importance of this conserved intron sequence has been demonstrated genetically; mutations at positions 1, 2, and 5 significantly impair splicing (*22*) (mutations at position 4 have not been tested). Furthermore, while few mammalian 5′ junctions fit their consensus perfectly (*23*), Zhuang and Weiner demonstrated that increasing the degree of complementarity from 5 to 6 bp (of 9 bp possible) between U1 and a mammalian 5′ splice site resulted in more efficient use of that junction (*3*).

On the other hand, perfect (nine out of nine) complementarity may be detrimental to optimal splicing efficiency. It is also possible that position 4 of the intron is not needed to form a canonical base pair with snR19, but may be involved in a separate recognition event. In addition, there may be a structural requirement for a U at position 5 (and 10) of snR19, perhaps for tertiary interactions. Construction of mutations in conserved sequences of snR19 should provide a genetic test of their function. Moreover, making the complementary changes in the intron will allow us to assess the specific contribution of Watson-Crick complementarity to 5′ splice junction recognition by snR19.

REFERENCES AND NOTES

1. T. Maniatis and R. Reed, *Nature (London)* **325**, 673 (1987).
2. M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin, J. A. Steitz, *ibid.* **283**, 220 (1980); J. Rogers and R. Wall, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1877 (1980). The original model included a more extensive interaction between U1 and the intron, which also included base pairing to the 3′ acceptor; the latter interaction has been ruled out on the basis of phylogenetic comparison (*25*).
3. Y. Zhuang and A. Weiner, *Cell* **46**, 827 (1986).
4. N. Riedel, J. A. Wise, H. Swerdlow, A. Mak, C. Guthrie, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8097 (1986).
5. J. A. Wise *et al.*, *Cell* **35**, 743 (1983).
6. D. Tollervey, J. A. Wise, C. Guthrie, *ibid.* p. 753; D. Tollervey and C. Guthrie, *EMBO J.* **4**, 3873 (1985); R. Parker *et al.*, in preparation.
7. M. R. Lerner, J. A. Boyle, S. M. Mount, S. L. Wolin, J. A. Steitz, *Nature (London)* **283**, 220 (1980).
8. N. Riedel, S. Wolin, C. Guthrie, *Science* **235**, 328 (1987).
9. E. M. De Robertis, S. Lienhard, R. F. Parisot, *Nature (London)* **295**, 572 (1982); D. J. Forbes, T. B. Kornberg, M. W. Kirschner, *J. Cell Biol.* **97**, 62 (1983).
10. C. Branlant *et al.*, *EMBO J.* **1**, 1259 (1982); J.-P. Liautard, J. Sri-Wadada, C. Brunel, P. Jeanteur, *J. Mol. Biol.* **162**, 623 (1982); I. Mattaj and E. De Robertis, *Cell* **40**, 111 (1985).
11. D. Tollervey and I. Mattaj, *EMBO J.* **6**, 469 (1987).
12. B. Patterson and C. Guthrie, *Cell* **49**, 613 (1987). The single copy gene for snR7 encodes two forms of the RNA, designated L (214 nt) and S (179 nt), which differ at their 3′ ends.
13. P. Siliciano, D. Brow, H. Roiha, C. Guthrie, *ibid.*, in press.
14. C. W. Pikielny, B. C. Rymond, M. Rosbash, *Nature (London)* **324**, 341 (1986); N. Riedel *et al.*, in preparation.
15. M. Ares, Jr., *Cell* **47**, 49 (1986).
16. D. Tollervey and I. Mattaj (*11*) have reported anti-Sm precipitation of snR7 and snR14. Precipitation of other, higher molecular weight RNAs was not observed.
17. C. Branlant, A. Krol, J.-P. Ebel, *Nucleic Acids Res.* **9**, 841 (1981).
18. A. Weiner, personal communication.
19. J. A. Wise, personal communication.
20. B. Rymond, D. Torrey, M. Rosbash, *Genes Dev.* **1**, 238 (1987); N. Riedel *et al.*, in preparation.
21. J. L. Teem *et al.*, *Nucleic Acids Res.* **12**, 8295 (1984).
22. R. Parker and C. Guthrie, *Cell* **41**, 107 (1985); A. Jacquier, J. R. Rodriguez, M. Rosbash, *ibid.* **43**, 423 (1985); L. A. Fouser and J. D. Friesen, *ibid.* **45**, 81 (1986); U. Vijayraghavan *et al.*, *EMBO J.* **5**, 1683 (1986).
23. S. M. Mount, *Nucleic Acids Res.* **10**, 459 (1982).
24. S. L. Wolin and J. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1996 (1984).
25. R. J. Lin, A. J. Newman, S. Cheng, J. Abelson, *J. Biol. Chem.* **260**, 14780 (1985).
26. T. E. England and O. C. Uhlenbeck, *Nature (London)* **275**, 560 (1978).
27. S.-C. Cheng and J. Abelson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2387 (1986).
28. D. McPheeters, A. Christensen, E. Young, G. Stormo, L. Gold, *Nucleic Acids Res.* **14**, 5813 (1986).
29. S. M. Mount and J. A. Steitz, *Nucleic Acids Res.* **9**, 6351 (1981).
30. The *SNR19* gene has also been cloned and sequenced by L. Kretzner, B. Rymond and M. Rosbash (*Cell*, in press).
31. We thank J. Steitz for giving us antibodies; A. Weiner and J. A. Wise for communication of results prior to publication; J. Tamm for help with the RNA sequencing; N. Riedel and C. Desplan for discussion and advice; and T. Simmons, B. Patterson, and K. Yamamoto for comments on the manuscript. Supported by NIH grant GM21119 and NSF grant DCB8603926 (C.G.), by a postdoctoral fellowship from the Leukemia Society (M.H.J.), and by a Damon Runyon–Walter Winchell Cancer Fund Fellowship, DRG-872 (P.G.S.)

