

Endo16, a Large Multidomain Protein Found on the Surface and ECM of Endodermal Cells during Sea Urchin Gastrulation, Binds Calcium

MARGARET SOLTYSIK-ESPAÑOLA, DAVID C. KLINZING, KENNETH PFARR, ROBERT D. BURKE,* AND SUSAN G. ERNST¹

*Department of Biology, Tufts University, Medford, Massachusetts 02155; and *Department of Biology, University of Victoria, Victoria, British Columbia, Canada V8W 2Y2*

Accepted June 8, 1994

Endo16 encodes a developmentally regulated protein restricted to cells participating in the formation of the archenteron during sea urchin gastrulation and to the stomach of the pluteus. The 4680-nt coding region of the Endo16 gene has been sequenced from overlapping cDNAs. Sequence analysis revealed that Endo16 is a large multidomain protein starting with a putative signal sequence at its amino terminus which is followed by a cysteine-rich region, two potential heparin-binding regions, an acidic domain of 5 clustered repeats, an RGD cell binding motif, and a group of 12 additional acidic repeats. Immunolocalization by confocal and electron microscopy demonstrate that the Endo16 protein is in the extracellular matrix and associated with the surface of endodermal cells in the mid and hindgut of the archenteron. The two distinct acidic repeat regions are similar to known calcium-binding sequences. A recombinant Endo16 protein containing both putative calcium-binding repeat regions has been shown to bind radioactive calcium. Tryptic digests of gastrula stage protein extracts in the presence and the absence of calcium have established that calcium stabilizes Endo16 protein against proteolytic degradation. © 1994 Academic Press, Inc.

INTRODUCTION

In the sea urchin, gastrulation is initiated at the vegetal pole by ingression of the skeletogenic primary mesenchyme cells into the blastocoel. Then cells in the vegetal region of the hollow blastula elongate, producing a flattened vegetal plate. The vegetal plate invaginates in three distinct phases forming the archenteron, which extends across the blastocoel and subsequently differentiates into the foregut, midgut, and hindgut. This progression from a flattened sheet of cells to a short rudimentary invagination that subsequently elongates into a tube and differentiates into the digestive tract is accompanied by major cell rearrangements and numerous cell-cell and cell-matrix interactions. The morphologi-

cal changes and the behavioral activities of cells during sea urchin gastrulation have been carefully documented (reviewed: Ettensohn and Ingersoll, 1992; McClay *et al.*, 1992), yet very little is known about the molecules which orchestrate this process.

It is clear from other systems that changes in cell recognition and adhesive interactions between cells and between cells and the extracellular matrix (ECM) induce changes in gene expression and cell differentiation, thereby regulating morphogenesis (reviewed: Juliano and Haskill, 1993; Adams and Watt, 1993). Since sea urchin gastrulation is arrested by various inhibitors of ECM synthesis or assembly (Karp and Solursh, 1974; Heifetz and Lennarz, 1979; McCarthy *et al.*, 1987; Butler *et al.*, 1987; Wessel and McClay, 1987; Benson *et al.*, 1991), it follows that molecules regulating cell-cell and cell-matrix adhesive interactions are likely to be instrumental in sea urchin gastrulation, an idea originally proposed by Gustafson and Wolpert in 1967.

In a previous report we described the initial isolation and characterization of a cDNA coding for a gene designated Endo16, since its expression is specific to the endodermal cell lineage (Nocente-McGrath *et al.*, 1989). Sequence analysis of the original cDNA clone, representing less than 20% of the entire coding region of the most abundant Endo16 mRNA, revealed the presence of an RGD tripeptide, demonstrated to be a cell binding site in many extracellular proteins (reviewed: D'Souza *et al.*, 1991). *In situ* hybridization to sectioned and whole mount embryos and indirect immunofluorescence confirmed that Endo16 has a spatial and temporal pattern of expression that are consistent with a role in gastrulation (Nocente-McGrath *et al.*, 1989; Ransick *et al.*, 1993). Therefore, we proposed that the protein product of the Endo16 gene may play a role in cell adhesion during gastrulation in the sea urchin embryo (Nocente-McGrath *et al.*, 1989).

In this study, we report the sequence of overlapping cDNAs coding for the 1560 amino acid coding region of the major Endo16 transcript. Analysis of the sequence

¹ To whom correspondence should be addressed. Fax: (617) 627-3805.

revealed several discrete structural domains and suggests that Endo16 is a large, potentially post-translationally modified protein with several functional domains. In addition, we have used confocal microscopy and immunogold electron microscopy and have localized Endo16 to the surface of endodermal cells and to the ECM. Sequence analysis predicts that Endo16 is a calcium-binding protein and experiments presented here confirm this prediction.

MATERIALS AND METHODS

Animals

Adult *Strongylocentrotus purpuratus* were purchased from Marinus, Inc. (Long Beach, CA). Spawning, fertilization, and culturing of embryos were as previously described (Smith *et al.*, 1974), with the modification that embryos were cultured at 14°C.

Microscopy

Confocal. Gastrula stage (36–38 hr) embryos were settled onto polysine-coated slides, fixed in ice-cold 90% methanol, and rehydrated through a graded methanol series to phosphate-buffered saline (PBS). Embryos were incubated in a 1:100 dilution of affinity-purified anti-Endo16 antibody in 3% bovine serum albumin in PBS, followed by incubation with a 1:100 dilution of fluorescein-conjugated goat anti-rabbit antisera, as previously described (Nocente-McGrath *et al.*, 1991). Stained slides were mounted in FITC-Guard mounting medium (Testog, Inc.). Optical sections were viewed with a Bio-Rad confocal laser scanning microscope at the Integrated Microscopy Resource for Biomedical Research at the University of Wisconsin with the assistance of Dr. Stephan Paddock.

Electron microscopy. Late gastrula stage embryos (48 hr) were prepared by freeze substitution following the methods described by Campbell and Crawford (1991). Propylene glycol (15%) in filtered seawater was used as a cryoprotectant and after dehydration embryos were embedded in LR White resin. Sections were incubated overnight with anti-Endo16, diluted 1/100, rinsed, and incubated with GAR-gold (5 nm), rinsed, and contrasted with uranyl acetate and lead citrate. Controls were similar sections treated with normal rabbit serum. Grain counts were carried out for randomly selected regions of the cytoplasm of endodermal and ectodermal cells.

Sequencing

cDNA library screen. A gastrula stage λ gt11 cDNA library enriched for endodermal and mesodermal sequences (gift of Drs. N. George and F. Wilt) was used as a source of all cDNAs. Random-primed probes were

derived from a previously characterized clone, pCNM4 (Nocente-McGrath *et al.*, 1989), or from appropriate cDNA fragments obtained during subsequent library screenings. Positives were plaque purified, amplified, and phage DNA was isolated using standard molecular biology protocols (Sambrook *et al.*, 1989). cDNA fragments were excised from λ gt11 DNA by *EcoRI* digestion and subcloned into the *EcoRI* site of pGEM 7zf(+) phagemid (Promega).

