Human Keratinocytes Express a New CD44 Core Protein (CD44E) as a Heparan–Sulfate Intrinsic Membrane Proteoglycan with Additional Exons

* Fred Hutchinson Cancer Research Center and § Department of Pathobiology, University of Washington, Seattle, Washington 98104; † Cytel Corporation, La Jolla, California 92037

Abstract. We previously identified a 90-kD (GP90), collagen-binding, membrane glycoprotein, termed extracellular matrix receptor III (ECMR III), that is homologous to the lymphocyte homing receptor and CD44 antigen (Gallatin, W. M., E. A. Wayner, P. A. Hoffman, T. St. John, E. C. Butcher, and W. G. Carter. 1989. Proc. Natl. Acad. Sci. USA. 86:4654–4658). CD44 is abundantly expressed in many epithelial tissues, and is localized predominantly to filopodia in cultured keratinocytes. Here we establish CD44 as a polymorphic family of related membrane proteoglycans and glycoproteins possessing extensive diversity in both glycosylation and core protein sequence. Human neonatal foreskin keratinocytes (HFKs) and QG56 lung squamous carcinoma cells express an alternatively spliced form of the CD44 core protein (termed CD44E) that contains an additional 132 amino acids in the carbohydrate attachment region of the extracellular domain. HFKs, HT1080 fibrosarcoma and QG56 cells, as well as many other human cells, contain varying ratios of GP90 and structurally related, higher molecular mass forms of CD44 that express the following characteristics: (a) each form reacted with anti-CD44 (mAbs) P1G12, P3H9, and P3H5. Each of these mAbs recognized a distinct, nonoverlapping epitope present on each CD44 form. (b) Differences in mass were due primarily to variation in carbohydrate moieties, including sulfated asparagine-linked glycopeptides (GP), chondroitin sulfate (CS), and heparan sulfate (HS) glycosaminoglycans, as well as O-linked mucin and polylactosamine structure(s). The major polymorphic forms were designated HT1080 GP90 and CS180, QG56 GP230, and HFK HS/CS250, based on dominant carbohydrate moieties and relative mass. (c) The polymorphic forms use CD44 and CD44E core proteins, each containing a unique set of potential attachment sites for O- and N-glycosides and glycosaminoglycans. (d) Immunofluorescence microscopy, differential extraction with Triton-X-114 detergent, and incorporation into liposomes indicated that all the forms were membrane bound glycoconjugates. These results define CD44 as a structurally diverse, but immunologically related, set of intrinsic membrane macromolecules, and suggests that these structurally varied forms might be expected to manifest multiple functions.

The polymorphic family of integral membrane glycoproteins CD44 (see Haynes et al., 1989 for review) is found on a wide variety of cells. The exact function(s) of CD44 has yet to be conclusively defined, although recent reports have implicated CD44 in lymphocyte homing (Jalkanen et al., 1987), lymphohemopoiesis (Miyake et al., 1990a) and T cell activation (Shimizu et al., 1989; Huet et al., 1989). The widespread distribution of CD44 in non-hemopoietic cell types (Carter and Wayner, 1988; Flanagan et al., 1989), however, argues for a more general function than those delegated to it in the hemopoietic lineage. Previously reported biochemical and cell biological properties are consistent with a role as an adhesive cell-surface glycoprotein. CD44 may function either by linking extracellular matrix proteins such as collagen (Carter and Wayner, 1988) or hyaluronic acid (Miyake et al., 1990b; Aruffo et al., 1990) with the cytoskeleton (Carter and Wayner, 1988; Kalomiris and Bourguignon, 1988), or may be involved with homotypic (St. John et al., 1990; Belittos et al., 1990) or heterotypic (Shimizu et al., 1989; King et al., 1990) intercellular adhesion.

Most studies have concentrated on the 80–90-kD form of CD44 in hemopoietic cells. Before the CD44 classification, it was referred to by several designations, including extracellular matrix receptor III (ECMR III; Wayner and Carter, 1988; Flanagan et al., 1989), the hyaluronate receptor (Alho and Underhill, 1989; Underhill, 1989), Pgp-1 (Isacke et al., 1986), the hyaluronate receptor (Alho and Underhill, 1989; Underhill, 1989), p80 (Haynes et al., 1983), p85 (LeTarte et al., 1985), or antigens of the mAbs F-10-44-2 (Dalchau et al., 1980), Hermes-1 (Jalkanen et al., 1989), and HSF-1 (Flanagan et al., 1989).
Materials and Methods

