Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 8521–8525, November 1989
Immunology

Isolation of mouse CD44 cDNA: Structural features are distinct from the primate cDNA

(Pgp-1/polymerase chain reaction/allotypes/poly(A)$^+$ RNA)

CAROL NOTTENBURG*, GEORGIA REES†, AND TOM ST. JOHN†‡

*Division of Clinical Research and †Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, WA 98104

Communicated by Irving L. Weissman, August 10, 1989 (received for review July 7, 1989)

ABSTRACT **CD44 is a glycoprotein that participates in adhesion of lymphocytes to high endothelial cells of lymph organs and likely functions in other intercellular adhesions as well. We have isolated a mouse CD44 cDNA by polymerase chain reaction amplification of cDNA synthesized from total cellular RNA isolated from 38C-13, a B-lymphocyte cell line. The oligonucleotide sequences were based at the 5′ end on the baboon CD44 sequence upstream of and including the translation initiator ATG and at the 3′ end on a mouse CD44 sequence that was determined from a rare, incomplete cDNA clone. The mouse CD44 DNA sequence is similar to the baboon CD44 sequence but has structural features that are distinct. The external domain has a region that is only 35% similar between the mouse and primate proteins, in contrast to 85%–90% similarity in the rest of the sequence. In addition, just upstream from the predicted cleavage site of the leader peptide, nucleotide insertions result in the addition of two or four amino acids, depending upon the mouse strain. An amino acid replacement between two strains carrying different CD44 (also called Pgp-1) allotypes is likely responsible for the Pgp-1 polymorphism. The most striking aspect of mouse CD44 is that the RNA does not fractionate with polyadenylylated RNA, unlike the CD44 RNA in human, baboon, rat, and chicken.**

The CD44 membrane protein has been defined by monoclonal antibodies that detect the mouse or the primate forms. Antibodies to the CD44 proteins, originally referred to as the Hermes-1 antigen (1) and ECMRIII (extracellular matrix receptor) (2) in human and Pgp-1 in both mouse and human (3, 4), have been shown to recognize the same molecule (5, 6). CD44 is found on the cell surface of many different cell types, including hematopoietic cells such as lymphocytes, monocytes, granulocytes and erythrocytes, and nonhematopoietic cells such as fibroblasts and epithelial cells (2, 7).

CD44 has been implicated in promoting adhesion of cells to various substrates. The ECMRIII molecule is found on essentially all nucleated cells and interacts with both type I and type VI collagen (2). The monoclonal antibodies Hermes-1 and -3 detected a gp90 molecule on lymphocytes capable of binding to specialized endothelial cells in postcapillary venules (HEV) (1, 8), the major site of entry of circulating lymphocytes into lymph nodes and Peyer's patches. Hermes-3 blocks the binding of lymphoid cells to HEVs of mucosal lymphoid organs (8), affirming a critical role of this gp90 in mediating adhesion of lymphocytes to HEV. In addition, both ECMRIII and Pgp-1 have been shown to be associated with the cytoskeleton (2, 9, 10). These widely distributed specificities argue that the CD44 molecule functions as a multispecific adhesion receptor, and HEV adhesion is but one facet of this activity.

Recently, CD44 cDNAs have been isolated from nonhuman primate and human cDNA libraries (11–13). The CD44 sequence is distinct from other adhesion molecules such as members of the integrin family or the MEL-14 antigen family of adhesion proteins (14). In concordance with the immunological data, mRNA encoding CD44 was detected in a variety of tissues, including those that are nonhematopoietic in origin (12, 31). Questions addressing how specific interactions of lymphocytes and HEV cells are mediated by a protein that is widely expressed are easily approached in a model system such as mouse, with the advantages of homozygous inbred strains, CD44 (Pgp-1) allotypes (15), and easily obtainable and manipulatable tissues. With this goal in mind, we have isolated a mouse CD44 cDNA.

Surprisingly, mouse CD44 sequences were absent in a variety of cDNA libraries, some derived from cell types expressing large quantities of the Pgp-1 protein. A single cDNA clone that did not contain the entire coding region was isolated from a cDNA library made from 14- to 15-day-old mouse embryos. Therefore, an alternate strategy was devised to clone the mouse CD44 cDNA. CD44 was amplified by the technique of polymerase chain reaction (PCR) from cDNA by using one oligonucleotide sequence derived from the baboon sequence and the other oligonucleotide sequence derived from the mouse sequence. The 5′ oligonucleotide, based on the baboon sequence, we expected to be an exact match at the 3′-most bases by including the DNA sequences encoding the initiator methionine codon, which is conserved for all eukaryotic genes. Overall, the mouse and primate CD44 sequences have similar sequences and features, but several exceptional disparities are apparent. The mouse CD44 cDNA sequence§ contains a region of little similarity to the primate species and an insertion of amino acids near the leader peptide cleavage site. In contrast to several species that have been examined, the mRNA for mouse CD44 does not fractionate with poly(A)$^+$ RNA.

