CELL STRUCTURE AND FUNCTION 16: 23–30 (1991)

Basigin, a New Member of the Immunoglobulin Superfamily: Genes in Different Mammalian Species, Glycosylation Changes in the Molecule from Adult Organs and Possible Variation in the N-terminal Sequences

Takuro Kanekura[*,**], **Teruo Miyauchi**[**,***], **Masaaki Tashiro**[*], **and Takashi Muramatsu**[**1]

[*]*Department of Dermatology and* [**]*Department of Biochemistry, Faculty of Medicine, Kagoshima University, 1208-1 Usukicho, Kagoshima 890, and* [***]*Japan Immunoresearch Laboratories, 17-5 Sakaecho, Takasaki, Gunma 370, Japan*

Key words: basigin/immunoglobulin superfamily/glycosylation/N-terminal heterogeneity/species specificity

ABSTRACT. **Basigin is a new member of the immunoglobulin superfamily with homology to both the immunoglobulin V domain and major histocompatibility complex class II antigen β-chain. Southern blot analysis indicated that the basigin gene was present as a single copy or as a few copies per mouse genome. Although a homologous gene was detected in the hamster and human, Southern and Northern blotting experiments indicated considerable species specificity in the basigin structure. The molecular weight of N-glycanase-treated basigin from embryonal carcinoma cells was about 32,000 and was close to the value of basigin polypeptide inferred from the cDNA sequence; the result confirmed the open reading frame of basigin. Upon Western blotting, large amounts of basigin were detected in the mouse kidney as a glycoprotein bound to *Ricinus communis* agglutinin (RCA)-I and as a glycoprotein bound to concanavalin A; the molecular weight of the former was 38,000–43,000, and of the latter was 30,000. Basigin of the molecular weight of 48,000 was detected in RCA-I-binding glycoproteins of the liver, small intestine and spleen. Thus, different forms of basigin can be produced by different modes of glycosylation. Another source of heterogeneity of basigin may be differences in N-terminal sequences, since cDNA clones with different 5′ coding sequences were identified.**

Structural studies on cell-surface proteins have revealed that large numbers of them belong to several distinguishable superfamilies or families. Among them, the immunoglobulin (Ig) superfamily comprises the broadest ranges of molecules, such as immunoglobulins (Igs), T cell receptors, major histocompatibility complex (MHC) class I antigens, MHC class II antigens, Thy-1, N-CAM, CD4 and CD8 (6, 24). By screening a cDNA expression library constructed from embryonal carcinoma (EC) cells, we have recently found a new member of the Ig superfamily, basigin, which has strong homology to both the Ig variable (V) domain and MHC class II β-chain (12). Thus, basigin is an interesting molecule from the viewpoint of the molecular evolution of the Ig superfamily.

Basigin mRNA is expressed not only in EC cells but also in various organs of adult mice. This molecule may play important roles in the interaction of various adult and embryonic cells. This paper deals with further studies on basigin. We are particularly interested in resolving the following questions: 1) Are genes homologous to mouse basigin present in other mammalian species? If present, how closely are they related to mouse basigin? 2) Is basigin, which is present in organs of adult mice, identical to basigin in early embryonic cells?

[1] To whom correspondence should be addressed.

Abbreviations: Ig, immunoglobulin; MHC, major histocompatibility complex; N-CAM, neural cell adhesion molecule; EC, embryonal carcinoma; FCS, fetal calf serum; SSC, 0.15 M NaCl, 0.015 M sodium citrate; LTA, *Lotus tetragonolobus* agglutinin; RCA-I, *Ricinus communis* agglutinin-I; Con A, concanavalin A.

MATERIALS AND METHODS

Cells and tumors. Teratocarcinoma OTT6050 (22) and N4-1 EC cells (18) were propagated in 129/Sv mice as embryoid bodies and solid tumors, respectively. Mouse EC cell line F9 (1) was cultured in Dulbecco's modified Eagle's minimal essential medium containing 15% fetal calf serum (FCS) under 12% CO_2 at 37°C. Human normal diploid cell line WI-38 (4) and human gastric cancer cell line KATO-III (20) were cultured in Eagle's minimal essential medium with Earle's salts containing 10% FCS and RPMI-1640 containing 10% FCS, respectively. They were grown under 5% CO_2 at 37°C.

Probes. An EcoRI-AccI fragment (nucleotide number −24~792 in Fig. 9) of pFR27, one of basigin cDNA clones and PstI-AccI fragment (nucleotide number 102~792 in Fig.

9) of pFR1 (12) were labeled by [α-^{32}P]dCTP using a random oligonucleotide priming method (3) and used as probes for Southern blot hybridization and Northern blot hybridization, respectively. The EcoRI-PstI fragment (nucleotide number $-24 \sim 101$ in Fig. 9) and EcoRI-AccI fragment of pFR27 were labeled using the same method and used as probes for the screening of the λgt11 cDNA library of OTT6050.

