Amino Acid Sequence and Domain Structure of Entactin. Homology with Epidermal Growth Factor Precursor and Low Density Lipoprotein Receptor

Marian E. Durkin, Shukti Chakravarti, Barbara B. Bartos, Shu-Huang Liu, Robert L. Friedman, and Albert E. Chung

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Abstract. Entactin (nidogen), a 150-kD sulfated glycoprotein, is a major component of basement membranes and forms a highly stable noncovalent complex with laminin. The complete amino acid sequence of mouse entactin has been derived from sequencing of cDNA clones. The 5.9-kb cDNA contains a 3,735-bp open reading frame followed by a 3′-untranslated region of 2.2 kb. The open reading frame encodes a 1,245-residue polypeptide with an unglycosylated M_r of 136,500, a 28-residue signal peptide, two Asn-linked glycosylation sites, and two potential Ca^{2+}-binding sites. Analysis of the deduced amino acid sequence predicts that the molecule consists of two globular domains of 70 and 36 kD separated by a cysteine-rich domain of 28 kD. The COOH-terminal globular domain shows homology to the EGF precursor and the low density lipoprotein receptor. Entactin contains six EGF-type cysteine-rich repeat units and one copy of a cysteine-repeat motif found in thyroglobulin. The Arg-Gly-Asp cell recognition sequence is present in one of the EGF-type repeats, and a synthetic peptide from the putative cell-binding site of entactin was found to promote the attachment of mouse mammary tumor cells.

BASEMENT membranes are a type of extracellular matrix that form thin sheets separating epithelial, endothelial, muscle, fat, and nerve cells from connective tissue (51). Epithelial cells require contact with a basement membrane to maintain their morphology and differentiated phenotype, and this is a result of interactions between basement membrane molecules and cell surface receptors for them (26). The major components of basement membranes are type IV collagen, laminin, entactin, and heparan sulfate proteoglycan (reviewed in reference 48). Entactin is a 150-kD sulfated glycoprotein first identified as a product of a teratocarcinoma-derived parietal endoderm line (6). It is identical to nidogen, a polypeptide originally isolated as an 80-kD proteolytic fragment from the Engelbreth-Holm-Swarm tumor (49). Immunostaining has shown entactin to be a ubiquitous component of adult, fetal, and extraembryonic basement membranes (4, 6, 17, 29, 37, 48, 54), and it is also synthesized by early embryos (17, 54), embryonal carcinoma cells (7, 11, 17), and mesenchyme cells (28, 29, 52).

Laminin and entactin can be coextracted from cell culture media and basement membranes in the form of a highly stable noncovalent complex (7, 17, 32, 40). As visualized by electron microscopy, the complex consists of one entactin molecule bound to one of the short arms of laminin, near the center of the cross (40). Entactin also binds to fibronectin and type IV collagen, but has no affinity for heparan sulfate proteoglycan (18). Progress in elucidating the function of entactin has been slow, due to its susceptibility to proteolysis during extraction and the fact that denaturants required to inhibit proteases and dissociate the laminin–entactin complex cause a loss of binding activity (18, 40). We have used cDNA cloning to obtain the complete amino acid sequence of mouse entactin, and this has provided new insights into the structure, function, and evolution of the molecule. The entactin sequence has been found to contain EGF-like cysteine-rich repeats, segments showing homology to the EGF precursor, the low density lipoprotein (LDL)[1] receptor, and thyroglobulin, and possible sites for Ca^{2+}- and cell binding.

Materials and Methods

Purification and NH_2-terminal Amino Acid Sequencing of Entactin

The extracellular matrix proteins synthesized by the mouse parietal endoderm cell line M1536-B3 were extracted as previously described (10), and resolved by SDS-PAGE (35) on 5% polyacrylamide slab gels. The bands were visualized by soaking the gels in 1 M KCl (24). The entactin bands were cut out and the protein isolated by electroelution (30). The preparation was homogeneous when tested by SDS-PAGE. Samples were dialyzed against three 2,000-vol changes of H_2O over a 36-h period, and NH_2-terminal sequence analysis of purified entactin was performed on an amino acid sequencer (model 890M; Beckman Instruments, Inc., Palo Alto, CA).

1. *Abbreviation used in this paper*: LDL, low density lipoprotein.

Library Screening

Two M1536-B3 cDNA libraries in λgtll, one primed with oligo(dT) and one primed with a laminin B2 chain-specific oligonucleotide, were constructed and screened as described (14). λ611 (Fig. 1) was isolated from the oligo(dT)-primed library by screening with the ^{32}P-labeled insert of λ1E, a rat entactin cDNA clone isolated previously (15). A 950-bp Eco RI/Pvu II fragment from the 5′ end of the λ611 insert and a 700-bp Eco RI/Pvu II fragment from the 3′ end of the insert were then used to screen both libraries. Due to the high concentration of the laminin B2 primer used in constructing the specifically primed library, many entactin clones were obtained due to semirandom annealing of the primer.

Sequencing of cDNAs

Restriction fragments of the λ611 insert were subcloned into bacteriophage M13 mp18 or mp19 and sequenced by the dideoxy chain termination method (46) using [α-^{35}S]dATP and the Klenow fragment. The other cDNA inserts were subcloned into the Bluescript KS plasmid (Stratagene Cloning Systems, La Jolla, CA) and partially sequenced by the supercoiled plasmid dideoxy technique (9). The λ611 insert and the 2.2-kb 3′ Eco RI fragment of the λ107 insert were subcloned into Bluescript and sequenced in their entirety by constructing nested deletions using exonuclease III and S1 nuclease (27, 36). Both strands were completely sequenced, using specific synthetic oligonucleotide primers where necessary to fill in gaps. Secondary structure analysis of the derived amino acid sequence was performed by the method of Garnier et al. (20) using MacGene Plus software (Applied Genetic Technology, Inc., Fairview Park, OH). A search of the Protein Identification Resource and GenBank databases was accomplished using Bionet software (Intelligenetics, Inc., Palo Alto, CA).