8 July 1987; accepted 14 August 1987

A Sea Urchin Gene Encodes a Polypeptide Homologous to Epidermal Growth Factor

DEBORAH A. HURSH, MARY E. ANDREWS, RUDOLF A. RAFF

A sea urchin DNA clone complementary to an embryonic messenger RNA whose protein product bears striking homology to the epidermal growth factor family of proteins has been identified and characterized. The structure of the protein is similar to that of previously identified regulatory genes in *Drosophila* and *Caenorhabditis*. RNA gel blot hybridization showed a unique temporal pattern of expression of this gene during embryogenesis and transcript enrichment in the embryonic ectoderm. These results suggest that this member of the epidermal growth factor gene family plays a role in early development decisions in sea urchin embryos.

THE EPIDERMAL GROWTH FACTOR (EGF) family represents a rapidly expanding collection of proteins that have a variety of cellular functions, including differentiation, cell proliferation, and neoplastic transformation (*1*). Recently, the products of two developmental switch genes, the *notch* gene of *Drosophila* (*2, 3*) and the *lin 12* gene of *Caenorhabditis elegans* (*4*), were shown to be members of this gene family, heretofore thought to be exclusively mammalian. Our isolation of a complementary DNA (cDNA) clone encoding a protein resembling the products of these two genes suggests that the EGF peptide domain is widely used in metazoan development.

We isolated a 1.5-kb cDNA clone, uEGF-1, during a screen for cell lineage–specific genes in the sea urchin *Strongylocentrotus purpuratus*. This cDNA clone recognizes transcripts of 3.0 kb and 4.0 kb on RNA gel blots and is preferentially expressed in embryonic ectoderm, as determined by cell fractionation techniques (*5, 6*) (Fig. 1).

The sequence of this cDNA revealed a long open reading frame (ORF) of 1447 nucleotides ending with three in-frame stop codons (Fig. 2). Codon bias for this reading frame agrees with that of other sea urchin ORFs (*7*). A search of the Protein Information Resource (PIR) database by means of

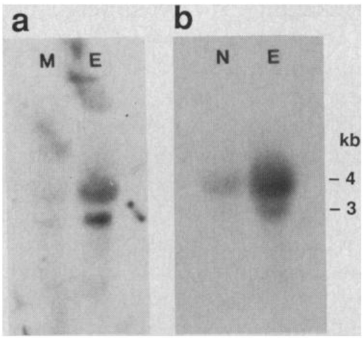


Fig. 1. RNA gel blot analysis of embryonic tissues. (**a**) Mesenchyme blastula embryos (27 hours) were fractionated into mesenchyme (M) and ectoderm (E) (*5*). (**b**) Early pluteus-stage embryos (68 hours) were fractioned into a mesenchyme-endoderm fraction (N) and ectoderm (E) (*6*). Total RNA was extracted from each fraction, electrophoresed in formaldehyde gels, and transferred to nitrocellulose. Ten micrograms of RNA was loaded from each fraction. Gels were stained with ethidium to monitor concentration. RNA extractions, blots, and hybridizations were as described (*26*).

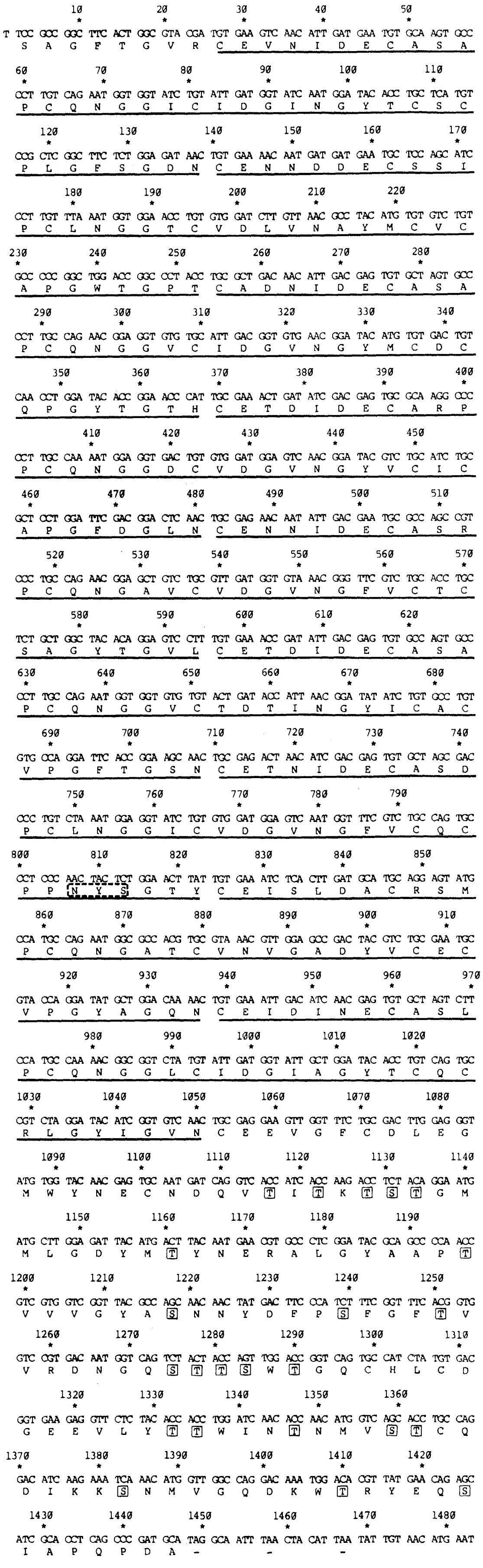
Institute for Molecular and Cellular Biology, Department of Biology, Indiana University, Bloomington, Indiana 47405.