Materials

PMSF, N-ethylmaleimide, 2-mercaptoethanol, BSA, Triton X-100 and Triton X-114 detergents, n-octyl-β-D-glucopyranoside detergent, ethylene diamine, aminooctanesoic acid, protein A-agarose, egg yolk phosphatidyicholine, were purchased from Sigma Chemical Co. (St. Louis, MO). Random hexamer and oligo dT25 were obtained from Pharmacia Fine Chemicals (Piscataway, NJ). Pronase and lactoperoxidase were from Calbiochem-Behring Corp. (La Jolla, CA). Fluorescein-conjugated (goat) anti-mouse IgG and IgM (H and L chains) and rhodamine-conjugated (goat) anti-rabbit IgG were purchased from Sigma Chemical Co. (St. Louis, MO). Protein concentration was determined by the fluorescamine method using BSA as standard (Udenfriend et al., 1974). Glycosidases and DNA-modifying enzymes, unless noted elsewhere, were obtained from Boehringer Mannheim Biochemicals (Indianapolis, IN). Anti-CD44 mAbs (P3H9, P3H5, and P1G12) were prepared as previously described (Carter et al., 1988). Rabbit polyclonal antibodies against human fibronectin were prepared as previously described (Carter, 1982). Rabbit polyclonal antibodies were prepared against mouse laminin isolated from mouse EHS carcinoma. Rabbit polyclonal antibodies against involucrin, a component of the keratinocyte envelope (Rice and Green, 1979), was a gift from Dr. Robert Rice (University of California, Davis, CA). Rabbit polyclonal antibodies (Ab33; Roman et al., 1989) against integrin α5 were from Dr. John McDonald (Washing- ton University, St. Louis, MO). Rabbit polyclonal antibodies against type 1 collagen was from Chemicon International, Inc. (El Segundo, CA).

Metabolic Labeling, Pulse-Chase Experiments, and PAGE

Metabolic labeling with 35SO42− or with [35S]methionine or pulse-chase studies using [35S]methionine were performed essentially as described previously (Carter and Weyer, 1988). Polycrylamide slab gels containing SDS (SDS-PAGE gels) were prepared following the basic stacking gel system of Laemmli (1970). Pre-stained protein standards for relative molecular mass estimation were obtained from either Bethesda Research Laboratories (Bethesda, MD) or Sigma Chemical Co. (St. Louis, MO). Protein concentration was determined by the fluorescent method using BSA as standard (Udenfriend et al., 1972).

Nitrous Acid Cleavage of Heparan Sulfate

Trition X-100 extracts of 35SO42−-labeled HKFs or QG56 cells were immune precipitated with P3H9 to isolate radiolabeled CD44. The selective removal of heparan sulfate by the method of Shively and Conrad (1976) was used to identify focal adhesions (FAs) in the same field as fluorescence, interference reflection or antibody exclusion microscopy as described (Shively et al., 1976) and was coupled to C4Br-activated Sepharose (Parikh et al., 1974). Glycosidases and DNA-modifying enzymes, unless noted elsewhere, were obtained from Boehringer Mannheim Biochemicals (Indianapolis, IN). Cells were adhered to coverslips as described (Carter et al., 1990a) and were incubated with combinations of mouse or rat mAbs and rabbit polyclonal primary antibodies and processed for visualization by two-color fluorescence, interference reflection or antibody exclusion microscopy as previously described (Carter et al., 1990a). Interference reflection microscopy (IRM) was performed basically as described (Lazarov and Loewer, 1982) and was used to identify focal adhesions (FAs) in the same field as the two color fluorescence. FAs were also localized by the antibody exclusion technique (Neyfakh et al., 1983; Carter et al., 1990a).

Fluorescence Microscopy of Tissue Sections

The distribution of receptors and ligands in tissue were determined by immunofluorescence microscopy of 6-μm cryostat sections as previously
described (Wayner et al., 1988). All sections were fixed in 4% paraformaldehyde in PBS before incubation in primary antibodies and peroxidase-conjugated secondary antibodies. The samples were examined with a Zeiss fluorescence microscope and photographed with Kodak T-Max film at ASA 800. In control experiments, no fluorescence of rhodamine was detected using the fluorescein filters or vice versa.

PCR Amplification

The isolation of poly-A⁺ RNA from primary HFKs, dermal fibroblasts (HFFs), and the human cell lines HT1080 or QG56 was as described by Badley et al. (1988). Two micrograms of poly-A⁺ RNA was transcribed into cDNA by reverse transcriptase using either random hexamer oligomers or oligo (dT)$_12$ as primers and cDNAs were purified by gel exclusion chromatography (BioGel A5m; Bio-Rad Laboratories, Richmond, CA). Purified cDNA was amplified by polymerase chain reaction (PCR) in a DNA thermal cycler (Perkin-Elmer Cetus Corp., Norwalk, CT) using *Thermus aquaticus* (Taq) DNA polymerase and oligonucleotide pairs as noted in the text (see Fig. 8 A for an oligonucleotide map). Other buffer components were as described by the enzyme supplier (Perkin-Elmer Cetus Corp.). Reaction conditions were as follows: 4 min at 94°C followed by 35 cycles of 1 min at 94°C, 2 min at 50°C and 4 min at 72°C. The reaction products obtained by PCR were separated by agarose gel electrophoresis; individual bands were excised and the cDNA isolated by the freeze/squeeze method (Tautz and Renz, 1983).

