and after digestion with neuraminidase alone or with neuraminidase and O-glycanase (double digestion) (Fig. 9B). When immunoblots were repeated with mAb 9.2, reactivity with both single and double digested CEACAM1a-4L proteins was also detected (Fig. 9B), but the
enzymes (lanes 1–4, CEACAM1a-4L and CEACAM1b-4S, respectively. Proteins are indicated on the left.

et al. rules proposed by Pisano 2.0 Prediction Server (double asterisk). Consensus sequences for double asterisk are noted with an asterisk from American Fischer hepatocytes. Purified isoforms were digested, or double digested CEACAM1b-4S (Fig. 9, lanes 1–3), neuraminidase of glycans on the former but not the latter cell type.

Fig. 9. Detection of O-linked glycans on CEACAM1a-4L. A, mapping of O-glycosylation site for O-glycosylation were determined by the rules proposed by Pisano et al. (43) (single asterisk) or by the NetOGlyc 2.0 Prediction Server (double asterisk). Consensus sequences for O-glycosylation are underlined with a dark solid line. The O-glycosylation sites (serine or threonine residues) unique to the CEACAM1 N-domain are noted with an asterisk, and the O-glycosylation sites shared with CEACAM1 are noted with a pound sign. N-Glycosylation sites are denoted by white N residues. Epitopes are indicated by shaded areas. B–D, O-glycanase digestion. CEACAM1a-4L and CEACAM1b-4S were isolated on immobilized pAb 669 from NBS-treated extracts of COS-1 cells transfected with the corresponding cDNA expression vectors and from American Fischer hepatocytes. Purified isoforms were digested with C. perfringens neuraminidase (N; lanes 1 and 5), neuraminidase and O-glycanase (N+O; lanes 3 and 6), or were incubated without enzymes (U; lanes 1 and 4). Immunoblots of digested CEACAM isoforms resolved by SDS-PAGE on 7.5% gels were performed with both pAb 669 (lanes 1–3) and mAb 9.2 (lanes 4–6). Sizes of intact and deglycosylated proteins are indicated on the left. B and C, immunoblot of purified CEACAM1a-4L and CEACAM1b-4S, respectively. D, an immunoblot of CEACAM1a-4L and CEACAM1b-4S splice variants purified from Fischer rat hepatocytes. Note that both CEACAM1a-4L and CEACAM1b-4S from COS-1 cells but not hepatocytes show a decrease in size following O-glycanase digestion, indicating the presence of O-linked glycans on the former but not the latter cell type.

level of binding was much less than with pAb 669. In contrast, mAb 9.2 showed no reactivity with undigested, neuraminidase-digested, or double digested CEACAM1a-4S (Fig. 9C). CEACAM1a-4L and CEACAM1b-4S (Fig. 9, B and C) visualized in blots with pAb 669 also showed an initial decrease in size following digestion with neuraminidase and a further reduction following digestion with O-glycanase indicating that the two isoforms were O-glycosylated when expressed in transfected COS cells. However, the size of CEACAM1a-4L and CEACAM1b-4S from isolated hepatocytes, which demonstrated a strong reactivity with pAb 669 AND mAb 9.2, was decreased in size by neuraminidase but did not show a further decrease following digestion with O-glycanase (Fig. 9D), a result consistent with previous studies identifying N- but no O-linked glycans on CEACAM1 (65). Decreases in size were also observed with single or double digested fetuin, a glycoprotein with O-linked sugars known to be susceptible to O-glycanase, confirming that the two enzymes were active under the conditions used for digestion (data not shown).

**DISCUSSION**

In this report, we present evidence that N-domains from members of the CEACAM1 branch of the CEA family from both human and rat contain two Ig-like V-regions. These V-regions were similar in size and location to those described previously by Kodelja et al. (30) for N-domains from the pregnancy-specific glycoprotein (PSG) branch of the CEA family. The overall similarity of the CEACAM1 V-regions in the low similarity groups of rat and human were also comparable with those reported for PSG V-regions but were much lower in sequence conservation than V-regions from the high similarity groups of rat and human N-domain groups. This can be seen by comparing the range of variability (y axis) in the 1st and 2nd panels in the top row and the 2nd and 3rd panels in the bottom row of Fig. 3. According to Zimmermann (3), the PSG family in the rat most likely arose after separation of rodent and primate orders and then underwent extensive expansion before mouse/rat speciation. Zimmermann (3) further suggests that Ceacam1a, Ceacam1b, and Ceacam10 are relatively recent arrivals that arose by gene duplication after mouse/rat speciation. These rat Ceacam1 family members may thus have had less time than the PSG family to diverge and undergo expansion. The same argument may also apply to the closely related human V-domains (hCEACAM1, -3, -5, and -6).

