The early fate specification of primary mesenchyme cells in sea urchin embryos makes them an attractive system for studying alterations in gene expression and protein synthesis during cell lineage determination and differentiation. To analyze the developmental regulation of gene expression in Strongylocentrotus purpuratus, we have isolated and sequenced genomic and cDNA clones encoding msp130, a mesenchyme-specific cell surface glycoprotein. We have located the transcription initiation site of the msp130 gene and sequenced several kilobases of the promoter region. The region of the gene that encodes the protein is divided into numerous small (less than 500 base pairs) exons. The msp130 protein possesses two novel glycine-rich domains and a signal peptide, but apparently lacks a transmembrane domain. The carboxyl-terminal sequence suggests that msp130 may be phosphatidylinositol-linked to the cell membrane, and experiments with phospholipases support this conclusion. The implications of the msp130 sequence for its possible functions are discussed.

The determination and differentiation of embryonic cell lineages are key events in the development of multicellular organisms. In sea urchin embryos, the earliest specification of cell fate occurs in the primary mesenchyme cell lineage. An unequal fourth cleavage produces a vegetal tier of four micromeres, which subsequently give rise to the primary mesenchyme cells. After ingressing into the blastocoele, primary mesenchyme cells migrate around the blastocoelic wall. They eventually form a ring, fuse into a syncytium, and secrete the larval calcareous skeleton. The fate of primary mesenchyme cells appears to be determined by the time of the fourth cleavage, as isolated micromeres will differentiate autonomously into mesenchyme cells that duplicate the behavior of their counterparts in intact embryos and produce skeletal spicules (Okazaki, 1975). The early fate specification and ease of culturing mesenchyme cells (e.g. Okazaki, 1975; Harkey and Whiteley, 1980; Ettensohn and McClay, 1987) make them an attractive model system for studying the alterations in gene expression that are responsible for cell lineage determination and differentiation.

In the primary mesenchyme cell lineage of the sea urchin Strongylocentrotus purpuratus, a variety of mRNAs accumulate with similar kinetics (Harkey et al., 1988). These transcripts are absent from the egg and only detectable at low levels prior to hatching of the blastula. They increase dramatically in abundance after hatching, reaching peak levels at gastrula or prism stages. It is likely that some of these mRNAs encode proteins that are responsible for the final differentiated state of primary mesenchyme cells. The synchronous transcript accumulation patterns suggest that coordinate gene regulation might occur in the primary mesenchyme cell lineage.

If the expression of a set of primary mesenchyme cell genes is coordinately regulated, the promoters of these genes should share some common control elements. The promoter region of one gene expressed in a primary mesenchyme-specific manner, SM50, has been well characterized, and sequences sufficient to direct proper expression of the gene have been identified (Sucov et al., 1987, 1988). The analysis of other genes that are expressed in primary mesenchyme cells is necessary to investigate possible mechanisms of coordinate gene regulation.

MATERIALS AND METHODS

Library Construction and Screening—The phage vector λgt10 was used to construct a cDNA library from S. purpuratus gastrula poly(A)⁺RNA (Gubler and Hoffman, 1983). A genomic library in the vector EMBl3 was kindly provided by Drs. William Klein and Paul L. Speicher. The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J05204.

*This research was supported by National Institutes of Health Grant RO1 HD21986. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

 Characterization of Genomic Clones—DNA from restriction digests of genomic clones was electrophoresed on 1% agarose gels, blotted to nitrocellulose (Southern, 1975), and probed with nick-translated or oligo-labeled (Feinberg and Vogelstein, 1983) cDNA restriction fragments. The blot was hybridized overnight at 37 °C in 50% formamide, 1 X SET (1 X = 150 mM NaCl, 2 mM EDTA, 25 mM Tris, pH 8.0), 1 X Denhardt's solution, 10% dextran sulfate, 0.1 mg/ml carrier DNA, and washed at a final stringency of 65 °C, 1 X SET. To locate additional transcribed regions upstream of the longest cDNA, labeled restriction fragments from genomic clones were hybridized to Northern blots containing 10 µg of RNA per lane. Hybridization and washing conditions were identical with those used for the Southern blots except that the overnight hybridizations were carried out at 42 °C.

 Subcloning and Sequencing—Restriction fragments from the original library clones were subcloned into the phagemid vectors pTZ (U.S. Biochemicals) or pBluescript (Stratagene). Single-stranded DNA templates were sequenced with the Sequenase kit (U.S. Biochemicals).