Sequencing methods. For each cDNA clone, nested deletion subclones were generated with Exonuclease III, using the Erase-a-base system (Promega). Both single-stranded and double-stranded DNA templates were used in dideoxy chain-termination reactions using a Sequenase kit (United States Biochemical). Both DNA strands were sequenced, except for regions where the sense strands of at least two overlapping, independent cDNA clones had identical sequences. To clarify any sequence ambiguities high-temperature sequencing of single-stranded DNA templates was done using *Taq* polymerase and modified deaza nucleotides provided in the *Taq*-traq sequencing system (Promega).

Sequence analysis. Nucleic acid and protein data banks (Genetics Computing Group/University of Wisconsin) were searched against Endo16 using programs WORDSEARCH and FASTA. Any matches found were checked by alignment using SEGMENTS program to determine the validity of the match. The DNA Strider software (Marck, 1988) was used in analyses of nucleotide-derived amino acid sequences.

Tryptic Digestion in the Presence of Divalent Cations or EGTA

Tryptic digestion of embryonic protein extracts including Endo16 were modeled on the procedure of Hyafil *et al.* (1981). Pelleted gastrula stage embryos were resuspended by gentle vortexing in 3–4 vol of NP-40 buffer (1% NP-40, 100 mM Tris, pH 7.4, 150 mM NaCl, 3 mM EDTA) with the protease inhibitors pepstatin (1 μ g/ml) and TPCK (100 μ g/ml). Proteins were extracted on ice for 30 min. Protein concentration was estimated by the BCA method (Smith *et al.*, 1985) using bovin serum albumin as a standard (MicroBCA Protein Assay kit, Pierce). Protein extracts were transferred into tubes to which CaCl_2 , MgCl_2 , MnCl_2 , or EGTA was added to a final concentration of 5 mM. After 10 min incubation in the presence of a divalent cation or EGTA, 50- μ g samples of total embryonic protein were pipetted into tubes and from 0 to 12.5 μ g of trypsin (from GIBCO, stock made to 5 mg/ml in 50 mM Tris, pH 8.8, 10 mM CaCl_2) was added to each sample. Trypsin-treated and control samples (no trypsin added) were incubated on ice for 1 hr. Limited tryptic digests were terminated by addition

of an equal volume of 2× Laemmli sample buffer (100 mM Tris-Cl, pH 6.8, 4% SDS, 20% glycerol, 0.002% bromophenol blue) with 0.1 M dithiothreitol and trypsin inhibitors and were boiled for 2–3 min. Samples were spun in a microcentrifuge to pellet nonsolubilized materials and loaded onto a SDS–polyacrylamide gel. As a control, purified serum albumin was digested in the presence of calcium or EGTA. Serum albumin was added to NP-40 buffer and pipetted into two tubes to which either CaCl₂ or EGTA was added to final a concentration of 5 mM. Samples containing 25 μg of serum albumin each were pipetted into tubes, to which 0 to 50 μg of trypsin was added. Limited tryptic digestion was carried out and then terminated as described for the embryonic protein extracts. Following SDS–polyacrylamide gel electrophoresis, the gel was stained with Coomassie brilliant blue.

Electrophoresis and Western Blotting

Embryonic protein extracts and serum albumin samples were resolved in a discontinuous gel system (Laemmli, 1970). The stacking gel contained 4% acrylamide and 0.14% bisacrylamide and the resolving gel contained either 12% acrylamide and 0.05% bisacrylamide or 6.7% acrylamide and 0.25% bisacrylamide. Proteins were transblotted onto a PVDF membrane (Millipore) according to the method of Towbin *et al.* (1979). PVDF membranes were blocked in 5% dry milk in TBST (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Tween-20), and then incubated with polyclonal anti-Endo16 antibodies (Nocente-McGrath *et al.*, 1989) at 1:3000 dilution. All incubations and washing steps were done with reagents diluted in TBST. Immunodetection was carried out with alkaline phosphatase-conjugated goat anti-rabbit secondary antibodies (Promega). The alkaline phosphatase color development with NBT (nitro blue tetrazolium) and BCIP (5-bromo-4-chloro-3-indolylphosphate) (Sigma) was carried out in 100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM MgCl₂.

Generation of the pET 5.1 + 13 Fusion Construct

The pET 5.1 + 13 expression construct was produced from existing clones in the pET-His-Tag System from Novagen (Klinzing and Ernst, unpublished). First a 3087-bp fragment of clone 5.1 was excised from the pGEM 7Zf(+) vector with *EcoRI* and *SpeI* and the fragment was blunt-ended and inserted into the blunt-ended *HindIII* site of the pET-20b vector from Novagen. Cloning junctions were sequenced to check for the maintenance of reading frames. A 1478-nt fragment of clone 13 was excised from the pGEM 7Zf(+) vector by partial digestion with *SspI* and *NsiI* and isolated. This 1478-nt fragment of clone 13 was inserted directly 5' of the 5.1

fragment in the pET vector. This produced a construct with 4565 bp of Endo16 linked 5', in proper reading frame, to the pel B signal sequence of the vector and 3' to the histidine-tag. The plasmid with Endo16 5.1 + 13 under control of the T7 promoter was used to transform *Escherichia coli* strain BL21 (DE3) from Novagen. A control construct contained β-galactosidase in the pET 15b vector with no signal sequence under the control of the T7 promoter and followed by the histidine-tag (Novagen). This construct was used to transform BL26 (DE3) cells which lack endogenous β-galactosidase expression.

Production of Fusion Proteins

Bacterial cultures bearing fusion constructs were grown at 37°C in 100 ml of LB media with 100 μg/ml ampicillin, until the OD₆₀₀ was 0.4–0.6. The culture with pET 5.1 + 13 was then transferred to a 25°C bath while the pET β-galactosidase culture was kept at 37°C, and both cultures were induced with 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG). After 3 hr, bacteria were pelleted and resuspended in 2 ml TE buffer. Because recombinant pET 5.1 + 13 protein was insoluble, bacteria were lysed with equal volumes of 2× Laemmli buffer with 0.1 M DTT and protease inhibitors and boiled for 5 min. Bacterial DNA was sheered by passage through a syringe needle and insoluble components were pelleted by centrifugation. Proteins were resolved by 8% SDS-PAGE, and following brief staining with 4 M Na acetate (Higgins and Dahmus, 1979), a gel slice containing the pET 5.1 + 13 protein was excised from the gel and rinsed in distilled H₂O for 5–10 min. The fusion protein was eluted by finely chopping the gel slice into pieces and soaking the pieces in Tris NaCl buffer (50 mM Tris-HCl, pH 6.5; 150 mM NaCl) with 0.1% SDS. The protein was precipitated with 4 vol of acetone and dissolved in Tris NaCl buffer with 0.1% SDS. The pETβ-gal fusion protein was isolated using the His-Bind Purification System, according to the manufacturer's specifications (Novagen). Protein concentrations were estimated by comparison to known amounts of purified BSA, on SDS-polyacrylamide gels stained with Coomassie brilliant blue.

