Carboxy-terminal sequence of entactin deduced from a cDNA clone
(basement membrane)

MARIAN E. DURKIN*, BARRY E. CARLIN†, JANE VERNIGES*, BARBARA BARTOS*, JOHN MERLIE†, AND ALBERT E. CHUNG‡

*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260; †Department of Pharmacology, Washington University School of Medicine, St. Louis, MO 63110

Communicated by Klaus Hofmann, November 7, 1986 (received for review July 23, 1986)

ABSTRACT Entactin is a widely distributed basement membrane sulfated glycoprotein of ~150 kDa. The entactin gene is expressed early in mouse embryogenesis. Two cDNA clones complementary to rat entactin mRNA were isolated by antibody screening of an oligo(dT)-primed cDNA library constructed in the Agt11 expression vector. One of the clones, AIE, was subcloned into plasmid pBR322 and further characterized. The clone contained sequences complementary to an mRNA species 6 kilobases in length. This mRNA was translated in rabbit reticulocyte lysates to yield a polypeptide of 143 kDa that was precipitated with anti-entactin antiserum. The cDNA insert, 1328 base pairs long, was sequenced and found to contain an open reading frame of 729 base pairs that coded for 243 amino acids at the carboxy terminus of entactin. Analysis of the peptide revealed no extended α-helical or β-sheet secondary structures. Radial-labeled probes prepared by nick-translation of pAIE were used to monitor the steady-state levels of entactin mRNA in F9 embryonal carcinoma cells that were induced to differentiate by exposure to retinoic acid and dibutyryl cyclic AMP. The increase in steady-state levels of entactin mRNA lagged behind the increase in mRNA for the B2 chain of laminin, suggesting that laminin and entactin are independently regulated rather than coordinately regulated.

The sulfated glycoprotein entactin is widely distributed in basement membranes (1-4). It is synthesized in culture by mouse embryonal carcinoma-derived endodermal cells (1, 5), mouse parietal yolk sac cells (6), rat yolk sac carcinoma (7), and rat mammary fibroblasts (8). A similar, if not identical, protein, nidogen, has been isolated from mouse EHS tumors (9). The molecule, which has an apparent molecular weight of ~150,000, is usually found in tight association with laminin (1, 10). The dissociation of the two proteins can only be effected with detergents or other potent denaturing agents (10, 11). Immunological techniques have demonstrated the colocalization of laminin and entactin in Reichert's membrane (a thick extraembryonic basement membrane that surrounds rodent embryos) and in basement membranes of the kidney, gut, vascular tissues, eye, and muscle (3, 12); however, there are subtle differences in their quantitative and qualitative distribution. The recently described matrismal gel that supports cell differentiation and growth consists of a complex that includes laminin, entactin, type IV collagen, and heparan sulfate (13). Furthermore, treatment of F9 embryonal carcinoma cells with retinoic acid and N6,0'-dibutyryl cyclic AMP (Bt2cAMP) resulted in the enhanced synthesis and secretion of laminin, type IV collagen, and entactin (10, 14, 15). The intimate association of entactin with the basement membrane and its components argues that it plays an important, although as yet undefined, role in the integrity and function of basement membranes.

The biochemical characterization of entactin has been hampered by the unavailability of substantial quantities of purified material and its susceptibility to proteolytic degradation. We have used recombinant DNA techniques to isolate a cDNA clone for rat entactin. The carboxy terminus of entactin has been deduced from the DNA sequence of the clone. In addition, the steady-state levels of entactin mRNA have been compared with those of laminin mRNA during the induction of F9 cells with retinoic acid and Bt2cAMP.

MATERIALS AND METHODS

Cell Culture. Skeletal muscle cells were dissociated from the forelimbs of 20-day embryonic rats and cultured as described (16). The mouse parietal endoderm line M1536-B3 and the mouse embryonal carcinoma line F9 were cultured as described (17).