Nuclease S1 Protection Assays—A 312-bp² restriction fragment spanning the 5'-end of the gene was gel-purified, dephosphorylated, and end-labeled with [γ-32P]ATP by T4 kinase. Unincorporated nucleotides were removed by passing the probe over a Bio-Gel A-1.5m column. Single-stranded probe was precipitated with 50 µg of gastrula or prism stage total RNA and resuspended in 30 µl of hybridization buffer (40 mM PIPES, pH 6.4, 1 mM EDTA, 400 mM NaCl, 80% formamide). The hybridization mixture was incubated at 75 °C for 15 min, then at 50 °C overnight. 300 ul of S1 buffer (50 mM sodium acetate, pH 4.6, 280 mM NaCl, 4.5 mM ZnSO₄) containing 400 units of nuclease S1 (Boehringer Mannheim) was added, and nuclease digestion was carried out at 37 °C for 1 h. The products were analyzed on 8% acrylamide, 8 M urea sequencing gels.

Primer Extensions—A 19-nucleotide oligomer complementary to the mspl30 mRNA (Indiana University Institute for Molecular and Cellular Biology DNA Synthesis Facility) was end-labeled with [γ-32P]ATP by T4 kinase, and the unincorporated nucleotides were removed by passage through a Sep-Pak cartridge (Waters Associates). 100 µg of labeled primer (1-2 x 10⁸ cpm/µg) was incubated with 3 µg of gastrula poly(A) RNA in annealing buffer (10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 1 mM EDTA) at 85 °C for 10 min, then at 55 °C for 6 h. The primer was extended by 10 units of reverse transcriptase (Seikagaku) in reverse transcriptase buffer (50 mM Tris, pH 8.2, 10 mM dithiothreitol, 0.5 mM concentration of each dNTP, 6 mM MgCl₂) at 42 °C for 1 h. When dideoxy sequencing was performed to verify that the primer was recognizing only mspl30 RNA, the extension reaction was split into four tubes, with a 0.05 mM concentration of the appropriate ddNTP present in each tube. Primer extension products were run on 8% acrylamide, 8 M urea sequencing gels.

Sequence Analysis Nucleotide and amino acid sequences were analyzed, and protein secondary structure was predicted by the IBI Pustell and University of Wisconsin Genetics Computer Group (Devereux et al., 1984) programs.

Phospholipase Experiments—S. purpuratus late gastrulae (48 to 58 h) were dissociated by washing embryos two to three times in calcium, magnesium-free seawater (27 g/liter of NaCl, 0.8 g/liter of KCl, 1.0 g/liter of NaSO₄, 0.2 g/liter of NaHCO₃, pH 8.0), followed by two washes in buffer A (0.35 M glycine, 0.1 M Hepes, pH 7.8). The final pellet was resuspended in buffer A plus protease inhibitors (0.005 mM PMSF, 0.01 mg/ml leupeptin, 0.01 mg/ml pepstatin A, 0.001 M phenylmethylsulfonyl fluoride). Embryos were gently pipetted until approximately 90% dissociated and stored at -20 °C. Phospholipases were incubated with aliquots of the above cell preparation at 37 °C for 60-75 min using the following buffers: phospholipase A₂ (bee venom, Sigma), buffer A plus 10 mM CaCl₂, phospholipase C (Bacillus cereus, type III, Sigma), buffer A plus 10 mM ZnCl₂, phospholipase D (cabbage, type IV, Sigma), 50 mM sodium acetate, pH 6.0, 50 mM CaCl₂, 30 mM NaCl, 0.35 mM glycine, phosphatidylinositol-specific phospholipase C (C. cereus; Boehringer Mannheim), buffer A. Following incubation, cells were pelleted, and pellets were resuspended in 2 X Laemmli sample buffer (20% glycerol, 10% β-mercaptoethanol, 4.6% SDS, 0.2% Triton-X-100, pH 6.8), vortexed, and boiled for 2.5 min. Supernatant proteins were precipitated with 2-3 volumes of acetone for 4 h at overnight at -20 °C and raised in 2 X Laemmli sample buffer as above. Proteins were separated on SDS-polyacrylamide gels according to Laemmli (1970) and transferred to nitrocellulose according to Towbin et al. (1979). Blots were immunostained using hybridoma supernatant of the monoclonal antibody B2C2, which specifically recognizes a carbohydrate moiety on the mspl30 protein (Anstrom et al., 1987). Primary antibody was detected using a goat anti-mouse peroxidase-conjugated antibody (HyClone) with 4-chloro-1-naphthol as the chromagen.