MATERIALS AND METHODS

PCR Amplification. Fifty micrograms of total cellular RNA from 38C-13, L cell, and both BALB/c and C57Bl/J mouse spleens were transcribed into cDNA by reverse transcriptase from random hexamer oligomers (unpublished data). Purified cDNA was amplified by *Thermus aquaticus* (*Taq*) DNA polymerase by extension of the 5′ oligonucleotide AATGAATTCTGCGCCCTCGGTTGGCTCCGGACGCCATGG and the 3′ oligonucleotide GCACGATTGGAAACGTCATTGAATTCCATT, present at a concentration of 1 μM each in 50 mM KCl/10 mM Tris, pH 8.3/1.5 mM $MgCl_2$/200 μM each dATP, dCTP, dGTP, and dTTP/0.1% gelatin/40 units of *Taq*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HEV, high endothelial cells in postcapillary venules; PCR, polymerase chain reaction.

‡To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M27129 and M27130).

DNA polymerase per ml. The 25-μl reaction mixture was denatured for 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 in a DNA thermal cycler (Perkin–Elmer). Some amplifications used an annealing step of 2 min at 60°C with no appreciable difference in outcome. The reaction mixture was electrophoresed on a 1.5% low-melting-point agarose gel, and the major fragment that migrated at 1.5 kb was excised and eluted. After *Eco*RI digestion, the cDNA–PCR fragment was ligated into *Eco*RI-digested and dephosphorylated pBSM13$^+$ vector. The identity of the clones was confirmed by hybridization with a *Bam*HI/*Apa* I fragment from the cDNA clone MM100.

RNA Blots. The RNA in Fig. 3*A* was made by lysing 38C-13, 594S, and HL-60 cells in 500 μg of proteinase K per ml/0.15 M NaCl/10 mM Tris, pH 7.4/1% sodium dodecyl sulfate (SDS) (16). The solution was adjusted to 0.5 M NaCl and applied to a poly(U)-Sepharose matrix. The nonbinding fraction (containing poly(A)$^-$ RNA) was collected and precipitated with ethanol. The bound fraction (containing poly(A)$^+$ RNA) was eluted with water, adjusted to 0.3 M NaOAc (pH 7.0), and precipitated with ethanol. One microgram of each poly(A)$^+$ RNA was treated with glyoxal (17), electrophoresed on a 1.2% agarose gel, and transferred to nylon. Postmeiotic round spermatids were purified from adult NIH Swiss Webster mice (18). Total RNA was prepared by lysing the cells in 5 M guanidinium isothiocyanate/25 mM sodium citrate, pH 7.2/0.5% sarcosyl/0.1 M 2-mercaptoethanol and was centrifuged through a 5.7 M CsCl/100 mM EDTA cushion (19). The RNA pellet was suspended in water, extracted with phenol/chloroform, 1:1 (vol/vol), precipitated, and resuspended in water. This RNA solution was adjusted to 20 mM Tris, pH 7.6/0.5 M NaCl/1 mM Na_2EDTA/0.1% SDS, heated to 65°C, cooled, and applied to an oligo(dT)-cellulose matrix. The bound fraction was eluted with 10 mM Tris, pH 7.5/1 mM Na_2EDTA/0.05% SDS. Bound and unbound fractions were precipitated with ethanol. Eighteen micrograms of total or poly(A)$^-$ RNA and 1 μg of poly(A)$^+$ RNA were treated with formaldehyde and electrophoresed on a 1.2% agarose gel. The RNA was transferred to nitrocellulose membrane. The RNA gel transfer in Fig. 3 *A* and *B* was subsequently hybridized with a ^{32}P-labeled 1.5-kb fragment generated from the C3H CD44 clone by PCR amplification.

Library Screening. Approximately 30,000 recombinant phages from λgt11-38C-13, λgt10-EL4, λgt10-C57Bl/J thymocyte, λ345-EL4, and λgt10-BALB/c embryo libraries were plated on each of 100 150-mm L agar plates. Nylon filter replicas were prepared. Two baboon *Bam*HI fragments containing 600 base pairs (bp) of the 5′ portion of the cDNA (11) were ^{32}P-labeled by the extension of random hexamers by DNA polymerase I in the presence of [^{32}P]dCTP and were hybridized at 2×10^6 dpm per ml in 35% formamide/5× SSPE/0.1% Ficoll/0.1% polyvinylpyrrolidone/0.1% bovine serum albumin/0.5% SDS at 42°C for 18 hr (1× SSPE is 10 mM $H_{1.5}Na_{1.5}PO_4$, pH 7.0/0.18 mM NaCl/1 mM Na_2EDTA).