RNA Purification. Total RNA was isolated from 10-day cultures of rat skeletal muscle cells by a modification of the method of Chirgwin et al. (18). After three cycles of extraction with 7.5 M guanidine hydrochloride and precipitation with ethanol, the RNA was dissolved in 3.75 M guanidine hydrochloride and extracted with an equal volume of 1:1 phenol/chloroform. Poly(A)+ RNA was selected by passage over oligo(dT)-cellulose as described by Sebbane et al. (19). RNA was purified from other cell lines as described (20).

cDNA Synthesis and Library Construction. Ten micrograms of poly(A)+ RNA from embryonic rat skeletal muscle cultures was used to prepare cDNA in a reaction employing cloned Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories) according to the supplier's protocol. The second strand was synthesized exactly as described by Gubler and Hoffman (21), except that the cDNA was treated with nuclease S1 (22) to increase the efficiency of linker addition. After the ends were made flush by means of DNA polymerase 1 Kl Entrophy fragment and synthetic EcoRI linkers were added, the cDNA was fractionated on a Bio-Gel A-50m column. Molecules longer than 1.5 kilobases (kb) were pooled and ligated to Agt11 arms and the resulting mixture was packaged using Gigapack (Vector Cloning Systems, San Diego, CA) according to the supplier's protocol.

Immunoscreening of the Agt11 Library. Two hundred thousand recombinant plaques were grown on Escherichia coli Y1088 at a density of 20,000 plaques per 90-mm Petri dish and immunoscreened with entactin antiserum according to the method of Young and Davis (23, 24), with the following modifications. The filters were incubated overnight at room temperature in a 1:200 dilution of entactin antiserum in Tris-buffered saline (TBS: 0.15 M NaCl/0.05 M Tris Cl, pH 7.4) containing 0.5% Tween 20 (TBS/Tween). Antibody binding was detected by soaking for 1 hr in a 1:3000 dilution of alkaline phosphatase-conjugated goat anti-rabbit IgG (Boehringer).

Abbreviation: Bt2cAMP, N6,0'-dibutyryl cyclic AMP.
† To whom reprint requests should be addressed.

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

1570
ringer Mannheim) in TBS/Tween, and the color was developed for 5-10 min at room temperature in 200 mM Tris CI (pH 8.0) containing fast red TR salts (0.6 mg/ml) and naphthol AS-MX phosphate (0.2 mg/ml). Positive plaques were purified by three cycles of screening with entactin antiserum.

**Fusion Protein Analysis.** *E. coli* Y1089 cells were lysogenized with either λI1 or wild-type λgt11, and freeze-thaw lysates were prepared following induction of β-galactosidase fusion-protein synthesis with isopropyl β-D-thiogalactoside (25). Lysate proteins were separated in 8% polyacrylamide gels (26) and electroblotted onto nitrocellulose (27). The blots were blocked in 5% non-fat dry milk in phosphate-buffered saline (PBS: 137 mM NaCl/2.7 mM KCl/8.1 mM Na2HPO4/1.5 mM KH2PO4) and incubated for 1 hr in antiserum diluted into PBS. Bound antibodies were visualized using biotinylated horseradish peroxidase (Vectastain ABC Kit, Vector Laboratories, Burlingame, CA) and H2O2 and 4-chloro-1-naphthol (28).

**In Vitro Translation of Hybridization-Selected RNA.** Fifteen micrograms of linearized pα1E DNA was denatured and bound to a 0.8-cm square of DPT paper (Schleicher & Schuell), and hybridization-selection was performed as described (20). The selected RNA was translated in a rabbit reticulocyte lysate system (Bethesda Research Laboratories), and immunoprecipitation of the radiolabeled translation products was performed with anti-entactin antiserum (20). The immunoprecipitates were analyzed by NaDodSO4/PAGE and fluorography (30).

**RNA Dot Blot Analysis.** Hybridization of RNA dot blots to nick-translated cDNA probes (31) was performed as described (20).

**Sequencing of DNA.** DNA was sequenced by the chemical-cleavage method of Maxam and Gilbert (32) or by a modification of the dideoxynucleotide chain-termination method of Sanger et al. (33).

### RESULTS

**Identification of Entactin cDNA Clones.** Two cDNA clones, λI1 and λI14E, were identified by screening the rat muscle cDNA library in the expression vector λgt11 with rabbit anti-entactin serum, and λI1 was selected for further characterization. The β-galactosidase fusion protein produced by isopropyl β-D-thiogalactoside-induced lysogens was studied by immunoblotting. Staining of cell lysate proteins with anti-β-galactosidase monoclonal antibody (Fig. 1A) revealed a single band at 116 kDa in cells lysogenized with wild-type λgt11 and a series of bands in the extract of cells lysogenized with λI1. The largest of these bands was at ~145 kDa. Positive bands at lower molecular mass, presumably degradation products of the fusion protein, were also present. Anti-entactin serum (Fig. 1B) stained only a single band, at approximately 145 kDa, in the lane containing the fusion protein; there were no visible bands in the control lane. These results indicated the presence of entactin sequences in λI1.

