J. Biochem. 107, 316-323 (1990)

Basigin, a New, Broadly Distributed Member of the Immunoglobulin Superfamily, Has Strong Homology with Both the Immunoglobulin V Domain and the β-Chain of Major Histocompatibility Complex Class II Antigen

Teruo Miyauchi,*,** Takuro Kanekura,**,*** Akihiro Yamaoka,** Masayuki Ozawa,** Sanzo Miyazawa,**** and Takashi Muramatsu**

*Japan Immunoresearch Laboratories, Takasaki, Gumma 370; **Department of Biochemistry and ***Department of Dermatology, Faculty of Medicine, Kagoshima University, Kagoshima, Kagoshima 890; and ****Center for Genetic Information Research, National Institute of Genetics, Mishima, Shizuoka 411

Received for publication, September 11, 1989

***Lotus tetragonolobus* agglutinin (LTA) binds preferentially to early embryonic cells in the mouse. The affinity-purified antibody raised against LTA receptors from embryonal carcinoma cells were used to screen a λgt11 expression library of F9 embryonal carcinoma cells, resulting in detection of a cDNA clone specifying a new glycoprotein termed "basigin." The glycoprotein has been suggested to be a transmembrane one, and was found to be a new member of the immunoglobulin (Ig) superfamily. The molecular weight of basigin was largely in the range between 43,000 and 66,000, while that of the peptide portion with a putative signal sequence was inferred to be about 30,000. Significant levels of basigin mRNA were detected not only in embryonal carcinoma cells, but also in mouse embryos at 9-15 days of gestation and in various organs of the adult mouse. The Ig-like domain of basigin is unique, since it has strong homology to both the β-chain of major histocompatibility class II antigen and the Ig V domain. The number of amino acids between the two conserved cysteine residues is intermediate between those of the Ig V and C domains. Therefore, basigin is an interesting protein in connection with the molecular evolution of the superfamily.**

A large number of cell-surface molecules involved in intercellular recognition belong to a group of proteins called the immunoglobulin (Ig) superfamily (*1, 2*), which includes Igs, T cell receptors, CD3, CD4, CD8, Thy-1, N-CAM, and major histocompatibility complex (MHC) antigens (class I and II). Functions of the members of the superfamily have been best clarified in lymphoid cells and nerve cells. For example, in T cell recognition, T cell receptors are associated with CD3, which is probably involved in signal transduction (*3*). Furthermore, MHC antigen class I molecules on target cells are recognized by CD8 molecules on cytotoxic/suppressor T cells (*4*). Likewise, MHC class II antigen on macrophages interacts with CD4 on helper T cells (*5*). N-CAM (*6*) and L1 (*7*) are involved in nerve cell adhesion by homophilic association to the same molecule in neighboring cells.

Generally speaking, most of the members of this superfamily are not expressed in early embryonic cells. However, we have recently identified Gp70 (*8*), a new member of the Ig superfamily expressed in embryonal carcinoma (EC) cells, which are teratocarcinoma stem cells and resemble multipotential cells of early embryos (*9*). In this communication, we describe another new member of the Ig superfamily expressed in EC cells.

This glycoprotein was found during the study of a developmentally regulated cell-surface marker, namely receptors for *Lotus tetragonolobus* agglutinin (LTA) (*10*). In the mouse, this marker is expressed in EC cells, but disappears from most of the cells differentiated from them. The receptor found in EC cells was determined to have the structure Galβ1→4(Fucα1→3)GlcNAc (*10*), which is identical to that of a developmentally regulated antigen, SSEA-1 (*11*). Apparently, glycoproteins carrying LTA binding sites or SSEA-1 consist of many molecules including the one reported herein. In contrast to Gp 70 which is preferentially expressed in early embryonic cells (*8*), the present molecule was expressed equally well in adult tissues. Because of its broad distribution, we have termed this molecule "basigin" (basic immunoglobulin superfamily).

MATERIALS AND METHODS

Cells and Tumors—EC cell lines, F9 (*12*), N4-1 (*13*), and PCC4 (*14*), parietal endoderm cell line PYS-2 (*15*) and fibroblast cell line L (*16*) were cultured in Dulbecco's modified Eagle's minimal essential medium containing an appropriate concentration of serum (15% fetal calf serum for F9, N4-1, and PCC4 cells, 10% fetal calf serum for PYS-2 cells and L cells) under 12% CO_2 at 37°C. For *in vitro*

Abbreviations: LTA, *Lotus tetragonolobus* agglutinin; Ig, immunoglobulin; MHC, major histocompatibility complex; N-CAM, neural cell adhesion molecule; EC, embryonal carcinoma; PBS(+) or PBS(−), Dulbecco's phosphate-buffered saline with or without Ca^{2+} and Mg^{2+}; SSC, 0.15 M NaCl, 0.015 M sodium citrate; SDS, sodium dodecyl sulfate.

differentiation of F9 cells, cells were plated at a concentration of 5×10^5/ml and treated with 10^{-7} M retinoic acid and 10^{-3} M dibutyryl cyclic AMP (*17*). Teratocarcinoma OTT6050 (*18*) was grown in the ascitic form as embryoid bodies in 129/Sv mice by intraperitoneal passage. N4-1 cells were also grown as solid tumors by subcutaneous inoculation into 129/Sv mice.

Embryos and Organs—Embryos and organs were obtained from ICR mice. The day on which a vaginal plug was observed was determined as day 0 of pregnancy.

Preparation of Antibodies against LTA Receptors—LTA receptors were prepared from the particulate fraction of N4-1 tumors using an agarose column conjugated with LTA as described by Kamada *et al.* (*10*). A New Zealand White rabbit was immunized subcutaneously with the lectin receptors (400 μg of protein) in Freund's complete adjuvant by injection into the foot pads and back skin. The animal received three booster injections at 2-week intervals using the same amount of the antigen. Antiserum was collected 10 days after the last immunization and stored at −70°C. Four milliliters of the antiserum was applied to a column of LTA receptors coupled to Affi-Gel 10 (Bio-Rad, 1.3×1.5 cm; 4 mg protein of LTA receptors/ml) equilibrated with Dulbecco's phosphate-buffered saline containing Ca^{2+} and Mg^{2+} [PBS(+)] (*19*). After washing of the column with 30 ml of the same buffer, antibodies were eluted stepwise with 2 ml each of 0.1 M glycine-HCl buffer (pH 2.5) containing 0.15 M NaCl and 0.1% sodium azide, and immediately neutralized with 1 ml of 1 M Tris-HCl buffer (pH 8.2) containing 0.1% sodium azide. The peak fractions of eluates were pooled and dialyzed against three changes of 1 liter of PBS(+). The purified antibodies were mixed with bovine serum albumin to a final concentration of 1 mg/ml and stored at −70°C.