Northern Blot Analysis of Entactin RNA

Methods for extraction of total RNA, formaldehyde-agarose gel electrophoresis, blotting to Gene Screen, hybridization of blots to ^{32}P-labeled probes, and subsequent washing steps have been described (16). The hybridization probe was the 240-bp internal Eco RI fragment of λ104, labeled with ^{32}P by nick translation. The blot was exposed to x-ray film at −70°C overnight with an intensifying screen.

Cell Attachment Assay

The synthetic peptide SIGFRGDGQTC was prepared by Dr. Ming F. Tam, Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan. The peptide was conjugated to BSA-coated 35-mm plastic petri dishes as described by Pytela et al. (43). The peptide-coated plates were blocked with 3% BSA in PBS overnight at 4°C before plating the cells. Mouse mammary tumor cells (MMT 060562, American Type Culture Collection, Rockville, MD) were labeled for 12 h with 2.5 μCi/ml [^{3}H]thymidine in DME + 10% fetal calf serum. The cells were harvested by trypsinization, and 0.5–1.0 × 10^5 cells per dish were plated in serum-free DME on dishes coated with peptide-BSA conjugate or BSA alone. The plates were incubated at 37°C for 4 h in an atmosphere of 100% humidity/5% CO_2 in air, then washed twice with 1 ml PBS and once with 0.5 ml 0.05% trypsin to remove unattached cells. The cells were removed by incubating with 1 ml trypsin, transferred to scintillation vials, mixed with scintillation fluid, and counted in a scintillation counter. Percent attachment was calculated as follows:

$$\frac{[\text{cpm in peptide-coated plate} - \text{cpm in BSA-coated plate}]}{\text{total cpm per plate.}} \times 100$$

Results

Isolation of Mouse Entactin cDNA Clones

Previously we had characterized a rat entactin cDNA clone, λ1E, obtained by screening a λgtl1 library with anti-entactin antiserum (15). To isolate mouse entactin cDNA clones, an oligo(dT)-primed M1536-B3 library was probed with the nick-translated λ1E insert, yielding a phage carrying a 3.4-kb insert, λ611. The oligo(dT)-primed library and a specifically primed library were then screened with nick-translated restriction fragments from the 5′ and 3′ ends of the λ611 insert, and a series of clones covering a total of 5,959 bp was isolated (Fig. 1). The size of entactin mRNA was estimated to be 6-kb by Northern blot analysis (Fig. 2), indicating that the cDNA is nearly full length. Entactin mRNA is considerably more abundant in basement membrane-secreting M1536-B3 cells (Fig. 2, lane *2*) than in F9 embryonal carcinoma cells (lane *1*), which produce much less of the protein.

Features of the Entactin cDNA and Protein Sequence

The entactin cDNA sequence and the deduced amino acid sequence are shown in Fig. 3. The cDNA contains a 3,735-bp open reading frame beginning at the ATG codon at nucleotide 12, which lies in a favorable context for translation initiation (33) and terminating at a TGA stop codon at nucleotide 3,747. The 5′ untranslated region is very short (11 bp), and it is likely that we are missing most of the 5′ leader. After the open reading frame is a 2.2-kb 3′ untranslated region, an ATTAAA polyadenylation signal, and a poly(A) tail. While

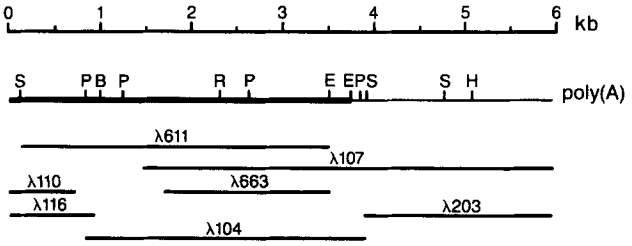

Figure 1. Map of mouse entactin cDNA clones. The thick line represents the coding region and the thin line represents the 3′ untranslated region. Restriction sites for Sac I (*S*), Pst I (*P*), Bam HI (*B*), Eco RV (*R*), Eco RI (*E*), and Hind III (*H*) are indicated. λ611 and λ663 were obtained from the oligo (dT)-primed library; the 3′ ends were lost due to cleavage at the natural Eco RI sites during construction of the library. λ110, λ116, λ104, and λ107 were isolated from the laminin B2 oligonucleotide-primed library. λ203 was isolated from the oligo (dT)-primed library by screening with a restriction fragment from the 3′ end of λ107. All cDNAs were characterized by restriction endonuclease mapping and partial sequencing. The λ611 insert and the 3′-most Eco RI fragment of λ107 were completely sequenced.

Figure 2. Northern blot analysis of entactin gene expression. Aliquots of total RNA (5 μg per lane) from mouse F9 embryonal carcinoma cells (lane *1*) and M1536-B3 cells grown in suspension culture for 6 d (lane *2*) were separated on a 0.75% agarose/2.2 M formaldehyde gel, blotted onto Gene Screen, and hybridized to the nick-translated 240-bp internal Eco RI fragment of λ104. The positions of the 18 S and 28 S rRNAs in the gel are indicated. The size of entactin mRNA (6 kb) was estimated by comparison to Hind III fragments of bacteriophage λ DNA (not shown).