⁴⁵*Calcium-Binding Assay*

Equal amounts of the Endo16 fusion protein, pET 5.1 + 13 and β-galactosidase and an equimolar amount of calmodulin (Sigma) were prepared in Tris NaCl buffer. The pET 5.1 + 13 and β-galactosidase samples also contained 0.1% SDS. Proteins were applied to a PVDF membrane using a slot blot manifold. The pET 5.1 + 13 and β-galactosidase samples were washed several times with Tris NaCl buffer. The membrane was incubated in


```

1054 TTTTGGGTGCAACTGTAGATGAAGCATGTGCAGCCCTTCCTACTACCTGATTCTTCTCTGCTACGTTACTGCGCTGGAATATGCTTCAGAACTTGCA
1019 F W V Q L L D E A (C) A A L P Y L P D S S S A T S T G L E Y A S E L A
1156 CAAGCCGATAATAATGTGCTCGCCAGTCCATTGACGTTTGGAGGACGGGAGACATTGAACTCTAAAGCAGTAAAGTCAATCTCAGAACTACAGTC
1053 Q A D N N V L A P A P L T F G G (R) (G) (D) I R T S K A V K S I S E S T V
1158 GTTCGAATTATGATGAAGCAACTCCGATCATGCTGTGCTCTTTGGGGAAGCCCAACTGAAATGCTTCCAAATGCACTATAGTCTTAAATGCTCTAGC
1087 V S N Y D E D N F R S (C) (C) A L F G E A Q L K (C) F Q M H Y S P N A (P) S
1160 TCACAAATGGAAGTGAAGGATCTCCGAAATCCGTCAACGATGGCAGTGTGTCCGAGGATATGATCTGAAAGCCAGGGCAAGAGTGAAGATCTGAA
1121 S Q M E S E G S A N P S T I G S V S E G Y V S E S Q G Q (E) S E R S E
1162 AGTGAAGCTTCTGAAGGGGAGGATCCGAAAGAGACAGAGGGCGCTGAAGAACCACATGAGCAAAATGGAAGTGAAGGGGAGATCAGGAAGAGTCAAGGGC
1155 S E G S E G E D S E E T E G A E E (P) H E Q M E S E G G D Q E E S E G
1164 GCTGAACAACCACAGGAGCAAAATGGAAGTGAAGACAAGAACCAGTCAAGGAGAGTGTGATCTGGAGAAATAACTGATGGATCGGACATGAACATG
1189 A E Q (P) Q E Q M E S E G Q E P E S E E D V D S G E I T D G S D M N M
1166 CCAGGACCGGATCATCTGACGGAGGTGTCCACTAGGACTAACCGAAGTCCGATGAACAACATCCGATGGAGAGGAGCCCGAGGAACCAAGAGCCAGGA
1223 P G P G S S D G G V P L G L T E V D E Q T S D G E E P E E P E P G
1168 GGAGAACAAGGACAGCAAGAGTGGAGAGGGGAGAAGAAATGAATCTGAAGAAGAGGAGTTGAGGAACCCAGGAAGTGAAGAAAGTGAAGAAAGTGAAG
1257 G E Q G (P) H E Q M E S E G E G N E S E E E E V E E (P) Q E Q M E S I E S E G
1170 CAACAGTCTGAATCAGAAGACCCAGAATTTGAAGGAGAAGAACCAGGAAGTCAAGGAGAGAACAAGAACCCGCAAGACAGATGGAAAGTATAGTACAGAA
1291 Q Q S E S E D P E F E G E E P E E S E G E Q E (P) Q E Q M E S I V Q E
1172 CCGAGTCCGGGGCAGATGTGATCTGAGAGAAGTAACTGACGGATCCGACATGAACATGCCAGGACAGGGATCGTCTGACGGAGCAGGAGGTGTCCACTA
1325 P E S G A D V D S G E V T D G S D M N M P G Q G S S D G A G G V P L
1174 GCACTAGTGGAAATCAAAACATCCAACCAAGAGGGGTCTGAAGTACAAGAAGAAATCCGAGGACCCACAGGAACAAATGGAAGGTGAAGGTCAAGGAAGTGA
1359 G L V E V K T S N Q E G S E V Q E E S Q D (P) Q E Q M E G E G Q S E
1176 TCGGAGGAGAGAACCCAGAAGAACAGGAGAACAGGACCCACAAGAAGCAGTGGAGAGCGAAGGAGAAGAAATGAATCTGAAGAAGAGGAGGTGAA
1393 S E G E E P E E P E G E Q G (P) Q E Q V E S E G E E N E S E E E V E
1178 GAACCCAGGGAGTGAAGAAAGTGAAGGACAACAGTCTGAATCAGAAGAACCAGAATTTGAAGGAGAAGAACCCGGAAGAAATCAGAAGGAGAACAAGAACCA
1427 E (P) Q G V M E S E G Q Q S E S E E P E F E G E E P E E S E G E Q E (P)
1180 CAGGACAGATGGAAGTGAAGGACAAGAACCAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCAAGGAGTCA
1461 Q D Q M E S E G Q E P E S E E D V D S G E V T D G S D M N M P G Q G
1182 TCGTCTGACGGAGCAGGAGGTGTCCACTAGGACTAGTGGAAATCAAAACATCCAACCAAGAGGGGTCTGAAGTACAAGAAGAAATCCGAGGACCCACAGGAA
1495 S S D G A G G V P L G L V E V K T S N Q E G S E V Q E E S Q D (P) Q E
1184 CAAACGGAAGTGAAGTCAAGGAAAGTGAATCGGAGGAGAAGAACCAGAAGAACCAGAAAAAGAAAAAAGATCACTCAGGCTGATGA
1529 Q T E G E G S E S E E E P E E P E K E K K K D H S * *

```

FIG. 1. Nucleotide and nucleotide-derived amino acid sequences of Endo16. The open reading frame lies between nucleotides 1 and 4680, coding for 1560 amino acids. The putative signal sequence is underlined, and the potential signal sequence cleavage site is indicated with an arrowhead. Cysteine residues are circled. Putative heparin-binding regions are marked with dashed lines. Potential N-linked glycan attachment sites are double underlined. The first clustered repeat domain (A) of five repeats has the beginning of each repeat marked with a diamond. The putative cell adhesion motif, RGD, is boxed. The second clustered repeat domain (B) of 12 repeats has the start of each repeat indicated with a square. Accession No. L34680.

$^{45}\text{Ca}^{2+}$ -binding buffer (90 mM imidazole-HCl, pH 6.5; 60 mM KCl; 5 mM MgCl_2) for 1 hr with three buffer changes and then incubated in the same buffer with 1 $\mu\text{Ci}/\text{ml}$ of $^{45}\text{Ca}^{2+}$ (New England Nuclear) for 15 min, washed with 30% ethanol for 8 min, and autoradiographed for 5 days, following a modified Maruyama *et al.* (1984) procedure.

RESULTS

Sequence Analysis and General Nature of Endo16

The major Endo16 transcript is approximately 6.6 kb in length (Nocente-McGrath *et al.*, 1989) and at most developmental stages there is also a less abundant larger transcript (Godin and Ernst, unpublished). Figure 1 presents the nucleotide and nucleotide-derived amino acid sequence of Endo16 obtained from several overlapping cDNA clones, which are displayed under the schematic representation of Endo16 in Fig. 2A. There is a 4680-nt open reading frame which, when translated into

an amino acid sequence, starts with two methionines, which are immediately followed by a consensus signal sequence, and ends with two consecutive stop codons.