Screening of λgt11 cDNA Library—The affinity-purified antibodies against LTA receptors were used to screen a λgt11 cDNA library according to the procedure of Young and Davis (*20*) as modified by Schuh *et al.* (*21*). The library constructed from mRNA of F9 embryonal carcinoma cells (*21*) was kindly provided by Dr. R. Kemler.

Subcloning and DNA Sequencing—Phage DNA from positive clones was digested with *Eco*RI/*Sac*I, and the cDNA inserts were subcloned into the *Eco*RI/*Sac*I site of plasmid vector pUC18 since the 5′ end of the *Eco*RI linker sites was converted to a form which could not be digested with *Eco*RI. The inserts were further digested with various restriction enzymes, and the resulting fragments were subcloned into the same vector. DNA nucleotide sequences of the fragments were determined by the dideoxy chain-termination method (*22*).

Primer Extension—A 30-mer primer (2.5 ng) which was complementary to nucleotides 203-232 of basigin cDNA and was synthesized using a Gene Assembler (Pharmacia), was hybridized with 5 μg of poly(A)$^+$RNA from F9 cells and extended with 100 units of reverse transcriptase, and the second strand was synthesized using ribonuclease H and DNA polymerase I (*23*). Phosphorylated *Eco*RI linkers were ligated to the blunt-ended and methylated cDNAs, which were then digested with *Eco*RI and ligated to λgt10 arms. The phage DNA was packaged *in vitro*. *Mva*I-*Pst*I fragment of pFR1 was [α-^{32}P]dCTP-labeled by using a random oligonucleotide primer (Amersham) (*24*) and was used as a probe for screening the λgt10 library by plaque hybridization.

RNA Analysis—Total cellular RNA was extracted by the guanidium isothiocyanate/cesium chloride method (*25*). Poly(A)$^+$RNA was prepared on an oligo(dT)cellulose column (*25*). RNA was denatured with glyoxal, separated on a 1% agarose gel and transferred to nitrocellulose membrane in 20×SSC (*25*). 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 in 50% deionized formamide, 5×SSC, 50 mM sodium phosphate, pH 6.5, 200 μg/ml of heat-denatured salmon sperm DNA, 10×Denhart's solution and 0.1% SDS at 42°C and hybridized for 20 h with [α-^{32}P]-dCTP-labeled probes prepared by a random hexanucleotide priming method at $0.5\text{-}2.0\times10^7$ cpm/ml under conditions identical to those used for prehybridization. The membrane was washed with several changes of 2×SSC, 0.1% SDS at room temperature followed by several successive washes in 0.1×SSC, 0.1% SDS at 56°C. RNA slot blotting was performed using an apparatus from Schleicher and Schuell. Hybridization conditions were the same as in the case of Northern blotting. β-Actin cDNA (donated by Dr. K. Shimada) was used as a control.

Preparation of Anti-Fusion Protein Antibodies—The fusion protein produced by λFR1, a recombinant λgt11 clone containing a cDNA insert of basigin gene was prepared by the method of Young and Davis (*20*). BNN103 (λFR1) lysogen was grown in LB medium at 30°C until the optical density at 600 nm of the cell culture reached 0.5, and was induced for phage production by a temperature shift to 42°C for 20 min. Cells were grown at 37°C for 2 h and harvested by centrifugation. The cell pellet was suspended in PBS(−) and disrupted by sonication. After centrifugation of the sonicate at 40,000 rpm for 30 min, the pellet, which will subsequently be called membrane fraction, was solubilized in SDS sample buffer (*26*) and purified by preparative SDS-polyacrylamide gel electrophoresis on a 9% gel since a large amount of the fusion protein produced by the insert cDNA was associated with the bacterial cell surface fraction. The gel bands containing the fusion protein were cut out and washed in five changes of PBS(+) for 1 day at 4°C. Gel pieces containing around 100 μg of the fusion protein were homogenized and emulsified with an equal volume of Freund's complete adjuvant. New Zealand White rabbits were immunized subcutaneously with the emulsion. After six booster injections at 2-week intervals using the same amount of the antigen, the rabbits were bled and antisera were obtained.

Electrophoretic Aanalysis of LTA Receptors—LTA receptors were separated by SDS-polyacrylamide gel electrophoresis on a 9% gel (*26*) and examined by Western blotting (*27*). The staining of the blots with antisera was performed as previously described (*8*).

Computer Analysis of Sequence Data—Homology was searched by the program from Software Development, Tokyo, employing the data base of EMBL-GDB, LASL-GDB (Gen Bank), and NBRF-PDB. The algorithm of the program was that of Lipman and Pearson (*28*); the condition employed was $K_{tup}=2$ (*28*). The significance of homology was tested in two ways: 1) alignment by LFASA and examination of significance by RDF2 (*29*) with $K_{tup}=1$ and a penalty for deletion which is equal to −12 for the first residue in a gap and −4 for each subsequent residue,

2) alignment by SEQHP and examination by SEQDP (*30*) with a break penalty of 8. The algorithm used by the SEQHP provides a somewhat more complete search of homology than that by the FASTA and LFASTA. The log odds matrix for 250 PAMs calculated by Dayhoff *et al.* (*31*) was used as a similarity matrix for amino acid mismatches with these local homology search programs. The significance of alignments was expressed as SD away from the mean score of 140 random alignments.

Secondary structure of a polypeptide was predicted according to the principle of Chou and Fasman (*32*); the calculation was aided with a program from Software Development, Tokyo.