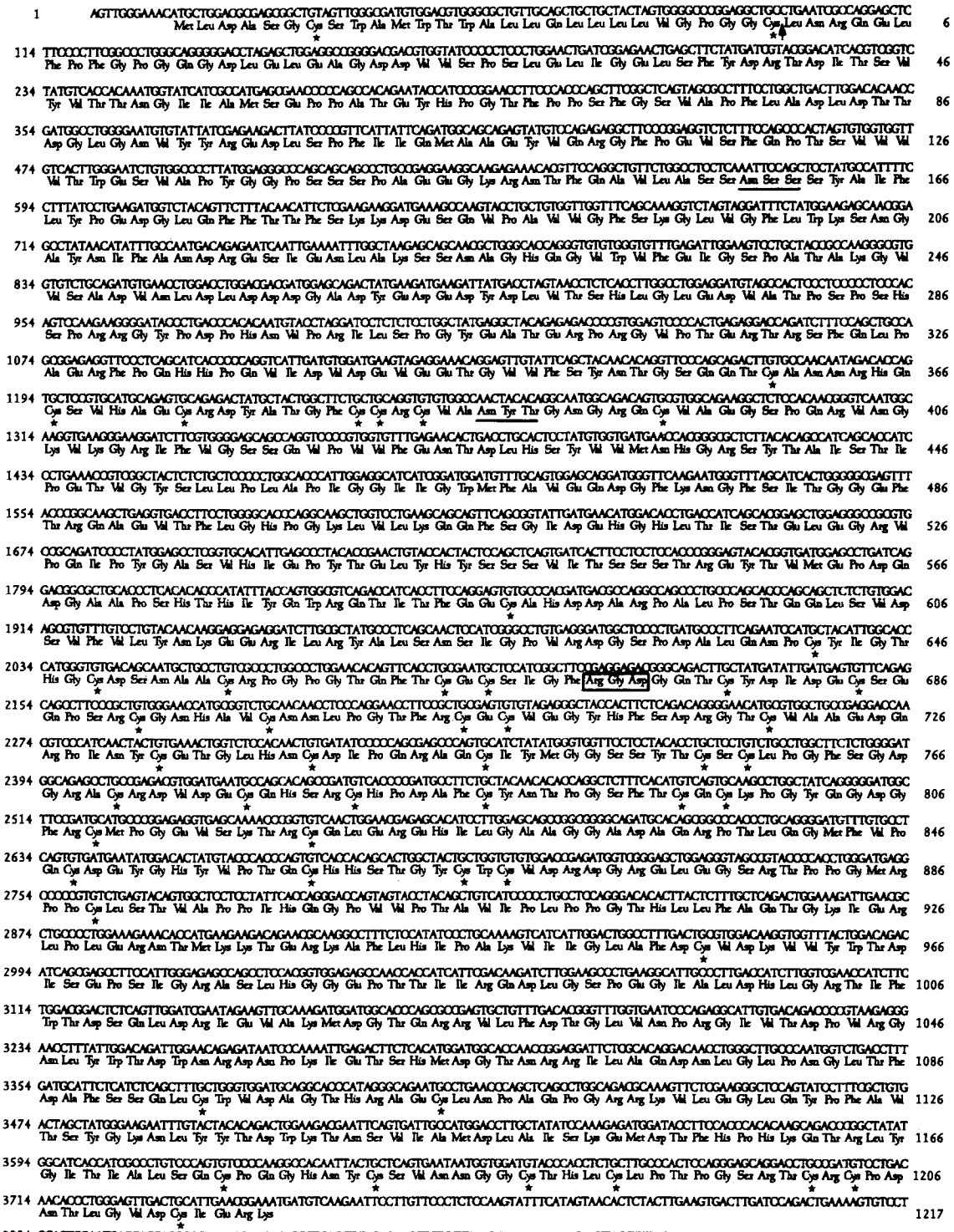

Figure 3. Nucleotide sequence of the composite entactin cDNA and the derived amino acid sequence. Numbering of the amino acid sequence begins at the NH_2 terminus of mature entactin, as determined by direct protein sequencing. An arrow indicates the signal peptide cleavage site. The two potential N-linked glycosylation sites are underlined and cysteine residues are marked with stars. The ATTAAA sequence that may serve as the polyadenylation signal is underlined with a double line, and the RGD tripeptide is boxed.

the sequence ATTAAA was found to be much less efficient than the canonical AATAAA motif in promoting cleavage and polyadenylation in one set of experiments, it occurs upstream of the poly(A) addition site in 12% of mRNAs (53).

The open reading frame encodes a 1,245-amino acid polypeptide with an unglycosylated M_r of 136,500; this is close to the M_r of 143,000 estimated for the in vitro translation product of entactin mRNA (15). The NH_2 terminus of mature entactin was determined to be LNXQELFPFGPG by Edman degradation, and this agrees with the sequence of residues 1–12. The sequence of the first 28 residues (−28 to −1) is characteristic of a signal peptide (50), which predicts that the mature entactin polypeptide has an unglycosylated M_r of 133,500 and consists of 1,217 amino acids. The sequences of residues 1–10, 204–212, 298–307, 351–360, and 621–630, with the exception of two amino acids, match the NH_2-terminal sequences of intact nidogen and its 130-, 100-, 80-, and 40-kD proteolytic fragments, respectively (41), confirming that entactin and nidogen are identical proteins. Secondary structure analysis of the deduced entactin amino acid sequence predicts that the polypeptide has no extended α-helical structure and consists primarily of β-sheet, β-turn, and random coil structures, consistent with circular dichroism measurements (41).

Entactin possesses both N- and O-linked oligosaccharides (29, 49), and two potential sites for N-linked glycosylation (N–X–S or T) are found in the entactin sequence. Sites for the addition of *N*-acetylgalactosamine to serine or threonine residues do not appear to have a simple consensus sequence, and it is believed that O-glycosylation occurs on clustered serine and threonine residues in exposed, proline-rich regions of the polypeptide (25). A number of sites that meet these criteria are present in entactin (for example, residues 57–72, 138–142, 281–288, 547–556, and 965–984). Another posttranslational modification of entactin is tyrosine *O*-sulfation (42). Tyrosines at residues 262 and 267 appear to be the most likely sulfate acceptor sites since they are surrounded by acidic residues, a common feature of tyrosine sulfation sites (31).

When the sequence of the 1,326-bp rat λ1E insert was compared with that of the mouse cDNA, it was found that the overlap began at nucleotide 282 in the mouse sequence and diverged after nucleotide 1,250. The final 354 nucleotides of the rat clone apparently represent either an intron or a cloning artifact, and the first 972 nucleotides are derived from near the NH_2 terminus of entactin, not the COOH terminus as reported originally (15). The overlapping mouse and rat sequences are 94% identical at the nucleotide level (not shown) and 93% identical at the amino acid level (Fig. 4).

```
rat     EFHPGTFPPSFGSVAPFLADLDTTDGLGNV
mouse   -Y----------------------------   92

rat     YYREDLSPFIIQMAAEYVQRGFPEVSFQPT
mouse   ------------------------------  122

rat     SVVVVTWESMAPYGGPSGSLVEEGKRNTFQ
mouse   ---------V-------S-PA---------  152

rat     AVLASSNSSSYAIFLYPDDGLQFFTTFSKK
mouse   ----------------E-------------  182

rat     DENQVPAVVGFSKGLEGFLWKSNGAYNIFA
mouse   --S------------V--------------  212

rat     NDRESIENLAKSSNAGHQGVWVFEIGSPAT
mouse   ------------------------------  242

rat     AKGVVPADVNLDVDDDGADYEDEDYDLQTS
mouse   -----S------L--------------V--  272

rat     HLGLEDVATQPFPSHSPRRGYPDPHNVPRT
mouse   ---------*-S-----------------I  301

rat     LAPSYEATERPHGIPTERTKSFQLPVERFP
mouse   -S-G-------R-V-----R-----A----  331

rat     QKHPQVIDVDEVEETGVVFSYNTGSQQTCA
mouse   -H----------------------------  361

rat     NNRHQCSVHAECRDYATGFCCRCV
mouse   ------------------------   385
```