the algorithm of Lipman and Pearson (*8*) indicated strong homology to EGF and other EGF-related proteins. The homology is restricted to an approximately 38–amino acid cysteine-rich repeat common to these proteins. The characteristic feature of this repeat consists of the distribution of six cysteine residues in the pattern $CX_4CX_5CX_8CXCX_8CX_6$ (*2*). One such repeat comprises most of the mature EGF polypeptide and is repeated nine more times in the EGF precursor (*9*).

Other members of the EGF family contain variable numbers of this repeat. Vaccinia virus protein 19 (VVGF) (*10*), transforming growth factor–α (TGF-α) (*11*), and tissue-type plasminogen activator (tPA) (*12*) each contain one copy, while the mammalian clotting factors IX and X have two (*13, 14*), the low-density lipoprotein receptor (LDLR) has three (*15*), and the ***notch*** gene product has 36 (*2*).

Figure 2 shows the nucleotide sequence of uEGF-1 with its predicted protein sequence. The 1100 nucleotides nearest the 5′ end encode nine 38–amino acid repeats of the EGF-like motif. In addition we have identified one potential site of *N*-linked glycosylation (Asn-Tyr-Ser) at nucleotide position 803. There are no obvious sites of proteolytic cleavage surrounding any of the repeats, unlike the situation in EGF (*9*) or TGF-α (*11*). Within uEGF-1 the repeat structure is highly conserved, the spacing of the cysteine residues being identical in all nine sequenced uEGF-1 repeats. The identities of most of the other repeat residues are also highly conserved. Figure 3 shows the alignment of the nine repeats of uEGF-1. It appears that only three positions in the 38–amino acid repeat readily tolerate amino acid substitutions. Overall, very little variation exists among the cysteine repeats of the uEGF-1 protein.

A comparison of the cysteine repeat of uEGF-1 with those of other members of the EGF family is shown in Fig. 4. The register of the cysteines in the uEGF-1 repeat is most homologous to ***notch*** and coagulation factors IX and X, although amino acid similarities between uEGF-1 and all the listed members can be seen. The similarity to coagulation factors extends to the positioning of the repeats with respect to one another. Two recently reported sequences of the blood coagulation factors thrombomodulin (*16*) and protein S (*17*) contain short tandem repeats of the EGF-like domain. Uninterrupted tandem repetition is not seen in any other members of the EGF family from vertebrates, and resembles the structure of the cysteine repeats found in ***notch***, ***lin 12***, and uEGF. The members of the EGF family found in invertebrates form a distinct sub-