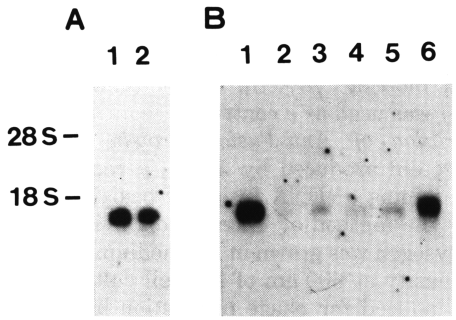

Fig. 1. **Northern blot analysis of basigin mRNA.** Poly(A)$^+$ RNAs (10 μg) or total RNAs (20 μg) were denatured, separated by 1.0% agarose gel electrophoresis, transferred to a nitrocellulose membrane and hybridized with an [α-^{32}P]dCTP-labeled probe (*Pst*I-*Acc*I restriction fragment of basigin cDNA) as described in "MATERIALS AND METHODS." The sources of poly(A)$^+$RNAs or total RNAs were as follows: A), poly(A)$^+$RNAs from 1) F9 cells, 2) teratocarcinoma OTT6050; B), 1) poly(A)$^+$RNAs from F9 cells, total RNAs from 2) the liver, 3) the kidney, 4) the brain, 5) the spleen, 6) the testis.

RESULTS

Cloning of a Gene Specifying a Carrier of LTA Binding Sites in EC Cells—The λgt11 expression library constructed from RNA of F9 EC cells was screened by the use of affinity-purified antibodies against LTA receptors isolated from N4-1 EC cells. After screening 1.4×10^6 colonies, we obtained 4 positive clones. Analysis by restriction endonucleases revealed that the 4 clones were identical. Thus, we analyzed one of the clones, λFR1 in detail.

By Northern blot analysis, the FR1 probe (*Pst*I-*Acc*I fragment of pFR1) detected a single band of 1.5 kb (Fig. 1). Since the λFR1 insert was 1.2 kb, we analyzed the 5′ end sequences by screening the library constructed by primer extension and could extend the sequences about 0.1 kb (Fig. 2A).

LTA receptors comprise many glycoproteins of different molecular weights. We examined what molecular species corresponded to the protein specified by the cDNA clone. For that purpose, we isolated the fusion protein produced by λFR1, and immunized it into a rabbit. When LTA-binding glycoproteins from N4-1 tumors were separated by SDS gel electrophoresis, transferred to nitrocellulose membranes, and stained with anti-fusion protein serum, broad bands with a molecular weight range of 43,000-66,000 were detected (Fig. 3). In addition, faint bands in the molecular weight range of 66,000-140,000 were present. The staining of these bands was abolished by incubating the antiserum with membrane fraction of *E. coli* BNN103 (λFR1) lysogen, but not with that of *E. coli* lysogenized with λgt11 lacking the insert (BNN97). Therefore, it is certain that λFR1 specifies the core portion of the glycoprotein of molecular weight 43,000-66,000. The faint staining of high-molecular-weight bands may be significant; it is possible that all bands share the protein portion while the large ones are difficult to stain due to interference by large amounts of covalently bound carbohydrates. Further studies are needed to clarify this point.

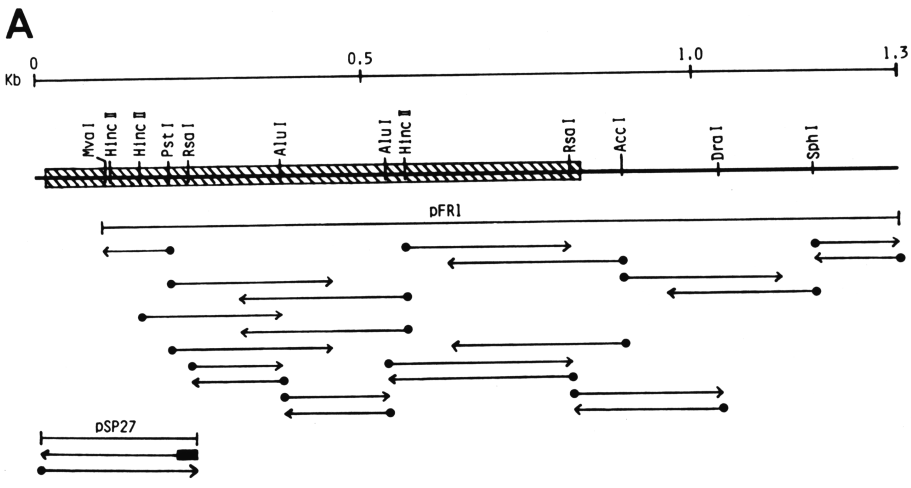

Fig. 2. **A: Restriction endonuclease map and sequencing strategy of the basigin cDNA clone.** The scale indicates nucleotide positions (in kilobases) beginning from the first base of the cDNA. Beneath the scale basigin mRNA is represented with a thick black line delineating the coding region. The restriction endonuclease sites indicated above this line were used to generate cDNA fragments for sequencing. Basigin cDNA clone pFR1 is represented below the restriction endonuclease map. pSP27 clone is a cDNA clone isolated by primer extension using a synthetic primer complementary to nucleotides 203-232 of basigin cDNA (indicated as ■). The arrows illustrate the direction and extent of DNA sequences obtained in each experiment. **B: Sequence of cDNA encoding basigin and the deduced protein sequence.** ——, putative signal and transmembrane sequences; ▲, potential Asn-glycosylation sites; ═══, polyadenylation signals.