DNA and RNA analysis. DNA was isolated from the liver of BALB/C mouse and the liver of hamster as described by Blin and Stafford (2). Human DNA was isolated from peripheral leucocytes of a 31-year-old healthy male according to the method of Kunkel *et al.* (8)

Southern blotting was performed as described by Southern (21). Southern blots were prehybridized for 3 h in prehybridization solution [50% deionized formamide, 5 × SSC, 50 mM sodium phosphate, pH 6.5, 200 μg/ml of heat-denatured salmon sperm DNA, 10 × Denhardt's solution and 0.1% SDS] at 42°C and hybridized for 20 h with the probe described above at $0.5\text{–}2.0 \times 10^7$ cpm/ml under conditions identical to those used for prehybridization. The membrane was washed with three changes of 2 × SSC, 0.1% SDS at room temperature followed by three successive washes in 0.1 × SSC, 0.1% SDS at 55°C. RNA was prepared by the guanidium isothiocyanate/cesium chloride method (11). For Northern blotting analysis, RNA was denatured with glyoxal, separated on a 1% agarose gel and transferred to nitrocellulose membrane in 20 × SSC. The membrane was baked for 2 h under vacuum at 80°C and treated in 20 mM Tris-HCl buffer, pH 8.0 at 100°C for 5 min. RNA blots were prehybridized for 3 h at 35°C in the same prehybridization solution used in the Southern blot hybridization and hybridized for 20 h at 35°C in the prehybridization solution containing the probe described above. The membrane was washed with three changes of 2 × SSC, 0.1% SDS at room temperature followed by three successive washes in the same solution at 50°C. The DNA sequence was determined by the dideoxy chain termination method (17) after subcloning into pUC8 or pUC18.

Isolation and Analysis of Glycoproteins. Glycoproteins binding to *Lotus tetragonolobus* agglutinin (LTA) were isolated by affinity chromatography from N4-1 EC cells as described by Kamada *et al.* (7) Glycoproteins binding to *Ricinus communis* agglutinin-I (RCA-I)-agarose and those binding to concanavalin A (Con A)-agarose were isolated from the liver, kidney, small intestine, brain, testis and spleen of 129/Sv mice, as described by Ozawa *et al.* (15) Protein content was determined by the method of Lowry *et al.* (10) using bovine serum albumin as a standard. SDS-polyacrylamide gel electrophoresis was performed according to the method of Laemmli (9) using 9.0% running gel. Proteins in the gel were transferred to nitrocellulose membranes according to Towbin *et al.* (23) The membranes were stained by anti-basigin rabbit antiserum as described by Ozawa *et al.* (16) Anti-basigin antiserum was prepared by immunization with basigin-β-galactosidase fusion protein, and absorbed by the membrane fraction of *Escherichia coli* BNN97 (25) to remove anti-β-galactosidase anti-bodies, as described by Miyauchi *et al.* (12) The specificity of the immunological reaction was confirmed as before (12): the reaction was inhibited by the membrane fraction from *E. coli* BNN103 (λFR1) which has a basigin insert, but not by the membrane fraction from *E. coli* BNN97.

Deglycosylation of glycoproteins. LTA receptors prepared from N4-1 tumors were boiled for 3 min in the presence of 0.5% SDS and 0.1 M β-mercaptoethanol. The denatured glycoproteins were diluted with 0.2 M sodium phosphate buffer, pH 8.6, 10 mM 1, 10-phenanthroline and 1.25% NP-40. N-Glycanase (Genzyme Co.) (5) and several drops of toluene were added to the reaction mixture and incubated at 37°C for 24 h. The reaction mixture was subjected to trichloroacetic acid precipitation followed by ethanol precipitation. The deglycosylated glycoproteins were analyzed electrophoretically, as described above, with the exception of using a 12% running gel.

Screening of λgt11 cDNA library. The λgt11 library constructed from mRNA of teratocarcinoma OTT6050 (16) was screened by the plaque hybridization method (11) using the probes described above.

Primer extension. A 30-mer primer (2.5 ng), which was complementary to nucleotides 106–135 of the common sequence of basigin cDNA (Fig. 9) and was synthesized using a gene assembler (Pharmacia), was hybridized with 5 μg of poly(A)$^+$RNA from teratocarcinoma OTT6050 and extended with 100 units of reverse transcriptase, and the second strand was synthesized using ribonuclease H and DNA polymerase I (14). Phosphorylated EcoRI linkers were ligated to the blunt-ended and methylated cDNAs, which were then digested with EcoRI and ligated to λgt10 arms. The phage DNA was packaged *in vitro*. The EcoRI-AccI fragment of pFR27 cDNA was used as a probe for screening of the λgt10 library by plaque hybridization.

RESULTS

Southern blot analysis of basigin gene. DNA from BALB/C mice was digested with restriction enzymes and the fragments were analyzed by Southern blotting (Fig. 1). Three bands (30 kb, 13 kb and 3.0 kb) reacting with the basigin probe were detected in EcoRI digest. A broad band (5.2 kb) and a faint band (8.3 kb) were revealed in SacI digest; a broad band (12 kb) was present in AccI digest. These relatively simple profiles of restriction fragments of the basigin gene imply that the basigin gene is present as a single copy or a few copies per genome.

By Southern blot analysis, we examined whether a gene homologous to the basigin gene is present in other mammalian species. EcoRI digest of hamster DNA yielded a band of 9.5 kb reacting with the basigin probe. Two bands (14 kb, 4.8 kb) were detected in SacI digest. Since the single band detected in EcoRI digest was a sharp one, it is likely that a basiginlike gene in

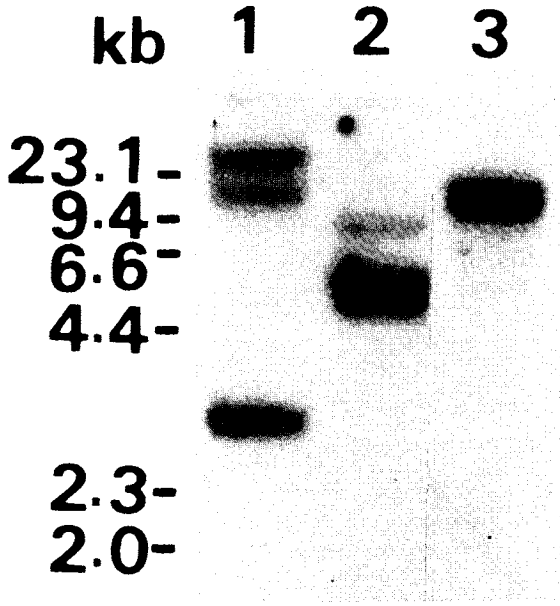

Fig. 1. Southern blot analysis of mouse DNA using basigin cDNA probe. Each lane contained the digest from 20 μg of DNA. Restriction endonucleases used were as follows: lane 1, EcoRI; lane 2, SacI; lane 3, AccI.

hamster is present as a single copy (Fig. 2).