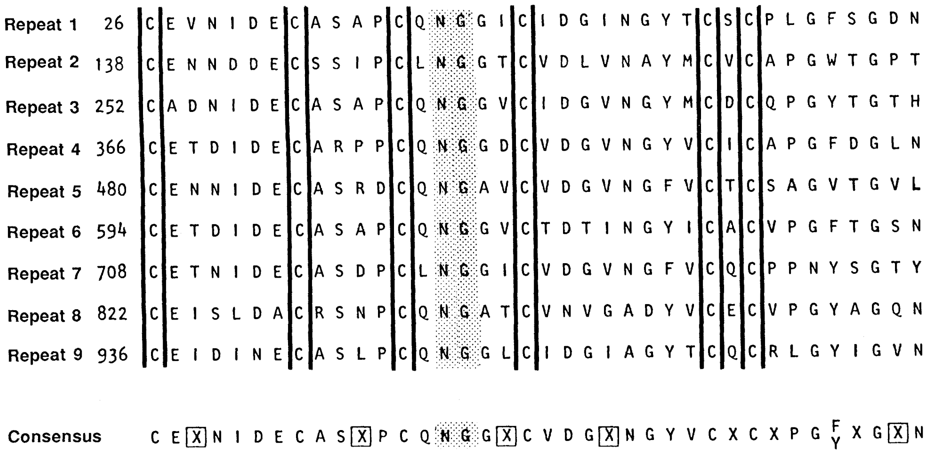


```
           10            20           30           40            50
            *             *            *            *             *
T TCC GCC GGC TTC ACT GGC GTA CGA TGT GAA GTC AAC ATT GAT GAA TGT GCA AGT GCC
   S   A   G   F   T   G   V   R   C   E   V   N   I   D   E   C   A   S   A

60           70            80           90           100           110
 *            *             *            *            *             *
CCT TGT CAG AAT GGT GGT ATC TGT ATT GAT GGT ATC AAT GGA TAC ACC TGC TCA TGT
 P   C   Q   N   G   G   I   C   I   D   G   I   N   G   Y   T   C   S   C

    120          130           140          150          160           170
     *            *             *            *            *             *
CCG CTC GGC TTC TCT GGA GAT AAC TGT GAA AAC AAT GAT GAT GAA TGC TCC AGC ATC
 P   L   G   F   S   G   D   N   C   E   N   N   D   D   E   C   S   S   I

        180          190           200          210          220
         *            *             *            *            *
CCT TGT TTA AAT GGT GGA ACC TGT GTG GAT CTT GTT AAC GCC TAC ATG TGT GTC TGT
 P   C   L   N   G   G   T   C   V   D   L   V   N   A   Y   M   C   V   C

230         240          250           260          270          280
*            *            *             *            *            *
GCC CCC GGC TGG ACC GGC CCT ACC TGC GCT GAC AAC ATT GAC GAG TGT GCT AGT GCC
 A   P   G   W   T   G   P   T   C   A   D   N   I   D   E   C   A   S   A

   290          300          310           320          330          340
    *            *            *             *            *            *
CCT TGC CAG AAC GGA GGT GTG TGC ATT GAC GGT GTG AAC GGA TAC ATG TGT GAC TGT
 P   C   Q   N   G   G   V   C   I   D   G   V   N   G   Y   M   C   D   C

       350          360          370           380          390          400
        *            *            *             *            *            *
CAA CCT GGA TAC ACC GGA ACC CAT TGC GAA ACT GAT ATC GAC GAG TGC GCA AGG CCC
 Q   P   G   Y   T   G   T   H   C   E   T   D   I   D   E   C   A   R   P

           410          420          430           440          450
            *            *            *             *            *
CCT TGC CAA AAT GGA GGT GAC TGT GTG GAT GGA GTC AAC GGA TAC GTC TGC ATC TGC
 P   C   Q   N   G   G   D   C   V   D   G   V   N   G   Y   V   C   I   C

 460           470          480          490           500          510
  *             *            *            *             *            *
GCT CCT GGA TTC GAC GGA CTC AAC TGC GAG AAC AAT ATT GAC GAA TGC GCC AGC CGT
 A   P   G   F   D   G   L   N   C   E   N   N   I   D   E   C   A   S   R

     520           530          540          550           560          570
      *             *            *            *             *            *
CCC TGC CAG AAC GGA GCT GTC TGC GTT GAT GGT GTA AAC GGG TTC GTC TGC ACC TGC
 P   C   Q   N   G   A   V   C   V   D   G   V   N   G   F   V   C   T   C

         580           590          600          610           620
          *             *            *            *             *
TCT GCT GGC TAC ACA GGA GTC CTT TGT GAA ACC GAT ATT GAC GAG TGT GCC AGT GCC
 S   A   G   Y   T   G   V   L   C   E   T   D   I   D   E   C   A   S   A

630          640           650          660          670           680
 *            *             *            *            *             *
CCT TGC CAG AAT GGT GGT GTG TGT ACT GAT ACC ATT AAC GGA TAT ATC TGT GCC TGT
 P   C   Q   N   G   G   V   C   T   D   T   I   N   G   Y   I   C   A   C

    690          700           710          720          730           740
     *            *             *            *            *             *
GTG CCA GGA TTC ACC GGA AGC AAC TGC GAG ACT AAC ATC GAC GAG TGT GCT AGC GAC
 V   P   G   F   T   G   S   N   C   E   T   N   I   D   E   C   A   S   D

        750          760           770          780          790
         *            *             *            *            *
CCC TGT CTA AAT GGA GGT ATC TGT GTG GAT GGA GTC AAT GGT TTC GTC TGC CAG TGC
 P   C   L   N   G   G   I   C   V   D   G   V   N   G   F   V   C   Q   C

800         810          820           830          840          850
*            *            *             *            *            *
CCT CCC AAC TAC TCT GGA ACT TAT TGT GAA ATC TCA CTT GAT GCA TGC AGG AGT ATG
 P   P   N   Y   S   G   T   Y   C   E   I   S   L   D   A   C   R   S   M

   860          870          880           890          900          910
    *            *            *             *            *            *
CCA TGC CAG AAT GGC GCC ACG TGC GTA AAC GTT GGA GCC GAC TAC GTC TGC GAA TGC
 P   C   Q   N   G   A   T   C   V   N   V   G   A   D   Y   V   C   E   C

       920          930          940           950          960          970
        *            *            *             *            *            *
GTA CCA GGA TAT GCT GGA CAA AAC TGT GAA ATT GAC ATC AAC GAG TGT GCT AGT CTT
 V   P   G   Y   A   G   Q   N   C   E   I   D   I   N   E   C   A   S   L

           980          990         1000          1010         1020
            *            *            *             *            *
CCA TGC CAA AAC GGC GGT CTA TGT ATT GAT GGT ATT GCT GGA TAC ACC TGT CAG TGC
 P   C   Q   N   G   G   L   C   I   D   G   I   A   G   Y   T   C   Q   C

1030          1040         1050         1060          1070         1080
  *             *            *            *             *            *
CGT CTA GGA TAC ATC GGT GTC AAC TGC GAG GAA GTT GGT TTC TGC GAC TTG GAG GGT
 R   L   G   Y   I   G   V   N   C   E   E   V   G   F   C   D   L   E   G

    1090          1100         1110         1120          1130         1140
      *             *            *            *             *            *
ATG TGG TAC AAC GAG TGC AAT GAT CAG GTC ACC ATC ACC AAG ACC TCT ACA GGA ATG
 M   W   Y   N   E   C   N   D   Q   V   T   I   T   K   T   S   T   G   M

        1150          1160         1170         1180          1190
          *             *            *            *             *
ATG CTT GGA GAT TAC ATG ACT TAC AAT GAA CGT GCC CTC GGA TAC GCA GCC CCA ACC
 M   L   G   D   Y   M   T   Y   N   E   R   A   L   G   Y   A   A   P   T

1200        1210          1220         1230         1240          1250
 *            *             *            *            *             *
GTC GTG GTC GGT TAC GCC AGC AAC AAC TAT GAC TTC CCA TCT TTC GGT TTC ACG GTG
 V   V   V   G   Y   A   S   N   N   Y   D   F   P   S   F   G   F   T   V

   1260         1270          1280         1290         1300          1310
     *            *             *            *            *             *
GTC CGT GAC AAT GGT CAG TCT ACT ACC AGT TGG ACC GGT CAG TGC CAT CTA TGT GAC
 V   R   D   N   G   Q   S   T   T   S   W   T   G   Q   C   H   L   C   D

       1320         1330          1340         1350         1360
         *            *             *            *            *
GGT GAA GAG GTT CTC TAC ACC ACC TGG ATC AAC ACC AAC ATG GTC AGC ACC TGC CAG
 G   E   E   V   L   Y   T   T   W   I   N   T   N   M   V   S   T   C   Q

1370       1380         1390          1400         1410         1420
*            *            *             *            *            *
GAC ATC AAG AAA TCA AAC ATG GTT GGC CAG GAC AAA TGG ACA CGT TAT GAA CAG AGC
 D   I   K   K   S   N   M   V   G   Q   D   K   W   T   R   Y   E   Q   S

  1430         1440         1450          1460         1470         1480
    *            *            *             *            *            *
ATC GCA CCT CAG CCC GAT GCA TAG GCA ATT TAA CTA CAT TAA TAT TGT AAC ATG AAT
 I   A   P   Q   P   D   A   -           -           -
```