B

```
                                                                ACGAGGCGAC      10

11   ATG GCG GCG GCG CTG CTG CTG GCG CTG GCC TTC ACG CTC TTG AGC GGC CAA GGC GCC TGC      70
     Met Ala Ala Ala Leu Leu Leu Ala Leu Ala Phe Thr Leu Leu Ser Gly Gln Gly Ala Cys

                                        10                                      20

71   GCG GCG GCG GGC ACC ATC CAA ACC TCT GTC CAG GAA GTC AAC TCC AAA ACA CAG CTT ACC     130
     Ala Ala Ala Gly Thr Ile Gln Thr Ser Val Gln Glu Val Asn Ser Lys Thr Gln Leu Thr

                                        30                                      40

131  TGC TCT TTG AAC AGC AGT GGC GTT GAC ATC GTT GGC CAC CGC TGG ATG AGA GGT GGC AAG     190
     Cys Ser Leu Asn Ser Ser Gly Val Asp Ile Val Gly His Arg Trp Met Arg Gly Gly Lys
                 ▲
                                        50                                      60

191  GTA CTG CAG GAG GAC ACT CTG CCC GAC CTG CAT ACG AAG TAC ATA GTG GAC GCA GAT GAC     250
     Val Leu Gln Glu Asp Thr Leu Pro Asp Leu His Thr Lys Tyr Ile Val Asp Ala Asp Asp

                                        70                                      80

251  CGC TCT GGG GAA TAT TCC TGC ATC TTC CTT CCT GAG CCT GTG GGC AGA AGC GAG ATC AAT     310
     Arg Ser Gly Glu Tyr Ser Cys Ile Phe Leu Pro Glu Pro Val Gly Arg Ser Glu Ile Asn

                                        90                                     100

311  GTG GAA GGG CCA CCC AGG ATC AAG GTC GGA AAG AAA TCA GAG CAT TCC AGT GAG GGA GAG     370
     Val Glu Gly Pro Pro Arg Ile Lys Val Gly Lys Lys Ser Glu His Ser Ser Glu Gly Glu

                                       110                                     120

371  CTT GCG AAA CTG GTC TGC AAG TCC GAT GCA TCC TAC CCT CCT ATT ACA GAT TGG TTC TGG     430
     Leu Ala Lys Leu Val Cys Lys Ser Asp Ala Ser Tyr Pro Pro Ile Thr Asp Trp Phe Trp

                                       130                                     140

431  TTT AAG ACC TCT GAC ACT GGG GAA GAA GAG GCA ATC ACC AAT AGC ACT GAA GCC AAT GGC     490
     Phe Lys Thr Ser Asp Thr Gly Glu Glu Glu Ala Ile Thr Asn Ser Thr Glu Ala Asn Gly
                                                    ▲
                                       150                                     160

491  AAG TAT GTG GTG GTA TCC ACG CCT GAG AAG TCA CAG CTG ACC ATC AGC AAC CTT GAC GTA     550
     Lys Tyr Val Val Val Ser Thr Pro Glu Lys Ser Gln Leu Thr Ile Ser Asn Leu Asp Val

                                       170                                     180

551  AAT GTT GAC CCT GGC ACC TAC GTG TGT AAT GCC ACC AAC GCC CAG GGC ACT ACT CGG GAA     610
     Asn Val Asp Pro Gly Thr Tyr Val Cys Asn Ala Thr Asn Ala Gln Gly Thr Thr Arg Glu
                                          ▲
                                       190                                     200

611  ACC ATC TCA CTG CGT GTG CGG AGC CGC ATG GCA GCC CTC TGG CCC TTC CTA GGC ATC GTG     670
     Thr Ile Ser Leu Arg Val Arg Ser Arg Met Ala Ala Leu Trp Pro Phe Leu Gly Ile Val

                                       210                                     220

671  GCT GAG GTC CTG GTG TTG GTT ACC ATC ATC TTT ATC TAT GAG AAG AGG CGG AAG CCA GAC     730
     Ala Glu Val Leu Val Leu Val Thr Ile Ile Phe Ile Tyr Glu Lys Arg Arg Lys Pro Asp

                                       230                                     240

731  CAG ACC CTG GAC GAG GAT GAC CCT GGC GCC GCC CCA CTG AAG GGC AGT GGA ACT CAC ATG     790
     Gln Thr Leu Asp Glu Asp Asp Pro Gly Ala Ala Pro Leu Lys Gly Ser Gly Thr His Met

                                       250                                     260

791  AAT GAC AAG GAC AAG AAT GTA CGC CAG AGG AAC GCC ACC TGA GTGGTGGGGCAGGCGGGGGAGGG     855
     Asn Asp Lys Asp Lys Asn Val Arg Gln Arg Asn Ala Thr ***
                                             ▲
                                       270

856  GAGGTGCCCAGGGTGCCTGACCCCAGGCCAGCGTCTACCTCCACTCCAGTATCCCATCCTGTCCCGATTTGAACCTACC     934

935  CAACCCAACCTATCCCAACCCAAGTGAAGACAGAGCCTTACCTTACAGAAAACCCACCTGGAAGAAGCAAGCCACTTGC    1013

1014 AGCCCCTGTTTCTAATTTAAACTAAATGAGGTTTCTATGCAGACAATCCATTCCTTAGGGGTTTATGTTTTTATTTTTC    1092

1093 CTTCCCTTCTGAAGTGTTGTCACTACAGCCCTGTGGAGTGGGGGAATGGGGCCTTGTCCTTGGTCAGGAGGGAAGGCCA    1171

1172 GTGCATGCTCTGACTTACTGTTGGAGGGGGCTGGGCCTGCTGGAAACCCCCCAAATAAAACCTAACCCACCAAAAAAAA    1250

1251 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA                            1302
```

Fig. 2B.

Structure of cDNA Coding Basigin—Sequencing of the cDNAs (Fig. 2A) gave a composite nucleotide sequence comprising 1,302 bases (Fig. 2B). In all of the cases, the sequence was read from both strands.

A single open reading frame, extending from the ATG at nucleotide 11 to the TGA stop codon at base 830 encoded a polypeptide with a predicted molecular weight about 30,000. For the following reasons we have assigned the methionine codon as the initiator, although no in-frame stop codon was detected upstream from the ATG: 1) Sequences flanking the ATG codon, namely, CGACATGG-CG were homologous to the consensus sequence flanking the initiator ATG codon, namely CAA(C)CATGGCT (*33*); 2) the methionine residue was followed by a stretch of 21 hydrophobic amino acid residues. This stretch is highly likely to be a signal sequence (*34*). Considering that Ala prefers the −1 and −3 positions (taking the cleavage site as −1 and +1) but not the −2 position, and that as the +1 amino acid, Ala, Asn, Gln, and charged amino acids are preferred (*34*), the preferable cleavage site for the signal sequence appears to be between Ala_{21} and Ala_{22}. There is another stretch of 24 hydrophobic amino acid residues starting from Met at position 210. This sequence is most likely to be a transmembrane domain. Directly downstream from the putative transmembrane domain, there is a cluster of charged amino acids, which is usually found downstream from transmembrane domains. In the putative extracellular domain, there are 3 potential asparagine glycosylation sites.