The basigin probe also reacted with DNA fragments released from human DNA, although the bands were broad and smears were detected (Fig. 3). Thus, we have concluded that genes homologous to the basigin gene are present both in the hamster and the human. The unrelatedness of the Southern blot pattern and faint bands in the human DNA fragments indicate considerable species specificity in the basigin structure.

Detection of basigin mRNA in human cells. By Northern blot analysis, we examined whether mRNA homologous to basigin cDNA is really expressed in human cells. The basigin cDNA probe detected a single band of 1.5 kb in F9 cells, as described previously (12), but a single band of 5.6 kb in human normal diploid cell line, WI-38 and human gastric cancer cell line, KATO-III under relaxed condition of hybridization (Fig. 4). In human cells, only a faint band with the same size of 5.6 kb was detected when hybridization was performed at 42°C in the presence of 50% formamide and washing was carried out at 55°C in 0.1 × SSC, 0.1% SDS (data not shown). These results show that mRNA homologous to basigin cDNA is really expressed in the

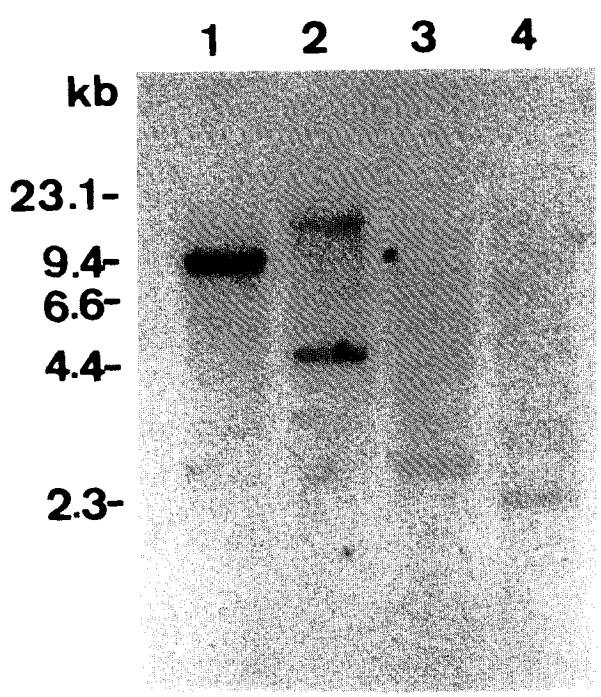

Fig. 2. Southern blot analysis of hamster DNA using basigin cDNA probe. Each lane contained the digest from 20 μg of DNA. Restriction endonucleases used were as follows: lane 1, EcoRI; lane 2, SacI; lane 3, PstI; lane 4, AccI.

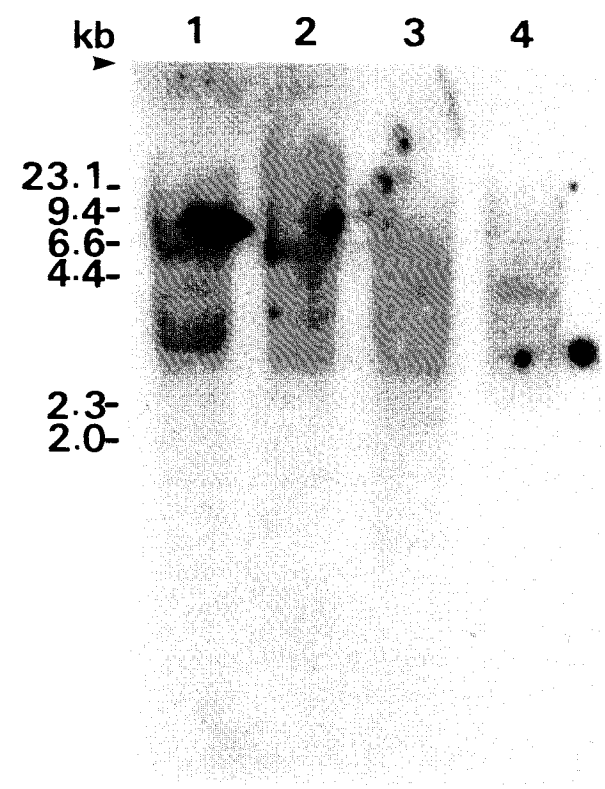

Fig. 3. Southern blot analysis of human DNA using basigin cDNA probe. Each lane contained the digest from 20 μg of DNA. Restriction endonucleases used were as follows: lane 1, EcoRI; lane 2, SacI; lane 3, PstI; lane 4, AccI.