Figure 4. Comparison of rat and mouse entactin sequences. The deduced amino acid sequence of nucleotides 1–972 of the rat λ1E cDNA (15) are compared with the sequence of residues 63–385 of mouse entactin. Dashes indicate residues that are identical in both species, and an asterisk marks an extra amino acid present in rat entactin. The rat sequence includes changes found after publication of the original sequence.

Possible Cell- and Ca^{2+}-binding Properties of Entactin

Inspection of the entactin sequence revealed the presence of the tripeptide RGD (residues 672–674), located in a cysteine-rich domain of the molecule (see below). The RGD sequence is the cell recognition site of a number of adhesive proteins (45), which led us to investigate whether this sequence in entactin possessed cell-binding activity. Cell attachment assays were performed using the synthetic undecapeptide SIGFRGDGQTC, corresponding to residues 668–678. When coupled to BSA-coated petri dishes the peptide promoted the attachment of MMT cells in a dose-dependent manner, and 90–100% attachment was obtained at concentrations >5 μg per plate (Fig. 5). The entactin sequence also contains two potential Ca^{2+}-binding sites, DLELEAGDDVVS (residues 15–26) and DVNLDLDDDGAD (residues 250–261). The first, third, fifth, ninth, and twelfth positions have oxygen-containing residues that can coordinate with Ca^{2+} (34). Binding of $^{45}Ca^{2+}$ to nitrocellulose-blotted entactin has been observed (unpublished observations). Secondary structure analysis predicts that the Ca^{2+}-binding sites are not in the EF-hand conformation typical of the Ca^{2+}-binding sites of intracellular Ca^{2+}-binding proteins (34), but nonEF-hand Ca^{2+}-binding sites are found in extracellular proteins such as thrombospondin (36).

Cysteine-rich Repeats in Entactin

Mature entactin contains 48 cysteine residues located mainly in three clusters, and most of the cysteines are organized into an EGF-like repeat pattern of ~40 residues, with six cysteines in conserved positions in each repeat (Fig. 6 *A*). The sequence of 12 other positions in the repeat unit is also conserved, and a consensus sequence can be derived. The six cysteines in each repeat probably form three disulfide-bonded loops, as in EGF (8). Single copies of the EGF-type repeat occur between residues 358–395 and at the COOH terminus, and four copies, one of which contains the RGD tripeptide, occur in tandem between residues 640–816. Immediately after there is a stretch of 75 residues (817–891) containing six cysteines that do not conform to the EGF pattern. A computer search of the GenBank and Protein

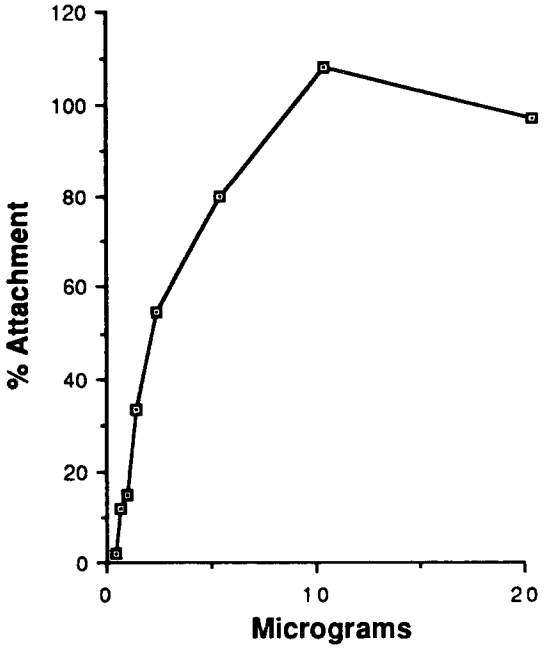

Figure 5. Attachment of MMT cells to a synthetic peptide containing the RGD sequence of entactin. MMT cells were plated on petri dishes coated with 0.2–20 μg of the synthetic peptide SIGFRGDGQTC (residues 668–678), and percent attachment was measured as described in Materials and Methods. Each point is the average of duplicate dishes. The data shown are from a representative experiment. Greater than 100% attachment is sometimes observed, due to experimental variability.

Identification Resource databases revealed that this segment is similar to a cysteine-rich repeat motif found in thyroglobulin, the precursor of thyroid hormone (Fig. 6 *B*). Residues 844–891 of entactin are 51% identical to residues 29–75 of bovine thyroglobulin (38).

Homology to the LDL Receptor and the EGF Precursor

In addition to containing cysteine-rich repeats similar to those found in EGF, a region near the COOH terminus of entactin shows homology to a cysteine-poor segment of both the EGF precursor (23) and the LDL receptor (55). Over a stretch of 192 amino acids (residues 953–1,144) entactin is 31% identical to residues 523–714 of the mouse EGF precursor and 32% identical to residues 411–608 of the human LDL receptor (Fig. 7). There are also sequence similarities in the EGF-type repeat units after this segment in all three proteins (not shown). Besides the sequence homologies in this region, the three polypeptides display some similarities in the organization of their EGF-type repeats (Fig. 8).