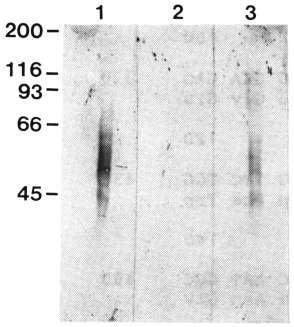

Fig. 3. **Detection of basigin by antibodies against a fusion protein produced by λFR1.** LTA receptors (40 μg as proteins) isolated from N4-1 tumors were separated by SDS gel electrophoresis on a 9% gel under reducing conditions, and proteins were electrophoretically transferred to a nitrocellulose membrane and reacted with antiserum against the fusion protein produced by λFR1. Lane 1: reaction with the antiserum; lane 2: reaction with the antiserum, which was pre-absorbed with the membrane fraction of BNN103 (λFR1) lysogen; lane 3: reaction with the antiserum, which was pre-absorbed with that of BNN97, a lysogen with λgt11 lacking insert. The membrane fraction used for absorption was prepared as described in "MATERIALS AND METHODS" (section entitled *Preparation of Anti-Fusion Protein Antibodies*). Molecular weight markers used were myosin (200,000), β-galactosidase (116,250), phosphorylase *b* (92,500), bovine serum albumin (66,200), and ovalbumin (45,000).

Basigin as a Member of the Ig Superfamily—When the sequence of basigin was compared with each sequence of the entire data base, we found homology with the Eβ-chain of MHC class II antigen (*35*) and the variable (V) domain of Igκ chain (*36*) (Fig. 4). In the former case, the homologous region comprised 125 amino acid residues with an overall homology of 22.4% identity. The significance of these homologies was examined by LFASA and RDF2. The SD values obtained were 7.69 between basigin and Eβ, and 6.40 between basigin and Igκ; if alignment scores of random sequences obey the Gaussian distribution, those SD values correspond to the probabilities of similarity by chance of 7.4×10^{-15} and 7.8×10^{-11}, respectively. By the SEQHP program 177 amino acid residues of basigin could be aligned with 171 residues of MHC class II Eβ-chain (data not shown). The SD value obtained by the SEQDP was 7.51, indicating that the probability of similarity by chance was 3.0×10^{-14}. This difference of alignment between the two method results from the difference in the gap penalty used by them; the LFASTA with the gap penalty of −8 generated the essentially same alignment. On the other hand, 84 residues of basigin were aligned with 86 residues of Igκ chain by the LFASTA and also the SEQHP. The SD value obtained by SEQDP was 7.3, indicating that the probability of similarity by chance was 1.4×10^{-13}. It seems reasonable to say that the two proteins belong to the same superfamily, because the probability of coincidental

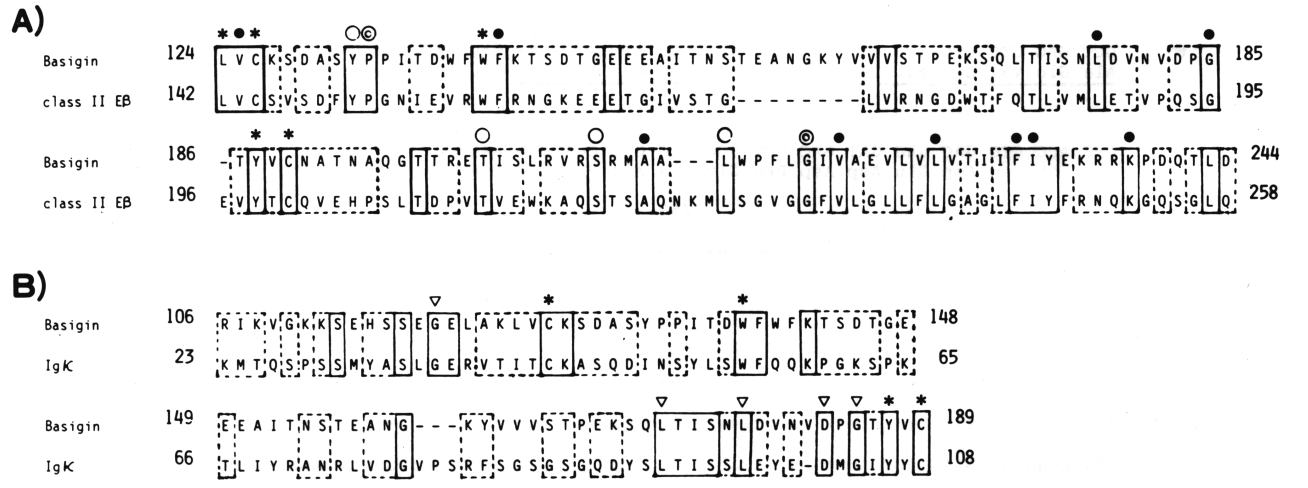

Fig. 4. **Homology of basigin with MHC class II Eβ-chain (A) (*35*) and Ig V_κ (B) (*36*).** Numbers indicate the position of the amino acid residue. ∗, residues conserved in many members of the Ig superfamily; ●, residues conserved in many MHC class I, class II α- and class II β-chains; ○, residues conserved in many MHC class I and class II β- chains; ●, residues conserved in many class II β-chains; ▽, residues conserved in many V region-like domains. ☐ means identical amino acids and ⋮⋮ means conservative amino acid substitutions.