Overall the Endo16 protein is acidic (24%) and hydrophilic in nature. The acidic/basic profile and the hydrophilicity plot of the Endo16 protein are shown in Figs. 2B and 2C, respectively. The carboxyl half of the molecule is extremely acidic, while the rest of the protein is more balanced in charge, with interspersed regions of acidic and basic residues. Hopp and Woods' (1983) analysis revealed that, excluding the signal sequence, the Endo16 protein is predominantly hydrophilic and is devoid of any hydrophobic regions capable of spanning a membrane (Fig. 2C).

Endo16 Exhibits a Modular Composition

Computer-assisted searches of nucleotide and protein data bases have not identified an Endo16 homologue.

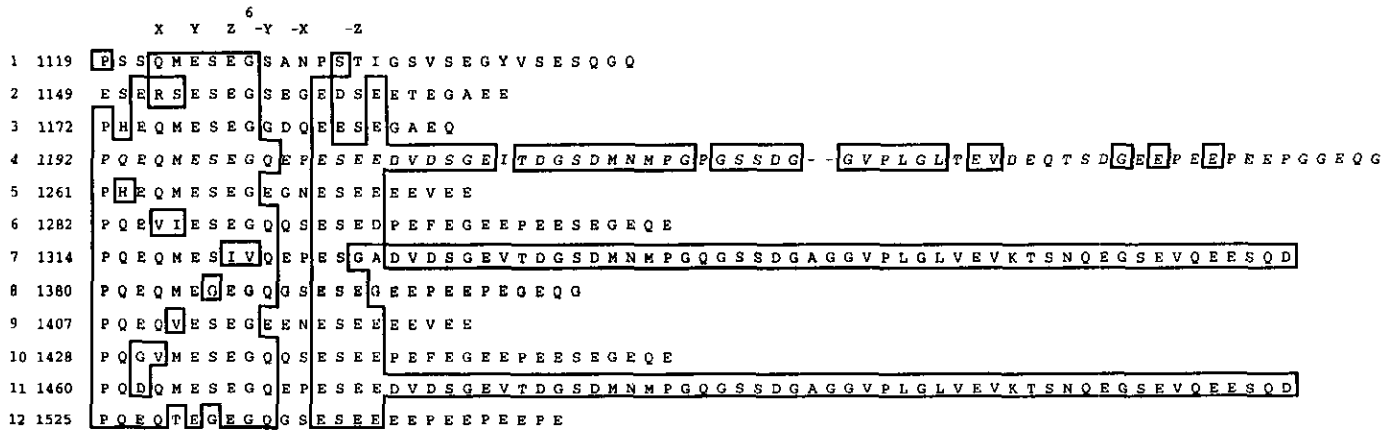


FIG. 5. The second acidic domain (B) of 12 potential calcium-binding repeats. The coordinates X, Y, Z, -Y, -X, -Z, where X = position 1, and the conserved glycine in the 6th position, indicated above repeats, correspond to the conserved residues which coordinate Ca^{2+} ions in the Ca^{2+} -binding loops of the EF-hand motif (da Silva and Reinach *et al.*, 1991). Identical amino acids found in 7 or more of the 20 amino acid repeats are boxed. Repeats 6 and 10 are followed by 12 identical amino acids and repeats 7 and 11 are followed by 46 identical amino acids which are similar to those following repeat 4 (also boxed). The first number to the left of each repeat indicates the repeat number and the second number indicates the amino acid at the beginning of each repeat.

Endo16 Binds ^{45}Ca Calcium

Since Endo16 contains two regions of clustered repeats, characteristic of calcium-binding motifs, experiments were carried out to test the protein's ability to bind calcium. A bacterially produced, Endo16 fusion protein, designated pET 5.1 + 13, was used in these studies. The fusion protein started at the signal sequence and contained all of region A and 7 of the 12 potential calcium-binding repeats of region B (Fig. 2). Calmodulin was used as a positive control, and bacterially produced β -galactosidase served as the negative control. β -Galactosidase was selected to control for the histidine-tag at the end of the Endo16 pET 5.1 + 13 fusion protein since both genes are in the pET-His-Tag vector system and both bacterially produced proteins have 6 histidines at their carboxyl ends. When approximately equimolar amounts of Endo16 and calmodulin were slot-blotted onto a PVDF membrane, they both bound ^{45}Ca calcium strongly, while there was no detectable calcium binding by β -galactosidase (Fig. 6).

Endo16 Is Protected against Proteolysis in Embryonic Extracts Supplemented with Calcium

To determine if calcium binding causes a conformational change, as demonstrated by protection against proteolysis, total protein extracts from gastrula stage embryos were treated with varying concentrations of trypsin in the presence and absence of calcium (EGTA added), as has been done to demonstrate calcium-binding conformational changes of purified thrombospondin (Lawler *et al.*, 1985) and uvomorulin (E-CAM) (Hyafil *et al.*, 1981; Ringwald *et al.*, 1987). The tryptic digestion

products were then separated on a SDS-polyacrylamide gel and detected by Western blotting, using a polyclonal anti-Endo16 antibody. Results presented in Fig. 7A demonstrate that in the presence of 5 mM EGTA even without trypsin added (lane labeled 0), Endo16 is digested into numerous proteolytic degradation products by endogenous proteases. Increasing the amounts of trypsin further degraded the protein. However, when samples of the same protein extract were exposed to the same amounts of trypsin in the presence of 5 mM CaCl_2 , in all but the sample with the highest trypsin concentration, the majority of Endo16 is not digested and is detected as a broad band with an estimated molecular weight greater than 300 kDa (Fig. 7A). Furthermore, the effect is calcium-specific, since neither magnesium nor manganese protect Endo16 from proteolytic degradation (Fig. 7B).

Results of treatment of albumin with trypsin under the same conditions that were used for the embryonic

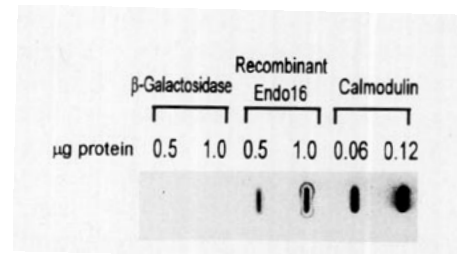


FIG. 6. ^{45}Ca calcium-binding by Endo16. Approximately equimolar amounts of recombinant β -galactosidase with the histidine tag, recombinant Endo16, and calmodulin were applied to a PVDF membrane and washed with Tris-HCl buffer. The membrane was incubated with 1 $\mu\text{Ci/ml}$ ^{45}Ca calcium for 15 min in binding buffer, pH 6.5, washed with 30% ethanol, and autoradiographed for 5 days.

