Gene, 141 (1994) 243–248

GENE 07679

Cloning and sequencing of a cell surface protein-encoding gene conserved in sea urchin species

(Recombinant DNA; *Paracentrotus lividus*; multigene family; maternal RNA; promoter region)

Marta Di Carlo, Salvatore Perriera, Giovanna Montana, Daniele Paolo Romancino, Stefano Reale

Istituto di Biologia dello Sviluppo, C.N.R., 90123 Palermo, Italy

Received by G. Bernardi: 27 April 1993; Revised/Accepted: 28 June 1993/2 July 1993; Received at publishers: 19 October 1993

SUMMARY

We report the nucleotide sequence of a fragment of DNA derived from a sea urchin genomic clone containing the cell surface Bep4 (butanol-extracted protein 4)-encoding gene. The structural gene is interrupted by four introns and the promoter region contains TATA and CAAT consensus motifs. The transcription start point (*tsp*) was also determined. Remarkable homologies, between Bep4 and other proteins known to be involved in cell interactions, were observed regarding two poteintial Ca^{2+}-binding sites and the corresponding DNA consensus sequences. We also report the conservation of the *bep4* gene and its corresponding Bep4 protein between various sea urchin species by way of Southern and Western blotting.

INTRODUCTION

Normal development of an organism is dependent upon cell-to-cell interactions. These interactions play an integral role during development as they mediate the assembly and organization of cells during morphogenesis. Dynamic changes in cell adhesion and cell migration occur in the course of sea urchin embryogenesis (Gustafson and Wolpert, 1967), i.e., proteins employed in cell to cell interaction in early embryonic cells are replaced as development and differentiation proceeds (McClay et al., 1977; Giudice, 1973; 1984). One likely explanation for this observation is that the different cell interactions during the early sea urchin morphogenesis are mediated by the synthesis of different cell interaction proteins during development. In order to test this hypothesis genes encoding for cell-surface proteins from various organisms have been cloned to study expression of proteins in different tissues and during development (reviewed in Edelman, 1986). Several genes of particular interest have been cloned from sea urchins, including a gene encoding a protein involved in cell adhesion of endoderm and secondary mesenchyme cells, designated as *Endo 16*, which is expressed between blastula and pluteus stages (Nocente-McGrath et al., 1989) and a gene coding a cell surface glycoprotein, called *msp 130*, which produces transcripts that can be detected in premesenchyme-blastula embryos and that is significantly accumulated after primary mesenchyme cell ingression (Leaf et al., 1987; Parr et al., 1989). No genes, however, coding for cell surface proteins, transcribed by maternal messengers, have been cloned yet.

We have previously isolated from an ovary cDNA library two clones, called *bep1* and *bep4* (butanol extracted proteins) corresponding to two cell surface proteins involved in cell interactions (Di Carlo et al., 1990; Romancino et al., 1992). Reaggregation of dissociated *P.*

Correspondence to: Dr. M. Di Carlo, Istituto di Biologia dello Sviluppo, C.N.R., via Archirafi 20, 90123 Palermo, Italy. Tel. (39-91) 616-4501; Fax (39-91) 616-5665.

Abbreviations: aa, amino acid(s); bp, base pair(s); Bep4, butanol-extracted protein 4; *bep4*, gene encoding Bep4; cDNA, DNA complementary to mRNA; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; *P.*, *Paracentrotus*; PA, polyacrylamide; PAGE, PA-gel electrophoresis; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na_3-citrate pH 7.6; *tsp*, transcription start point(s).

SSDI 0378-1119(93)E0675-4

lividus blastula embryos is inhibited by Fab fragments raised against these proteins.These two clones are single-copy genes and are members of a multigene family. By Northern blot analysis we found that they are transcribed during oogenesis, stored in the egg and utilized until gastrulation. By their deduced amino acid sequences they show two well-conserved domains flanking single specific domains. The aim of present study was to characterize and sequence the *bep4* cell-surface gene and analyse its promoter region.