Fig. 2. Nucleotide sequence and predicted amino acid sequence of the uEGF-1 cDNA. Nucleotides are numbered above the sequence. A single asterisk denotes each ten amino acids. The termination codons are indicated by a hyphen. The nine repeated units of the EGF-like domain are underlined. A potential *N*-linked glycosylation site is indicated by the dashed box. The region of 110 residues rich in serines and threonines to which *O*-linked sugars could be attached is indicated with boxes surrounding those residues. The single letter amino acid codes used are: A, alanine; R, arginine; N, asparagine; D, aspartic acid; C, cysteine; Q, glutamine; E, glutamic acid; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; F, phenylalanine; P, proline; S, serine; T, threonine; W, tryptophan; Y, tyrosine; and V, valine. The uEGF-1 cDNA clone was isolated from a λgt11 mesenchyme blastula cDNA library by standard protocols. The probe was a cDNA obtained by a differential screen designed to isolate transcripts restricted to the embryonic lineages of the sea urchin (*31*). The uEGF-1 cDNA clone was subcloned into M13 (*32*) for sequencing by the dideoxy chain termination method (*33, 34*) as modified by Biggin *et al.* (*35*). Sequences were analyzed by use of the Pustell programs (International Biotechnologies, New Haven, Connecticut).

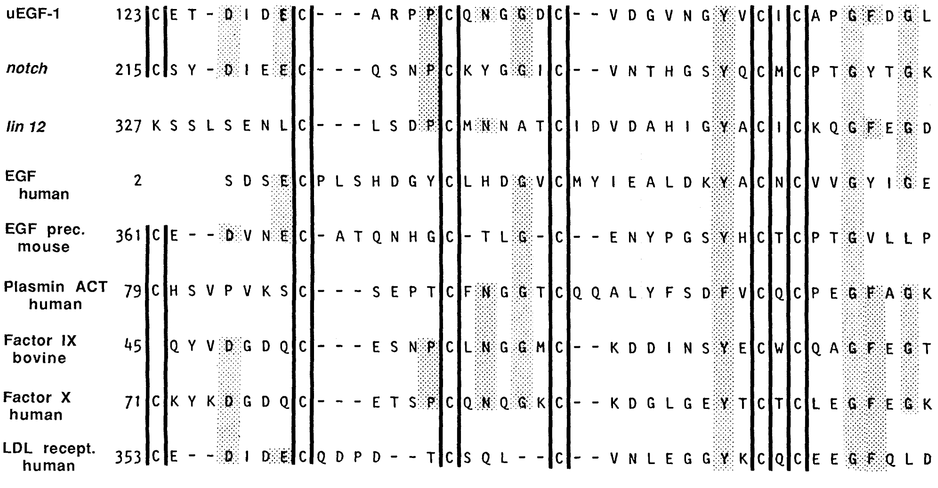
group, as judged by the lengths of tandemly repeated EGF-like domains. The *notch* gene contains 36 repeats, *lin 12* contains 11, and a partial sequence of a longer uEGF cDNA clone indicates that it repeats this pattern for at least another 1 kb, suggesting that the uEGF protein may contain a minimum of 18 contiguous repeats. In comparison with *notch* and *lin 12*, however, the uEGF-1 repeats are more rigidly conserved. Each uEGF-1 repeat contains exactly 38 amino acids. Furthermore, the register of the six cysteines within the repeat remains perfect throughout the nine sequenced repeats, unlike the situation in *notch* and *lin 12*. Similarly the uEGF-1 repeats are less divergent in the noncysteine residues. The basis of such constraint is unclear, although Stenflo *et al.* (*18*) suggested that the post-translational hydroxylation of aspartic acid and asparagine residues within the EGF-like repeat of various plasma proteins may require a specific cysteine register and the conservation of particular amino acids around the hydroxylated residues. Figure 3 shows that uEGF-1 contains many conserved aspartic acid and asparagine residues that might be sites for hydroxylation.