DNA Sequencing and Analysis

Double-stranded cDNA for sequencing was obtained by a second round of PCR amplification using the purified cDNA from the first round as template. DNA was purified by gel electrophoresis, as described above, and used directly for sequencing by the dideoxy chain termination method (Sanger et al., 1977), using a modified T7 DNA polymerase (Sequenase; U. S. Biochemical Corp., Cleveland, OH) and the modification of Winship (1989). Primers were annealed to the DNA by snap cooling from 100°C into dry ice, and included D1, B7-6, HEI, and SJI oligonucleotides (see Fig. 8 A). Both strands from at least three independent, primary PCR reactions were amplified and sequenced to check and correct for possible spontaneous mutations caused by Taq DNA polymerase (Ennis et al., 1990). DNA sequence analysis was performed with Genepro software (Riverside Scientific, Seattle, WA).

Northern Blot Analysis

Northern blot analysis of formaldehyde-treated, size fractionated poly-A⁺ RNA was, with slight modification, as described by Selden (1989). Hybridization probes were either derived from cDNA clones or synthesized by PCR, tailored for use (if necessary) by restriction digestion and subsequently purified by elution of the appropriate band following agarose gel electrophoresis. Probes were labeled with $^{32}$P by extension of random hexamer primers using Klenow fragment of DNA polymerase I in the presence of [alpha-$^{32}$P]dCTP (Feinberg and Vogelstein, 1983). Molecular sizes were determined relative to ethidium bromide-stained molecular mass markers (Bethesda Research Laboratories).

Results

Expression of CD44 in Epithelial Tissue and Cultured Cells

Localisation of CD44 in cryostat sections of human tissue, particularly epithelium, identified extensive expression in many nonhematopoietic cells. In neonatal foreskin (Fig. 1), CD44 was preferentially expressed around the entire periphery of epidermal cells in the basal and spinous layers. CD44 was differentially expressed and decreased in the stratified layers subsequent to expression of involucrin, a
Figure 2. Localization of CD44 (HS/CS250) in filopodia and cell-cell contacts of cultured HFKs. HFKs were adhered to fibronectin- or type I collagen-coated coverslips and then the indicated receptors were localized by immunofluorescence microscopy. Where indicated, focal adhesions were identified by interference reflection microscopy or antibody exclusion techniques (see Materials and Methods). (A) CD44 localized in Ca^{2+}-induced HFKs with mAb P3H9. Arrows indicate accumulations of CD44 at cell-cell contacts. (B) HFKs adhered to fibronectin. CD44 accumulates in filopodia on apical surface and at points of cell contact with the substrate (white arrows). (C, same field as B) Integrin α5β1 (fibronectin receptor) localized in focal adhesions. Arrows in B and C designate the same points and distinguish CD44 in filopodia and α5β1 in focal adhesions. (D) CD44 in filopodia at periphery of HFKs adhered to type I collagen. (E, same field as D) Focal adhesions (dark hole designated by arrows) localized by staining with anti-type I collagen (antibody exclusion technique). (F, same field as D and E) Focal adhesions identified by interference reflection microscopy: CD44 locates near but not in the focal adhesions. Bars: (A) 100 μm; (B-F) 10 μm.
component of the keratinocyte envelope (Rice and Green, 1979; compare Fig. 1, A and B). In fetal skin (Fig. 1 C), CD44 was expressed in both the epidermis and dermis, whereas fibronectin was expressed in only the dermis (Fig. 1 D). In human fetal lung (Fig. 1 E), CD44 was also expressed around the entire periphery of cells in developing bronchial epithelium surrounded by the laminin containing basement membrane (Fig. 1 F).

CD44 antigens, detected by immunofluorescent microscopy in cultured keratinocytes, were localized to the cell surface in HFKs (Fig. 2). Although CD44 was distributed over the entire cell surface, it was clearly increased in concentration in filopodia present on the apical surface of cells and at the periphery of cells where they contact the substrate (Fig. 2 B). The concentration in filopodia, however, was distinct from stable cell attachment sites such as focal adhesions (Burridge et al., 1988; Carter et al., 1990a). CD44 was not detected around the entire periphery of cells in developing bronchial epithelium surrounded by the laminin containing basement membrane (Fig. 1 F).

CD44 antigens detected by immunofluorescent microscopy in cultured keratinocytes were localized to the cell surface in HFKs (Fig. 2). Although CD44 was distributed over the entire cell surface, it was clearly increased in concentration in filopodia present on the apical surface of cells and at the periphery of cells where they contact the substrate (Fig. 2 B). The concentration in filopodia, however, was distinct from stable cell attachment sites such as focal adhesions (Burridge et al., 1988; Carter et al., 1990a). CD44 was not detected in focal adhesions of cells attached to collagen, laminin or fibronectin. In contrast, integrins α5β1 (Fig. 2 C) and α2β1, α3β1 (Carter et al., 1990a,b) were readily detected in elevated concentrations in focal adhesions of many cells. Negligible CD44 was found in the extracellular matrix.

In cultures of HFK cells that had been induced to aggregate by growth in the presence of Ca²⁺, CD44 concentrated in sites of cell–cell contact (Fig. 2 A), suggesting that it may be involved in cell–cell adhesion. This function would be consistent with the known relationship between CD44 and homotypic cell cell aggregation.