RESULTS

mspl30 Gene Structure—The initial characterization of mspl30 cDNAs (18C1-6) has been reported (Leaf et al., 1987). To obtain longer clones, we constructed a cDNA library from S. purpuratus gastrula poly(A)⁺ RNA and screened it with the 5' most fragment of 18C4. Sequence analysis of the 5'-end of our longest clone (2C) indicated that we had not located the sequences coding for the beginning of the protein. Therefore, we used a 330-bp restriction fragment from the 5'-end of 2C to screen a genomic library. We recovered clones covering a total of 27 kb of DNA. The upstream clones extend 10 kb beyond the transcription start site of mspl30 and contain part of another gene. This second gene is transcribed in the opposite direction from mspl30 and is not expressed in a mesenchyme-specific fashion (data not shown).

Fig. 1 shows the mspl30 gene structure as determined by a combination of restriction mapping, Southern blots, and sequencing. We have not completely mapped the 3'-end of the gene. Consensus splice signals (Mount, 1982) are found at junctions where genomic and cDNA sequences diverge. The mspl30 gene consists of small exons (generally <500 bp) broken up by a minimum of 7 introns.

Characterization of the 5'-End of the Gene—Fig. 2 displays the nucleotide sequence of the promoter region and first exon of the mspl30 gene. The first exon is 492 bp long and includes 197 bp of nontranslated leader sequence. We performed primer extension and nuclease S1 protection assays to locate the transcription initiation site. For the S1 experiment, a 224-bp fragment was protected when a 312-bp genomic clone restriction fragment spanning the putative transcription start site was end-labeled with ϕ-32P, hybridized to either gastrula or prism RNA, and the hybrid was digested with nuclease S1 (Fig. 3). In the primer extension reactions, a 19-bp primer was hybridized to gastrula poly(A)⁺ RNA and extended to two products of 124 bp and 128 bp by reverse transcriptase (Fig. 3). The longer and more abundant primer extension product confirms the transcription start site predicted by the S1 experiments. We do not know the significance (if any) of the shorter primer extension product. We have performed dideoxy sequencing during the primer extension reactions to

---

3 The abbreviations used are: bp, base pair(s); kb, kilobase(s); PIPES, 1,4-piperazinediethanesulfonic acid; Hepes, 1 (2-hydroxyethyl)-1-piperazineethanesulfonic acid; SDS, sodium dodecyl sulfate.

---

FIG. 1. Structure of the mspl30 gene. Arrows indicate regions of genomic clone HG 2A that have been sequenced. Dotted boxes represent the 5'-nontranslated region. The Sall sites at the end of the exons are from the vector. Abbreviations: B = BamHI; S = Sall; Xb = XhoI; Xh = XhoI.
confirm that the primer hybridizes to only the msp130 transcript (data not shown).

The promoter region of msp130 contains a possible TATA box (TATTA) starting at position -27 (relative to the start of transcription). The closest approximations to a CCAAT box (TATTA) boxes, CGCTCAT heptamers and variants (boxes), and GCCCCT hexamer (triple underline). A polymorphic hexamer that is present in two cDNA clones, but absent from the genomic clone HG4B, is bracketed (see text). Several kilobases of additional upstream sequence have been determined but are not shown in this figure (available upon request).

The msp130 promoter region and first exon. The transcription (position +1) and translation (+198) initiation sites are marked by arrows. Putative regulatory elements in the promoter region are also noted (see text): possible CCAAT (underline) and TATA (double underline) boxes, CGCTCAT heptamers and variants (boxes), and GCCCCT hexamer (triple underline). A polymorphic hexamer that is present in two cDNA clones, but absent from the genomic clone HG4B, is bracketed (see text). Several kilobases of additional upstream sequence have been determined but are not shown in this figure (available upon request).

Fig. 2. Nucleotide sequence of the msp130 promoter and first exon. The transcription (position +1) and translation (+198) initiation sites are marked by arrows. Putative regulatory elements in the promoter region are also noted (see text): possible CCAAT (underline) and TATA (double underline) boxes, CGCTCAT heptamers and variants (boxes), and GCCCCT hexamer (triple underline). A polymorphic hexamer that is present in two cDNA clones, but absent from the genomic clone HG4B, is bracketed (see text). Several kilobases of additional upstream sequence have been determined but are not shown in this figure (available upon request).