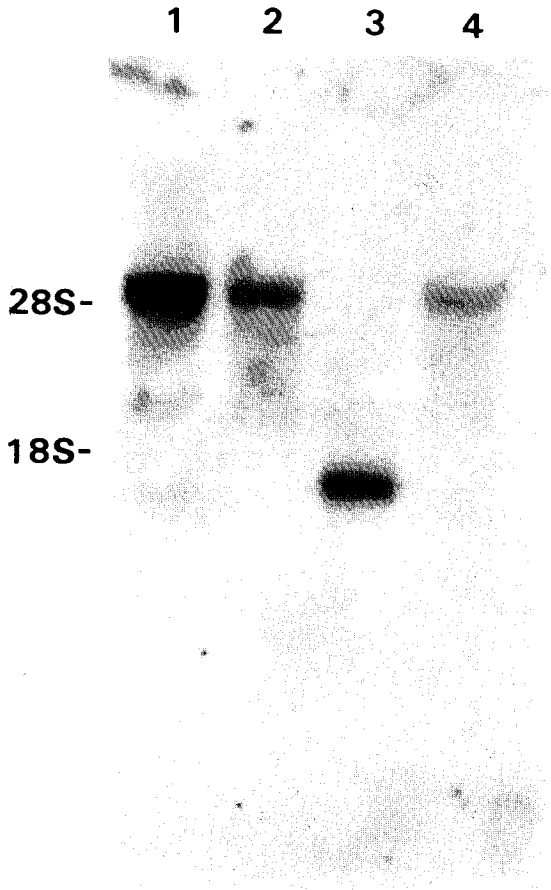

Fig. 4. Northern blot analysis of human mRNA homologous to basigin mRNA. Total RNAs (20 μg) or poly(A)$^+$RNAs were denatured, separated by 1.0% agarose gel electrophoresis, transferred to a nitrocellulose membrane and hybridized with an [α-^{32}P]dCTP-labeled probe (PstI-AccI fragment of basigin cDNA) as described in "MATERIALS AND METHODS". The sources of total RNAs or poly(A)$^+$RNAs were as follows: total RNAs from 1) human normal diploid cells WI-38, 2) human gastric cancer cells KATO-III; poly(A)$^+$RNAs from 3) F9 cells, 4) KATO-III cells.

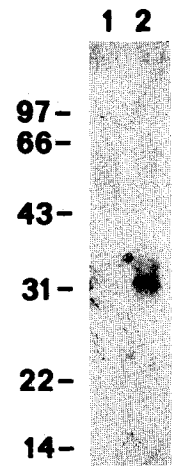

Fig. 5. Effects of N-glycanase digestion on molecular weight of basigin. LTA receptors (100 μg as proteins) isolated from N4-1 tumors were denatured by boiling in the presence of 0.5% SDS and 0.1 M β-mercaptoethanol followed by incubation in the reaction mixture with or without 40 units/ml of N-glycanase at 37°C for 24 h. The reaction mixture was subjected to trichloroacetic acid precipitation followed by ethanol precipitation, then separated by SDS-PAGE on a 12% gel under reducing conditions. Separated proteins were transferred to a nitrocellulose membrane and reacted with antiserum against the fusion protein produced by λFR1. The antiserum was preabsorbed with the membrane fraction of BNN97, as described by Miyauchi *et al.* (12) Lane 1: reaction without N-glycanase; lane 2: reaction with N-glycanase. Molecular weight markers used were rabbit muscle phosphorylase b (97,400), bovine serum albumin (66,200), hen egg white ovalbumin (42,699), bovine carbonic anhydrase (31,000), soybean trypsin inhibitor (21,500) and hen egg white lysozyme (14,400).

human cell, but its molecular size and structure are different from those of mouse teratocarcinoma cells or adult organs of mouse, and that the size of basigin mRNA from teratocarcinoma cells is identical to that from adult organs (the liver, kidney, brain, spleen and testis), as shown previously (12). Thus, the data of Northern blot analysis also indicate considerable species specificity in the basigin structure.

The size of deglycosylated basigin. The molecular weight of basigin predicted from the cDNA was 30,000 including the size of the putative signal sequence. However, basigin in N4-1 EC cells largely migrated as a broad band of apparent molecular weight 43,000–66,000 (Fig. 5, lane 1). This difference in molecular weight was previously considered to be glycosylation in the basigin molecule. We tested the possibility directly by deglycosylating basigin. After treatment with N-glycanase, which removes asparagine-linked carbohydrates, the molecular weight of basigin was indeed reduced to 32,000 (Fig. 5, lane 2). There are still differences of molecular weight of about 4,000 between the observed molecular weight of deglycosylated basigin and the predicted molecular weight of basigin polypeptide devoid of the signal sequence. However, this small difference can be easily explained by O-glycosylation or phosphorylation. Thus, the present result confirmed the open reading frame of basigin. Furthermore, the enormous difference between the size of intact basigin and the size of the deglycosylated one indicates that basigin is a heavily glycosylated molecule. Although the predicted N-glycosylation site in the extracellular domain of basigin is 3, the glycan portion with developmentally regulated markers is large with a poly-N-acetyllactosamine backbone.

Basigin molecules from organs of adult mice. Basigin mRNA is detected in various organs of adult mice. However, the expression of mRNA does not always lead to intense expression of the polypeptide. In

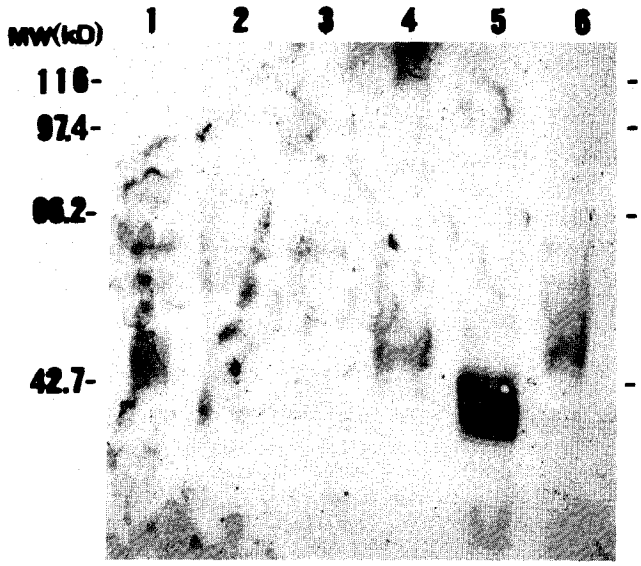

Fig. 6. Detection of basigin in RCA-I-binding glycoproteins isolated from mouse organs. The glycoproteins (400 μg) were separated by SDS-polyacrylamide gel electrophoresis, transferred to nitrocellulose membrane and reacted with anti-basigin antibody followed by HRP-anti-rabbit IgG. Lane 1, spleen; lane 2, testis; lane 3, brain; lane 4, small intestine; lane 5, kidney; lane 6, liver. Molecular weight markers used were the same ones as described in Fig. 5.