EXPERIMENTAL AND DISCUSSION

(a) Characterization of an isolated clone

A restriction fragment of the *bep4* clone, generated by *Sal*I + *Stu*I, containing the specific region of the cDNA, was utilized to screen a genomic library of *P. lividus* sperm DNA in λEMBL3 phage (Di Carlo et al., 1990). Three clones were isolated and one of these, called λb45A was rigorously analysed. By hybridization with two fragments containing sequences present at the 5′ end and 3′ end of *bep4* cDNA, we found that λb45A is interrupted at the 3′ end. The same results were obtained with the others clones isolated. The restriction map of λb45A is shown in Fig. 1.

(b) Sequencing of λb45A

Restriction fragments obtained according to the restriction map shown in Fig. 1 were subcloned in pUC18 and M13 derived vectors. The entire nt sequences of the cloned fragments are shown in Fig. 2. The size and the positions of the introns (IVS) I, II, III, IV, are shown in Fig. 2. The IVS interrupt the cDNA sequence, at the positions immediately following nt 571, 1425, 2105, 3083, respectively. The λb45A clone is truncated in the fourth intron of at least 2.7-kb long, i.e., the last 361 bp of *bep4* cDNA are absent. We are sure that this clone is not a pseudogene because *bep4* is a single-copy gene as demonstrated in our previous paper (Di Carlo et al., 1990). Canonic consensus of splicing sites in the λb45A clone, shown in Table I, are not always well conserved. This lack of conservation of consensus sequences has been previously reported in other organisms (Mount et al., 1982).

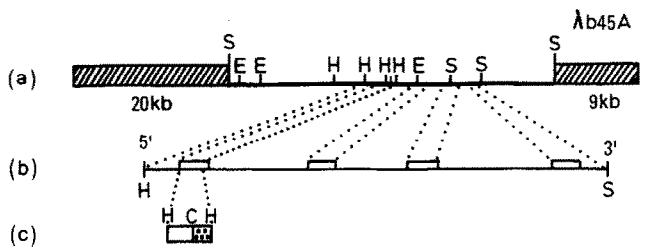

Fig. 1. Representation of the λb45A genomic clone. **(a)** Restriction map of λb45A; S (*Sal*I), E (*Eco*RI), H (*Hin*dIII), C (*Cla*I). **(b)** Structure of the sequenced region. Open boxes and lines indicate exons and introns, respectively. **(c)** Location of the *Hin*dIII-*Hin*dIII and *Cla*I-*Hin*dIII fragments utilized for S1 mapping and as probe for rescreening of the cDNA library is shown as an open box and a dotted box, respectively.

From the deduced aa sequence we found two possible Ca^{2+}-binding motifs, corresponding to nt 343 and 460 (Fig. 3A). The cluster of the type DXD(N)XD(N) is very frequent in many EF-hand domains (Tufty and Kretsinger, 1975) and has been observed in α-lactalbumin (Stuart et al., 1986) and in Ca^{2+}-dependent membrane proteins such as endonexin, calectrin (Geisow et al., 1986) and uvomorulin (Ringwald et al., 1987). This cluster is necessary for the coordination of Ca^{2+} ions by side chain oxygen atoms in the Ca^{2+}-binding sites. The aa of these motifs form a loop structure in which the ligands residues are located (Tufty and Kretsinger, 1975). The analysis of secondary structure prediction according to the Chou and Fasman (1979) method of the two sequence motifs assigned as putative Ca^{2+} binding sites of BEP4, shows that these sites are located in regions for which β-sheet turn or coils are predicted (Fig. 3B).

The possibility that the Bep4 protein can bind Ca^{2+} ions may be important for its function, which has been demonstrated to play a role in cell-to-cell interaction (Romancino et al., 1992). The Bep4 protein could interact via Ca^{2+} sites either with a receptor or with other cell adhesion molecules such as cadherin-like proteins which are present from early stages of sea urchin development (Ghersi and Vittorelli, 1990; Ghersi et al., 1993).