Fig. 4 presents the complete amino acid sequence of msp130. The sequence predicts a protein of relative molecular weight 81,000, which contrasts with the observed Mr = 136,000 on standard polyacrylamide gels (e.g. Anstrom et al., 1987). msp130 is a sulfated glycoprotein, and a precursor of 110 kDa has been observed in pulse-chase experiments (Anstrom et al., 1987). There are six predicted sites of N-linked glycosylation, and a position for O-linked glycosylation also exists. Some glycoproteins migrate substantially slower in SDS-polyacrylamide gels than nonglycosylated proteins of equivalent size (Segrest and Jackson, 1972), and the post-translational modifications to the msp130 protein apparently have a similar effect on its gel mobility.

The most striking features of the msp130 sequence are two glycine-rich regions toward the amino-terminal end of the protein (Fig. 4). Twenty six of the thirty-two amino acids from residues 39 to 70 are glycines, and the sequence Gly-Ala is repeated four times at the beginning of this segment. We have detected a polymorphism at the end of this run of glycines, as the final two glycines in the cDNA sequence apparently have a similar effect on its gel mobility.

The msp130 Protein—The methionine codon (ATG) at nucleotides 198-200 in Fig. 2 marks the beginning of the protein coding region of the gene. There is an in-frame stop four codons upstream of this ATG. The context for translation initiation (CAATAAT) resembles the Drosophila consensus (C/AAA/ATG, Cavener, 1987) more than the vertebrate consensus (CA/GCA/ATG, Kozak, 1987), but does contain a purine at position -3, the most conserved nucleotide in vertebrates (Kozak, 1987). The translation initiation sites of S. purpuratus transcripts are most conserved at positions -4 (C/A) and -3 (A), and these preferences conform to both vertebrate and Drosophila consensuses (data not shown).
**msp130 Gene and Protein**

**A.**

ACGT SI

-312 bp SI probe

ACGT PE.

-224 bp S1 protected fragment

128 bp 124 bp extension products

**FIG. 3. Determination of the msp130 transcription initiation site.** Nuclease SI protection assays (A) and primer extensions (B) were used to locate the msp130 transcription start site. A 312-bp genomic clone restriction fragment (spanning the putative initiation site) was used as an end-labeled probe in a nuclease SI protection assay. Following hybridization to pristm RNA and nuclease SI digestion, a 224-bp fragment was protected. For the primer extensions, a 19-nucleotide primer complementary to the msp130 mRNA was end-labeled with $^{32}$P and hybridized to gastrula poly(A$^+$) RNA. The primer was extended with reverse transcriptase to full length products of 128 and 124 base pairs. The longer primer extension product confirmed the transcription start site (position +1 in Fig. 2) predicted by the nuclease SI assay. The sequences in lanes A through T were employed as size markers.

**B.**

signal peptide 20 20 glycosyl-rich domain I

<table>
<thead>
<tr>
<th>60</th>
<th>90</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>180</td>
<td>210</td>
</tr>
<tr>
<td>240</td>
<td>270</td>
<td>300</td>
</tr>
<tr>
<td>330</td>
<td>360</td>
<td>390</td>
</tr>
</tbody>
</table>

**DISCUSSION**

We have established the complete amino acid sequence of msp130. Homology searches have revealed no significant similarities to msp130 among known proteins except in the glycine repeat regions. Many other proteins with glycine-enriched stretches have been identified, including collagens (Yamada et al., 1980), keratins (Tyner et al., 1985), elastin (Raju and Anwar, 1987), the neurosecretory vesicle protein synaptophysin (Leube et al., 1987), the Drosophila "pen" repeat proteins (Haynes et al., 1987), the Acanthamoeba myosin IB heavy chain (Jung et al., 1987), and the product of the Drosophila homeotic gene spalt (Frei et al., 1988). It is interesting to note that the glycine-rich domains often contain a relatively large number of alanine, glutamine, and proline residues as well. These proteins generally bear no functional or evolutionary similarity to each other. In most cases, the role of the glycine motifs is unknown. They are usually...
trifugation to retrieve the cell membranes. The location of \( \text{msp130} \)
protein from dissociated \textit{Strongylocentrotus purpuratus} gastrotricha were incubated with various phospholipases followed by cen-
tored by immunoblot analysis with an anti-\text{msp130} antibody. The
release of \text{msp130} from the pellet to the supernatant is only seen
following the addition of phosphatidylinositol-specific phospholipase
\text{C} (\text{PI-C}). On some blots, minor nonspecific bands were detected
that do not respond to phospholipase treatment. \( \text{C} \), no phospholipase
added; \( \text{PI} \), 0.5 unit/ml of phosphatidylinositol-specific phospholipase
\text{C}; \( \text{PC} \), 1 unit/ml of phospholipase \text{C}; \( \text{A} \), 1 unit/ml of phospholipase
\text{A2}.