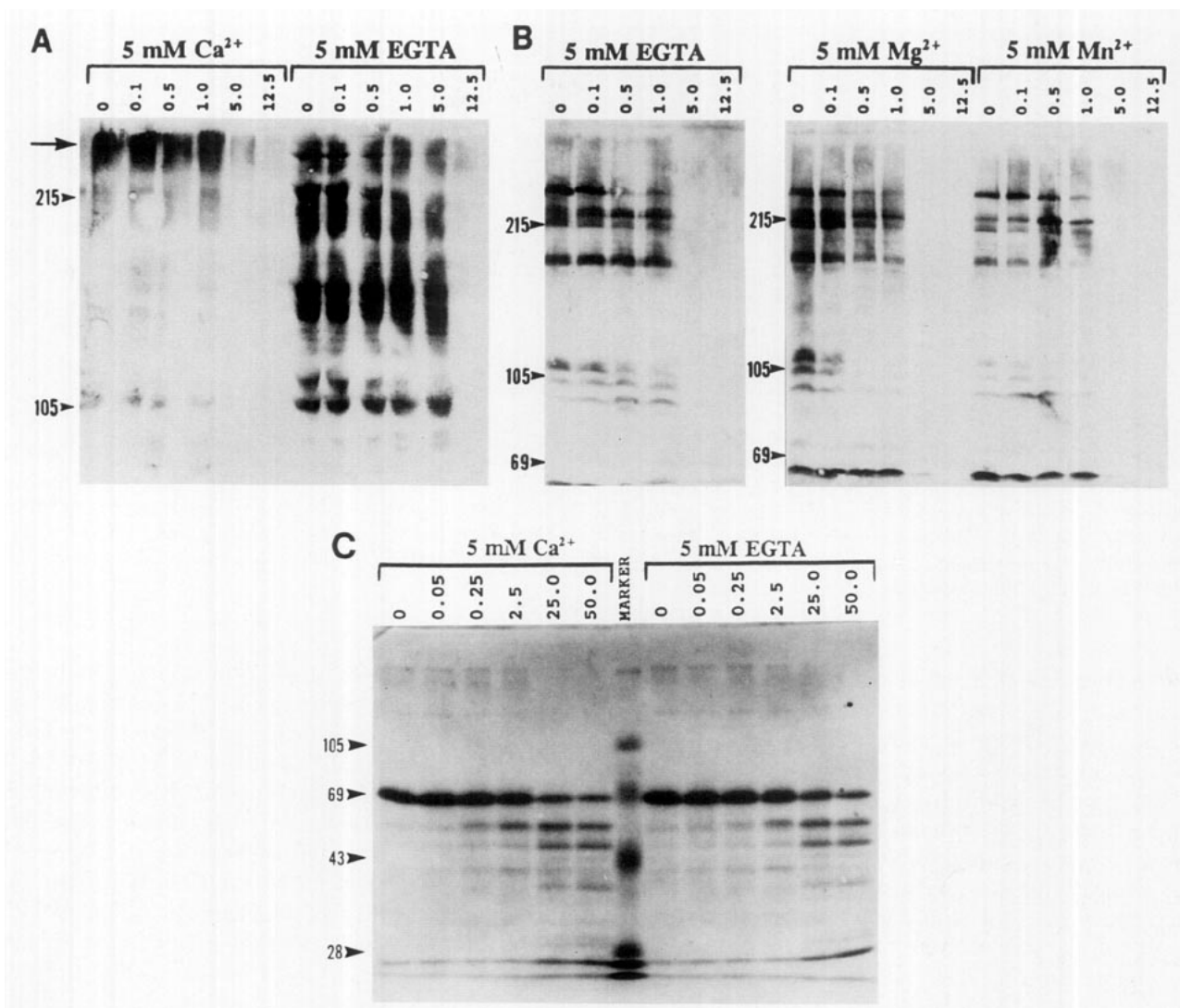


FIG. 7. Western blots of tryptic digestion products of Endo16 and albumin carried out in the presence of 5 mM CaCl₂, MgCl₂, MnCl₂, or EGTA. Each sample contained 50 μg of total embryonic protein. Different amounts of trypsin were added to each sample. The number above each lane indicate the number of micrograms of trypsin added. (A) In samples to which 5 mM CaCl₂ was added, the large molecular weight form of Endo16, estimated to be above 300 kDa, is indicated by arrow. The stacking gel contained 4% acrylamide and 0.14% bisacrylamide and the resolving gel contained 6.7% acrylamide and 0.25% bisacrylamide. (B) The proteolytic degradation patterns of Endo16 generated in the presence of MgCl₂ and MnCl₂ are similar to the pattern generated in the presence of EGTA. Sizes of molecular weight marker proteins in kDa are indicated. The stacking gel contained 4% acrylamide and 0.14% bisacrylamide and the resolving gel contained 12% acrylamide and 0.05% bisacrylamide. (C) Tryptic digestion of serum albumin in the presence of 5 mM Ca²⁺ and 5 mM EGTA. Each sample contained 25 μg albumin. Ratios of trypsin to serum albumin are indicated above the gel; 0 indicates a control sample which was not treated with trypsin, followed by samples with increasing ratios of trypsin added. The peptide patterns of serum albumin generated by tryptic digestion in the presence of Ca²⁺ and EGTA are virtually identical. Sizes of molecular weight marker proteins in kDa are indicated. The gel composition is the same as that in B.

protein extracts in the presence and the absence of calcium supplies two controls. Albumin is the major calcium carrier in blood; however, this calcium binding does not significantly alter the conformation of albumin to protect it from proteolysis (Hyafil *et al.*, 1981). In Fig. 6C, the tryptic digestion products of albumin generated in the presence of 5 mM CaCl₂ or 5 mM EGTA are virtually identical at every trypsin concentration, thus demonstrating that: (1) trypsin is active under these condi-

tions both in the presence and the absence of calcium and (2) calcium binding without a concomitant conformational change is not sufficient to protect a protein from digestion by trypsin.

Endo16 Is Found on the Blastocoelar Surfaces of Endodermal Cells and in the ECM

Using an anti-Endo16 antibody and a fluorescently labeled secondary antibody, Endo16 was localized to the