Identification of Polymorphic Forms of CD44

We have previously identified three mAbs, P1G12, P3H9, and P3H5 that react with distinct epitopes on the GP90 form of CD44 antigen from HT1080, human fibrosarcoma cells (Carter and Wayner, 1988). We have used these mAbs in immunoblotting experiments to establish that different cell populations express one or more distinct molecular mass forms of CD44, ranging from 80 to >250 kD (Fig. 3). All three mAbs identified the same forms of CD44 in each cell population, indicating that all the forms expressed the same nonoverlapping epitopes.

All the forms detected by immunoblotting or in subsequent immunoprecipitation experiments migrated as broad diffuse bands on SDS-PAGE suggesting a high degree of posttranslational modification, such as phosphorylation and/or glycosylation (Carter and Wayner, 1988): (a) HT1080 fibrosarcoma, U937 monocyte and ST-1 Epstein–Barr virus-transformed B lymphocytes all expressed at least two forms of CD44 migrating with relative molecular masses of ~90 and 180 kD. In general the 180-kD form was always present in minor quantities relative to the 90-kD form. (b) QG56 lung carcinoma (230 kD), neonatal HFKs (180 and 250 kD), and BT-20 breast carcinoma (120 kD), expressed the indicated higher molecular mass forms as the predominant species, with relatively lower quantities of the 90-kD forms detected. In contrast some cells such as Molt 4, T cell lymphoma, and Daudi, a Burkitt lymphoma expressed no detectable CD44 under these conditions.

Characterization of Sulfated and Nonsulfated Carbohydrates on CD44

We investigated carbohydrate diversity as a possible source of CD44 polymorphism. First, we examined the CD44 anti-
Figure 4. Metabolic labeling of HT1080, HFK, and QG56 cells with \( ^{35}S \)O\(_2\), immunoprecipitation and digestion of polymorphic CD44 antigens with carbohydrate degrading enzymes. QG56, HFK, and HT1080 cells were metabolically labeled with \( ^{35}S \)O\(_2\), extracted with Triton X-100 detergent and CD44 was immunoprecipitated from the extracts with mAb P1G12 yielding GP90 and CS180 from HT1080 cells, HS/CS250 from HFKs, and GP230 from QG56 cells and conditioned culture medium of QG56 cells. The purified polymorphic forms of CD44 were digested with protease-free carbohydrate degrading enzymes: chondroitin ABC lyase (CHON, from Proteus vulgaris, degrades chondroitin sulfate or hyaluronic acid); keratan sulfate, 1,4-\( \beta \)-D-galactanohydrolase (KERA, from Pseudomonas); heparitinase (HEPA, from Flavobacterium heparinum); glycopeptide-N-glycosidase (GLYC, from Flavobacterium meningosepticum); neuraminidase (NEUR, from Vibrio cholera); and neuraminidase, followed by glycopeptide-\( \alpha \)-glycosidase (GLYCO, from Diplococcus pneumoniae). The digested antigens were fractionated by SDS-PAGE (8%) and detected by fluorography. Migration of molecular mass standards are indicated. Major degradation products exhibiting greater mobility, but retaining \( ^{35}S \)O\(_2\), are identified with asterisks.
Figure 5. Effects of neuraminidase digestion on lectin binding to HT1080 GP90 and QG56 GP230. Purified GP90 from HT1080 cells and GP230 from QG56 cells were fractionated by SDS-PAGE (8%) then electrophoretically transferred to nitrocellulose. The nitrocellulose-bound glycoproteins were incubated with (+) or without (−) neuraminidase, then incubated with ^125I-labeled lectins or antibodies followed by ^125I-labeled protein A. Antibodies and lectins bound to the glycoproteins were detected by autoradiography.
Figure 6. Pulse-chase metabolic labeling of HT1080 GP90 and QG56 GP230, and HFK HS/CS250 with [35S]Met: identification of the core proteins. HT1080, HFK, and QG56 cells were pulse-labeled with [35S]Met for 30 min then chased in the presence of unlabeled Met. At the end of the indicated time periods (0, 15, 30 min, etc.), samples of the cells were extracted with Triton X-100. The extracts were immune-precipitated with either control culture supernatant (SP) or culture supernatant containing P1G12 mAb. The immune-precipitates were fractionated on SDS-PAGE gels (7%, HT1080 and QG56; 10%, HFK) followed by fluorography. Migration of HT1080 GP90, HFK HS/CS 250, and QG56 GP230 along with the major precursor core proteins, preGP90 (65 kD), HSA, b, c, and d, and GP a, b, c, and d (90–135 kD), respectively, are indicated. Analysis of duplicate samples on 10% acrylamide gels also failed to detect any lower molecular mass precursors for either GP90 or GP230.
Northern blot analysis with different CD44 probes using poly-A\(^+\) RNA from four cell populations. Poly-A\(^+\) RNA from HT1080, QG56, HFK, and HFF cells (lanes 1–4, respectively) was size fractionated by agarose gel electrophoresis, transferred to nitrocellulose and probed sequentially with a full-length lymphoid CD44 cDNA from baboon (93\% homologous to the human lymphoid sequence, CD44), a hamster actin probe (ACTIN), an EACR-specific cDNA (EACR), and a CD44 cytoplasmic domain cDNA (CYTO). The EACR-specific cDNA was a 372-bp fragment corresponding to nucleotides 803–1172 of the reported sequence. It was derived from an Rsa I digest of HFK cDNA amplified by PCR using the SJ1 and HE1 oligonucleotides. The cytoplasmic domain probe was a 140-bp fragment corresponding to nucleotides 1022–1162 of the human lymphoid sequence (Stamenkovic et al., 1989). It was derived by PCR amplification of HFF cDNA using the SJ2 and B7-2 oligonucleotides. Actin hybridization is shown for quantitative comparison of the total poly-A\(^+\) RNA loaded in each lane.