(c) Characterization of the promoter region

In order to determine the nt sequence of the 5′ end of *bep4* cDNA and analyse the promoter region of the gene, we rescreened an ovary cDNA library. For this screening we utilized, as a probe, the *Cla*I-*Hin*dIII fragment of the *bep4* gene (Fig. 1). This fragment is contained in the coding region of the gene as detected by hybridization to ovarian RNA (data not shown). In this way we isolated three clones, one of which corresponded to the full length cDNA. By comparison of the sequences of the new cDNA clone with the genomic clone we mapped an ATG codon at nt 277, indicating that a protein of 364 aa is translated, in agreement with the size seen in the Western blot of total sea urchin proteins incubated with polyclonal antibodies against Bep4 protein (Romancino et al., 1992). To determinate the exact location of *tsp* a primer extention experiment was performed. A 30-mer oligo, shown in Fig. 2, was labelled and utilized as a primer for reverse transcriptase (Fig. 4A). The *tsp* of the *bep4* gene was mapped at nt 215 of the *bep4* sequence, as shown in Fig. 2 producing transcripts with a 5′ untranslated region of

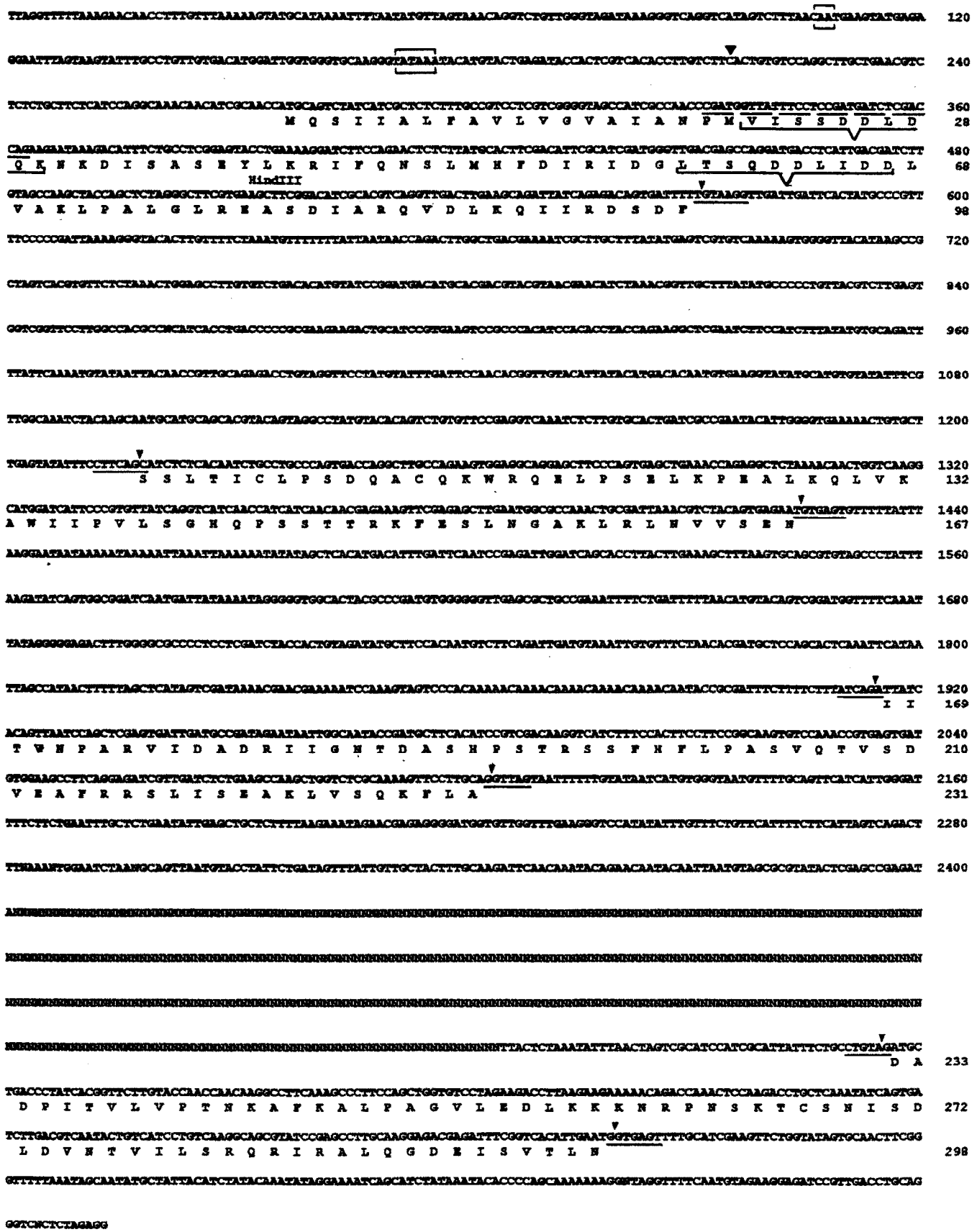

Fig. 2. The nt sequence of the *bep4* gene. The sequence was determinated by the dideoxy chain-termination method (Sanger et al., 1977; 1980) of single-stranded templates from M13 derived vectors using T7 DNA polymerase (Tabor and Richardson, 1987). The aa sequence deduced from the nt sequences is shown below. Chains of N indicate undetermined intron sequence. The TATA and CAAT consensus sequences are boxed. The wide black triangle indicates the *tsp*. The arrowheads mark the splice sites and the consensus sequences are underlined. The sequence utilized for the synthesis of the oligo primer is underlined with a broken line. Horizontal brackets represent potential Ca^{2+}-binding sites. The *Hin*dIII site of the fragment utilized as probe for S1 mapping is indicated. The nt sequence of *bep4* gene has been submitted to GenBank Nucleotide Sequence Database (accession No. L11925).