Model for the Structure of Entactin

Electron microscopy shows that entactin has an asymmetric dumbbell shape, consisting of two globules of 38 and 85 kD, separated by a 17-nm stalk of 27 kD (40). On the basis of this and analysis of the amino acid sequence, one can predict the existence of at least three domains in entactin, diagramed in Fig. 9. The estimated positions of the major proteolytic fragments in the intact molecular are also indicated. Domain I (residues 1–639, 70 kD) probably corresponds to the larger globular domain, and it contains one of the EGF-type repeats, the Ca^{2+}-binding sites, and the tyrosine sulfation site. Upon further structural and functional analysis, this domain may be divided into several subdomains. The sequence surrounding the NH_2 terminus of the 100-kD proteolytic fragment is rich in prolines and charged residues, and this segment may be more exposed and accessible to proteases. The cysteine-rich domain II (residues 640–889, 28 kD) contains four EGF-like repeats, one thyroglobulin-like repeat, and the putative cell-binding site, and it forms a disulfide-bonded stalk connecting the globular domains. The smaller globule, domain III (residues 890–1,217, 36 kD) has the EGF precursor/LDL receptor-homologous region, and one EGF-type

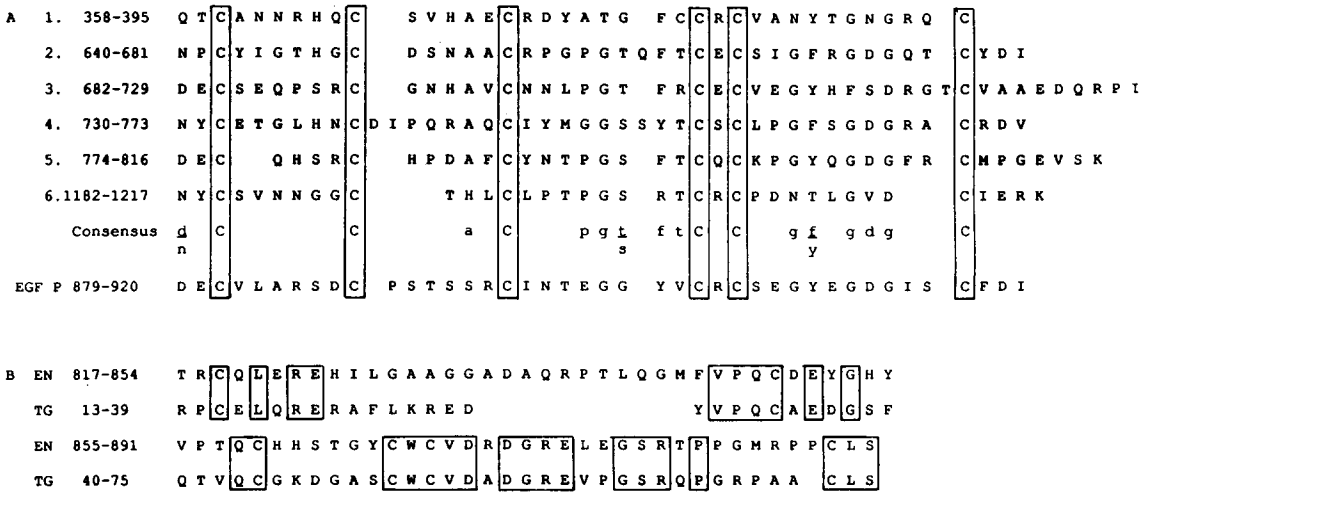

Figure 6. Cysteine-rich repeats in entactin. (*A*) EGF-type repeats. The sequences of the six EGF-type repeats were aligned by inspection, and the six cysteine residues in each repeat are boxed. Below is a consensus sequence that includes residues found in the same position in at least three of the repeats. The sequence of one of the EGF-like repeats in the mouse EGF precursor (23) is shown for comparison. (*B*) Thyroglobulin-like repeat. Residues 817–891 of entactin were aligned with residues 13–79 of bovine thyroglobulin (*TG*) (38). Residues that are identical in both proteins are boxed.

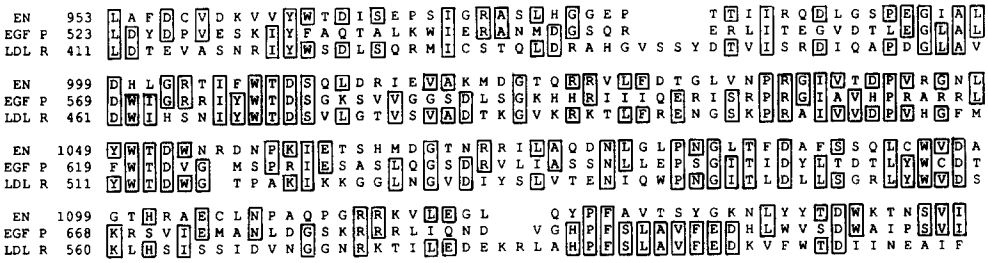

Figure 7. Homology of entactin to the EGF precursor and the LDL receptor. Residues 953–1,144 of entactin were aligned with residues 523–714 of the mouse EGF precursor (23) and residues 411–608 of the human LDL receptor (55) by inspection, after the alignment of the latter two sequences by Yamamoto et al. (55), by introducing only one additional gap. Identical residues are boxed. This segment appears to contain four copies of a repeat unit with the consensus sequence I Y W T D X_{12} G X_2 R X $_{10}$ P X G I X_2 D X_5.

repeat. The disulfide-bonded regions appear to confer protease resistance, as the NH_2 termini of the 80 and 40 kD fragments occur just before two of the cysteine-repeat units.

Discussion

The complete primary structure of mouse entactin has been determined from cDNA sequencing, and many of the features predicted from the deduced amino acid sequence are consistent with data obtained from other experimental approaches. The results presented in this report suggest new properties for entactin, such as cell- and Ca^{2+}-binding, and reveal the multidomain organization of the molecule. The availability of the entactin sequence should stimulate further investigations on its contribution to basement membrane function and assembly.