hypothesized to provide flexibility to a protein and perhaps act as linkers between other domains. Predictions of \text{msp130}
secondary structure suggest that its glycine repeats also form
act as linkers between other domains. Predictions of \text{msp130}
hypothesized to provide flexibility to a protein and perhaps
are not ruled out by the available data. The late onset of
\text{msp130} expression (after primary mesenchyme cells have
ceased migrating) in some cidaroids and sand dollars (\text{Wray}
and McClay, 1989) argues against a role in mesenchyme cell
migration. The association of \text{msp130} with both larval and
adult skeletal elements (\text{Parks} \text{et al.}, 1988), structures of very
different sizes and shapes, indicates that it is not involved in
pattern formation \textit{per se}. Finally, there is no direct evidence
bearing on the possible role of \text{msp130} in primary mesen-
chyme cell adhesion, fusion, or cell-cell communication.

\text{msp130} is the second gene exhibiting mesenchyme-specific
expression to be well characterized. It is interesting to com-
pare the promoter region of \text{msp130} with that of \text{SM50} (\text{Sucov}
\text{et al.}, 1987). SM50 and \text{msp130} are both terminal differentia-
tion products of the mesenchyme cell lineage, whose mRNAs
accumulate with approximately the same kinetics (\text{Benson}
\text{et al.}, 1987; \text{Leaf} \text{et al.}, 1987). The promoters of the two genes
share some sequence elements, the GCCCCGT and GCCTCAT
motifs serving as examples with possible functional significa-
cance. However, there are some significant differences in
promoter structure as well. For example, SM50 possesses a
CCAAT element but lacks a TATA box, whereas the \text{msp130}
promoter appears to contain TATA but not CCAAT.

Future studies will determine whether the promoters of
different primary mesenchyme-specific genes are truly inter-
changeable. The cloning and analysis of lineage-specific genes
such as \text{msp130} will help decipher the mechanisms by which
embryonic cell lineages reach their final differentiated phe-
notype.

Acknowledgments—We thank David Leaf for the initial character-
ization and sequencing of \text{msp130} cDNAs, Mary Jozwiak for South-
erns, Lois Hunt for a homology search, and Deborah Hursh for
sharing unpublished results.

REFERENCES

\text{Anstrom}, J. A., \text{Chin}, J. E., \text{Leaf}, D. S., \text{Parks}, A. L., and \text{Raff}, R. A.
(1987) \text{Development} \text{101}, 255-265

\text{Dev. Biol.} \text{120}, 499-506

\text{Cell} \text{41}, 639-648


\text{12}, 387-395


\text{57}, 285-320

Chem.} \text{260}, 4983-4988

\text{Frei}, E., \text{Schuh}, R., \text{Baumgartner}, S., \text{Buri}, M., \text{Noll}, M., \text{Jurgens},
G., \text{Seifert}, E., \text{Nauber}, U., and \text{Jackle}, H. (1988) \text{EMBO J} \text{7}, 197-
204

\text{Futerman}, H., \text{Fiorini}, R.-M., \text{Roth}, E., \text{Low}, M. G., and \text{Silman}, I.
(1985) \text{Biochem J} \text{226}, 369-377

\text{Gubler}, U., and \text{Hoffman}, B. J. (1983) \text{Gene (Amst.)} \text{25}, 263-269

\text{Hardin}, S. H., \text{Carpenter}, C. D., \text{Hardin}, P. E., \text{Bruskin}, A. M., and

Biol.} \text{189}, 111-122

\text{195}, 381-395

\text{Haynes}, S. R., \text{Rebert}, M. L., \text{Mozer}, B. A., \text{Forquignon}, F., and

\text{He}, H.-T., \text{Barbet}, J., \text{Chaix}, J.-C., and \text{Gordis}, C. (1986) \text{EMBO J.}
\text{5}, 2489-2494


Promoter structure and protein sequence of msp130, a lipid-anchored sea urchin glycoprotein.

B A Parr, A L Parks and R A Raff


Access the most updated version of this article at http://www.jbc.org/content/265/3/1408

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/265/3/1408.full.html#ref-list-1