TABLE I

Introns and exons of the *bep4* gene[a]

Exon No.	Exon size (nt)	Donor	Intron size (nt)	Acceptor
1	356	T:GTAAGG	647	CTTCAG:C
2	206	T:GTGAGT	489	TATCAG:A
3	189	G:GTTAGT	770	CTGTAG:A
4	204	G:GTGAGT	174+n	-------

[a]Canonic consensus of splicing sites in the sea urchin *bep4* gene. S sizes of the four exons and introns are shown. Colons indicate the splicing points; +n indicates those nt which are not included in λb45A clone; the broken line indicates the next unknown acceptor site.

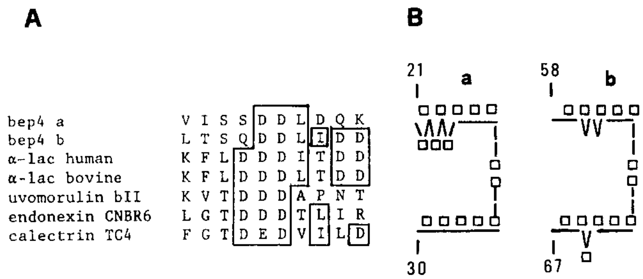

Fig. 3. Ca^{2+}-binding sites. (**A**) Alignment of two putative Ca^{2+}-binding sites of Bep4 with Ca^{2+}-binding domains of uvomorulin, human and bovine α-lactalbumin, endonexin and calectrin. (**B**) Schematic representation of two putative Ca^{2+} units **a** and **b** predicted by Chou and Fasman methods. Residues (□) predicted to be in β-structure = ΛV, turn = □, coils = □.

63 bp. This conclusion was confirmed by nuclease-S1 mapping of the *bep4* RNA (Fig. 4B). A set of promoter elements are situated in the region immediately upstream of the *tsp*, such as the classical TATA box at nt 171 and the CCAAT box at nt 106, i.e., 44 and 109 bp upstream of the *tsp*, respectively; no Spl-binding site is present in this sequence. In summary, the transcription from *bep4* DNA starting at nt 215 produces a 1318-nt messenger, which is in agreement with the RNA size predicted by Northern blotting (Di Carlo et al., 1990). It would be interesting to compare the sequences upstream of the *tsp* with the corresponding ones of the same gene family. It would be hoped to find some common sequences that are important for the regulation of some members of the *bep* gene family and perhaps of other maternal messengers. We cannot, in fact, rule out the possibility that some maternal regulatory elements are shared among the members of the gene family.