The 120 amino acids at the carboxyl terminal of the uEGF-1 ORF represent a distinct domain of the protein. This region is rich in threonine; 45% of the threonine residues of the predicted uEGF-1 protein appear in this second domain, which comprises only 25% of the ORF. The threonine residues are found in clusters of two or more, often interspersed with serine residues. Clusters such as this are thought to be sites of *O*-linked glycosylation (*13*, *19*). A scan of the PIR database with this domain of uEGF-1 did not reveal any significant homologies to known proteins.



Repeat 1	26	C E V N I D E C A S A P C Q N G G I C I D G I N G Y T C S C P L G F S G D N
Repeat 2	138	C E N N D D E C S S I P C L N G G T C V D L V N A Y M C V C A P G W T G P T
Repeat 3	252	C A D N I D E C A S A P C Q N G G V C I D G V N G Y M C D C Q P G Y T G T H
Repeat 4	366	C E T D I D E C A R P P C Q N G G D C V D G V N G Y V C I C A P G F D G L N
Repeat 5	480	C E N N I D E C A S R D C Q N G A V C V D G V N G F V C T C S A G V T G V L
Repeat 6	594	C E T D I D E C A S A P C Q N G G V C T D T I N G Y I C A C V P G F T G S N
Repeat 7	708	C E T N I D E C A S D P C L N G G I C V D G V N G F V C Q C P P N Y S G T Y
Repeat 8	822	C E I S L D A C R S N P C Q N G A T C V N V G A D Y V C E C V P G Y A G Q N
Repeat 9	936	C E I D I N E C A S L P C Q N G G L C I D G I A G Y T C Q C R L G Y I G V N
Consensus		C E $\boxed{X}$ N I D E C A S $\boxed{X}$ P C Q N G G $\boxed{X}$ C V D G $\boxed{X}$ N G Y V C X C X P G $^{F}_{Y}$ X G $\boxed{X}$ N

Fig. 3. Alignment of repeated sequence within uEGF-1. The cysteine repeat register is indicated by heavy bars surrounding those residues. Invariant noncysteine amino acids are shaded. A consensus sequence is presented at the bottom and is based on the presence of an amino acid in ≥50% of the sequenced repeats. Boxed X's indicate substitutions limited to nonpolar amino acids. X's without boxes indicate nonconservative substitution at that position. The numbers at the beginning of each repeat indicate the first nucleotide of that repeat.



uEGF-1	123	C E T - D I D E C - - - A R P P C Q N G G D C - - V D G V N G Y V C I C A P G F D G L
notch	215	C S Y - D I E E C - - - Q S N P C K Y G G I C - - V N T H G S Y Q C M C P T G Y T G K
lin 12	327	K S S L S E N L C - - - L S D P C M N N A T C I D V D A H I G Y A C I C K Q G F E G D
EGF human	2	S D S E C P L S H D G Y C L H D G V C M Y I E A L D K Y A C N C V V G Y I G E
EGF prec. mouse	361	C E - - D V N E C - A T Q N H G C - T L G - C - - E N Y P G S Y H C T C P T G V L L P
Plasmin ACT human	79	C H S V P V K S C - - - S E P T C F N G G T C Q Q A L Y F S D F V C Q C P E G F A G K
Factor IX bovine	45	Q Y V D G D Q C - - - E S N P C L N G G M C - - K D D I N S Y E C W C Q A G F E G T
Factor X human	71	C K Y K D G D Q C - - - E T S P C Q N Q G K C - - K D G L G E Y T C T C L E G F E G K
LDL recept. human	353	C E - - D I D E C Q D P D - - T C S Q L - - C - - V N L E G G Y K C Q C E E G F Q L D

Fig. 4. Alignment of uEGF-1 with other EGF-like proteins. With the exception of *notch* and *lin 12*, all protein sequences depicted were obtained by searching the PIR database (fastP program) with the fourth repeat of uEGF-1. The fourth repeat was chosen as a representative repeat to avoid using unspecified amino acids in the analysis. A *notch* repeat was also chosen because of its similarity to the *notch* consensus (*2*). The *lin 12* repeat used here was the repeat analyzed (*4*). Hyphens indicate gaps introduced to maximize homology. Cysteines are indicated with heavy bars. Amino acids which occur in ≥50% of the sequences are shaded. Numbers at the beginning of every sequence indicate the number of that amino acid residue within the total protein sequence.