Entactin appears to be a mosaic protein that may have evolved by "exon-shuffling" and acquired segments from other genes. The cysteine-rich EGF-type repeats found in entactin occur in a wide variety of extracellular proteins such as growth factors, receptors, developmental gene products, extracellular matrix proteins, and proteins of the coagulation, fibrinolytic, and complement systems (5, 13, 44). The B1 and B2 chains of laminin possess cysteine-rich repeats that show some similarity to EGF, but they contain eight cysteines per repeat instead of the usual six (14, 47). The EGF-like units are believed to be involved in receptor-ligand interactions; both EGF and the EGF-like domain of urokinase-type plasminogen activator bind to cell surface receptors (2, 8). One of the EGF-type repeats in entactin may be involved in cell-binding and the functions of the other five remain to be determined. Entactin also contains one copy of a cysteine-repeat motif that occurs 10 times in each 330-kD thyroglobulin monomer (38); the significance of this is not known.

The COOH-terminal globular domain of entactin shows homology to a region found in both the LDL receptor and the EGF precursor, indicating that these three functionally dissimilar proteins share a common ancestral precursor. In vitro mutagenesis was used to delete this domain in the LDL receptor, along with the three flanking EGF-like repeats, and the mutant receptor was unable to bind LDL, to recycle after binding β-very low density lipoprotein (β-VLDL), and to release β-VLDL at acid pH (12). Since the ligand-binding site of the LDL receptor is located in the NH_2 terminus (see Fig. 8), one function of this domain in entactin may be to modulate the binding properties of other segments of the molecule.

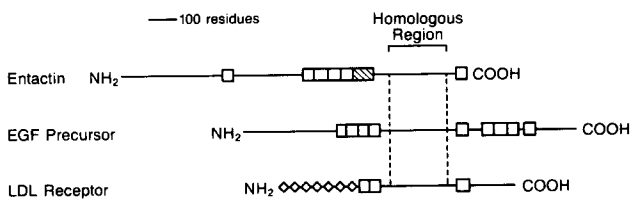

Figure 8. Organization of the cysteine-rich repeats in entactin, the EGF precursor, and the LDL receptor. Open boxes represent the EGF-type repeats. The hatched box represents the thyroglobulin-type repeat found in entactin. The diamonds represent the ligand-binding cysteine-rich repeats in the LDL receptor; these have a motif that is also found in complement component C9 (13).

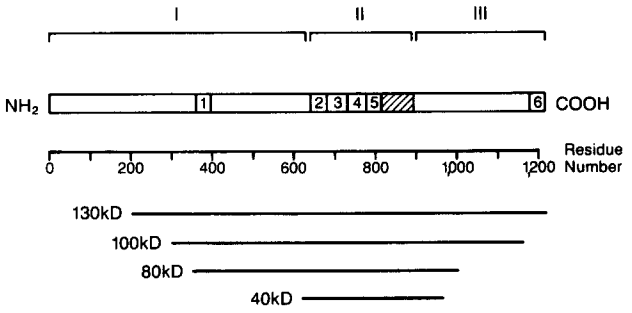

Figure 9. Structural domains in entactin. The numbered boxes represent the six EGF-like repeats (Fig. 6) and the hatched box represents the thyroglobulin-like repeat. Domains I and III probably comprise the large and small globules, respectively, seen in the electron microscope (40) and domain II forms the stalk connecting them. The positions of the 130-, 100-, 80-, and 40-kD proteolytic fragments of nidogen are shown. The precise NH_2 termini are known from protein sequencing (41); the COOH termini were estimated from the sizes of the fragments as determined by SDS-PAGE.

A role for entactin in cell–extracellular matrix interactions is suggested by the discovery of the RGD sequence in one of the EGF-type repeats, and we have demonstrated that a synthetic peptide derived from the RGD site of entactin has cell-binding activity. Although it is possible that cells may attach to the peptide via receptors for other RGD-containing proteins such as fibronectin and vitronectin, we have found that entactin itself promotes cell adhesion, and attachment to entactin is inhibited by RGD-containing peptides (Chakravarti, S., and A. E. Chung, manuscript in preparation). In another study, anti-entactin antiserum partially inhibited the attachment of epidermal cells to the M1536-B3 matrix (1). Laminin has two distinct cell-binding sites apparently recognized by different receptors, one located in a proteolytic fragment consisting of the intersection of the three short arms of the cross (fragment 1), and another located in a fragment derived from the long arm (fragment 8) (3, 21). The fragment 1 cell-binding site has been mapped to a sequence located in one of the cysteine-rich repeats in the B1 chain (22). Binding of entactin to laminin may provide a third cell-binding site in the complex, or it may mask the fragment 1 site; cells possessing only fragment 1–specific receptors did not attach to the laminin–entactin complex (3). Adhesion of cells to RGD-containing proteins is mediated by a class of heterodimeric cell-surface receptors known as integrins (45), and an entactin-specific receptor might possibly be a member of the integrin family.

The laminin–entactin complex has been reported to bind 16 Ca^{2+} ions (39), and our results indicate that at least two of the Ca^{2+}-binding sites may be present in entactin. In addition to the calmodulin-type Ca^{2+}-binding loops identified in the entactin sequence, other possible sites for Ca^{2+}-binding are the EGF-type repeats. Several of the blood coagulation proteins have EGF-like domains containing β-hydroxylated aspartate or asparagine residues that have been correlated with Ca^{2+}-binding (44). The third and fifth EGF-type repeats in entactin fall into the type C group of EGF-like sequences described by Rees et al. (44), as they contain asparagine residues (residues 699 and 789, respectively) located in a postulated consensus site for β-hydroxylation. Ca^{2+} promotes the self-aggregation of laminin (39, 56), and the laminin–entactin complex can be efficiently extracted from basement membranes using EDTA-containing buffers (40). Another Ca^{2+}-binding protein, SPARC/osteonectin/BM-40, is a component of basement membranes (19). These observations suggest that the assembly and stabilization of basement membranes involve Ca^{2+}-dependent interactions between the various protein constituents.

Knowledge of the entactin sequence will facilitate attempts to map the sites on the molecule involved in binding to other extracellular matrix components. By electron microscopy the laminin-binding site of entactin was localized to one of the globular domains, but it was not possible to determine which one (40). The 100-kD fragment of entactin, but not the 80-kD, was found to possess binding activity for laminin, fibronectin, and type IV collagen (18). Fig. 9 shows that the 100-kD fragment loses mass at both ends to generate the 80-kD fragment, so the binding site cannot be determined more accurately. By analogy with the LDL receptor, however, the ligand binding activity of entactin may be located in the NH_2-terminal domain. During the purification of entactin, exposure to denaturants causes a loss of affinity for laminin (18, 40), and studies on the interaction of entactin with laminin and other matrix molecules will require entactin in its native conformation.

