A differential display strategy identifies Cryptic, a novel EGF-related gene expressed in the axial and lateral mesoderm during mouse gastrulation

Michael M. Shen¹,²*, Hongyu Wang² and Philip Leder¹

¹Department of Genetics, Howard Hughes Medical Institute, Harvard Medical School, 200 Longwood Avenue, Boston, MA, USA
²Center for Advanced Biotechnology and Medicine, Department of Pediatrics, University of Medicine and Dentistry of New Jersey - Robert Wood Johnson Medical School, Piscataway, NJ 08854, USA

*Author for correspondence at address 2 (e-mail: mshen@mbcl.rutgers.edu)

SUMMARY

We have developed a differential display screening approach to identify mesoderm-specific genes, relying upon the differentiation of embryonic stem (ES) cells in vitro. Using this strategy, we have isolated a novel murine gene that encodes a secreted molecule containing a variant epidermal growth factor-like (EGF) motif. We named this gene Cryptic, based on its predicted protein sequence similarity with Cripto, which encodes an EGF-related growth factor. Based on their strong sequence similarities, we propose that Cryptic, Cripto, and the Xenopus FRL-1 gene define a new family of growth factor-like molecules, which we name the ‘CFC’ (Cripto, Frl-1, and Cryptic) family. Analysis of Cryptic expression by in situ hybridization shows that it is expressed during gastrulation in two spatial domains that correspond to the axial and lateral mesoderm. In the first domain of expression, Cryptic expression is progressively localized to the anterior primitive streak, the head process, and the node and notochordal plate. In the second domain, Cryptic expression is initially concentrated in the lateral region of the egg cylinder, and is later found circumferentially in the intermediate and lateral plate mesoderm. Furthermore, Cryptic expression can also be detected at the early head-fold stage in the midline neuroectoderm, and consequently is an early marker for the prospective floor plate of the neural tube. Expression of Cryptic ceases at the end of gastrulation, and has not been observed in later embryonic stages or in adult tissues. Thus, Cryptic encodes a putative signaling molecule whose expression suggests potential roles in mesoderm and/or neural patterning during gastrulation.

Key words: ES cell, embryoid body, epidermal growth factor motif, node, notochord, floor plate

INTRODUCTION

In the vertebrate embryo, the processes of pattern formation, differentiation, and morphogenesis that occur during development are regulated by cell-cell interactions, which are primarily mediated by secreted growth factor-like molecules. Among the signaling molecules that have been implicated in vertebrate developmental pathways are several members of the epidermal growth factor (EGF) superfamily. These EGF-related growth factors include EGF, TGF-α, neuregulin, amphiregulin, betacellulin, and heparin-binding EGF, all of which represent ligands for members of the erbB receptor family (Shoyab et al., 1989; Higashiyama et al., 1991; Prigent and Lemoine, 1992; Shing et al., 1993; Groenen et al., 1994; Wen et al., 1994). These genes encode secreted proteins that are often expressed as transmembrane precursors, and contain one or more copies of the EGF motif, a cysteine-rich domain that is found in many extracellular proteins and is implicated in protein-protein interactions (Davis, 1990; Prigent and Lemoine, 1992; Groenen et al., 1994). In addition, the EGF superfamily also includes less well-characterized members such as the distantly-related growth factor Cripto (Ciccocicola et al., 1989; Dono et al., 1993), whose receptor is unlikely to be formed by members of the erbB family (Brandt et al., 1994).

Gene targeting and other approaches have confirmed the importance of several of these genes for proper embryonic development. For example, neuregulin and the ErbB2 and ErbB4 receptors have been shown to be required for correct development of the heart and cranial sensory ganglia (Gassmann et al., 1995; Lee et al., 1995; Meyer and Birchmeier, 1995). Moreover, targeted gene disruption of the EGF receptor results in strain-dependent phenotypes ranging from peri-implantation to neonatal lethality (Sibilia and Wagner, 1995; Threadgill et al., 1995). In addition, the Xenopus FRL-1 gene encodes a mesoderm- and neural-inducing factor that was recently identified by its ability to activate fibroblast growth factor (FGF) receptors, thereby providing an unexpected link between EGF-related signaling molecules and FGF receptor signal transduction (Kinosita et al., 1995).