DNA Sequencing. DNA sequencing was performed with a Sequenase or Taquence kit (United States Biochemical). Sequences were obtained from the universal primer, and specific mouse and baboon oligonucleotides were spaced approximately 250 bp apart across the entire clone.

RESULTS

Isolation of Mouse CD44 Sequences. In attempting to isolate a mouse CD44 cDNA clone, we hybridized several radiolabeled fragments and 30 base oligomers containing baboon CD44 sequences to a total of approximately 30×10^6 cDNA clones from seven different cDNA libraries. Only one clone, MM100, present in a library prepared from 14- to 15-day-old BALB/c embryos, hybridized to baboon CD44. The MM100 cDNA clone is 2.2 kilobases (kb) in length, of which 1.6 kb was 3′ noncoding sequence. The DNA sequence of this clone revealed that the 5′ end of MM100 corresponded to mouse coordinate 588 (Fig. 1).

The CD44 protein is abundant on the cell surface of several of the cell types used to make the cDNA libraries (unpublished observations), but the DNA sequences were essentially absent in cDNA libraries. One hypothesis to explain these observations is that CD44 mRNA did not fractionate with the poly(A)$^+$ RNA used to generate the cDNA libraries.

```
          gaattctgcgccctcggttggctccggacgcc ATG GAC AAG TTT T          45
                                           M   D   K   F   W
GG TGG CAC ACA GCT TGG GGA CTT TGC CTC TTG CAG TTG AGC CTG G          90
  W   H   T   A   W   G   L   C   L   L   Q   L   S   L   A
CA CAT CCA CAT CAG CAG ATC GAT TTG AAT GTA ACC TGC CGC TAC G         135
  H   P   H   Q   Q   I   D   L  [N]  V   T   C   R   Y   A
CA GGT GTA TTC CAT GTG GAG AAA AAT GGC CGC TAC AGT ATC TCC C         180
  G   V   F   H   V   E   K   N   G   R   Y   S   I   S   R
GG ACT GAG GCA GCT GAC CTC TGC CAG GCT TTC AAC AGT ACC TTA C         225
  T   E   A   A   D   L   C   Q   A   F  [N]  S   T   L   P
CC ACC ATG GAC CAA ATG AAG TTG GCC CTG AGC AAG GGT TTT GAA A         270
  T   M   D   Q   M   K   L   A   L   S   K   G   F   E   T
CA TGC AGG TAT GGG TTC ATA GAA GGA AAT GTG GTA ATT CCG AGG A         315
  C   R   Y   G   F   I   E   G   N   V   V   I   P   R   I
TT CAT CCC AAC GCT ATC TGT GCA GCC AAC CAC ACA GGA GTA TAT A         360
  H   P   N   A   I   C   A   A  [N]  H   T   G   V   Y   I
TC CTC GTC ACG TCC AAC ACC TCC CAC TAT GAC ACA TAT TGC TTC A         405
  L   V   T   S  [N]  T   S   H   Y   D   T   Y   C   F  [N]
AT GCC TCA GCC CCT CCT GAA GAA GAC TGT ACA TCA GTC ACA GAC C         450
  A   S   A   P   P   E   E   D   C   T   S   V   T   D   L
TA CCC AAT TCC TTC GAT GGA CCG GTT ACC ATA ACT ATT GTC AAC C         495
  P   N   S   F   D   G   P   V   T   I   T   I   V   N   R
GT GAT GGT ACT CGC TAC AGC AAG AAG GGC GAG TAT AGA ACA CAC C         540
  D   G   T   R   Y   S   K   K   G   E   Y   R   T   H   Q
AA GAA GAC ATC GAT GCT TCA AAC ATT ATA GAT GAC GAT GTC AGC A         585
  E   D   I   D   A   S   N   I   I   D   D   D   V   S   S
GC GGC TCC ACC ATC GAG AAG AGC ACC CCA GAA AGC TAC ATT TTG C         630
  G   S   T   I   E   K   S   T   P   E  [S]  Y   I   L   H
AC ACC TAC CTT CCT ACT GAA CAG CCT ACT GGA GAT CAG GAT GAC T         675
  T   Y   L   P   T   E   Q   P   T   G   D   Q   D   D   S
CC TTC TTT ATC CGG AGC ACC TTG GCC ACC AGA GAT CGA GAC TCA T         720
  F   F   I   R   S   T   L   A   T   R   D   R   D   S   S
CC AAG GAC TCC AGG GGG AGT TCC CGC ACT GTG ACT CAT GGA TCC G         765
  K   D   S   R   G   S   S   R   T   V   T   H   G   S   E
AA TTA GCT GGA CAC TCA AGT GCG AAC CAG GAC AGT GGA GTG ACC A         810
  L   A   G   H   S   S   A   N   Q   D   S   G   V   T   T
CA ACT TCT GGT CCT ATG AGG AGA CCT CAG ATT CCA GAA TGG CTC A         855
  T   S   G   P   M   R   R   P   Q   I   P   E   W   L   I
TC ATC TTG GCA TCT CTC CTG GCA CTG GCT CTG ATT CTT GCC GTC T         900
  I   L   A   S   L   L   A   L   A   L   I   L   A   V   C
GC ATC GCG GTC AAT AGT AGG AGA AGG TGT GGG CAG AAG AAA AAG C         945
  I   A   V   N   S   R   R   R   C   G   Q   K   K   K   L
TG GTG ATC AAC GGT GGC AAT GGG ACG GTG GAA GAC AGG AAA CCC A         990
  V   I   N   G   G  [N]  G   T   V   E   D   R   K   P   S
GT GAG CTC AAC GGG GAG GCC AGC AAG TCT CAG GAA ATG GTG CAT T        1035
  E   L   N   G   E   A   S   K   S   Q   E   M   V   H   L
TG GTG AAC AAG GAA CCA TCA GAG ACC CCG GAC CAG TGT ATG ACA G        1080
  V   N   K   E   P   S   E   T   P   D   Q   C   M   T   A
CT GAC GAG ACC CGG AAT CTG CAG AGT GTG GAC ATG AAG ATT GGG G        1125
  D   E   T   R   N   L   Q   S   V   D   M   K   I   G   V
TG tagtgcctacgccattaacttgaaaagacagcacgattggaaacgtcattgaattc         1183
```