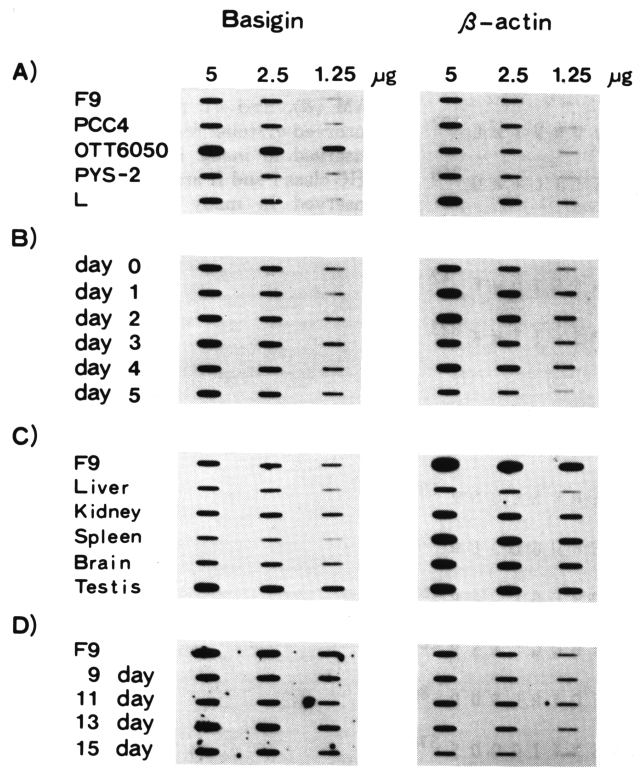

Fig. 5. **Comparison of basigin mRNA levels in several cells, organs, and embryos.** Total RNA isolated from the specimen (5-1.25 μg) was applied to nitrocellulose membranes and was hybridized with an [α-^{32}P]dCTP-labeled basigin probe (*PstI-AccI* fragment of the cDNA) (left side) or with a β-actin probe (right side). A) Teratocarcinoma cells and L fibroblasts. B) F9 cells induced to differentiate by treatment with retinoic acid and dibutyryl cyclic AMP. Day 0, no treament; day *n*, *n* days after addition of retinoic acid and dibutyryl cyclic AMP. C) Organs of adult mice. D) Mouse embryos. The number of days indicates the days of gestation.

identity is less than 10^{-6} (*37*). Since both the κ chain and the Eβ-chain belong to the Ig superfamily, we have concluded that basigin is a new member of the Ig superfamily. This point has been further confirmed as follows.

In the V and C (constant) domains of Igs, two cysteine residues form an intra-chain -S-S- bridge (*38*); this appears to be the case also in Ig-like domains of other members of the superfamily. These cysteine residues as well as certain amino acids in the N-terminal region of the cysteine residues are highly conserved (*2, 8, 39*) in most members of the superfamily. Furthermore, a tryptophan residue is usually present at about the 12th-14th amino acid residue position in the C-terminal region of the first conserved cysteine residue (*2, 38*). We found that these conserved features were also present in basigin (* in Fig. 4).

We examined the secondary structure of the Ig-related domain from its primary structure. Six segments (amino acids 123-125, 135-141, 160-165, 171-175, 177-181, and 185-188) were predicted to form β-sheets, while no α-helix was predicted in the region. Therefore, it is likely that the Ig-related domain of basigin is folded into multiple β-sheets as in the case of Ig V and C domains (*40*).

From all these results, it is certain that basigin is a new member of the Ig superfamily. Since members of the Ig superfamily often have multiple domains, we searched for other Ig-like domains in other parts of the molecule. Indeed, Cys_{41} and Cys_{87} appear to form another Ig-like domain. Leu_{39} and Trp_{55} residues are in conserved positions in relation to Cys_{41}. Gly_{83} and Tyr_{85} are also in conserved positions in relation to Cys_{87}. Therefore, there appears to be another Ig-like domain in the more N-terminal region.

Mode of Basigin Expression—We examined the mode of basigin expression by means of RNA blotting experiments (Fig. 5). A β-actin probe was simultaneously used to compare the basigin mRNA level not only on the basis of total RNA but also by determining the ratio to β-actin mRNA. The samples examined were F9 and PCC4 EC cells, OTT6050 teratocarcinoma cells, PYS-2 parietal endoderm cells, L fibroblasts (Fig. 5A), F9 EC cells induced to differentiate by treatment with retinoic acid and dibutyryl cyclic AMP (Fig. 5B), organs of adult mice *i.e.* the liver, kidney, spleen, brain, and testis (Fig. 5C), and mouse embryos (9-15 days of gestation, Fig. 5D). Basigin mRNA was expressed at significant levels in all the specimens mentioned above. Dramatic changes in the mRNA levels did not occur during mouse embryogenesis or differentiation of F9 cells. However, the testis and teratocarcinoma OTT6050 expressed considerably more basigin mRNA than other specimens; the level of β-actin mRNA was not significantly higher in the specimens. Upon Northern blot analysis, basigin mRNA from adult organs and F9 cells gave a single band with the same mobility (Fig. 1B).

DISCUSSION

Basigin, a newly found member of the Ig superfamily has two unique characteristics. Firstly, it was detectable to a similar degree, in adult organs, mouse embryos and EC cells. While MHC class I antigen, a member of the Ig superfamily, is present in many adult cells, it is not expressed in EC cells (*41*). Furthermore, mouse embryos before 10 days of gestation poorly express class I antigen (*42*).

Secondly, basigin showed homology to both the Ig V domain and MHC class II antigen β-chain. We have examined whether the amino acids conserved in basigin and MHC class II antigen β-chain are also conserved in MHC class II antigen α-chain and MHC class I antigen, since both also belong to the Ig superfamily and contain C domain-like regions (*35, 43-48*) (Figs. 4 and 6). The homologous amino acid positions between basigin and MHC class II antigen β-chain could be classified as follows (Fig. 4). In addition to amino acid residues conserved in diverse members of the Ig superfamily (*), 2 amino acids marked with ◎ were conserved in basigin, class I antigen and class II antigen α- and β-chains; 4 amino acids marked with ○ were conserved between basigin, class I antigen and class II antigen β-chain but not in class II antigen α-chain. On the other hand, 10 amino acids (●) were conserved between basigin and many of the class II antigen β-chains, which included both Eβ- and Aβ-chains from both the mouse and human (*35, 45-48*). This consideration reinforces the significance of the homology between basigin and MHC class II antigen β-chain.