Identification of a Distinct Form of CD44 Core Protein, CD44E

We initially used pulse-chase labeling with \[^{35}S\]methionine to identify underglycosylated precursors of the fully glycosylated glycoproteins (Fig. 6). Two types of precursors were identified for the four characterized forms of CD44. HT1080 cells exhibit a single 65-kD precursor to GP90 (termed Pre-GP90). In contrast, four precursors to both QG56 GP230 and HFK HS/CS250 migrated with similar relative molecular masses of 90–135 kD (designated GPa-d or HSa-d, respectively), despite the differences in final glycosylation for these two forms.

Northern blot analysis provided further evidence that the core proteins for different CD44 forms may be similar, but not identical. A full-length cDNA probe to lymphoid CD44 detected at least three mRNA species in all cell populations examined (Fig. 7). However, epithelial cell types (i.e., HFKs and QG56 cells) exhibited major mRNA species between 1.0 and 1.6 kb longer (2.8, 3.6, and 6.7 kb) than what appeared to be their counterparts in fibroblasts (1.8, 2.4, and 5.1 kb). Upon longer exposure, the latter set of mRNA species was also apparent as minor species in the epithelial cells. It was possible that the larger set of mRNA species in these epithelial cells contained a larger coding region with unique sequences. Therefore, we amplified specific regions of cDNA
from these cells using the PCR thermocycler with oligonucleotides specific for various regions of the lymphoid cDNA (Fig. 8 A). The reactions enlisted a common downstream or 3' oligonucleotide (i.e., B7-2 in the cytoplasmic protein domain, Fig. 8 A) coupled with various upstream (i.e., towards the extracellular protein domain) oligonucleotides (Fig. 8 B). Using the SJ2 oligonucleotide in conjunction with B7-2, only one cDNA of predicted length was amplified from both fibroblasts and epithelial cells (lanes 1), suggesting that this region is similar in all transcripts that contain the cytoplasmic domain. When the upstream oligonucleotide B7-6, B7-4, or SJ1 was used in conjunction with B7-2, cDNAs amplified from HT1080 cells or HFFS were the same sizes as predicted for the lymphoid cDNA sequence. This suggested that the fibroblast CD44 cDNA is identical to the lymphoid CD44 sequence. In contrast, amplification of cDNA from HFKs (Fig. 8 B) and QG56 cells (not shown) using the same combinations of oligonucleotides resulted in not only the predicted cDNA products from the lymphoid sequence, but also several larger cDNA products. As the upstream oligonucleotide was shifted more towards the 5' end, the epithelial-specific cDNA product obtained was uniformly larger. This indicated that the additional cDNA was inserted between the SJ1 and B7-2 oligonucleotides.

To identify the gel bands containing additional inserted cDNA sequences from HFKs and QG56 cells, individual DNA bands were excised from the gels, the cDNA isolated and used as templates in a subsequent, otherwise identical
form), resulting in a glycine substitution for a glutamine in the amino acid sequence.

To determine which of the different CD44 mRNA transcripts contained the EACR sequence, Northern blots were probed with a 372-bp fragment contained within the EACR sequence (Fig. 7, EACR). All of the major mRNA species identified by the full-length lymphoid probe in QG56 cells and HFKs hybridized with the EACR-specific probe. In contrast, none of the major mRNA species in HT1080 cells or HFFs hybridized with the EACR probe. Using a cytoplasmic domain cDNA probe, we detected in the epithelial cells not only three major CD44E mRNA species, but also the three major lymphoid-associated CD44 transcripts (Fig. 7, CY70). In addition, all four cell populations express a 1.3-kb transcript similar in size to the cloned CD44 sequence reported by Stamenkovic et al. (1989).

Figure 9. cDNA and deduced protein sequences of the EACR from HFKs and QG56 cells. The EACR, spliced in immediately before nucleotide 783 of the human lymphoid sequence (Stamenkovic et al., 1989; the sequence numbers used are relative to this splice site), is composed of 396 nucleotides and shown in uppercase. The cDNA flanking the EACR (118 bp, 5' and 67 bp, 3') that was sequenced is shown in lowercase. Potential O-glycosylation (underline), N-glycosylation (bold underline) and GAG attachment sites (*) are indicated. The cDNA sequence obtained from both cell populations was identical and is available from EMBL/Genbank/DDBJ under accession number X55938.