FIG. 1. DNA sequence of the mouse CD44 cDNA. Nucleotides and predicted amino acids (in single-letter code) are shown. Underlined nucleotides at the ends of the sequence are the two oligonucleotide sequences used to amplify the CD44 coding region. The oligonucleotide sequences presented are not the full-length nucleotides used; only the portion of the oligonucleotides contained within the *Eco*RI PCR fragment is shown. Potential N-linked carbohydrate attachment sites are boxed, cysteines are underlined, the putative transmembrane and leader peptide sequences are shaded, and the single amino acid that is strain specific is in a shaded box.

Indeed, ^{32}P-labeled MM100 sequences did not detect an RNA in four mouse cell lines and mouse embryo cells by blot-hybridization analysis of poly(A)$^+$ RNAs (not shown). Therefore, an alternate strategy was pursued to isolate a cDNA.

We amplified the sequences encoding mouse CD44 by PCR techniques (20). Total RNA from the 38C-13 cell line, an abundant expresser of CD44, was used to generate cDNA by extension of random hexamer primers by reverse transcriptase. This cDNA provided the template for PCR amplification. DNA sequence analysis of the MM100 clone allowed us to choose one oligonucleotide for amplification in the 3′ untranslated region just downstream of the termination codon (Fig. 1). The other oligonucleotide was selected from baboon sequences upstream of the initiator methionine codon. Because of the absence of 3′→5′ nucleolytic activity, *Taq* DNA polymerase can only extend a primer if the 3′-most bases of the primer are annealed to the template. The 3′ end of the chosen 5′ oligonucleotide sequence was ATGG-3′ (Fig. 1), the ATG being the methionine used to initiate translation in virtually all eukaryotic proteins and the second G being found 46% of the time in the +4 position (21).

A fragment of the expected size, approximately 1500 bp, based on homology to the baboon CD44 sequence, was the major fragment generated. Each oligonucleotide also contained an *Eco*RI restriction site that was used to clone the fragment into an appropriately prepared plasmid vector. Hybridization of a MM100 radiolabeled fragment confirmed that we had indeed cloned mouse CD44.

Sequence of Mouse CD44 cDNA. The DNA sequence and deduced amino acid sequence of mouse CD44 is presented in Fig. 1. This DNA sequence is a consensus of three sequences derived from independent PCR amplifications from 38C-13 (a C3H strain B-lymphoma line) and two sequences derived from independent amplifications from L cells (a C3H fibroblast cell line). Multiple complete sequences were obtained to identify misincorporated bases occurring either during cDNA synthesis by reverse transcriptase or during amplification by *Taq* polymerase, which has been shown to have a significant error rate (1 in 400 bases after 30 cycles) (22). We identified two base discrepancies out of 5915 bases sequenced, establishing the overall error rate as approximately 1 in 3000 bases after 30 cycles. The sequence in Fig. 1 contains the consensus base at these two positions. An additional independent and complete sequence was obtained from a C57Bl/J, a mouse strain with the same Pgp-1 allotype as C3H (Pgp-1.2) (15). The sequence of this clone was identical to the C3H sequence in Fig. 1.