The members of the EGF family are thought to execute their function extracellularly either as membrane-bound proteins (EGF, TGF-α, VVGF, LDLR, and *notch*) or as secreted proteins (factor X and mature EGF). A computer-aided analysis of uEGF-1, in which the hydropathy algorithm of Kyte and Doolittle (*20*) was performed by the programs of Pustell (International Biotechnologies), does not reveal a classic transmembrane domain as defined by those workers. Because we do not have complete cDNA clones for the uEGF-1 protein, we do not know whether the uEGF protein contains a signal sequence. The presence of so many potential *O*-linked glycosylation sites clustered in a discrete region is suggestive of an extracellular protein, but the cellular location of uEGF is currently unknown.

The temporal pattern of expression of transcripts recognized by the uEGF-1 cDNA clone is consistent with a function during early sea urchin development, a time when cell fates are known to be assigned (*21*). Figure 5a represents an RNA gel blot of staged embryonic RNAs probed with the uEGF-1 cDNA clone. Restriction fragments from all portions of the clone recognize both the 3.0- and the 4.0-kb transcripts. Whether these transcripts emanate from different genes or are alternatively spliced messages from a single gene cannot be determined at present. Figure 5b shows a quantitation of the RNA gel blot data. The transcripts recognized by uEGF-1 are at their highest level in the unfertilized egg, decline in prevalence during early cleavage, and increase in prevalence at the blastula stage before tapering off after gastrulation. This is an unusual transcript accumulation pattern for a sea urchin messenger RNA (mRNA). Studies of randomly chosen cDNAs have shown that many prevalent messages accumulate early, decay, and are replaced later in development (*22*), but none have been reported to do so within the first 6 hours of development. The rate of transcription per cell in sea urchins is virtually unchanged throughout embryogenesis (*23*). Thus uEGF transcript accumulation cannot be explained by changes in the transcriptional ability of zygotic nuclei, as would be the case in *Xenopus* or *Drosophila*. The accumulation of histone message is also shown in Fig. 5b. The

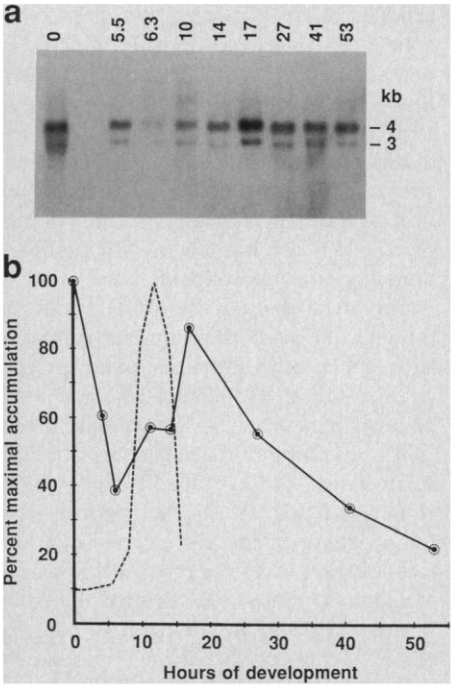


Fig. 5. Time course of uEGF transcript accumulation. (**a**) RNA gel blot analysis, carried out as described in Fig. 1. Numbers above the lanes indicate hours of development at 13°C. These times represent the following developmental stages; 0 hours, unfertilized egg; 5.5 hours, 8-cell stage; 6.3 hours, 16-cell stage; 10 hours, late-cleavage stage; 14 hours, prehatching blastula; 17 hours, late-hatched blastula; 27 hours, mesenchyme blastula; 41 hours, early gastrula; 53 hours, late gastrula. (**b**) Quantification by densitometry of RNA gel blot data. The ordinate represents hours of development at 13°C. The abscissa represents arbitrary values relative to the maximal accumulation level. (●) The behavior of the uEGF transcript; the results shown are the average of two blots, one of which is shown in (a). The dashed line represents the accumulation of the α histones and was plotted from (*36*).

synthesis of histones in this embryo correlates with the rate of cell division (*24*). Because the second period of uEGF message accumulation occurs after the rapid cell divisions of early cleavage, we do not expect uEGF to be operating solely as a stimulator of cell proliferation in this system. The transcripts of many lineage-specific genes in sea urchins first appear at the blastula stage; they accumulate asynchronously, but their midpoints of accumulation all occur after the midpoint of accumulation of the uEGF transcript (*25–27*).

Some mammalian members of the EGF family are thought to act as regulators of cell proliferation and differentiation (*1, 28*). The EGF-like molecules encoded by ***notch*** and ***lin 12*** are involved in the specification of embryonic cell fates (*29, 30*). The behavior of the uEGF transcript, whose accumulation precedes known molecular indices of cell differentiation, as well as the structural similarity of the uEGF gene to other invertebrate binary switch genes, suggests that it too may play an important role in embryogenesis. Whatever its function, the existence of this highly conserved sequence in the echinoderms indicates that the EGF peptide domain predates the radiation of coelomate animals and has been retained in a number of diverse evolutionary lineages.