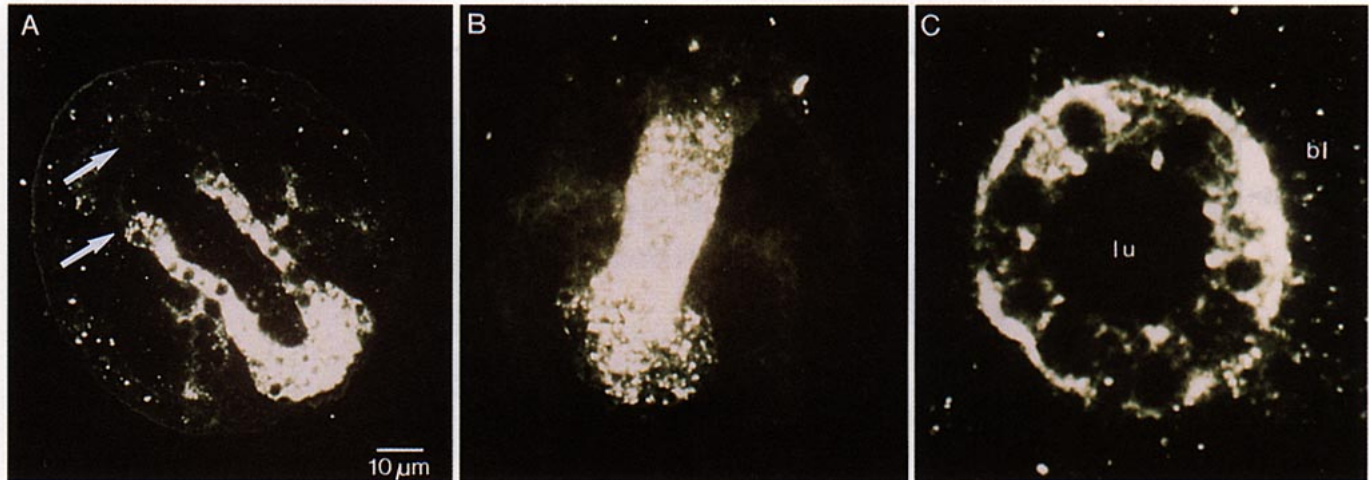


FIG. 8. Anti-Endo16 staining of 36- to 38-hr gastrula stage embryos viewed with a laser confocal microscope. (A) Optical section through the middle of the archenteron. Staining is found in the midgut and hindgut, but not the foregut. The region of the foregut is indicated with arrows. (B) Computer-reconstructed composite of optical sections showing the entire embryo. Note sharp borders between stained and unstained cells. (C) Optical cross section through the single cell layered archenteron. Staining is more intense on cell surfaces facing the blastocoel (bl); less staining is seen on the luminal surface of the gut (lu). Bar, 10 μ m.

surfaces of endodermal cells in fixed whole mount early and mid gastrula stage embryos (Nocente-McGrath *et al.*, 1989, 1991). However, resolution was not sufficient to establish whether Endo16 is extracellular, spans the membrane, or is found in the cortex of the cytoplasm. Here we have used confocal immunofluorescence microscopy and determined that in late stage gastrula embryos Endo16 appears to be cell surface-associated or extracellular with the signal being more intense on the blastocoelar surface compared to the luminal surface of the archenteron (Fig. 8). In addition, in agreement with the distribution of Endo16 mRNA (Nocente-McGrath *et al.*, 1989; Ransick *et al.*, 1993) in late stage gastrula, the protein is found on cells of the midgut and hindgut, but is absent from the foregut (Fig. 8).

The presence of a signal sequence and the lack of any transmembrane domains strongly suggest that Endo16 is secreted. Immunogold electron microscopy localization of Endo16 has confirmed this prediction. In immunogold preparations of late gastrulae, gold particles are associated with the surfaces of endodermal cells and the extracellular matrix (Fig. 9). On the surface of the cells the gold particles are in clusters over flocculent material associated with the plasmalemma. There are clusters of gold particles on both lateral and blastocoelar surfaces of the endodermal cells, but the clusters are more common on the blastocoelar surfaces (Fig. 9).

DISCUSSION

We present here the characterization of a sea urchin gene, Endo16, which encodes a large, extracellular, multidomain protein with similarities to regions of several

vertebrate extracellular molecules known to be important during embryogenesis and morphogenesis. The presence of a signal sequence at the amino terminus of Endo16 and immunolocalization by confocal and electron microscopy demonstrate that Endo16 is an extracellular protein associated with the surfaces and the ECM of the invaginating archenteron during gastrulation. Immunolocalization of the Endo16 protein is limited to the cells expressing Endo16 mRNA (Nocente-McGrath *et al.*, 1989; Ransick *et al.*, 1993). At the mesenchyme blastula stage, the Endo16 mRNA and protein are found in the vegetal plate, with a precise expression boundary between the cells derived from the Veg1-Veg2 layers. Endo16 expression is also excluded from primary mesenchyme cells. This pattern of expression is interesting when compared to the expression of two other endoderm-specific genes, Endo1 and LvN1.2 (reviewed: Wessel, 1993). The Endo1 and LvN1.2 gene products are also entirely absent from the foregut of the fully extended archenteron. However, in contrast to Endo16 expression, the products of these genes are excluded from the anterior third of the archenteron throughout invagination. This suggests that the Endo16 protein is needed in the foregut during the elongation phase, but not for differentiation, and is therefore turned over rapidly in these cells.

Confocal and electron microscopy suggest that there is an unequal distribution of Endo16 on the surfaces of the endodermal cells, with most of the protein concentrated on the blastocoelar (basal) surfaces and between the endodermal cells, and with significantly less Endo16 localized to the luminal (apical) surfaces of the endoder-

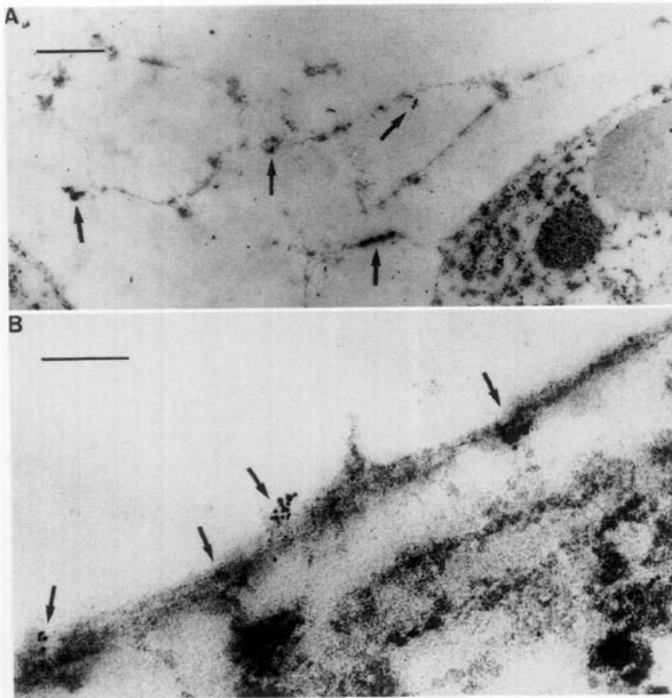


FIG. 9. Immunogold localization of *Endo16* in late gastrula stage *S. purpuratus* embryos. (A) Portion of the blastocoelar surface of an endodermal cell and the ECM. Arrows indicate clusters of gold particles in the ECM. Bar, 0.2 μm . (B) Portion of the blastocoelar surface of an endodermal cell; patches of gold particles are found on the surface of endodermal cells (arrows). Gold particles are 5 nm in diameter. Bar, 0.1 μm . Randomly selected regions of cytoplasm for endodermal cells has a mean of 20.4 gold particles per unit area (SEM = 3.08, $n = 5$) and ectodermal cells has a mean of 26.6 particles per unit area (SEM = 3.906, $n = 5$). The two-tailed P value is 0.247, considered not significant. Control rabbit serum had a similar pattern of nonspecific binding to the cytoplasmic regions of cells with a mean of 28.3 particles (SEM = 5.23, $n = 5$).

mal cells. Although the relationship between the *Endo16* molecules on the surfaces of endodermal cells and those that appear to be a component of the ECM is not presently known, the presence of an RGD tripeptide suggests that some *Endo16* molecules may interact with endodermal cells through an integrin-like receptor. Recently sequences coding for β -integrin (Marsden *et al.*, 1993) and α -integrin (Susan and Lennarz, 1993) have been identified in sea urchins by PCR amplification of cDNAs from gastrula stage embryos using degenerate oligonucleotide primers.