Out of a total of 26 altered codons in Ceacam1 N-domains, 88% (23/26) had base changes in the first and second nucleotide. Moreover, 92% of the base changes resulted in amino acid replacement (nonsynonymous substitutions). Quantification of patterns of synonymous and nonsynonymous differences further showed that the amino acid altering (nonsynonymous) mutations were becoming established in the population more frequently than “silent” (synonymous) mutations. If this locus were evolving under a simple model of purifying selection, one would expect that amino acid mutations would have a negative impact on the functional aspects of the gene product and that synonymous mutations would be more common. An alternative model that is more consistent with the data is one where amino acid changes have some positive effect, a conclusion similar to that of Hughes and Nei (66) and more recent studies of proteins involved in antigen recognition (67). Significant in this regard is the observation that a number of the amino acids involved in the Ig fold (67) were evolving under a simple model of purifying selection, one would expect that amino acid mutations would have a negative impact on the functional aspects of the gene product and that synonymous mutations would be more common. An alternative model that is more consistent with the data is one where amino acid changes have some positive effect, a conclusion similar to that of Hughes and Nei (66) and more recent studies of proteins involved in antigen recognition (67). Significant in this regard is the observation that a number of the amino acids involved in the Ig fold (67) were conserved in all of the N-domains. In addition, the β-strand locations in the N-domain of CEACAM11 and CEACAM1, which have only a 48% sequence similarity, are virtually identical based on secondary structure predictions using the PSAX Sequence Analysis Server from the Biomolecular Engineering Research Center. Similar findings obtained from a comparison of human and rat N-domains led Rudert et al. (68) to conclude that there is little functional constraint on the primary amino acid sequence except for key amino acid needed for the Ig fold.

Positive change is generated in antibodies by somatic hypermutation in V-domains or untemplated base changes during VDJ recombination that modify antibody specificity or affinity, thereby allowing the immune system to respond to mutations in infectious agents (31). It seems likely that naturally occurring mutations in expressed CEACAM1 N-domains would have a similar effect-modification but not loss of adhesive activity.
The fact that the N-domains for CEACAM1\(^a\), \(-1\), and \(-10\) (69) all have adhesion activity is consistent with this idea. Carrying the analogy to antibodies a step further, amino acid sequence diversification of CEACAM1 N-domains should be manifested by functional modifications similar to those that occur in antibodies, namely alterations in affinity or specificity that is advantageous for a particular kind of epithelium. It is noteworthy that all of the rat CEACAM1 proteins with adhesive activity are in the group of high similarity N-domains with amino acid sequence similarities relative to CEACAM1\(^a\) ranging from 96 to 95\%, values much higher than those for CEACAM9 and -11 (52 and 61\%, respectively). Although Ceacam9 is highly conserved between mouse and rat, its does not appear to play an essential role in development because its loss in Ceacam9\(^{-/-}\) mice had no effect on placental, embryonic, or postnatal development (70). Even less is known about the functional activity of rat CEACAM11 except by inference to the structurally identical mouse CEACAM11, the loss of which had no discernible consequence except for a small reduction in litter size (71).

Whether the amino acid sequence variations in Nx, Ny, and Nz lead to altered binding properties remains to be determined using the approach described by Lin et al. (69) for CEACAM10. However, this type of analysis would only be warranted if expression of Nx, Ny, Nz at the RNA or protein level can be demonstrated in future studies by carrying out a more exhaustive tissue analysis.

The same relationship between sequence similarity and functional activity held true for 3/4 members of the high similarity group (88\% similarity) in human, all three of which mediate cell aggregation. No cell-cell adhesive activity has been reported for CEACAM3, another member of the human high similarity group (89\%). However, the close match of its N-domain to CEACAM1 makes it highly likely that the CEACAM3 N-domain has cell adhesion activity, a supposition supported by its ability to bind Neisseria gonorrhoeae via its opra receptors (72)

Although human Ceacam8, a neutrophil CAM in the low similarity group (73\%), retains adhesive activity, it adhesion partners are limited to CEACAM6, a widely expressed neutrophil CAM that binds to human CEACAM1, -5, -8 (73), opra proteins (74), and itself. In contrast, CEACAM7 and CEACAM4, with similarities of 64 and 58\%, cannot bind opra receptors (74) and have no documented cell-cell adhesion activity.