Although laminin and entactin are present as an equimolar complex in basement membranes (17), their synthesis is not coordinately regulated (11, 15, 17, 54). Moreover, the level of entactin protein does not always correlate with the level of its mRNA; during the retinoic acid induced differentiation of F9 cells the amount of entactin increases modestly (7, 11, 17) while entactin mRNA levels increase to a much greater extent (15). Isolation of a nearly full-length entactin cDNA is the first step in analyzing the structure of its gene and identifying the transcriptional regulatory elements. The contribution of posttranscriptional, translational, and posttranslational mechanisms to the control of entactin gene expression will also need to be explored.

We thank Drs. Barry Carlin and John Merlie for providing us with the rat λ1E clone, Dr. Ru Chih Huang for facilitating the database searches, Dr. Ming F. Tam for synthesizing the entactin peptide, Dr. John Hempel for performing the amino acid sequencing, Dr. Jane Vergnes for help with the cDNA sequencing, Ms. Kathy Hoffman for typing the manuscript, and Mr. Bill Johnston for preparation of the figures.

This research was supported by National Institutes of Health grants GM25690 and CA21246.

Received for publication 9 June 1988, and in revised form 30 August 1988.

References

1. Alstadt, S. P., P. A. Hebda, A. E. Chung, and W. H. Eaglstein. 1987. The effect of basement membrane entactin on epidermal cell attachment and growth. *J. Invest. Dermatol.* 88:55–59.
2. Appella, E., E. A. Robinson, S. J. Ullrich, M. P. Stoppelli, A. Corti, G. Cassani, and F. Blasi. 1987. The receptor-binding sequence of urokinase. A biological function for the growth factor module of proteases. *J. Biol. Chem.* 262:4437–4440.
3. Aumailley, M., V. Nurcombe, D. Edgar, M. Paulsson, and R. Timpl. 1987. The cellular interactions of laminin fragments. Cell adhesion correlates with two fragment-specific high affinity binding sites. *J. Biol. Chem.* 262:11532–11538.
4. Bender, B. L., R. Jaffe, B. Carlin, and A. E. Chung. 1981. Immunolocalization of entactin, a sulfated basement membrane component, in rodent tissues, and comparison with GP-2 (laminin). *Am. J. Pathol.* 103:419–426.
5. Bender, W. 1985. Homeotic gene products as growth factors. *Cell.* 43:559–560.
6. Carlin, B., R. Jaffe, B. Bender, and A. E. Chung. 1981. Entactin, a novel basal lamina-associated glycoprotein. *J. Biol. Chem.* 256:5209–5214.
7. Carlin, B. E., M. E. Durkin, B. Bender, R. Jaffe, and A. E. Chung. 1983. Synthesis of laminin and entactin by F9 cells induced with retinoic acid and dibutyryl cyclic AMP. *J. Biol. Chem.* 258:7729–7737.
8. Carpenter, G., and S. Cohen. 1979. Epidermal growth factor. *Annu. Rev. Biochem.* 48:193–216.
9. Chen, E.-Y., and P. Seeburg. 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA (NY).* 4:165–170.
10. Chung, A. E., R. Jaffe, I. L. Freeman, J.-P. Vergnes, J. E. Braginski, and B. Carlin. 1979. Properties of a basement membrane-related glycoprotein synthesized in culture by a mouse embryonal carcinoma-derived cell line. *Cell.* 16:277–287.
11. Cooper, A. R., A. Taylor, and B. L. M. Hogan. 1983. Changes in the rate of laminin and entactin synthesis in F9 embryonal carcinoma cells treated with retinoic acid and cyclic AMP. *Dev. Biol.* 99:510–516.
12. Davis, C. G., J. L. Goldstein, T. C. Sudhof, R. G. W. Anderson, D. W. Russell, and M. S. Brown. 1987. Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. *Nature (Lond.).* 326:760–765.
13. Doolittle, R. F. 1985. The genealogy of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* 10:233–237.
14. Durkin, M. E., B. B. Bartos, S.-H. Liu, S. L. Phillips, and A. E. Chung. 1988. The primary structure of the mouse laminin B2 chain and comparison with laminin B1. *Biochemistry.* 27:5198–5204.
15. Durkin, M. E., B. E. Carlin, J. Vergnes, B. Bartos, J. Merlie, and A. E. Chung. 1987. Carboxyl-terminal sequence of entactin deduced from a cDNA clone. *Proc. Natl. Acad. Sci. USA.* 84:1570–1574.
16. Durkin, M. E., S. L. Phillips, and A. E. Chung. 1986. Control of laminin