Potential Allotypic Differences of CD44. To examine the potential contribution of amino acid substitutions to the CD44 (Pgp-1) allotypes, we obtained a complete DNA sequence from the PCR product of BALB/c CD44. Six bases at nucleotide coordinates 96–101 were deleted relative to the C3H strain sequence. One other strain-dependent nucleotide change was found at nucleotide 618 and changes a serine residue in C3H to a glycine residue in BALB/c. This BALB/c-specific alteration was confirmed by sequence analysis of the independent MM100 cDNA clone. Two other nucleotide differences are silent, with no effect on the predicted protein sequence.

Comparison of the Mouse and Primate CD44 Gene Products. The mouse CD44 gene product has generally similar features to the primate CD44 gene product. The CD44 proteins have conserved the potential N-linked glycosylation sites, cysteine residues (with one additional cysteine in the cytoplasmic domain of the mouse protein), the location of the transmembrane segment, and the length of the cytoplasmic tail. Minor differences include several three-base insertions and deletions that do not alter the reading frame (Fig. 2).

Two regions display major differences between the mouse and primate sequences. There are different numbers of amino

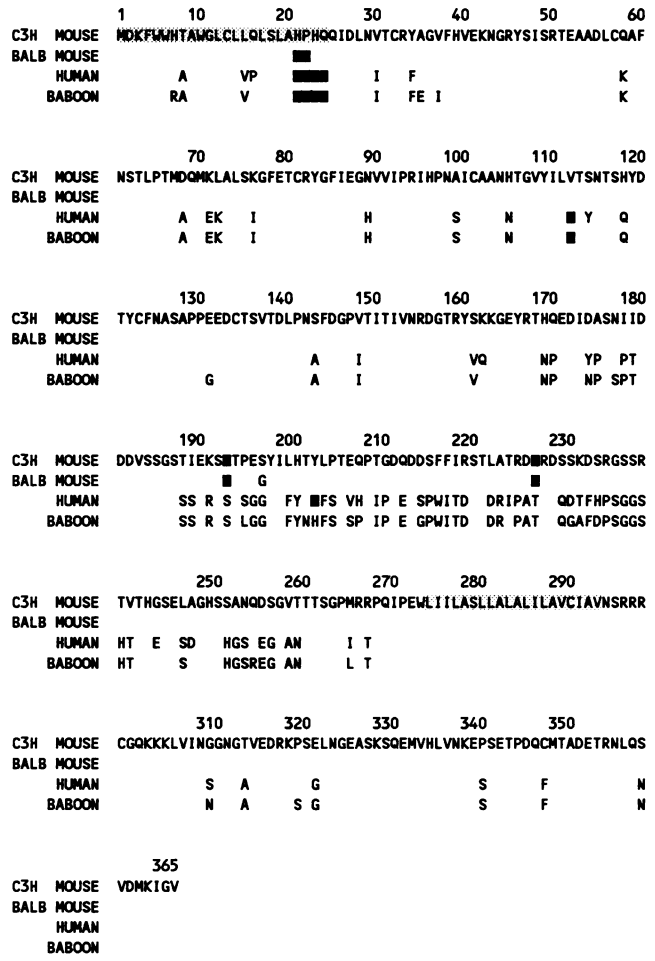

FIG. 2. Comparison of the protein sequences of CD44. The protein sequence determined from the nucleotides of the C3H strain sequence is shown in the top line. Amino acid differences found in BALB/c strain, human (12), and baboon (11) are presented. Filled boxes are deletions. The putative leader peptide and transmembrane segments are shaded.

acids preceding the probable cleavage site of the leader peptide. The predicted +1 amino acid for baboon and human CD44 is Gln-25 (mouse coordinate) (11–13, 23, 24). Immediately preceding Gln-25 are four additional amino acids in the C3H sequence (His-Pro-His-Gln) and two additional amino acids in the BALB/c sequence (His-Gln). These two or four amino acids of mouse sequence may alter the cleavage site of the mature peptide, although the precise site has not been determined for any CD44 protein.

The amino acid similarity for amino acids 1–187 is 85% between mouse and human (or baboon) and 95% between human and baboon; the transmembrane and cytoplasmic regions have even greater sequence similarities of 90–98% between the three species. However, a strikingly dissimilar region of protein sequence can be found from amino acid 187 to 259. In this region, the sequence between mouse and human (or baboon) is only 35% similar. Human and baboon are 83% similar in this region, making it the region of lowest similarity between these two very closely related species.