REFERENCES AND NOTES

1. G. Carpenter and S. Cohen, *Annu. Rev. Biochem.* **48**, 193 (1979).
2. K. A. Wharton, K. M. Johansen, T. Xu, S. Artavanis-Tsakonas, *Cell* **43**, 567 (1985).
3. S. Kidd, M. R. Kelley, M. W. Young, *Mol. Cell Biol.* **6**, 3094 (1986).
4. I. Greenwald, *Cell* **43**, 583 (1985).
5. M. A. Harkey and A. H. Whiteley, *Wilhelm Roux Arch. Dev. Biol.* **189**, 111 (1980).
6. D. McClay and A. F. Chambers, *Dev. Biol.* **63**, 179 (1978).
7. T. Maruyama, T. Gojobori, S.-I. Aota, T. Ikemura, *Nucleic Acids Res.* **14** (suppl.) r151 (1986).
8. D. J. Lipman and W. R. Pearson, *Science* **227**, 1435 (1985).
9. R. F. Doolittle, D. F. Feng, M. S. Johnson, *Nature (London)* **307**, 558 (1984).
10. J. P. Brown, D. R. Twardzik, H. Marquardt, G. J. Todaro, *ibid.* **313**, 491 (1985).
11. H. Marquardt, M. W. Hunkapiller, L. E. Hood, G. J. Todaro, *Science* **223**, 1079 (1984).
12. D. Pennica *et al.*, *Nature (London)* **301**, 214 (1983).
13. B. A. McMullen *et al.*, *Biochemistry* **22**, 2875 (1983).
14. D. L. Enfield *et al.*, *ibid.* **19**, 659 (1980).
15. D. W. Russell *et al.*, *Cell* **37**, 577 (1985).
16. R. W. Jackman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8834 (1986).
17. J. Hoskins, D. K. Norman, R. J. Beckman, G. L. Long, *ibid.* **84**, 349 (1987).
18. J. Stenflo, A. Lundwall, B. Dahlback, *ibid.*, p. 368.
19. T. Nikaido *et al.*, *Nature (London)* **311**, 631 (1985).
20. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
21. R. C. Angerer and E. H. Davidson, *Science* **226**, 1153 (1984).
22. C. N. Flytzanis, B. P. Brandhorst, R. J. Britten, E. H. Davidson, *Dev. Biol.* **91**, 27 (1982).
23. E. H. Davidson, *Gene Activity in Early Development* (Academic Press, New York, 1976).
24. B. Moav and M. Nemer, *Biochemistry* **10**, 881 (1971).
25. W. H. Klein *et al.*, in *Time, Space, and Pattern in Embryonic Development*, W. R. Jeffery and R. A. Raff, Eds. (Liss, New York, 1983), pp. 87–100.
26. R. J. Shott *et al.*, *Dev. Biol.* **101**, 295 (1984).
27. D. S. Leaf *et al.*, *ibid.* **121**, 29 (1987).
28. D. R. Twardzik, J. E. Ranchalis, G. J. Todaro, *Cancer Res.* **42**, 590 (1982).
29. I. S. Greenwald, P. W. Sternberg, H. R. Horvitz, *Cell* **34**, 435 (1983).
30. B. Yedvobnik *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **50**, 841 (1985).
31. D. A. Hursh and R. A. Raff, in preparation.
32. J. Messing, *Methods Enzymol.* **101**, 20 (1983).
33. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
34. A. J. H. Smith, *Methods Enzymol.* **65**, 560 (1980).
35. M. D. Biggin, T. J. Gibson, G. F. Hong, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963 (1983).
36. D. E. Wells *et al.*, *Wilhelm Roux Arch. Dev. Biol.* **195**, 252 (1986).

37. We thank T. Thomas for providing the cDNA libraries used in this study. We thank E. Raff, B. Parr, D. Leaf, P. Cherbas, and M. Muskavitch for critical comments regarding the manuscript. Supported by NIH grant R01 HD21986. D.A.H. was supported by an NIH predoctoral training grant.

23 March 1987; accepted 26 June 1987

Acantharian Fluxes and Strontium to Chlorinity Ratios in the North Pacific Ocean

R. E. Bernstein, P. R. Betzer, R. A. Feely, R. H. Byrne,
M. F. Lamb, A. F. Michaels

Data on particulate strontium sulfate fluxes and strontium to chlorinity ratios were compared to provide insights into the strontium cycle of the North Pacific. Free-drifting sediment traps were used to derive large particle fluxes between depths of 100 and 3500 meters in the eastern and western North Pacific Ocean. Flux data revealed substantial quantities of acantharian skeletons and cysts (both made of strontium sulfate) settling through the upper kilometer of the water column. The greatest fluxes of celestite were detected at 400 meters. Minimal to nondetectable fluxes noted at and below 900 meters provide evidence that by this horizon, the majority of acantharian specimens had dissolved, thereby contributing to the pool of dissolved strontium. Growth and subsequent dissolution of acantharians in the upper kilometer are qualitatively consistent with the well-developed minimum and maximum strontium to chlorinity ratios that are consistently noted in these waters. These fluxes of particulate strontium and model calculations for fluxes of dissolved strontium indicate that acantharians play an important role in the ocean's strontium budget.

THE GLOBAL OCEANIC STRONTIUM budget may be dramatically affected by the only marine organisms to use Sr as a major skeletal component. These abundant organisms, the acantharians, are marine planktonic protists that secrete a

R. E. Bernstein, P. R. Betzer, R. H. Byrne, Department of Marine Science, University of South Florida, St. Petersburg, FL 33701.
R. A. Feely and M. F. Lamb, Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA 98115.
A. F. Michaels, Institute of Marine Sciences, University of California, Santa Cruz, CA 95060.