(d) Conservation of *bep4* gene

To investigate whether the *bep4* gene is conserved in other sea urchin species and in organisms phylogenetically distant, we isolated DNA from *Paracentrotus lividus*, *Arbacia lixula*, *Spherechinus granularis*, *Dictyostelium discoideum*, *Styela partita*, *Ceratitis capitata*, mouse and human placenta and digested them with *Eco*RI. The blot was hybridized with the *bep4* cDNA. Fig. 5A shows that the *bep4* gene is present in *S. granularis* but not in *A. lixula*, moreover a faint band is visible in *Dictyostelium* DNA, while it is absent in the other organisms. The same results were obtained when a fragment containing the specific region of *bep4* was hybridized to the same

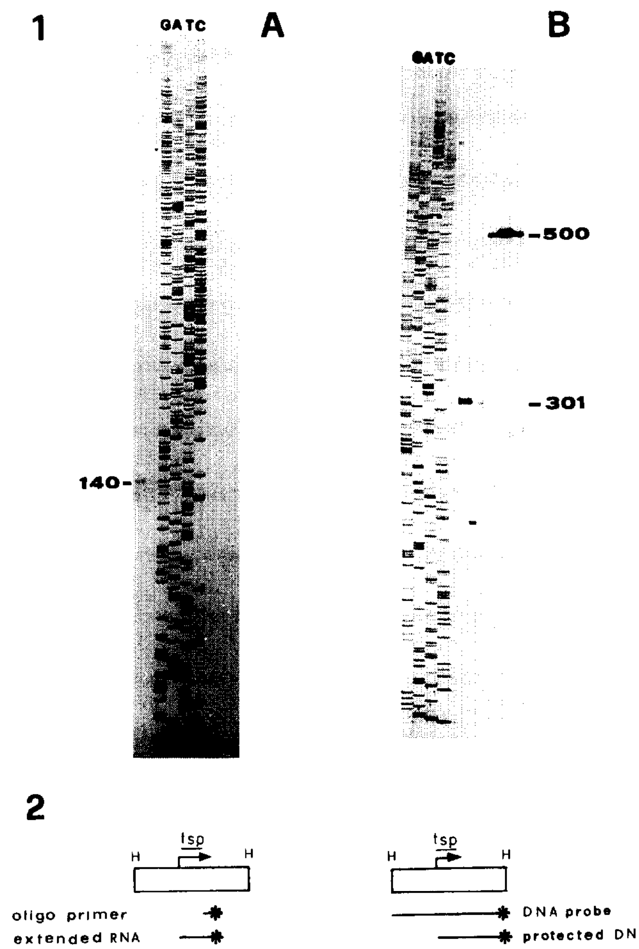

Fig. 4. Analysis (**1**) and schematic representation (**2**) of primer extention and S1 mapping. (**1**), Primer extention (**A**) and S1 mapping (**B**) of *bep4* gene. (**A**) A labeled oligo primer corresponding to sequences within exon I (Fig. 2) of the *bep4* gene was hybridized with egg poly(A)$^+$RNA (10 μg) at 30°C, extended with reverse transcriptase and the product electrophoresed on a denaturing 6% PA/8 M urea gel and the radioactive band detected by autoradiography. The lanes marked G A T C of panels (**A**) and (**B**) contain a known nt sequence ladder as a length marker. The length of the primer extention product is indicated. (**B**) Egg poly(A)$^+$RNA (10 μg) was analysed by nuclease S1. The *Hin*dIII fragment, shown in Fig. 1c, was labeled using a commercial '5' end labeling kit' (Boheringer, Mannheim) according to the vendor's recommendations. The annealing with egg poly(A)$^+$RNA (10 μg) was performed at 50°C. After treating with nuclease S1, under standard condition (Maniatis et al., 1982) the protected fragment was electrophoresed on a denaturing gel with a nt sequence ladder as marker. The length of the probe and the entire labeled fragment used are indicated. (**2**) The *tsp* is indicated by a bent arrow. The ^{32}P-labeled ends are indicated by asterisks.