Taken together, these observations suggest that a direct relationship exists between the degree of similarity to CEACAM1 and functionality, i.e. the lower the similarity, the lower the functionality. This relationship seems to hold true for the mouse hepatitis virus (MHV) receptor activity of mouse CEACAM1\(^a\), CEACAM1\(^b\), and CEACAM2 with sequence similarities of 100, 75, and 59\%, respectively. When tested for MHV binding, the efficiency of CEACAM2 as a receptor was found to be 10–100-fold lower than CEACAM1 (58, 75).

In virus neutralizing assays, the inhibition of MHV-A59 infectivity by CEACAM1\(^a\) was 4-fold and 1000-fold greater than CEACAM1\(^b\) and CEACAM2, respectively. Thus, the decrease in similarity relative to CEACAM1 was paralleled by a decrease in the efficiency of MHV binding.

Aside from direct effects resulting from differences in primary structure, sequence diversification could also alter binding properties of N-domains indirectly by creating, eliminating, or changing the usage of sites for post-translational modifications, the end result being subtle changes evidenced in the present study by loss of mAb binding. Diminished mAb binding to CEACAM1\(^b\), for example, could be caused in part by differential glycosylation. This would be consistent with the strong reactivity of both mAb 9.2 and pAb 669 with hepatocytes and the much weaker reactivity of mAb 9.2 relative to pAb 669 with immunoblots prepared from CEACAM1\(^a\)-positive COS-1 cells (Fig. 9B), a difference that could result from differential N-glycosylation or O-glycosylation at Thr-91 in CEACAM1.\(^b\)

Based on recent reports, it is clear that relatively small changes in O- or N-linked oligosaccharides are sufficient to alter conformation and mAb binding (76–79). NMR studies by Huang et al. (76) showed that addition of N- or O-linked sugars to a 24-residue peptide from the human immunodeficiency virus glycoprotein 120 envelope had major effects on local conformation, induced minor changes at more distance sites and enhanced or reduced mAb binding to an epitope on this peptide. Similar effects could account for the ability of CEACAM1\(^a\)-4S to mediate aggregation when expressed in COS-1 cells but not in SF9 cells where altered glycosylation shifted the molecular mass of CEACAM1\(^a\)-4S from 105 to 70 kDa (4).

In the present studies, glycosidase digestion unexpectedly revealed the presence of O-glycans on both forms of CEACAM1 in COS-1 cells but not in rat hepatocytes, the latter a finding consistent with previous reports (65). This suggested that either O-glycosylation was not involved in the loss of mAb binding or, alternatively, that differences in the pattern of O- and possibly N-glycosylation were primarily responsible, an alternative that will be examined in future studies. Marked cell type-specific differences in N-linked glycans on CEACAM1 have been described by Kannicht et al. (80), but we believe this is the first evidence for tissue-specific O-glycosylation of CEACAM1.

Of the three peptides tested for their ability to block aggregation, only peptide 9.2, containing epitope B, could completely inhibit adhesion (Table III) suggesting that epitope B was within or proximal to the adhesive domain. Further analysis with variants of peptide 9.2 containing C- and N-terminal deletions (Table III) defined the adhesion epitope for rat CEACAM1 to be PDSEARYIRS, a sequence encompassing the entire C' \(\beta\)-strand and part of (33) hCEACAM1 (28), and the docking sites for MHV (57–59) and opa proteins (81, 82) are located primarily in the C \(\beta\)-strand and the CC' loop domain (Fig. 2B). It seems likely that the shift in the rat adhesive domain to the C' \(\beta\)-strand stems at least in part from primary sequence-related differences in conformation. More important, despite the differences in their N-domains, CEACAM1\(^a\)-4L, CEACAM1\(^b\)-4S, and CEACAM10 were all capable of mediating cell adhesion, providing another example in the Ig family of diversification without loss of function. Indeed, the major adhesive epitopes for human, mouse, and rat were located in the middle of the first V-region, exactly the location one would predict if the object of diversification was to alter adhesive properties to produce changes in affinity or specificity that are favorable in certain tissues but have a neutral effect in others.