Typical C-domains found in Ig and MHC class I and class II antigens share, in most cases, the characteristic amino

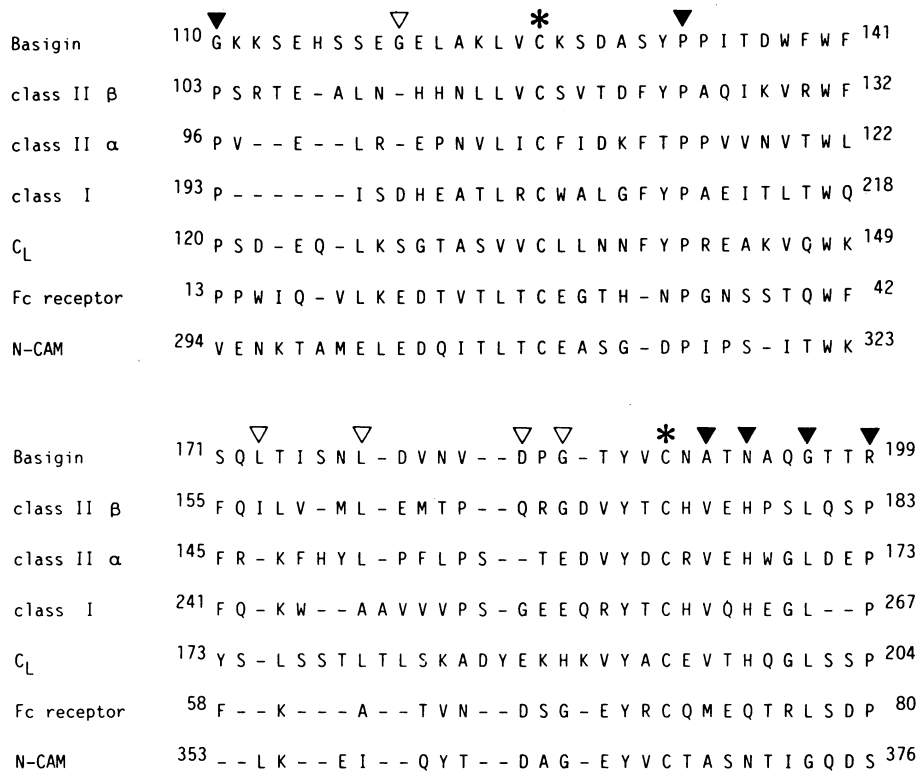

Fig. 6. **Comparison of the Ig-like domain of basigin with the C domain of human MHC antigens (*45*), an immunoglobulin light chain (*52*), N-CAM (*6*), and Fc receptor (*49*).** *, conserved cysteine residues; ▼, residues conserved in many Ig C domains and MHC class I and II antigens; ▽, residues conserved in many Ig V region-like domains.

acid residues marked with ▼ in Fig. 6. Among the 6 residues, only one, namely Pro_{133} was conserved in basigin. Thus, basigin does not belong to the typical C-domain group. The number of amino acids between the two conserved cysteines is 55–60 or less in the case of C domains, while the number is usually 65–70 in the case of V domains. The number of the amino acids was 62 in the case of basigin; in this respect also, basigin cannot be regarded as belonging to the C-domain group.

Recently, several members of the Ig superfamily have been shown to have domains somewhat related to C-domains (*2*). Of them the Fc receptor indeed has a domain similar to the C domain (*49*); of the 6 amino acids characteristic of typical C domains, 4 were found in the domain of Fc receptor (Fig. 6). However, N-CAM retains only one such amino acid (*6*) (Fig. 6). The number of amino acid residues between the conserved cysteines is the same as that of C-domain in both cases. These molecules are now classified as C_2 sets, and typical C-domains in Ig and MHC antigens as C_1 sets (*2*). Molecules of C_2 sets such as N-CAM share certain amino acids with the members of V sets (*2*) (Fig. 6). From the sequence characteristics mentioned above, basigin might be regarded as belonging to C_2 sets. However, in the size of the Ig domain, it is intermediate between C and V sets. Furthermore, the homology between basigin and the Ig V-domain was extensive, and all features of the V set sequences were present in basigin (Fig. 4). To the best of our knowledge, no member of the Ig superfamily has a high degree of homology both to a V domain and a C domain as in the case of basigin. For example, the homology score between N-CAM and V_κ is much less than that between basigin and V_κ (*50*). No significant local homology can be found between Igκ chain and MHC class II Eβ-chain, when the LFASTA (*29*) and also the SEQHP program (*30*) are used with the same set of parameters used for the present study. Therefore, basigin is an interesting molecule in connection with the molecular evolution of the Ig superfamily. It is tempting to speculate that a primordial molecule of the Ig superfamily is closely related to basigin.

Antibodies against fusion protein of basigin react with a portion of LTA receptors of molecular weight between 43,000 and 66,000. Therefore, there appears to be little doubt that basigin belongs to the category of LTA receptors. However, it should be noted that we have not yet measured the degree of affinity between basigin and LTA. Although basigin is believed to be a core protein carrying a developmentally regulated carbohydrate marker, basigin expression itself was not strongly influenced by developmental status. This observation is not surprising, when we consider that developmental changes of cell-surface carbohydrates are principally accomplished by alterations in glycosyltransferase activities (*51*).

Members of the Ig superfamily play critical roles in intercellular recognition. This concept has been established in lymphoid cells and nerve cells. Although most of the members of the superfamily are not expressed in early embryonic cells including EC cells, we have been able to demonstrate two new members of the superfamily, namely basigin and Gp 70, in EC cells simply by surveying core proteins carrying developmentally regulated cell-surface markers expressed in early embryonic cells. Thus, it is likely that there are still other members of the superfamily expressed in early embryonic cells and that they interact with each other and are involved in critical stages of embryogenesis, just as the members expressed in lymphoid

cells and nervous cells do in immunological reactions and nerve cell interactions. Furthermore, the widespread distribution of basigin in adult organs as well as in EC and embryonic cells implies a fundamental role in both adult and embryonic cells.

We thank Dr. Rolf Kemler for providing the λgt11 expression library constructed from F9 cells, Mr. Yasushi Masuzawa for cooperation in sequencing analysis, Mr. Daniel Mrozek for editing the manuscript and Miss Kumiko Sato for her expert secretarial assistance.