Expression of the Mouse CD44 Gene. Direct analysis of CD44 in unfractionated, poly(A)$^+$, and poly(A)$^-$ fractionated RNA was pursued to examine whether mouse CD44 sequences are essentially absent in cDNA libraries because the CD44 mRNA is not included in polyadenylylated mRNA preparations. Fig. 3*A* presents poly(A)$^+$ RNA samples from 38C-13, 594S (a baboon lymphoid cell line), and HL-60 (a

human cell line), which had been hybridized with the entire mouse CD44 cDNA. In contrast to the two RNAs that are present in poly(A)$^+$ RNA from the human and baboon cell lines, no RNA was detectable in mouse poly(A)$^+$ RNA from 38C-13, which expresses large amounts of CD44. The hybridizations were performed under stringent conditions. Cross-species hybridization in these conditions is generally reduced, yet strong human and baboon signals were still apparent. Hybridization of the same RNAs with a mouse ubiquitin sequence (T.S., unpublished data) showed that the mouse poly(A)$^+$ mRNA was intact (Fig. 3*A Lower*). In a similar experiment, mouse CD44 hybridized to comparably sized poly(A)$^+$ RNAs from rat embryo fibroblasts and a chicken bursal lymphoma (not shown).

Hybridization of CD44 to RNAs from round spermatids showed that two CD44 RNAs were present in unfractionated (total cellular RNA) and the poly(A)$^-$ fractions but were not present in the poly(A)$^+$ fractionated mRNA (Fig. 3*B*). Hybridization of this blot for mouse protamine 2 RNA sequences revealed no degradation of the RNAs. Although the sizes of CD44 RNAs were similar in size to ribosomal RNAs, the RNAs detected in these hybridizations were not ribosomal RNAs. In a separate experiment, no hybridization of CD44 occurred to mouse brain and muscle total RNA, while there was hybridization to mouse spleen and liver total RNA.

DISCUSSION

Isolation of Mouse CD44. PCR amplification was exploited to isolate a mouse CD44 cDNA sequence. Specific oligonucleotide sequences were designed based on DNA sequences at the 5′ end of the baboon cDNA and at the 3′ end of the mouse cDNA. The assumption was made that homology between the baboon and mouse sequences would be sufficient to allow annealing of the baboon oligonucleotides to mouse cDNA. Enzymatic extension of the baboon oligonucleotide could proceed from the ATG codon used in eukaryotic RNAs to initiate translation; our oligonucleotide sequence ended with ATGG-3′; the 3′-most G is found in almost half of all eukaryotic messages (21). This method is of general utility to clone genes from an organism when the only DNA sequence data are from a different species. In this case, the 3′ end of an oligonucleotide at the 3′ end of a gene could contain a mixture of the three terminator codons.

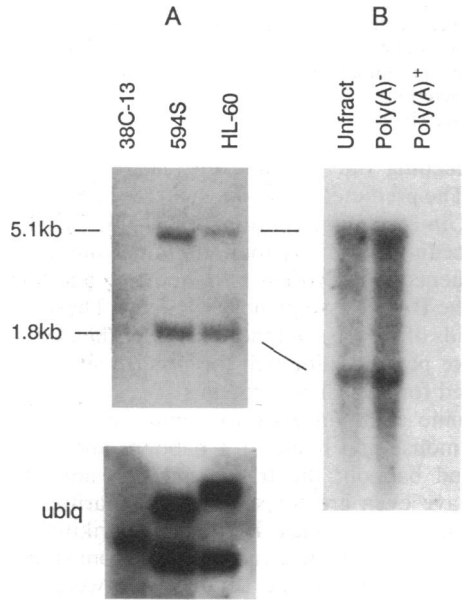

FIG. 3. RNA analysis of the CD44 gene. (*A*) Poly(A)$^+$ RNAs from 38C-13 (mouse), 594S (baboon), and HL-60 (human) were electrophoresed, transferred to a nylon membrane, and hybridized with a ^{32}P-labeled mouse CD44 fragment (*Upper*) and hybridized with a ^{32}P-labeled ubiquitin fragment (*Lower*). (*B*) Total cellular RNA from round spermatids was fractionated into poly(A)$^-$ and poly(A)$^+$ fractions by selection on an oligo(dT)-cellulose matrix; 18 μg of unfractionated RNA, 18 μg of poly(A)$^-$ RNA, and 1 μg of poly(A)$^+$ RNA were electrophoresed, transferred to a nitrocellulose membrane, and hybridized with a ^{32}P-labeled CD44 fragment as in *A*.