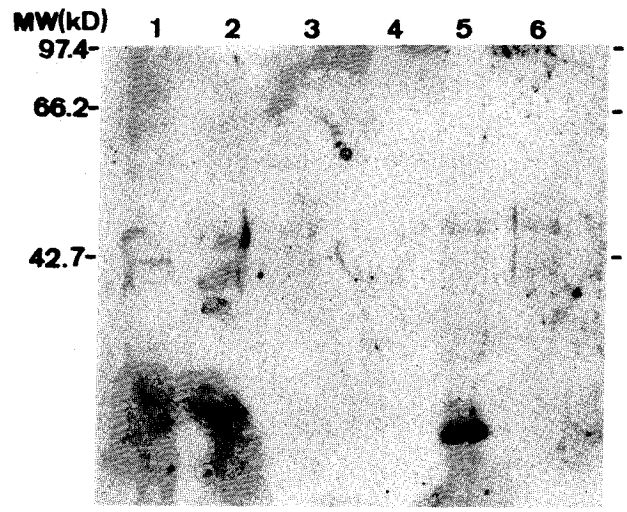

Fig. 7. Detection of basigin in Con A-binding glycoproteins isolated from mouse organs. The glycoproteins (400 μg) were separated by SDS-polyacrylamide gel electrophoresis, transferred to nitrocellulose membrane and reacted with anti-basigin antibody followed by HRP-anti-rabbit IgG. Lane 1, spleen; lane 2, testis; lane 3, brain; lane 4, small intestine; lane 5, kidney; lane 6, liver. Molecular weight markers used were the same ones as described in Fig. 5.

the case of carcinoembryonic antigen, which is also a member of the immunoglobulin superfamily, its mRNA is expressed in various normal tissues. The stability of the gene product determines its preferential expression in cancer tissues. Thus, we examined whether high levels of the basigin molecule are expressed in organs of adult mice. For that purpose, glycoproteins, which were enriched by affinity chromatography on RCA-I agarose or on Con A agarose, were analyzed by SDS-polyacrylamide gel electrophoresis and stained by anti-basigin antibodies. RCA-I-binding glycoproteins from the kidney were found to contain large amounts of basigin of molecular weight from 38,000–43,000 (Fig. 6, lane 5). RCA-I-binding glycoproteins from the spleen, small intestine and liver (Fig. 6, lane 1, 4, 6) contained considerable amounts of basigin of molecular weight around 48,000. The glycoproteins from the testis and brain (Fig. 6, lane 2, 3) contained trace amounts of the bands of molecular weight around 59,000. When we analyzed Con A-binding glycoproteins, large amounts of the basigin band was detected again in the kidney (Fig. 7, lane 5). In this case, the molecular weight of basigin

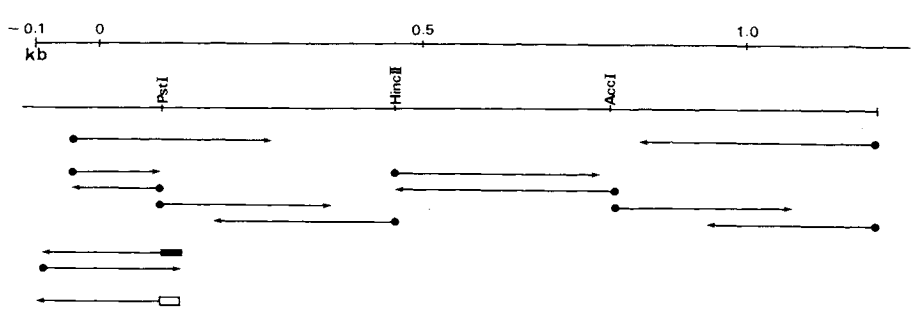

Fig. 8. Sequence strategy. The scale on the top indicates nucleotide position (in kilobases) beginning from the first base of the common sequence (cf. Fig. 9). Beneath the scale, a restriction map of the clone is represented. Restriction enzymes indicated on the map were used to prepare the fragments for sequencing. The direction and extent of sequence determination are shown by horizontal arrows beneath each fragment. Black box and open box indicate synthetic primer complementary to nucleotides 106–135. Each primer was used to construct a λgt10 library whose template was poly(A)$^+$RNA of F9 and that of OTT6050, respectively. Horizontal arrows which follow the black box and white box represent basigin type I and type III, respectively. Basigin type II sequence from OTT6050 is indicated by bidirectional arrows upstream of PstI site.

was around 30,000. As above the basigin polypeptide is certainly present in the kidney, testis, small intestine and liver; other organs are also likely to express small amounts of the basigin molecule. A Con A-binding glycoprotein from the testis (42 K, Fig. 7, lane 2) was also reactive with the anti-basigin. Faint bands were also detected in the glycoproteins from other organs. Thus, the kidney has two forms of basigin with different glycosylation profiles. When glycoproteins from F9 embryonal carcinoma cells were analyzed by the same approach, basigin was detected in RCA-I-binding glycoproteins, but not in Con A-binding glycoproteins (data not shown). These data also clearly indicate glycosylation differences in basigin molecules from different sources.