REFERENCES

1. Hood, L., Kronenberg, M., & Hunkapiller, T. (1985) *Cell* **40**, 225-229
2. Williams, A.F. & Barclay, A.N. (1988) *Annu. Rev. Immunol.* **6**, 381-405
3. Stewart, S.J., Prpic, V., Powers, F.S., Bocckino, S.B., Isaacks, R.E., & Exton, J.H. (1986) *Proc. Natl. Acad. Sci. U.S.* **83**, 6098-6102
4. Sarmiento, M., Glasebrook, A.L., & Fitch, F.W. (1980) *J. Immunol.* **125**, 2665-2672
5. Doyle, C. & Strominger, J.L. (1987) *Nature* **330**, 256-259
6. Cunningham, B.A., Hemperly, J.J., Murray, B.A., Prediger, E.A., Brackenbury, R., & Edelman, G.M. (1987) *Science* **236**, 799-806
7. Moos, M., Tacke, R., Scherer, H., Teplow, D., Früh, K., & Schachner, M. (1988) *Nature* **334**, 701-703
8. Ozawa, M., Huang, R.-P., Furukawa, T., & Muramatsu, T. (1988) *J. Biol. Chem.* **263**, 3059-3062
9. Martin, G.R. (1980) *Science* **209**, 768-776
10. Kamada, Y., Arita, Y., Ogata, S., Muramatsu, H., & Muramatsu, T. (1987) *Eur. J. Biochem.* **163**, 497-502
11. Gooi, H.C., Feizi, T., Kapadia, A., Knowles, B.B., Solter, D., & Evans, M.J. (1981) *Nature* **292**, 156-158
12. Bernstine, E.G., Hooper, M.L., Grandchamp, S., & Ephrussi, B. (1973) *Proc. Natl. Acad. Sci. U.S.* **70**, 3899-3903
13. Sato, M., Ogata, S., Ueda, R., Namikawa, R., Takahashi, T., Nakamura, T., Sato, E., & Muramatsu, T. (1983) *Dev. Growth Differ.* **25**, 333-344
14. Jakob, H., Boon, T., Gaillard, J., Nicolas, J.-F., & Jacob, F. (1973) *Ann. Microbiol. (Inst. Pasteur)* **124B**, 269-282
15. Lehman, J.M., Speers, W.C., Swartzendruber, D.E., & Pierce, G.B. (1974) *J. Cell. Physiol.* **84**, 13-28
16. Earle, W.R., Shilling, E.L., Stark, T.H., Straus, N.P., Brown, M.F., & Shelton, E. (1943) *J. Natl. Cancer Inst.* **4**, 165-212
17. Strickland, S., Smith, K.K., & Marotti, K.R. (1980) *Cell* **21**, 347-355
18. Stevens, L.C. (1970) *Dev. Biol.* **21**, 364-382
19. Dulbecco, R. & Vogt, M. (1954) *J. Exp. Med.* **99**, 167-182
20. Young, R.A. & Davis, R.W. (1983) *Proc. Natl. Acad. Sci. U.S.* **80**, 1194-1198
21. Schuh, R., Vestweber, D., Riede, I., Ringwald, M., Rosenberg, U.B., Jäckle, H., & Kemler, R. (1986) *Proc. Natl. Acad. Sci. U.S.* **83**, 1364-1368
22. Sanger, F., Nicklen, S., & Coulsen, A.R. (1977) *Proc. Natl. Acad. Sci. U.S.* **74**, 5463-5467
23. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 161-170
24. Feinberg, A.P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6-13
25. Maniatis, T., Fritsch, E.F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. Laemmli, U.K. (1970) *Nature* **227**, 680-685
27. Towbin, H., Staehelin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.* **76**, 4350-4354
28. Lipman, D.J. & Pearson, W.R. (1985) *Science* **227**, 1435-1441
29. Pearson, W.R. & Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. U.S.* **85**, 2444-2448
30. Goad, W.B. & Kenehisa, M.I. (1982) *Nucleic Acids Res.* **10**, 247-263
31. Dayhoff, M.O., Schwartz, R.M., & Orcutt, B.C. (1978) in *Atlas of Protein Sequence and Structure* Volume 5, Supplement 3, pp. 345-352, National Biomedical Research Foundation, Washington, D.C.
32. Chou, P.Y. & Fasman, G.D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**, 45-148
33. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857-872
34. von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243-251
35. Saito, H., Maki, R.A., Clayton, L.K., & Tonegawa, S. (1983) *Proc. Natl. Acad. Sci. U.S.* **80**, 5520-5524
36. Pech, M., Höchtl, J., Schnell, H., & Zachau, H.G. (1981) *Nature* **291**, 668-670
37. Dayhoff, M.O., Barker, W.C., & Hunt, L.T. (1983) *Methods Enzymol.* **91**, 524-545
38. Edelman, G.M., Cunningham, B.A., Gall, W.E., Gottlieb, P.D., Rutishauser, U., & Waxdal, M.J. (1969) *Proc. Natl. Acad. Sci. U.S.* **63**, 78-85
39. Williams, A.F. & Gagnon, J. (1982) *Science* **216**, 696-703
40. Saul, F.A., Amzel, L.M., & Poljak, R.J. (1978) *J. Biol. Chem.* **253**, 585-597
41. Morello, D., Daniel, F., Baldacci, P., Cayre, Y., Gachelin, G., & Kourilsky, P. (1982) *Nature* **296**, 260-262
42. Flavell, R.A., Allen, H., Burkly, L.C., Sherman, D.H., Waneck, G.L., & Widera, G. (1986) *Science* **233**, 437-443
43. Coligan, J.E., Kindt, T.J., Uehara, H., Martinko, J., & Nathenson, S.G. (1981) *Nature* **291**, 35-39
44. Lõpez de Castro, J.A., Bragado, R., Strong, D.M., & Strominger, J.L. (1983) *Biochemistry* **22**, 3961-3969
45. Larhammar, D., Gustafsson, K., Claesson, L., Bill, P., Wiman, K., Schenning, L., Sundelin, J., Widmark, E., Peterson, P.A., & Rask, L. (1982) *Cell* **30**, 153-161
46. McNicholas, J., Steinmetz, M., Hunkapiller, T., Jones, P., & Hood, L. (1982) *Science* **218**, 1229-1232
47. Yokoyama, K. (1984) in *Immunological Sciences* (in Japanese) (Tada, T. & Sasazuki, K., eds.) Vol. 6, pp. 93-186, Iwanami Publishers, Tokyo
48. Widera, G. & Flavell, R.A. (1984) *EMBO J.* **3**, 1221-1225
49. Lewis, V.A., Koch, T., Plutner, H., & Mellman, I. (1986) *Nature* **324**, 372-375
50. Matsunaga, T. & Mori, N. (1987) *Scand. J. Immunol.* **25**, 485-495
51. Muramatsu, T. (1988) *Biochimie* **70**, 1587-1596
52. Edelman, G.M. (1970) *Biochemistry* **9**, 3197-3205