PCR amplification. The largest PCR product (arrow, Fig. 8 B), resulted in much greater amplification of itself than the smaller lymphoid-associated product. This suggested that the largest product was a new epithelial cDNA product contaminated with minute amounts of the smaller lymphoid type cDNA (data not shown).

Both strands of the largest band from both HFKs and QG56 cells was sequenced directly using several of the oligonucleotides in Fig. 8 A as primers. The sequence was identical for both cell populations (Fig. 9), and contained an extra 396 bp between the 5' and 3' flanking lymphoid CD44 cDNA. We designated this extra cDNA the epithelial-associated coding region (EACR). Those transcripts that contain the EACR were designated as CD44E. The EACR was spliced into the CD44 mRNA before nucleotide 783 of the human sequence (Stamenkovic et al., 1989), the second nucleotide of the codon. This changed the arginine in the lymphoid form into an asparagine in the epithelial form. The EACR was unique in that it showed no significant homology to other sequences in the Genbank or EMBL databases (releases 63 and 23, respectively) and coded for an amino acid sequence composed of >30% serine and threonine residues.

The cDNA sequence derived from both flanking regions agreed exactly with one of the reported human lymphoid CD44 sequences (Goldstein et al., 1989), and varied by only a single base from the other (Stamenkovic et al., 1989). In the latter a G was substituted for an A at nucleotide 831 of the lymphoid sequence (nucleotide 1227 of the epithelial sequence), resulting in a glycine substitution for a glutamine in the amino acid sequence.

To determine which of the different CD44 mRNA transcripts contained the EACR sequence, Northern blots were probed with a 372-bp fragment contained within the EACR sequence (Fig. 7, EACR). All of the major mRNA species identified by the full-length lymphoid probe in QG56 cells and HFKs hybridized with the EACR-specific probe. In contrast, none of the major mRNA species in HT1080 cells or HFFs hybridized with the EACR probe. Using a cytoplasmic domain cDNA probe, we detected in the epithelial cells not only three major CD44E mRNA species, but also the three major lymphoid-associated CD44 transcripts (Fig. 7, CY70). In addition, all four cell populations express a 1.3-kb transcript similar in size to the cloned CD44 sequence reported by Stamenkovic et al. (1989).

Figure 10. Differential extraction, phase partitioning with Triton X-114 detergent and purification of HT1080 GP90 and QG56 GP230. HT1080 and QG56 cells were 125I-surface labeled, and sequentially extracted as previously described (Carter et al., 1982) first, with 2 M urea in 1 M NaCl to remove peripherally cell surface components, and second, with Triton X-114 (PARTITION) or X-100 (PURIFY) detergent at 4°C to solubilize membrane components followed by phase partitioning at 37°C for the Triton X-114 fractions (Bordier, 1981). The temperature increase results in formation of a detergent phase that contains membrane components with dominant hydrophobic character and an aqueous phase with dominant hydrophilic character. Each Triton X-114-derived phase was immune precipitated with anti-CD44 mAb (P1G12) and the antigens fractionated by SDS-PAGE followed by autoradiography. Unlabeled Triton X-100 extracts were used to affinity purify both HT1080 GP90 and QG56 GP230 by chromatography on WGA-Sepharose to purify all sialylated membrane glycoproteins followed by chromatography of the purified glycoproteins on PIG12-Sepharose. The bound and eluted GP90 and GP230 were labeled with 125I Iodobcad and fractionated on SDS-PAGE (GP90 on 10%, GP230 on 8%) followed by autoradiography. Neither protein stained appreciably with Coomassie blue, possibly due to extensive glycosylation.
Intrinsic Membrane Character of CD44

The GP90 form of CD44 has been previously characterized as a transmembrane glycoprotein (Carter and Wayner, 1988) and a corresponding hydrophobic transmembrane sequence has been identified in the amino acid sequence of the core protein (Idzerda et al., 1989). No such information was available for the related higher molecular mass forms of CD44. We have used PCR amplification of CD44E cDNA from HFKs (Fig. 8 C) and QG56 cells (not shown) to demonstrate that at least one CD44E transcript includes very similar coding sequences for the transmembrane and cytoplasmic domains. Moreover, the 2.8- and 3.6-kb CD44E transcripts comigrate with a subset of the CD44 transcripts detected with a probe from the cytoplasmic domain (Fig. 7, EACR and CYTO). This suggested that these CD44E transcripts express cytoplasmic, and presumably hydrophobic transmembrane, domains similar to those in CD44. To analyze the polymorphic forms of CD44, we surface labeled HT1080 and QG56 cells with 125I and subjected them to a sequential extraction as described in the legend to Fig. 10. Immunoprecipitation of the labeled cell extracts or immunoblotting of the unlabeled cell extracts with anti-CD44 mAbs detected the CD44 forms primarily in the Triton X-114 extract. As seen in Fig. 10 (Partition), temperature-induced phase partitioning of the Triton X-114 extract into detergent and aqueous phases further distinguished HT1080 GP90 and QG56 GP230; GP230 was recovered in the aqueous phase and as previously reported, GP90 was recovered in the detergent phase. In results not shown HT1080 CS180 and HFK HS/CS250 were recovered primarily in the Triton X-114 aqueous phase. Thus, all the CD44 forms required detergent for solubilization. However, the additional glycosylation and other possible modifications altered the partitioning character of the higher molecular mass forms.