Differences in 5′-region sequences of basigin. In addition to the glycosylation difference, the structure of N-terminal sequences may be another source of heterogeneity in the basigin molecule. During screening of basigin cDNA, we noted two types of cDNA clones whose 5′ sequences were different from the published sequence of basigin cDNA (12).

```
sequence of basigin
 -97 **********************************ACGAGGCGAC ATG GCG GCG GCG CTG CTG CTG GCG CTG  -61
                                                 Met Ala Ala Ala Leu Leu Leu Ala Leu
sequence of basigin typeII
     ********************************************************************************

sequence of basigin typeIII
 -93 **********************************TTC CAA CAG CAG ACC TCA CAG AGT GGG GCT GAC  -61
                                       Phe Gln Gln Gln Thr Ser Gln Ser Gly Ala Asp

(basigin)
 -60 GCC TTC ACG CTC TTG AGC GGC CAA GGC GCC TGC GCG GCG GCG GGC ACC ATC CAA ACC TCT   -1
     Ala Phe Thr Leu Leu Ser Gly Gln Gly Ala Cys Ala Ala Ala Gly Thr Ile Gln Thr Ser
(typeII)
 -39 ****************************AGG CCA GCG CCA GCA GCA GCG CCG CCG CCA TGT CGC CTC   -1
                                   Arg Pro Ala Pro Ala Ala Ala Pro Pro Pro Cys Arg Leu
(typeIII)
 -60 TAC TAT GCT GAG CTT ATA CTC TTT CTC CCT GTG GCT TCA GCG GGC ACC ATC CAA ACC TCT   -1
     Tyr Tyr Ala Glu Leu Ile Leu Phe Leu Pro Val Ala Ser Ala Gly Thr Ile Gln Thr Ser

--(common sequence)---------------------------------------------------------------
   1 GTC CAG GAA GTC AAC TCC AAA ACA CAG CTT ACC TGC TCT TTG AAC AGC AGT GGC GTT GAC   60
     Val Gln Glu Val Asn Ser Lys Thr Gln Leu Thr Cys Ser Leu Asn Ser Ser Gly Val Asp

  61 ATC GTT GGC CAC CGC TGG ATG AGA GGT GGC AAG GTA CTG CAG GAG GAC ACT CTG CCC GAC  120
     Ile Val Gly His Arg Trp Met Arg Gly Gly Lys Val Leu Gln Glu Asp Thr Leu Pro Asp

 121 CTG CAT ACG AAG TAC ATA GTG GAC GCA GAT GAC CGC TCT GGG GAA TAT TCC TGC ATC TTC  180
     Leu His Thr Lys Tyr Ile Val Asp Ala Asp Asp Arg Ser Gly Glu Tyr Ser Cys Ile Phe

 181 CTT CCT GAG CCT GTG GGC AGA AGC GAG ATC AAT GTG GAA GGG CCA CCC AGG ATC AAG GTC  240
     Leu Pro Glu Pro Val Gly Arg Ser Glu Ile Asn Val Glu Gly Pro Pro Arg Ile Lys Val

 241 GGA AAG AAA TCA GAG CAT TCC AGT GAG GGA GAG CTT GCG AAA CTG GTC TGC AAG TCC GAT  300
     Gly Lys Lys Ser Glu His Ser Ser Glu Gly Glu Leu Ala Lys Leu Val Cys Lys Ser Asp

 301 GCA TCC TAC CCT CCT ATT ACA GAT TGG TTC TGG TTT AAG ACC TCT GAC ACT GGG GAA GAA  360
     Ala Ser Tyr Pro Pro Ile Thr Asp Trp Phe Trp Phe Lys Thr Ser Asp Thr Gly Glu Glu

 361 GAG GCA ATC ACC AAT AGC ACT GAA GCC AAT GGC AAG TAT GTG GTG GTA TCC ACG CCT GAG  420
     Glu Ala Ile Thr Asn Ser Thr Glu Ala Asn Gly Lys Tyr Val Val Val Ser Thr Pro Glu

 421 AAG TCA CAG CTG ACC ATC AGC AAC CTT GAC GTA AAT GTT GAC CCT GGC ACC TAC GTG TGT  480
     Lys Ser Gln Leu Thr Ile Ser Asn Leu Asp Val Asn Val Asp Pro Gly Thr Tyr Val Cys

 481 AAT GCC ACC AAC GCC CAG GGC ACT ACT CGG GAA ACC ATC TCA CTG CGT GTG CGG AGC CGC  540
     Asn Ala Thr Asn Ala Gln Gly Thr Thr Arg Glu Thr Ile Ser Leu Arg Val Arg Ser Arg

 541 ATG GCA GCC CTC TGG CCC TTC CTA GGC ATC GTG GCT GAG GTC CTG GTG TTG GTT ACC ATC  600
     Met Ala Ala Leu Trp Pro Phe Leu Gly Ile Val Ala Glu Val Leu Val Leu Val Thr Ile

 601 ATC TTT ATC TAT GAG AAG AGG CGG AAG CCA GAC CAG ACC CTG GAC GAG GAT GAC CCT GGC  660
     Ile Phe Ile Tyr Glu Lys Arg Arg Lys Pro Asp Gln Thr Leu Asp Glu Asp Asp Pro Gly

 661 GCC GCC CCA CTG AAG GGC AGT GGA ACT CAC ATG AAT GAC AAG GAC AAG AAT GTA CGC CAG  720
     Ala Ala Pro Leu Lys Gly Ser Gly Thr His Met Asn Asp Lys Asp Lys Asn Val Arg Gln

 721 AGG AAC GCC ACC TGA GTGGTGGGGCAGGCGGGGGAGGGGAGGTGCCCAGGGTGCCTGACCCCAGGCCAGCGTCT  794
     Arg Asn Ala Thr ***

 795 ACCTCCACTCCAGTATCCCATCCTGTCCCGATTTGAACCTACCCAACCCAACCTATCCCAACCCAAGTGAAGACAGAGC  873

 874 CTTACCTTACAGAAAACCCACCTGGAAGAAGCAAGCCACTTGCAGCCCCTGTTTCTAATTTAAACTAAATGAGGTTTCT  952

 953 ATGCAGACAATCCATTCCTTAGGGGTTTATGTTTTTATTTTTCCTTCCCTTCTGAAGTGTTGTCACTACAGCCCTGTGG 1031

1032 AGTGGGGGAATGGGGCCTTGTCCTTGGTCAGGAGGGAAGGCCAGTGCATGCTCTGACTTACTGTTGGAGGGGGCTGGGC 1110

1111 CTGCTGGAAACCCCCCAAATAAAACCTAACCCACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1189

1190 AAAAAAAAAAAAAAA                                                               1205
```