The extracellular location of *Endo16* and the large number of regularly distributed cysteines suggest that most of these cysteines form covalent disulfide bridges, which would result in a globular amino half of the molecule. The repeating pattern of cysteines is strikingly similar to the cysteine pattern in proteins of the serum albumin family which includes serum albumin, α -feto-protein, and vitamin D-binding protein (Yang *et al.*,

1985). All members of the serum albumin family have 7 or 8 cysteine pairs which are followed within 6–10 amino acids by a single cysteine. *Endo16* has 13 cysteine pairs which are separated from the following single cysteine by 9 or 10 amino acids. X-ray analysis of the three-dimensional structure of human serum albumin (Carter *et al.*, 1989) and vitamin D-binding protein (Vogelaar *et al.*, 1991) confirm the predicted disulfide-bond pattern where the second cysteine of each pair forms a disulfide bond with the preceding single cysteine and the first member of the pair forms a disulfide bond with the following single cysteine (Brown, 1976). At present, this disulfide-bond pattern seems to be the most likely model for this portion of *Endo16*.

The strong likelihood of a structural similarity between the amino half of *Endo16* and members of the serum albumin family invites speculation on a possible functional relationship. Members of the serum albumin family of proteins bind and carry a variety of molecules in the blood. Binding studies show that each of the six binding regions of albumin bind distinct ligands (reviewed: Kragh-Hansen, 1990). Ligand binding is a consequence of the unusual structure and the high portion of both acidic and basic amino acids in these proteins. Interestingly, albumin is composed of 16.6% acidic and 12.9% basic amino acid residues (Peters, 1975), compared to 17.8% acidic and 14% basic residues for *Endo16* (first 809 amino acids minus the signal sequence). Evidence from a number of different systems suggests that molecules of the ECM bind ligands important for growth and differentiation and present them to cells (reviewed: Schubert, 1992; Adams and Watt, 1993; Juliano and Haskill, 1993). Given that the archenteron forms and differentiates within the blastocoel, cell signaling molecules secreted by endodermal cells would be likely to diffuse if not trapped by appropriately positioned extracellular molecules. Therefore, we propose that the cysteine-rich region of *Endo16* functions as a ligand-binding region, positioned at the surface and in the ECM immediately surrounding the differentiating archenteron.

Two regions of repeats, designated A and B have been characterized from sequence analysis as potential calcium-binding domains of *Endo16*. These two regions appear unrelated, and based on their similarities to proteins with distinct mechanisms for calcium binding, most likely would have different mechanisms for binding calcium. The first potential calcium-binding domain overlaps a region that may possibly serve as a substrate for transglutaminase, which could covalently cross-link *Endo16* to other ECM proteins (Aeschlimann and Paulsson, 1991).

We have established that a bacterially produced *Endo16* protein, which contains the entire first putative

calcium-binding domain and 7 of the 12 potential calcium-binding loops of the second domain, binds radioactive calcium. We do not know which region is responsible for calcium binding or if both regions contribute. Tryptic digestion of gastrula stage protein extracts in the presence of Ca^{2+} , Mg^{2+} , Mn^{2+} , and EGTA have revealed that only Ca^{2+} stabilizes Endo16 against degradation. This protection is likely due to a significant conformational change that Endo16 undergoes upon calcium binding, as has been observed with other extracellular calcium-binding proteins (Lawler *et al.*, 1985; Hyafil *et al.*, 1981; Ringwald *et al.*, 1987). An alternative explanation is that Endo16 undergoes a calcium-dependent aggregation that protects it from proteolysis in these extracts.

Endo16 may play a role in cell adhesion and, as with many other cell adhesion proteins, calcium binding may modulate this adhesion (Ringwald *et al.*, 1987). When sea urchin embryos are placed in calcium- and magnesium-free seawater the blastomeres fall apart. In an investigation of the role of calcium in sea urchin embryos McClay and Matraga (1986) concluded that calcium has distinct roles in both the initial and secondary phases of adhesion. The occurrence of an RGD tripeptide, known to be a cell attachment site in many extracellular proteins, between two putative calcium-binding regions predicted to have different affinities for calcium suggests the possibility that if Endo16 functions in cell adhesion, these calcium-binding sites may play distinct roles in modulating these adhesive interactions.

Large extracellular molecules are frequently multifunctional due to the presence of separate structural and functional domains. In some cases the different domains may actually have conflicting functions expressed in different developmental contexts. We propose that Endo16 is a multifunctional extracellular protein and that two likely functions of Endo16 during the formation and differentiation of the sea urchin archenteron are cell adhesion, modulated by calcium binding, and binding of signaling ligands. Experiments to test these possibilities are underway.

We would like to thank Dr. Stephen Paddock at the Integrated Microscopy Resource for Biomedical Research at the University of Wisconsin for help with confocal microscopy and Drs. N. George and F. Wilt for the gastrula stage cDNA library. We thank Dr. Catherine Nocente-McGrath for the anti-Endo16 antibody and for assistance in preparing embryos for confocal microscopy. This work was supported in part by a Medical Research Council of Canada Operating Grant to R.D.B. and a Biomedical Research Support grant and NSF Grant DCB 8910316 to S.G.E.

REFERENCES

Adams, J. C., and Watt, F. M. (1993). Regulation of development and differentiation by the extracellular matrix. *Development* **117**, 1183-1198.