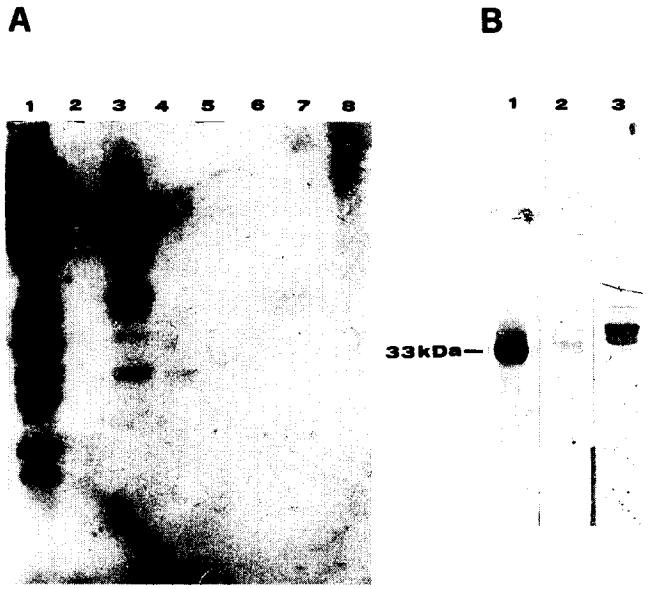

Fig. 5. Conservation of *bep4* gene. **(A)** DNAs (20 μg) isolated from (1) *Paracentrotus lividus*, (2) *Arbacia lixula*, (3) *Spherechinus granularis*, (4) *Dictyostelium discoideum*, (5) *Styela partita*, (6) *Ceratitis capitata*, (7) mouse, (8) human placenta, were digested with *Eco*RI, run in 0.8% agarose gel and blotted onto nylon filter (Hybond-N, Amersham). The blot was hybridized in 4 × SSC at 65°C, with the entire *bep4* fragment isolated according to Duro et al. (1991) and labeled by a random primer kit obtained from Promega, (Madison, WI USA). **(B)** Western blot of total proteins (50 μg) extracted from (1) *P. lividus*, (2) *A. lixula*, (3) *S. granularis* eggs run on 0.1% SDS-10% PAGE transferred onto nylon filter and incubated with anti-Bep4 protein polyclonal antibodies diluted 1:1000 as described previously (Romancino et al., 1992).

Southern blot (data not shown). Furthemore, to investigate if the Bep4 protein is conserved in different sea urchin species, we extracted total proteins from *P. lividus*, *A. lixula* and *S. granularis* eggs and a Western blot analysis was performed. Fig. 5B shows a 33-kDa band corresponding to the Bep4 protein in all the examined samples, demostrating conservation of at least a common epitope of the protein in the different sea urchin species.

(e) Conclusions

(*1*) The paper describes the structural analysis of a *P. lividus bep4* cell surface gene; this is the first sea urchin cell surface gene analysed, transcribed by a maternal messenger.

(*2*) The gene presents the typical structural organization of introns and exons and the regulatory region contains the TATA and CCAAT boxes present in other eukaryotic genes. Comparison with other genes of the same family will be important to establish the presence or not of common regulatory elements.

(*3*) The presence of two potential Ca^{2+} binding sites like those of other cell surface proteins, involved in cell interaction, might be important for Bep4 function.

(*4*) Lastly, this gene is well conserved in *S. granularis* but under the utilized condition, no hybridization is observed in *A. lixula* genome. One or more epitopes of the protein are instead detected with antibodies in all the sea urchin species.

ACKNOWLEDGEMENT

We thank Prof. Giovanni Giudice for suggestions in the preparation of the manuscript.

REFERENCES

Chou, P.Y. and Fasman, G.D.: Prediction of the secondary of proteins from their amino acid sequence. Annu. Rev. Biochem. 47 (1979) 251–276.