Fig. 9. Nucleotide sequence and the deduced amino acid sequence of the three types of basigin cDNA from murine teratocarcinoma cells. The nucleotide position numbers in base pairs begin from the first base of the common sequence. Different sequences on 5′-region are shown in 3 lines, and their nucleotide numbers are indicated in negative numbers. Underlined sequence is homologous with basigin. **; sequence was not determined.

When the λgt11 cDNA library constructed from teratocarcinoma OTT6050 was screened by the basigin probe, we obtained around 100 basigin cDNA clones from 1.2×10^6 plaque-forming units; among them, 7 clones were studied in detail. About 280 bases of the 5′-end of the cDNA inserts were sequenced, and among the 7 clones, 3 clones were found to have 5′ sequences different from that of the published basigin sequence. The remaining 4 clones possessed the sequence identical to the published sequence. We call the new-sequence basigin, type II, and the previous-sequence basigin type I. All of the type II sequence started from the same base suggesting that the type II sequence may be derived from a single clone. We determined the complete sequence of a representative type II clone (Fig. 8). The type II clone differed from the type I clone only in 39 base pairs in the 5′ region (Fig. 9). This nucleotide difference results in differences of the N-terminal amino acid sequence of the basigin molecule. Among 13 amino acids unique for the type II molecule, 10 were hydrophobic ones.

The cDNA sequence of basigin type II was incomplete and lacked the initiator ATG. Thus, we constructed a λgt10 library from teratocarcinoma OTT6050 using synthetic primer corresponding to nucleotides 106–135 of basigin cDNA (Fig. 8, Fig. 9). By screening 5×10^4 plaque-forming units, 12 cDNA clones of basigin were obtained. Among them, only one clone had an insert large enough to cover the additional 5′ sequence (250 base insert). The remainder had medium-sized inserts (220–230 base insert) or short inserts (140–150 base insert). Thus, we determined the sequences of the clone with the long insert, 2 clones with medium-size inserts and 2 clones with short inserts. Unexpectedly, the sequence of the long insert was different from both type I and type II, and was named type III (Fig. 9). Twenty-one nucleotides in the 3′-terminal side of the insert (nucleotide number -21 to -1) were common to those in type I, and the rest were different. The type III insert coded 31 amino acids, of which 15 were hydrophobic. The medium and small inserts had type I structure. As above, among the 12 cDNA clones sequenced, 4 clones had sequences different from the published basigin sequence (type I). Although there is little doubt that type I corresponds to the predominant basigin molecule, the occurrence of cDNAs with different 5′ sequences in significant frequency suggests that some basigin molecules have different N-terminal structures.

DISCUSSION

Basigin mRNA is detected in a variety of cells and organs (12). The result of the present investigation established the broad distribution of basigin at the protein level. Even though basigin is broadly distributed, the basigin structure is diverse. N-glycanase digestion has verified that basigin is a heavily glycosylated moleucle. Basigin from EC cells is polydisperse, and the major component has a molecular weight of 43,000–66,000 (12), while the deglycosylated molecule has a molecular weight of around 32,000. In the kidney, two types of basigin were detected: basigin reacting with RCA-I had a molecular weight of 38,000–43,000, while basigin reacting with Con A had a molecular weight of around 30,000. Basigin from the liver, small intestine and spleen had a molecular weight of around 48,000. Thus, glycosylation patterns are different in basigin from different sources. In other words, the glycosylation difference is a factor which can endow tissue specificity to basigin molecules. Further support for the above statement is that, except for renal tubules, the adult organs mentioned above do not express LTA-binding sites (19), while basigin was found as a carrier of LTA-binding sites in EC cells. Histochemical studies using lectins and monoclonal antibodies have shown that various tissues are differently glycosylated (13). It is quite possible that basigin carries some of tissue-specified carbohydrate markers in adult tissues, just as it carries a developmentally regulated carbohydrate marker in EC cells (12).

Another source of possible heterogeneity of the basigin molecule is N-terminal sequences. Among the 12 basigin cDNAs isolated from a library of teratocarcinoma OTT6050, 4 clones had 5′ structures different from the published basigin sequence. Including the previously reported one, there are 3 types of basigin cDNA: they are different only in the 5′ sequences. The frequency of the clones with the different structures was too high to regard the heterogeneity as a cloning artifact. The heterogeneity is probably caused by differential splicing of basigin mRNA at the 5′ region. However, it is appropriate to mention that at the present time, we do not know whether or not basigin protein with a different N-terminal sequence is actually present. The possibility that a part of the heterogeneity is due to unspliced mRNA is not completely eliminated, but is unlikely considering the frequency of cDNA clones with different structures.