We further establish that the larger polymorphic forms are intrinsic membrane glycoproteins with a hydrophobic domain that intercalates into the lipid bilayer of the plasma membrane. Unlabeled QG56 GP230 (and HT1080 GP90) was purified from Triton X-100 detergent extracts by affinity chromatography, and the purified proteins were labeled and incorporated into liposomes. Autoradiographs of the purified, labeled GP230 (and GP90) detected one major band in each preparation after SDS-PAGE (Fig. 10, PURIFY). Liposomes were prepared in the presence or absence of purified GP230. Only those liposomes containing GP230 would bind to immobilized anti-CD44 mAbs, indicating that GP230 incorporated into the liposomes in a proper configuration to interact with the antibodies (Table I).

Discussion

Peptide and Carbohydrate Polymorphism in CD44

We have identified and characterized the CD44 antigen as a polymorphic intrinsic membrane proteoglycan/glycoprotein with alternate splicing of the core peptide as well as extraordinary diversity in both the type and quantity of carbohydrate moieties. All the forms of CD44 contained at least one type of sulfated carbohydrate unit as well as nonsulfated carbohydrates. Previous reports relating to the CD44 have focused on hematopoietic cells (Jalkanen et al., 1988; Goldstein et al., 1989; Stamenkovic et al., 1989; Pals et al., 1989) and have identified chondroitin sulfate as a carbohydrate moiety of the relatively minor higher molecular mass forms. The present study has established that epithelial cells, in particular, express the higher molecular mass forms as the major species of CD44. Glycosylation of these forms can be unique for each cell type with heparan sulfate being the major carbohydrate derivative of epithelial cells. Glycosylation of the related core proteins may reflect major differences in the glycosylation pathways of the individual cells and/or functional differences for the polymorphic forms (Paulson, 1989). Others have reported cell-type specific polymorphic forms of cell surface proteoglycans such as syndecan (Sanderson and Bernfield, 1988). Based on cDNA sequences, however, CD44E and human syndecan (Mali et al., 1990) have unrelated core proteins.

Lymphoid and mesenchymal cells express three major mRNA species for CD44, all presumably coding for a 37-kD core protein (Idzerda et al., 1989; Stamenkovic et al., 1989). In contrast, HFKs and QG56 cells express mRNA species for an alternate core protein, CD44E, in addition to the three mRNAs for CD44. The CD44E mRNAs code for a core protein with an additional 14 kD inserted in the extracellular domain that we have designated the EACR. The DNA sequence flanking the EACR was identical to that previously reported for the lymphoid core protein. Thus, the CD44E mRNA species probably arise from alternative splicing of a common gene rather than the use of an alternate CD44E gene.

The extra 132 amino acids in the CD44E core protein are spliced in the region near the COOH-terminal portion of the extracellular domain. The sequence flanking the splice site contains four repeats of the SGXG motif (Stamenkovic et al., 1989; Goldstein et al., 1989), a presumptive attachment site for chondroitin sulfate chains (Bourdon et al., 1987). Interestingly, the EACR is inserted in between these SGXG repeats and contains only one additional SG dipeptide at the COOH terminus. This extra repeat is a poor candidate for a chondroitin sulfate attachment site, though, because it contains neither upstream acidic residues nor the complete SGXG motif described by Bourdon et al. (1987). However, five potential O-glycosylation sites composed of at least three consecutive serine or threonine residues (Tomita et al., 1978) are found in the EACR sequence, adding to three similar sites in the flanking regions. Thus, QG56 cells may use these sites for the attachment of O-linked carbohydrate chains to GP250. The presence of heparan sulfate chains on the HFK

<table>
<thead>
<tr>
<th>Adhesive surface</th>
<th>Liposomes + GP230</th>
<th>Liposomes - GP230</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSA</td>
<td>516</td>
<td>644</td>
</tr>
<tr>
<td>IgG</td>
<td>594</td>
<td>681</td>
</tr>
<tr>
<td>P3H9</td>
<td>7,612</td>
<td>58</td>
</tr>
</tbody>
</table>

GP230 was purified as described in Fig. 10 and then incorporated into liposomes prepared from [3H]phosphatidylcholine, unlabeled egg yolk phosphatidylcholine, and octylglucoside detergent by dialysis (Mimms et al., 1981; Pyeula et al., 1985). The labeled liposomes prepared with and without GP230 were incubated with plastic surfaces that had been coated with anti-CD44 antibody (P3H9) or control IgG and BSA (Carter and Wayner, 1988). The bound liposomes were solubilized and quantitated in a scintillation counter.