Sequence of the CD44. A region of 35% amino acid homology (amino acids 187–259) between primate and mouse CD44 proteins maps to the extracellular domain. This region is also most dissimilar between the human and the baboon sequences. Because of the great similarity in the rest of the molecule, we predict that a large fraction of existing monoclonal antibodies to CD44 detect epitopes in the dissimilar region. Indeed, at least three rat monoclonal antibodies to human CD44 detect a bacterial fusion protein containing sequences largely derived from this domain (13). This fusion protein contains human sequences corresponding to mouse amino acids 158–240. One antibody to an epitope in this region, the Hermes-3 epitope, is reported to block lymphocyte adhesion to mucosal HEVs. The precise location of the Hermes-3 epitope is undetermined. Hermes-3 may inhibit adhesion indirectly by steric hindrance or conformational changes of CD44. Nonetheless, in general, proteins with similar functions, like the human, baboon, and mouse CD44 proteins, have relatively similar sequences. If the adhesive function involves specific polypeptides, then the actual "adhesion" domain is more likely to map outside the dissimilar region. Alternatively, if the adhesion properties of CD44 are determined by carbohydrates (e.g., O-linked oligosaccharides or chondroitin sulfate), then the dissimilar region of these three molecules may be included in the adhesive domain. The endpoints of one CD44 exon have been mapped (31). This exon, found in a cDNA clone of unusual structure, encompasses mouse amino acids 152–227. The region of sequence dissimilarity between the two primates and mouse begins within this exon at amino acid 187 and extends well beyond it to amino acid 259.

CD44 Allotypic Differences. Allotypic differences have been defined for mouse CD44 (Pgp-1) by monoclonal antibodies (15). Sequences of mouse CD44 were determined for the Pgp1.2 mouse strains, C3H and C57Bl/J, and for the Pgp1.1 strain, BALB/c. Amino acid 198 in the protein sequence is either a serine (C3H) or a glycine (BALB/c). The number of amino acids just 5′ of the putative leader peptide cleavage site (23, 24) also differ in the two strains (and in primates). The presence of these amino acids may alter the site of signal peptide cleavage, making it less clear which alteration accounts for the allotypic difference.

Features of CD44 RNA. Mouse CD44 RNA does not fractionate with other polyadenylylated RNAs. This property accounts for its apparent virtual absence in several different cDNA libraries. Two likely possibilities are either that the CD44 RNA lacks a poly(A) extension or that it has a stable secondary structure that inhibits its binding to an affinity column. Failure to bind to oligo(dT)- or poly(U)-Sepharose is the working definition of a poly(A)-deficient RNA (25–28). RNAs with very short stretches of poly(A) (as few as 10 bases), such as β-actin, have been shown to be retained more efficiently on a poly(U)-Sepharose matrix (26) than on oligo(dT)-cellulose. The RNAs of Fig. 3*A* were prepared by proteinase K digestion and poly(U)-Sepharose chromatography. The poly(A)-containing RNA of Fig. 3*B* was prepared by guanidinium isothiocyanate/CsCl centrifugation methods and was heated prior to selection on oligo(dT)-cellulose. Heating total RNA before application to oligo(dT)-cellulose did not alter its binding behavior, although a stable secondary structure could be reestablished very rapidly. Similar results were obtained with both methods, indicating that this unusual

property of the mouse CD44 RNA is not technique dependent.

With the exception of replication-dependent histone RNAs that are usually not polyadenylylated (29), no other transcript population in adult cells is contained entirely within the poly(A)$^-$ fraction. Two transcripts, β-actin and c-myc, have been identified as being partly contained in the poly(A)$^-$ mRNA fraction. Some RNAs in oocytes lack poly(A), and generally these transcripts correlate with low translational activity (27). The function of the poly(A) tail on RNA is speculative, but correlations to translational activity and RNA stability have been observed (27, 28, 30). Whatever the function of poly(A), the long 3′ untranslated region (at minimum, 2 kb) of CD44 may compensate for a deficiency of poly(A).

The sizes of the CD44 mRNAs in Fig. 3 are approximately 5.1 kb and 1.8 kb. The contribution of these two transcripts to the expression of cell surface CD44 is undetermined, but only the larger transcript hybridizes with a mouse CD44 probe containing 3′ sequences derived from the MM100 cDNA clone (coordinate 758 to approximately 1358). Both transcripts hybridized with the amplified fragment containing the entire coding region (nucleotide coordinates 1–1183). The simplest interpretation is that the smaller transcript contains different 3′ sequences and does not extend into the known transmembrane sequence. Perhaps the protein derived from this sequence is secreted or linked by a phosphotidylinositol lipid in the membrane. CD44 cDNA clones containing truncated forms of the coding region have been reported (13, 31).