Southern blotting and Northern blotting experiments suggested the presence of basiginlike molecules in the human and the hamster. Since considerable species specificity was implicated from the profile of the banding patterns, it becomes necessary to examine how closely basiginlike molecules from different sources are related to mouse basigin. For example, analysis of the human cDNA clone of basigin is required to determine how the human and the mouse proteins are related.

The present report complements the previous one, and considerable information has accumulated on the structure of the newly found cell-surface molecule. The

future efforts should be directed to clarifying the function of this molecule. Preparation of large amounts of recombinant basigin is an important step toward this end, and work along this line is now in progress.

Acknowledgments. This work was supported in part by grants from the Ministry of Education, Science and Culture, Japan. We thank Miss Kumiko Sato for her expert secretarial assistance.

REFERENCES

1. Bernstine, E. G., Hooper, M. L., Grandchamp, S., and Ephrussi, B. (1973). Alkaline phosphatase activity in mouse teratoma. *Proc. Natl. Acad. Sci. USA*, **70**: 3899–3903.
2. Blin, N. and Stafford, D. W. (1976). A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res.*, **3**: 2303–2308.
3. Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**: 6–13
4. Hayflick, L. (1965). The limited *in vitro* lifetime of human diploid cell strains. *Exp. Cell Res.*, **37**: 614–636.
5. Hirani, S., Bernasconi, R. J., and Rasmussen, J. R. (1987). Use of N-glycanase to release asparagine-linked oligosaccharides for structural analysis. *Anal. Biochem.*, **162**: 485–492.
6. Hood, L., Kronenberg, M., and Hunkapiller, T. (1985). T cell antigen receptors and the immunoglobulin supergene family. *Cell*, **40**: 225–229.
7. Kamada, Y., Arita, Y., Ogata, S., Muramatsu, H., and Muramatsu, T. (1987). Receptors for fucose-binding proteins of *Lotus tetragonolobus* isolated from mouse embryonal carcinoma cells: Structural characteristics of the poly (N-acetyllactosamine)-type glycan. *Eur. J. Biochem.*, **163**: 497–502.
8. Kunkel, L. M., Smith, K. D., Boyer, S. H., Borgaonkar, D. S., Wachtel, S. S., Miller, O. J., Breg, W. R., Jones, H. W., and Rary, J. M. (1977). Analysis of human Y-chromosome-specific reiterated DNA in chromosome variants. *Proc. Natl. Acad. Sci. USA*, **74**: 1245–1249.
9. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**: 680–685.
10. Lowry, O. H., Rosebrough, N. J., Farr, A. L., and Randall, R. J. (1951). Protein measurement with the Folin phenol reagent. *J. Biol. Chem.*, **193**: 265–275.
11. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. Miyauchi, T., Kanekura, T., Yamaoka, A., Ozawa, M., Miyazawa, S., and Muramatsu, T. (1990). Basigin, a new broadly distributed member of the immunoglobulin super family, has strong homology with both the immunoglobulin V domain and the β-chain of major histocompatibility complex class II antigen. *J. Biochem.*, **107**: 316–323.
13. Muramatsu, T. (1988). Alterations of cell-surface carbohydrates during differentiation and development. *Biochemie*, **70**: 1587–1596.
14. Okayama, H. and Berg, P. (1982). High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.*, **2**: 161–170.
15. Ozawa, M., Fujiyama, J., Furukawa, T., Sato, K., Noguchi, T., and Muramatsu, T. (1981). Glycoproteins from teratocarcinomas and organs of adult mice: Analysis by affinity chromatography on lectin-conjugated agarose. *Dev. Growth Differ.*, **23**: 515–522.
16. Ozawa, M., Huang, R. -P., Furukawa, T., and Muramatsu, T. (1988). A teratocarcinoma glycoprotein carrying a developmentally regulated carbohydrate marker is a member of the immunoglobulin gene superfamily. *J. Biol. Chem.*, **263**: 3059–3062.
17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**: 5463–5467.
18. Sato, M., Ogata, S., Ueda, R., Namikawa, R., Takahashi, T., Nakamura, T., Sato, E., and Muramatsu, T. (1983). Reactivity of a monoclonal antibody raised against human leukemic cells to embryonic and adult tissues of the mouse and teratocarcinomas. *Dev. Growth Differ.*, **25**: 333–344.
19. Sato, M., Yonezawa, S., Uehara, H., Arita, Y., Sato, E., and Muramatsu, T. (1986). Differential distribution of receptors for two fucose-binding lectins in embryos and adult tissues of the mouse. *Differentiation*, **30**: 211–219.
20. Sekiguchi, M., Sakakibara, K., and Fujii, G. (1978). Establishment of cultured cell lines derived from a human gastric carcinoma. *Japan. J. Exp. Med.*, **48**: 61–68.
21. Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**: 503–517.
22. Stevens, L. C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev. Biol.*, **21**: 364–382.
23. Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA*, **76**: 4350–4354.
24. Williams, A. F. and Barclay, A. N. (1988). The immunoglobulin superfamily—Domains for cell surface recognition. *Ann. Rev. Immunol.*, **6**: 381–405.
25. Young, R. A. and Davis, R. W. (1983). Yeast RNA polymerase II genes: Isolation with antibody probes. *Science*, **222**: 778–782.

(Received for publication, September 20, 1990
and in revised form, December 28, 1990)