Di Carlo, M., Montana, G. and Bonura, A.: Analysis of the sequence and expression during sea urchin development of two members of a multigenic family, coding for butanol-extractable proteins. Mol. Rep. Dev. 25 (1990) 28–36.

Duro, G., Izzo, V., Barbieri, R., Cantone, M., Costa, M.A. and Giudice, G.: A method for eluting in a wide range of molecular weights from agarose gels. Anal. Biochem. 195 (1991) 111–115.

Edelmam, G.M.: Cell adhesion molecules in the regulation of animal form and tissues pattern. Annu. Rev. Cell Biol. 2 (1986) 81–116.

Geisow, M.J., Fritsche, U., Hexham, J.M., Dash, B. and Johnson, T.: A consensus amino-acid sequence repeat in *Torpedo* and mammalian Ca^{2+}-dependent membrane-binding proteins. Nature 320 (1986) 636–638.

Ghersi, G. and Vittorelli, M.L.: Immunological evidence for the presence in sea urchin embryos of a cell adhesion protein similar to mouse uvomorulin (E-cadherin). Cell Diff. Dev. 31 (1990) 67–75.

Ghersi, G., Salomone, M., Dolo, V., Levi, G. and Vittorelli, M.L.: Differential expression and function of cadherin-like proteins in the sea urchin embryo. Mech. Dev. 41 (1993) 47–55.

Giudice, G.: Developmental Biology of Sea Urchin Embryo. Academic Press, New York, 1977.

Giudice, G.: The Sea Urchin Embryo: A Developmental Biological System. Springer-Verlag, New York, 1986.

Gustafson, G. and Wolpert, L.: Cellular movement and contact in sea urchin embryos. Dev, Biol. 91 (1967) 27–37.

Leaf, D.S., Anstrom, J.A., Chin, J.E., Harkey, M.A., Showman, R.M. and Raff, R.: Antibodies to a fusion protein identify a cDNA clone encoding msp 130, a primary mesenchyme-specific cell surface protein of the sea urchin embryo. Dev. Biol. 121 (1987) 29–40.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

McClay, D.R., Chamber, A.F. and Warren, R.H.: Specificity of cell-cell interaction in sea urchin embryos. Dev. Biol. 56 (1977) 343–355.

Mount, S.M.: A catalogue of splice junction sequences. Nucleic Acids Res. 10 (1982) 459–472.

Nocente-McGrath, C., Brenner, C.A. and Ernst, S.G.: Endo 16, a lineage-specific protein of the sea urchin embryo, is first expressed just prior to gastrulation. Dev. Biol. 136 (1989) 264–272.

Parr, B.A., Parks, A.I. and Raff, R.: Promoter structure and protein sequence of msp 130, a sea urchin primary mesenchyme cell glycoprotein. J. Biol. Chem. 265 (1989) 1408–1413.

Ringwald, M., Schuh, R., Vestweber, D., Eistetter, H., Lottspeich, F., Engel, J., Dolz, R., Jahnig, F., Epplen, J., Mayer, S., Muller, C. and Kemler, R.: The structure of cell adhesion molecule uvomorulin. Insights into the molecular mechanism of Ca^{2+} dependent cell adhesion. EMBO J. 6 (1987) 3647–3653.

Romancino, D.P., Ghersi, G., Montana, G., Bonura, A., Perriera, S. and Di Carlo M.: Characterization of bep1 and bep4 antigens involved in cell interaction during P. lividus development. Differentiation 50 (1992) 67–74.

Sanger, F., Nicklen, S. and Coulson, A.R.: DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A.: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J.Mol. Biol. 143 (1980) 161–178.

Stuart, D.I., Acharya, K.R., Walker, N.P.C., Smith, S.G., Lewis, M. and Phillips, D.C.: α-Lactalbumin possesses a novel calcium binding loop. Nature 324 (1986) 84–87.

Tabor, S. and Richardson, C.C.: DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84 (1987) 4767–4771.

Tufty, R.M. and Kretsinger, R.H.: Troponin and parvalbumin calcium binding regions predicted in myosin light chain and T4 lysozyme. Science 187 (1975) 167–169.