synthesis during differentiation of F9 embryonal carcinoma cells. A study using cDNA clones complementary to the mRNA species for the A, B1, and B2 subunits. *Differentiation.* 32:260–266.
17. Dziadek, M., and R. Timpl. 1985. Expression of nidogen and laminin in basement membranes during mouse embryogenesis and in teratocarcinoma cells. *Dev. Biol.* 111:372–382.
18. Dziadek, M., M. Paulsson, and R. Timpl. 1985. Identification and interaction repertoire of the large forms of the basement membrane protein nidogen. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:2513–2518.
19. Engel, J., W. Taylor, M. Paulsson, H. Sage, and B. Hogan. 1987. Calcium binding domains and calcium-induced conformational transition of SPARC/BM-40/osteonectin, an extracellular glycoprotein expressed in mineralized and nonmineralized tissues. *Biochemistry.* 26:6958–6965.
20. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 120:97–120.
21. Goodman, S. L., R. Deutzmann, and K. von der Mark. 1987. Two distinct cell-binding domains in laminin can independently promote non-neuronal cell adhesion and spreading. *J. Cell Biol.* 105:589–598.
22. Graf, J., Y. Iwamoto, M. Sasaki, G. R. Martin, H. K. Kleinman, F. A. Robey, and Y. Yamada. 1987. Identification of an amino acid sequence in laminin mediating cell attachment, chemotaxis, and receptor binding. *Cell.* 48:989–996.
23. Gray, A., T. J. Dull, and A. Ullrich. 1983. Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor. *Nature (Lond.).* 303:722–725.
24. Hager, D. A., and R. R. Burgess. 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: results from sigma subnunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* 109:76–86.
25. Hanover, J. A., and W. J. Lennarz. 1981. Transmembrane assembly of membrane and secretory glycoproteins. *Arch. Biochem. Biophys.* 211: 1–19.
26. Hay, E. D. 1985. Matrix-cytoskeletal interactions in the developing eye. *J. Cell. Biochem.* 27:143–156.
27. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene (Amst.)* 28:351–359.
28. Hogan, B. L. M., A. Taylor, and A. R. Cooper. 1982. Murine parietal endoderm cells synthesize heparan sulphate and 170 K and 145 K sulphated glycoproteins as components of Reichert's membrane. *Dev. Biol.* 90:210–214.
29. Hogan, B. L. M., A. Taylor, M. Kurkinen, and J. R. Couchman. 1982. Synthesis and localization of two sulphated glycoproteins associated with basement membranes and the extracellular matrix. *J. Cell Biol.* 95:197–204.
30. Hunkapiller, M. W., E. Lujan, F. Ostrander, and L. E. Hood. 1983. Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91:227–236.
31. Huttner, W. B. 1987. Protein tyrosine sulfation. *Trends Biochem. Sci.* 12:361–363.
32. Kleinman, H. K., M. L. McGarvey,. J. R. Hassell, V. L. Star, F. B. Cannon, G. W. Laurie, and G. R. Martin. 1986. Basement membrane complexes with biological activity. *Biochemistry.* 25:312–318.
33. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15:8125–8148.
34. Kretsinger, R. H. 1980. Structure and evolution of calcium-modulated proteins. *CRC Crit. Rev. Biochem.* 8:119–174.
35. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680–685.
36. Lawler, J., and R. O. Hynes. 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. *J. Cell Biol.* 103:1635–1648.
37. Martinez-Hernandez, A., and A. E. Chung. 1984. The ultrastructural localization of two basement membrane components: entactin and laminin in rat tissues. *J. Histochem. Cytochem.* 32:289–298.
38. Mercken, L., M.-J. Simons, S. Swillens, M. Massaer, and G. Vassart. 1985. Primary structure of bovine thyroglobulin deduced from the sequence of its 8,431-base complementary DNA. *Nature (Lond.).* 316:647–651.
39. Paulsson, M. 1988. The role of Ca^{2+} binding in the self-aggregation of laminin-nidogen complexes. *J. Biol. Chem.* 263:5425–5430.
40. Paulsson, M., M. Aumailley, R. Deutzmann, R. Timpl, K. Beck, and J. Engel. 1987. Laminin-nidogen complex. Extraction with chelating agents and structural characterization. *Eur. J. Biochem.* 166:11–19.
41. Paulsson, M., R. Deutzmann, M. Dziadek, H. Nowack, R. Timpl, S. Weber, and J. Engel. 1986. Purification and structural characterization of intact and fragmented nidogen obtained from a tumor basement membrane. *Eur. J. Biochem.* 156:467–478.
42. Paulsson, M., M. Dziadek, C. Suchanek, W. B. Huttner, and R. Timpl. 1985. Nature of sulphated macromolecules in mouse Reichert's membrane. Evidence for tyrosine *O*-sulphate in basement membrane proteins. *Biochem. J.* 231:571–579.
43. Pytela, R., M. D. Pierschbacher, S. Argraves, S. Suzuki, and E. Ruoslahti. 1987. Arginine-glycine-aspartic acid adhesion receptors. *Methods Enzymol.* 144:475–489.
44. Rees, D. J. G., I. M. Jones, P. A. Handford, S. J. Walter, M. P. Esnouf, K. J. Smith, and G. G. Brownlee. 1988. The role of β-hydroxyaspartate and adjacent carboxylate residues in the first EGF domain of human factor IX. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2053–2061.
45. Ruoslahti, E., and M. D. Pierschbacher. 1987. New perspectives in cell adhesion: RGD and integrins. *Science (Wash. DC).* 238:491–497.
46. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.
47. Sasaki, M., S. Kato, K. Kohno, G. R. Martin, and Y. Yamada. 1987. Sequence of the cDNA encoding the laminin B1 chain reveals a multidomain protein containing cysteine-rich repeats. *Proc. Natl. Acad. Sci. USA.* 84:935–939.
48. Timpl, R., and M. Dziadek. 1986. Structure, development, and molecular pathology of basement membranes. *Int. Rev. Exp. Pathol.* 29:1–112.
49. Timpl, R., M. Dziadek, S. Fujiwara, H. Nowack, and G. Wick. 1983. Nidogen: a new, self-aggregating basement membrane protein. *Eur. J. Biochem.* 137:455–465.
50. Von Heijne, G. 1985. Signal sequences. The limits of variation. *J. Mol. Biol.* 184:99–105.
51. Vracko, R. 1982. The role of basal lamina in maintenance of orderly tissue structure. *In* New Trends in Basement Membrane Research. K. Kuehn, H.-H. Schoene, and R. Timpl, editors. Raven Press, NY. 1–7.
52. Warburton, M. J., P. Monaghan, S. A. Ferns, P. S. Rudland, N. Perusinghe, and A. E. Chung. 1984. Distribution of entactin in the basement membrane of the rat mammary gland. Evidence for a non-epithelial origin. *Exp. Cell. Res.* 152:240–254.
53. Wickens, M., and P. Stephenson. 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutations prevent mRNA 3' end formation. *Science (Wash. DC).* 226:1045–1051.
54. Wu, T.-C., Y.-J. Wan, A. E. Chung, and I. Damjanov. 1983. Immunohistochemical localization of entactin and laminin in mouse embryos and fetuses. *Dev. Biol.* 100:496–505.
55. Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell. 1984. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. *Cell.* 39:27–38.
56. Yurchenco, P. D., E. C. Tsilibary, A. S. Charonis, and H. Furthmayr. 1985. Laminin polymerization *in vitro*. Evidence for a two-step assembly with domain specificity. *J. Biol. Chem.* 260:7636–7644.