The cDNA insert of λI1 was subcloned into pBR322 to yield the new plasmid pα1E. Immobilized pα1E DNA was hybridized to total RNA from M1536-B3 cells, and the hybridization-selected RNA was translated in vitro. Entactin antiserum precipitated a 143-kDa polypeptide from the translation products of M1536-B3 poly(A)+ RNA (Fig. 2), which is slightly smaller than mature entactin (150 kDa) found in the extracellular matrix of M1536-B3 cells. Antiserum raised against mouse entactin also precipitated in vitro-synthesized rat entactin, indicating that considerable homology exists between rat and mouse entactins. The translation product of RNA hybridization-selected by pα1E was indistinguishable from entactin synthesized from M1536-B3 poly(A)+ RNA. No entactin was synthesized from RNA selected by pBR322 alone or by p16, a plasmid carrying a laminin B2 cDNA insert

![Fig. 1. Immunoblot analysis of λI1 fusion protein.](image)

**Expression of the Entactin Gene.** Entactin is synthesized constitutively by mouse parietal endoderm cells, and in addition, F9 embryonal carcinoma cells can be induced by treatment with retinoic acid and Bt2cAMP to synthesize substantial quantities of the protein. RNA blot analysis

![Fig. 2. In vitro translation of hybridization-selected RNA.](image)
demonstrated that this biosynthetic capacity correlated with the presence of entactin mRNA. Nick-translated 32P-labeled \( \lambda 1E \) insert hybridized to a 6-kb mRNA present in parietal endoderm and differentiated F9 cells but not in embryonal carcinoma cells (data not shown).

The steady-state levels of entactin mRNA were monitored during the course of retinoic acid-induced differentiation of F9 cells. 32P-labeled \( \lambda 1E \) insert was hybridized to dot blots of RNA extracted from cultures of F9 cells at predetermined intervals after the addition of retinoic acid and Bt2cAMP. The results (Fig. 3A) show a dramatic increase in the steady-state level of entactin mRNA between 72 and 96 hr. The steady-state levels of entactin mRNA observed are consistent with the observed increase in entactin production in induced cells (10).

The levels of laminin A-, B1-, and B2-chain RNAs were previously found to increase synchronously during the differentiation of F9 cells (20). A similar dot blot (Fig. 3B) hybridized to a laminin B2 cDNA probe shows that laminin mRNA levels increase \( \sim 24 \) hr before the induction of entactin mRNA. This suggests that the laminin and entactin genes are not coordinately expressed.

**Sequence of \( \lambda 1E \).** A partial restriction map of \( \lambda 1E \) is shown in Fig. 4. The entire nucleotide sequence was determined by a combination of the Maxam–Gilbert and dideoxynucleotide chain-termination procedures, as also noted in Fig. 4. The complete sequence of 1328 nucleotides is shown in Fig. 5. The sequence contains a polyadenylation signal at position 1295. An open reading frame (positions 1–729) encodes 243 amino acids at the carboxy-terminal end of entactin. There is a direct repeat of 21 base pairs containing three mismatches in the nontranslated segment. The first begins at nucleotide 1113 and the second at 1213. These sequences are followed by stretches of pyrimidines.

**Carboxyl-Terminal Amino Acid Sequence Deduced from \( \lambda 1E \).** The amino acid sequence deduced from the oligonucleotide sequence of \( \lambda 1E \) is shown in Fig. 5. There is one potential site for N-glycosylation, at residue 97, that contains the sequence Asn-Ser-Ser. The amino acid composition reveals a predominance of acidic vs. basic amino acids, resulting in a potential negative charge of \( \sim 20 \) at pH 7.0. A

![Fig. 3](image)

**DISCUSSION**

The biochemical characterization of the constituents of basement membranes and their organization and assembly pose a difficult challenge. The advent of recombinant DNA techniques and high-resolution electron microscopy, in combination with immunological and classical biochemical approaches, has rapidly generated new information, especially on the structures of laminin and type IV collagen. The work described here has defined more rigorously the nature of entactin and initiated studies on the regulation of its biosynthesis.