The partial loss of mAb 9.2 reactivity with mutants lacking the A1 domain suggested that epitope C in the C-terminal G \(\beta\)-strand was dependent on A1 for its correct presentation, a finding similar to that reported for the binding of virus to the MHV-A59 receptor in mice (57). Epitope C was the only one of the three epitopes that was recognized primarily by mAb 9.2 (Fig. 7), suggesting it was responsible, in part, for the stronger adhesion blocking activity of mAb 9.2 relative to mAb 362.50. At odds with this conclusion was the inability of the peptide containing the C epitope to block cell aggregation (Table III, peptide 9.2C). However, a peptide containing epitope C would be unable to destabilize intercellular homophilic adhesion if binding between B epitopes was much stronger than between C epitopes. This viewpoint is consistent with results from mAb binding studies with a series of truncated CEACAM1\(^a\) N-domain peptides which indicated that each of the three epitopes
was necessary but not sufficient for mAb binding. It can be surmised that deletion of any one of the epitope regions causes conformational changes in the remaining epitopes that greatly diminish antibody binding. The same may hold true for peptides where effective blocking of cell adhesion may depend on peptide-specific variations in conservation of tertiary structure, a variable that may be more critical for adhesion blocking than for antibody binding. Alternatively, the adhesive epitope may be upshifted from peptide 9.2, another possibility currently under investigation.

Another seeming inconsistency is the lack of reactivity of both mAbs with the CEACAM1b N-domain. Peptide array analysis showed that despite the differences in primary sequence, mAb 362.50 bound to the B epitope region from both CEACAM1a and 1b. Additionally, binding of mAb 9.2 with peptide C in the CEACAM1b and CEACAM1 N-domains should be virtually identical because the amino acid sequence for epitope C is exactly the same. A possible explanation for these results is that because of their proximity to V-domains, the small differences in conformation produced by minor differences in sequence may be sufficient to alter binding affinity, a possibility consistent with the subtle difference in the location of epitope B in the CEACAM1a and 1b isoforms. Regardless of the reason, these results strongly suggest that relatively minor differences in sequence in some may have significant effects on adhesive and antibody binding properties of naturally occurring CEACAM1 N-domain variants.

In summary, in this report we have identified three unique N-domains (N1, N2, and N3), which are present in the rat genome. In keeping with the current nomenclature, we propose that these should be designated as Cenacm12, Cenacm13, and Cenacm14. We have also shown that in the rat and human CEACAM1 families, there are subgroups with high and low sequence similarity. We present evidence that shows for the first time that recombination plays a significant role in the generation of diversity in the rat but not the human group of high similarity N-domains. We also show that CEACAM1 N-domains in both humans and rats harbor two variable regions that contain or are situated adjacent to adhesive domains. We have also delineated cell type-specific differences in O-glycosylation, a modification that could alter the binding characteristics of CEACAM1 N-domains. Finally, we have defined a primary cell-cell adhesive epitope in the rat CEACAM1 N-domain that differs from the principal site defined for mouse CEACAM1 and MVH. Moreover, a peptide corresponding to this epitope blocked adhesion mediated by both the 1a and 1b alleles, showing that sequence diversification did not significantly alter this domain.

REFERENCES

1. Beauchemin, D., Draper, R., Dveksler, G., Gold, P., Gray-Owen, S., Grunert, F., Hammarstrom, S., Holmes, K., Koeber, A., Korski, M., Lin, S. H., Lucka, L., Najjar, S. M., Neumaier, M., Obrink, B., Shively, J. E., Skokut, K. M., Stanners, C. P., Thomas, P., Thompson, J. A., Virji, M., von Kleist, S., Wagner, C., Watt, S., and Zimmermann, W. (1999) Exp. Cell Res. 252, 243–249

2. Hammarstrom, S., Olsen, A., Teglund, S., and Baranov, V. (1998) in Cell Adhesion and Communication Mediated by the CEA Family (Stanners, C. P., ed) Vol. 5, pp. 1–30. Hardwood Academic Publishers, Singapore

3. Zimmermann, W. (1998) in Cell Adhesion and Communication Mediated by the CEA Family (Stanners, C. P., ed) Vol. 5, pp. 31–56. Hardwood Academic Publishers, Singapore

4. Cheung, P., Thompson, N., Earley, K., Culic, O., Hixson, D., and Lin, S. H. (1993) J. Biol. Chem. 268, 6139–6146

5. Culic, O., Huang, G.-H., Flanagan, D., Hixson, D., and Lin, S. H. (1992) J. Biol. Chem. 267, 47–53

6. Lin, S. H., Culic, O., Flanagan, D., and Hixson, D. C. (1991) Biochem. J. 278, 155–161

7. Edlund, M., Gaardevdal, H., Bock, E., and Obrink, B. (1993) Eur. J. Biochem. 213, 1109–1116
CEACAM1 N-domain Variable Regions