In summary, we have isolated a mouse CD44 cDNA sequence by PCR amplification methods. This approach was necessary because CD44 mRNA does not purify with other polyadenylylated RNAs. This unusual feature of the mouse CD44 RNA distinguishes it from the CD44 of other species and mRNAs in general. The overall similarity of the CD44 products between primate and mouse species is very extensive. Conservation of potential N-linked glycosylation sites and cysteine residues suggests that the proteins may assume similar three-dimensional structures. However, there is a region of very low sequence similarity that may alter the protein folding pattern.

We thank Dr. H. K. Gershenfeld for the mouse thymocyte cDNA library and Dr. L. Graf for the embryo cDNA library. We are indebted to Dr. S. Moss for the generous loan of Northern strips, for discussions about polyadenylylation, and critical reading of the manuscript. M. Honda and N. Shough provided excellent technical assistance. We thank T. Loucks for assistance in preparing this manuscript. This work was supported by National Cancer Institute Grants CA42571 to T.S. and CA18221 to C.N.

1. Jalkenen, S., Bargatze, R. F., Herron, L. R. & Butcher, E. C. (1986) *Eur. J. Immunol.* **16,** 1195–1202.
2. Carter, W. G. & Wayner, E. A. (1988) *J. Biol. Chem.* **263,** 4193–4201.
3. Trowbridge, I. S., Lesley, J., Schulte, R., Hyman, R. & Trotter, J. (1982) *Immunogenetics* **15,** 299–312.
4. Isacke, C. M., Sauvage, C. A., Hyman, R., Lesley, J., Schulte, R. & Trowbridge, I. S. (1986) *Immunogenetics* **23,** 326–332.
5. Gallatin, W. M., Wayner, E. A., Hoffman, P. A., St. John, T., Butcher, E. C. & Carter, W. G. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4654–4658.
6. Picker, L. J., de los Toyos, J., Telen, M. J., Haynes, B. F. & Butcher, E. C. (1989) *J. Immunol.* **142,** 2046–2051.
7. Telen, M. J., Eisenbarth, G. S. & Haynes, B. F. (1983) *J. Clin. Invest.* **71,** 1878–1886.
8. Jalkenen, S., Bargatze, R. F., de los Toyos, J. & Butcher, E. C. (1987) *J. Cell Biol.* **105,** 983–990.
9. Jacobson, K., O'Dell, D. & August, J. T. (1984) *J. Cell Biol.* **99,** 1624–1633.
10. Kalomiris, E. L. & Bourguignon, L. Y. W. (1988) *J. Cell Biol.* **106,** 319–327.
11. Idzerda, R. L., Carter, W. G., Nottenburg, C., Wayner, E. A., Gallatin, W. M. & St. John, T. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4659–4663.
12. Stamekovic, I., Amiot, M., Pesando, J. M. & Seed, B. (1989) *Cell* **56,** 1057–1062.
13. Goldstein, L. A., Zhou, D. F. H., Picker, L. J., Minty, C. N., Bargatze, R. F., Ding, J. F. & Butcher, E. C. (1989) *Cell* **56,** 1063–1072.
14. Stoolman, L. M. (1989) *Cell* **56,** 907–910.
15. Lesley, J. & Trowbridge, I. S. (1982) *Immunogenetics* **15,** 313–320.
16. Badley, J. E., Bishop, G. A., St. John, T. & Frelinger, J. A. (1988) *Biotechniques* **6,** 114–116.
17. McMaster, G. K. & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4835–4838.
18. Bellve, A. R., Milletee, C. F., Bhathagar, Y. M. & O'Brian, D. A. (1977) *J. Histochem. Cytochem.* **25,** 480–494.
19. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
20. Saiki, R. K., Scharf, S. J., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. A. (1985) *Science* **230,** 1350–1354.
21. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148.
22. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
23. von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21.
24. von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
25. Milcarek, C., Price, R. & Penman, S. (1974) *Cell* **3,** 1–10.
26. Hunter, T. & Garrels, J. I. (1977) *Cell* **12,** 767–781.
27. Rosenthal, E. T., Tansey, T. R. & Ruderman, J. V. (1983) *J. Mol. Biol.* **166,** 309–327.
28. Swartwout, S. G., Preisler, H., Guan, W. & Kinniburgh, A. J. (1987) *Mol. Cell. Biol.* **7,** 2052–2058.
29. Hentschel, C. C. & Birnstiel, M. L. (1981) *Cell* **25,** 301–313.
30. Wilson, T. & Treisman, R. (1988) *Nature (London)* **326,** 396–399.
31. St. John, T., Meyer, J., Idzerda, R. & Gallatin, W. M. (1989) *Reg. Immunol.*, in press.