An entactin cDNA clone, \( \lambda 1E \), was isolated by immunoscreening a \( \lambda gt11 \) cDNA library. The fusion protein synthesized by \( \lambda 1E \) lysogens contained amino acid sequences recognized by anti-entactin serum. DNA from the \( \lambda 1E \) insert hybrid-selected RNA that programmed a rabbit reticulocyte lysate to synthesize entactin, confirming that it contained entactin sequences. RNA blot analysis revealed that entactin mRNA from mouse parietal endoderm cells is \( \sim 6 \) kb long. Since \( 4 \) kb is sufficient to code for a polypeptide of 150 kDa, entactin mRNA apparently has 2 kb of untranslated sequences.

The entire 1328-base sequence of the cDNA clone was determined. The open reading frame coded for a polypeptide of molecular weight 26,455. This represents somewhat less than one-fifth of the estimated molecular weight of unglycosylated entactin (\( M_0, 143,000 \)). Examination of the amino acid sequence revealed a single potential N-glycosylation site. Although the number of N-linked oligosaccharide chains in entactin is not known, the amino acid sequence rules out the presence of clusters of such chains at the carboxyl end of the molecule. There are only two cysteine residues in the derived sequence, out of \( \sim 45 \) for the intact molecule as determined by amino acid analysis (unpublished).

![Fig. 4](image)

---

Fig. 5. Nucleotide sequence of pX1E and the deduced carboxyl-terminal amino acid sequence of entactin. Nucleotides are numbered above the sequence and amino acids are numbered at right. Cysteine residues (positions 73 and 243) are underlined and the potential glycosylation site (residues 97–99) is shown in italics. The open reading frame extends from nucleotide 1 to 729. Direct repeats beginning at nucleotides 1113 and 1213 are underlined. Polyadenylation signal (AATAAA) is shown in bold face.

This is reminiscent of the distribution of cysteine residues in the B1 and B2 chains of laminin (35, 38), where most of the cysteine residues are located in the more amino-terminal domains. There is no evidence to suggest that
entactin is linked covalently to other basement membrane components or to itself; thus, it is likely that these cysteine residues form intramolecular disulfide bonds. In contrast to the B1 and B2 chains of laminin, the carboxyl-terminal region of entactin does not have a highly ordered secondary structure and is probably more like the globular NC-1 domain of type IV collagen, although no sequence homology with this protein was observed. The intimate association of the basement membrane glycoproteins to form functional structures argues for the tight control of their biosynthesis and assembly. During mouse embryogenesis, laminin can be detected as early as the eight-cell stage of development, whereas entactin was first observed at the hatched blastocyst stage (39). The induced differentiation of embryonal carcinoma F9 cells by treatment with retinoic acid and Bt-cAMP mimics the early differentiation of mouse embryonic cells to parietal endoderm. The synthesis and secretion of the basement membrane components laminin, type IV collagen, and entactin were markedly increased (10, 14, 15). The steady-state levels of mRNA for the A, B1, and B2 chains of laminin and type IV collagen were shown to increase in parallel with the enhanced biosynthesis of the molecules (20, 29). It was observed in the present studies that the enhanced synthesis of entactin previously reported (10) was accompanied by an increase in the steady-state levels of entactin mRNA. However, the kinetics indicated that the sharp rise in the steady-state level of entactin mRNA occurred between 72 and 96 hr of induction, ∼24 hr later than laminin A, B1, and B2 mRNA levels. These data thus suggest that the biosynthesis of laminin and entactin are regulated by separate mechanisms, as indicated by the in vivo studies on mouse embryos previously mentioned (39).

The isolation of a cDNA clone for entactin has allowed a more specific biochemical definition of this protein. The clone will permit further studies on the entactin gene to proceed at an accelerated pace to yield information on its structure and its programmed regulation during embryogenesis. Furthermore, extension of the primary protein structure deduced for entactin will enable the design of specific antibody probes to explore structure–function relationships.

We thank Dr. Roger Hendrix (Department of Biological Sciences, University of Pittsburgh) for analysis of the secondary structure of the polypeptide chain. This work was supported by grants from the National Institutes of Health to A.E.C. and J.M. and from the Muscular Dystrophy Association and Monsanto to J.M.