1941–1948

58. Rao, P. V., Kumari, S., and Gallagher, T. M. (1997) *Virology* **229**, 336–348
59. Tan, K., Zelus, B. D., Meijers, R., Liu, J. H., Bergelson, J. M., Duke, N., Zhang, R., Joachimiak, A., Holmes, K. V., and Wang, J. H. (2002) *EMBO J.* **21**, 2076–2086
60. Faris, R. A., McIntire, K. D., Thompson, N. L., and Hixson, D. C. (1990) *Cancer Res.* **50**, 4755–4763
61. Dveksler, G. S., Basile, A. A., Cardellichio, C. B., and Holmes, K. V. (1995) *J. Virol.* **69**, 543–546
62. Center for Biological Sequence Analysis. *NetOGlyc Software*, 3.0 Ed., Lyngby, Denmark
63. Bause, E. (1983) *Biochem. J.* **209**, 331–336
64. Pinter, A., and Honnen, W. J. (1988) *J. Virol.* **62**, 1016–1021
65. Odin, P., Tingstrom, A., and Obrink, B. (1986) *Biochem. J.* **236**, 559–568
66. Hughes, A. L., and Nei, M. (1988) *Nature* **335**, 167–170
67. Hughes, A. L., and Yeager, M. (1997) BioEssays **19**, 777–786
68. Rudert, F., Zimmermann, W., and Thompson, J. A. (1989) *J. Mol. Evol.* **29**, 126–134
69. Lin, S. H., Cheng, H., Earley, K., Luo, W., and Chou, J. (1998) *Biochem. Biophys. Res. Commun.* **245**, 472–477
70. Finkenzeller, D., Fischer, B., Lutz, S., Schrewel, H., Shimizu, T., and Zimmermann, W. (2000) *Mol. Cell. Biol.* **20**, 7140–7145
71. Finkenzeller, D., Fischer, B., Lutz, S., Schrewel, H., Shimizu, T., and Zimmermann, W. (2003) *Mol. Cell. Biol.* **23**, 272–279
72. McCaw, S. E., Schneider, J., Liao, E. H., Zimmermann, W., and Gray-Owen, S. D. (2003) *Mol. Microbiol.* **49**, 623–637
73. Kuroki, M., Abe, H., Imakirii, T., Liao, S., Uchida, H., Yamauchi, Y., and Oikawa, S. (2001) *J. Leukocyte Biol.* **70**, 543–550
74. Popp, A., Dehio, C., Grunert, F., Meyer, T. F., and Gray-Owen, S. D. (1999) *Cell. Microbiol.* **1**, 169–181
75. Nedellec, P., Dveksler, G. S., Daniels, E., Turbide, C., Chow, R., Basile, A. A., Holmes, K. V., and Beauchemin, N. (1994) *J. Virol.* **68**, 4525–4537
76. Huang, X., Barchi, J. J., Jr., Lung, F. D., Koller, P. P., Nara, P. L., Muschik, J., and Garrity, M. R. (1997) *Biochemistry* **36**, 10846–10856
77. Blochberger, T. C., Sabatine, J. M., Lee, Y. C., and Hughey, R. P. (1989) *J. Biol. Chem.* **264**, 20718–20722
78. Hayes, G. R., Enns, C. A., and Lucas, J. J. (1992) *Glycobiology* **2**, 355–359
79. Huang, Y., Jellies, J., Johansen, K. M., and Johansen, J. (1997) *J. Cell Biol.* **138**, 143–157
80. Kanzig, B., Lucka, L., Nuck, R., Reutter, W., and Gehlke, M. (1999) *Glycobiology* **9**, 897–906
81. Virji, M., Makepeace, K., Ferguson, D. J., and Watt, S. M. (1996) *Mol. Microbiol.* **22**, 941–950
82. Bos, M. P., Kau, D., Hogan, D. M., Grant, C. C., and Belland, R. J. (2002) *Infect. Immun.* **70**, 1715–1723
83. Wu, T. T., and Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211–250

Downloaded from http://www.jbc.org/ by guest on July 25, 2018
Two Variable Regions in Carcinoembryonic Antigen-related Cell Adhesion Molecule1 N-terminal Domains Located in or Next to Monoclonal Antibody and Adhesion Epitopes Show Evidence of Recombination in Rat but Not in Human
Meghan M. Comegys, Sue-Hwa Lin, David Rand, Deborah Britt, Donna Flanagan, Helen Callanan, Kate Brilliant and Douglas C. Hixson

J. Biol. Chem. 2004, 279:35063-35078.
doi: 10.1074/jbc.M404431200 originally published online June 7, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M404431200

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

This article cites 80 references, 43 of which can be accessed free at http://www.jbc.org/content/279/33/35063.full.html#ref-list-1