- Aeschlimann, D., and Paulsson, M. (1991). Cross-linking of laminin-nidogen complexes by tissue transglutaminase: A novel mechanism for basement membrane stabilization. *J. Biol. Chem.* **266**, 15308-15317.
- Benson, S., Rawson, R., Killian, C., and Wilt, F. (1991). Role of the extracellular matrix in tissue-specific gene expression in the sea urchin embryo. *Mol. Reprod. Dev.* **29**, 220-226.
- Brown, J. R. (1976). Structural origins of mammalian albumin. *Fed. Proc.* **35**, 2141-2144.
- Butler, E., Hardin, J., and Benson, S. (1987). The role of lysyl oxidase and collagen crosslinking during sea urchin development. *Exp. Cell Res.* **173**, 174-182.
- Campbell, S. S., and Crawford, B. J. (1991). Ultrastructural study of the hyaline layer of the starfish embryo, *Pisaster ochraceus*. *Anat. Rec.* **231**, 125-135.
- Cardin, A. D., and Weintraub, H. J. R. (1989). Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis* **9**, 21-32.
- Carter, D., He, X. M., Munson, S., Twigg, P., Gernert, K., Broom, M., and Miller, T. (1989). Three-dimensional structure of human serum albumin. *Science* **244**, 1195-1198.
- da Silva, A. C., and Reinach, F. C. (1991). Calcium binding induces conformational changes in muscle regulatory proteins. *Trends Biosci.* **16**, 53-57.
- D'Souza, S. E., Ginsberg, M. H., and Plow, E. F. (1991). Arginyl-glycyl-aspartic acid (RGD): A cell adhesion motif. *Trends Biosci.* **16**, 246-250.
- Eitensohn, C. A., and Ingersoll, E. P. (1992). Morphogenesis of the sea urchin embryo. In "Morphogenesis. An analysis of the Development of Biological Form" (Rossomando and Alexander, Eds.), pp. 189-262. Dekker, New York.
- Funk, S., and Sage, H. (1991). The Ca^{2+} binding glycoprotein SPARC modulates cell cycle progression in bovine aortic endothelial cells. *Proc. Natl. Acad. Sci. USA* **88**, 2648-2652.
- Gustafson, G., and Wolpert, L. (1967). Cellular movement and contact in sea urchin morphogenesis. *Biol. Rev.* **42**, 442-498.
- Heifeitz, A., and Lennarz, W. J. (1979). Biosynthesis of N-glycosidically linked glycoproteins during gastrulation of sea urchin embryos. *J. Biol. Chem.* **254**, 6119-6127.
- Hofmann, S., Goldstein, J., Orth, K., Moomaw, C., Slaughter, C., and Brown, M. (1989). Molecular cloning of a histidine-rich Ca^{2+} -binding protein of sarcoplasmic reticulum that contains highly conserved repeated elements. *J. Biol. Chem.* **264**, 18083-18090.
- Hopp, T. P., and Woods, K. R. (1983). A computer program for predicting protein antigenic determinants. *Mol. Immunol.* **20**, 483-489.
- Hyafil, F., Babinet, C., and Jacob, F. (1981). Cell-cell interactions in early embryogenesis: A molecular approach to the role of calcium. *Cell* **26**, 447-454.
- Juliano, R. L., and Haskill, S. (1993). Signal transduction from the extracellular matrix. *J. Cell Biol.* **120**, 577-585.
- Karp, G. C., and Solursh, M. (1974). Acid mucopolysaccharide metabolism, the cell surface, and primary mesenchyme cell activity in the sea urchin embryo. *Dev. Biol.* **41**, 110-123.
- Kragh-Hansen, U. (1990). Structure and ligand binding properties of human serum albumin. *Danish Med. Bull.* **37**, 57-84.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**, 680-685.
- Lawler, J., Derick, L., Connolly, J., Chen, J. H., and Chao, F. C. (1985). The structure of human platelet thrombospondin. *J. Biol. Chem.* **260**, 3762-3772.
- Lawler, J., Weinstein, R., and Hynes, R. (1988). Cell attachment to thrombospondin: The role of ARG-GLY-ASP, calcium and integrin receptors. *J. Cell Biol.* **107**, 2351-2361.
- Marck, C. (1988). "DNA Strider": A "C" program for the fast analysis

- of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* **16**, 1829-1836.
- Marsden, M., Wang, D., and Burke, R. D. (1993). Beta integrin expression during embryogenesis in the sea urchin. *Mol. Biol. Cell* (Suppl.) **4**, 378a.
- Maruyama, K., Mikawa, T., and Ebashi, S. (1984). Detection of calcium binding proteins by ^{45}Ca autoradiography on nitrocellulose membrane after sodium dodecyl sulfate gel electrophoresis. *J. Biochem.* **95**, 511-519.
- McCarthy, R., Beck, K., and Burger, M. (1987). Laminin is structurally conserved in the sea urchin basal lamina. *EMBO J.* **6**, 1587-1593.
- McClay, D., and Matrangola, V. (1986). On the role of calcium in the adhesion of embryonic sea urchin cells. *Exp. Cell Res.* **165**, 152-164.
- McClay, D. R., Armstrong, N. A., and Hardin, J. (1992). Pattern formation during gastrulation in the sea urchin embryo. In "Gastrulation" (Stern and Ingham, Eds.), pp. 33-41. The Co. of Biologists, Ltd., Cambridge.
- Michalak, M., Milner, R. E., Burns, K., and Opas, M. (1992). Calreticulin. *Biochem. J.* **285**, 681-692.
- Moncrief, N., Kretsinger, R., and Goodman, M. (1990). Evolution of EF-hand calcium-modulated proteins. *J. Mol. Evol.* **30**, 522-562.
- Nocente-McGrath, C., Brenner, C., and Ernst, S. G. (1989). Endo16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. *Dev. Biol.* **136**, 264-272.
- Nocente-McGrath, C., McIsaac, R., and Ernst, S. G. (1991). Altered cell fate is LiCl-treated sea urchin embryos. *Dev. Biol.* **147**, 445-450.
- Ono, T., Slaughter, G. R., Cook, R. G., and Means, A. R. (1989). Molecular cloning sequence and distribution of rat caldesmon, a high affinity calmodulin-binding protein. *J. Biol. Chem.* **264**, 2081-2087.
- Peters, T. (1975). Serum albumin. In "The Plasma Proteins (F. Putnam, Ed.), pp. 133-181. Academic Press, New York.
- Ransick, A., Ernst, S., Britten, R., and Davidson, E. (1993). Whole mount *in situ* hybridization shows Endo16 to be a marker for vegetal plate territory in sea urchin embryos. *Mech. Dev.* **42**, 117-124.
- Ringwald, M., Schuh, R., Vestweber, E., Eistetter, H., Lottspeich, F., Engel, J., Dölz, R., Jähmig, F., Epplen, J., Mayer, S., Müller, C., and Kemler, R. (1987). The structure of cell adhesion molecule uvomorulin: Insights into the molecular mechanism of Ca^{2+} dependent cell adhesion. *EMBO J.* **6**, 3647-3653.
- Sambrook, J., Fritsch, E., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual." 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Schubert, D. (1992). Collaborative interactions between growth factors and the extracellular matrix. *TICB* **2**, 63-66.
- Smith, M. J., Hough, B. R., Chamberlin, M. E., and Davidson, E. H. (1974). Repetitive and nonrepetitive sequence in sea urchin heterogeneous nuclear RNA. *J. Mol. Biol.* **85**, 103-126.
- Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985). Measurement of protein using bicinchonine acid. *Anal. Biochem.* **150**, 76-85.
- Susan, J. M., and Lennarz, W. J. (1993). Identification of an integrin α subunit in gastrula stage sea urchin embryos. *Mol. Biol. Cell* (Suppl.) **4**, 378a.
- Tobin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.
- Tuckwell, D. S., Bass, A., and Humphries, M. J. (1992). Homology modelling of integrin EF-hands. *Biochem. J.* **285**, 325-331.
- Vogelaar, N., Lindberg, U., and Schutt, C. (1991). Crystallization and preliminary X-ray analysis of GC, the vitamin D-binding protein in serum. *J. Mol. Biol.* **220**, 545-547.
- von Heijne, G. (1988). Transcending the impenetrable: How proteins come to terms with membranes. *Biochim. Biophys. Acta.* **947**, 307-333.
- Wessel, G. (1993). Differentiation of the endodermal epithelium during gastrulation in the sea urchin embryo. In "Molecular Basis of Morphogenesis; 51st Symposium of the Society of Developmental Biology." Wiley-Liss, New York.
- Wessel, G., and McClay, D. (1987). Gastrulation in the sea urchin embryo requires the deposition of crosslinked collagen within the extracellular matrix. *Dev. Biol.* **121**, 149-165.
- Wylie, D. C., and Vanaman, T. C. (1988). Structure and function of the calmodulin family of calcium regulatory proteins. In "Calmodulin" (P. Cohen and C. Biklee, Eds.). Elsevier, New York.
- Yang, F., Brune, J., Naylor, S., Cupples, R., Naberhaus, K., and Bowman, B. (1985). Human group-specific component (Gc) is a member of the albumin family. *Proc. Natl. Acad. Sci. USA* **82**, 7994-7998.