The Journal of Cell Biology, Volume 113, 1991

218
HS/CS250 and the QG56 "HEP" released form, but not on other forms of CD44, suggests that the EARC sequence may also contain as yet undescribed heparan sulfate attachment sites. Alternatively, the presence of the EARC may induce a new protein conformation exposing distant and previously buried heparan sulfate attachment sites in the flanking regions.

The expression of CD44E mRNAs by HFKs and QG56 cells correlated with the observation by pulse/chase labeling of higher mass precursors. QG56 GP230 and HFK HS/CS250 appear to originate from precursors that were distinct from the 65-kD HT1080 GP90 precursor. The QG56 and HFK precursors were similar, if not identical, despite the differences in final glycosylation. Nevertheless, we cannot conclude that the GP230 and HS/CS250 forms contain the differences in final glycosylation. Nevertheless, we can not conclude that the GP230 and HS/CS250 forms contain the CD44E core protein since mRNA species for both core proteins are expressed by these cells.

HT1080 GP90 contains a phosphorylated cytoplasmic domain that is linked to the glycosylated extracellular domain by a transmembrane hydrophobic sequence of amino acids based on both biochemical (Carter and Wayne, 1988) and cDNA sequence data (Idzerda et al., 1989; Stamenkovic et al., 1989). Similarly, the QG56 GP230 and HFK HS/CS250 both required detergent (but not high salt or urea) for solubilization and could be incorporated into liposomes, indicating the presence of a dominant hydrophobic domain. Moreover, at least two of the CD44E mRNA species codes for transmembrane and cytoplasmic domains very similar, if not identical, to the CD44 core proteins previously reported in lymphoid cells. This observation suggests that the CD44E core protein expressed by these mRNA species is not part of a form destined for secretion and may be used in the HFK HS/CS250 and QG56 GP230 forms. In addition, these CD44E core proteins would be expected to have similar associations with intracellular components, unlike the core protein reported by Goldstein et al. (1989) that lacks most of the cytoplasmic domain.

Functional Role for CD44 in Cell Adhesion
Filopodia have been suggested to be directly involved in initial cell–cell and cell–substratum contacts (Albrecht-Buehler et al., 1976). CD44 localizes in filopodia that contact with the substrate and in cell–cell contact sites suggesting that CD44 may form weak initial cell interactions. CD44 does not appear to be a cell–substrate adhesion receptor for collagen, fibronectin, or laminin (Wayner and Carter, 1987; Carter and Wayne, 1988). This conclusion is based on the absence of CD44 in focal adhesions, a major site of stable cell adhesion, and the inability of anti-CD44 antibodies to inhibit cell adhesion to ECM ligands. However, CD44 is an intrinsic membrane proteoglycan, a group of macromolecules that have been proposed as mediators of cell–substrate adhesion (Cole et al., 1985; Koda and Bernfield, 1984; Laterra et al., 1983; Woods et al., 1984, 1985; LeBaron et al., 1988; Ruoslahti, 1988). We have recently identified an alternate adhesion structure, termed a stable anchoring contact (SAC; Carter et al., 1990b). The possibility that CD44 interacts with SACs and/or other adhesion structures is currently under investigation. As a possible mechanism, CD44 expresses a partial sequence homology with mouse link protein (Idzerda et al., 1989; Goldstein et al., 1989; Stamenkovic et al., 1989). Link protein stabilizes the interaction of hyaluronate with the core protein of proteoglycans in cartilage (Neame et al., 1987). Recently, hyaluronate was shown to be a possible ligand for CD44 (Miyake et al., 1990b; Aruffo et al., 1990).

The possible role of CD44 in mediating cell adhesion may make use of the variability in carbohydrate content. The presence of major amounts of glycosaminoglycan on the core protein of HS/CS250 may modulate cell–cell interactions directly or via components of the extracellular matrix. Normal differentiation, stratification as well as abnormal lesions of epithelium have been reported to correlate with the differential expression of carbohydrate structures in skin (Dabelstein et al., 1986). CD44 is one of the major glycoconjugates of differentiating epidermis and exhibits an extraordinary sensitivity to regulation in carbohydrate content. This suggests a possible regulatory role for glycosylation in the function of CD44 in epidermis. Whether or not CD44 is primarily an adhesive glycoconjugate, its extensive polymorphism in both protein and carbohydrate dictate that care is required to identify the CD44 isoform responsible for a particular event when examining functional corollaries.

Support for this research was provided by grants to W. G. Carter from the National Institute of Health #RO1-CA49259.

Received for publication 17 September 1990 and in revised form 17 December 1990.

References


Jalkanen, P. D., A. H. Batchelor, P. Lawlor, and D. R. Katz. 1990. The role of CD44 contributes to T cell activation.


Mali, M., P. Jankkola, A. M. Arvilommi, and M. Jalkanen. 1990. Sequence of HLA-A, BcDNA by using the polymerase chain reaction: frequency and distribution


Tew, E. S., C. J. Crook, P. Lawlor, F. Hogervorst, G. D. Keizer, T. Thepen, E. Horst, and C. C. Fig

Today. 10:423-428.


Miyake, K., K. L. Medina, S. Hayashi, S. Ono, T. Hamaoka, and P. W. Kin