However, many of the cellular interactions associated with events during early mammalian development have not yet been associated with the activities of specific signaling factors. This has been particularly problematic in the mouse, in part due to the small amounts of tissue present in the gastrulating embryo, and the difficulties involved in isolating genes based on their
expression patterns. In this study, we describe a novel approach for the isolation of genes expressed in early embryogenesis, based on an in vitro system for the differentiation of ES cell-derived embryoid bodies. This strategy compares the differentiation of embryoid bodies cultured in the absence versus the presence of leukemia inhibitory factor (LIF), which inhibits mesoderm formation (Shen and Leder, 1992). By examining differential gene expression in these two cell populations with the differential display RT-PCR protocol, which can detect rare transcripts (Liang and Pardee, 1992), we can successfully identify novel genes that are expressed specifically during mesoderm differentiation in vitro.

Using this screening methodology, we have isolated a new EGF-related gene that encodes a putative secreted molecule. We have named this gene Cryptic, because its predicted product shares significant sequence similarities with that of murine and human Cryptic (Ciccodicola et al., 1989; Dono et al., 1993), as well as the Xenopus FRL-1 gene (Kinosita et al., 1995). We propose that Cryptic, Cripto, and FRL-1 define a distinct family of EGF-related genes, which we have named the ‘CFC’ (Cripto, FRL-1, and Cryptic) family. Moreover, our in situ hybridization analysis shows that Cryptic possesses a spatially and temporally restricted expression pattern that is consistent with potential signaling roles during mouse gastrulation. Specifically, Cryptic is expressed in the axial and lateral mesoderm during gastrulation, and also represents one of the earliest markers for the prospective floor plate of the neural tube. Furthermore, these findings demonstrate that our differential display screening approach may have significant utility for the identification of embryonically expressed genes that would be difficult to identify by other methods.

**MATERIALS AND METHODS**

**Cell culture**

Culture of ES cells and embryoid body formation was described previously (Shen and Leder, 1992). Briefly, D3 ES cells (Doetschman et al., 1985) were grown on mitomycin C-treated STO cell feeder layers in the presence of 1,000 U/ml recombinant LIF (Gibco). To initiate embryoid body formation, ES cells were grown with exogenous LIF for one passage in the absence of feeder cells, trypsinized to a single-cell suspension, and plated in bacteriological Petri dishes at a density of 1x10^5 cells/ml. Embryoid bodies were maintained in the absence or continual presence of 1,000 U/ml recombinant LIF, and aliquots were removed at appropriate time points for RNA extraction.

F9 cells were differentiated in monolayer cultures in the presence of 2x10^-7 M retinoic acid (Kodak), 0.5 mM dibutyryl cyclic AMP (Sigma), and 0.25 mM theophylline (Sigma), essentially as described (Dean et al., 1986). The F9 cells were plated at 1x10^6 cells/15 cm dish, and maintained in the presence of retinoic acid for four days. Differentiation into parietal endoderm was monitored by altered morphology and by ribonuclease protection analysis of marker genes at various time points (data not shown).

**Differential display RT-PCR**

We used conditions similar to those originally described (Liang and Pardee, 1992). Briefly, we primed first-strand cDNA synthesis from DNase-treated total RNA using an oligonucleotide of sequence 5'CTTCACTGGATTTTT; for fragment 31-24 (SCG10), the primers were AGTGGGATCA and TTTTTTTTTTCC; for fragment 34-25 (Cryptic), the primers were GCTGTTGTTA and TTTTTTTTTTTTA; and for fragment 35-3, the primers were AGAGCCTGCTT and TTTTTTTTTTTGGC. In control experiments, we utilized nine 5' primers with all possible combinations of ten 3' primers, for a total of 90 differential display comparisons. To amplify fragment 32-1, the primers used were: TTGGACATGC and TTTTTTTTTTTAA; for fragment 33-24 (SCG10), the primers were AGTGGGATCA and TTTTTTTTTTTCC; for fragment 34-25 (Cryptic), the primers were GCTGTTGTTA and TTTTTTTTTTTTA; and for fragment 35-3, the primers were AGAGCCTGCTT and TTTTTTTTTTTGGC. In control experiments, we utilized nine 5' primers corresponding to sequences in the 3' untranslated region of Brachyury, since this gene is differentially expressed between LIF-treated and untreated embryoid bodies (M. M. Shen, unpublished observations). We found that all nine primers amplified a differentially expressed fragment of the expected size using five different 3' end primers (data not shown).

Candidate differentially expressed fragments of interest were excised from the dried gel, eluted, and re-amplified as previously described (Liang and Pardee, 1992; Liang et al., 1993). These re-amplified fragments were used as probes for northern blots of poly(A)+ RNA from embryoid bodies. The positive re-amplified fragments were cloned into T-vectors (Marchuk et al., 1991) prepared from Bluescript IKS(+) vector (Stratagene) digested with EcoRV.

**cDNA library screening and sequence analysis**

Poly(A)+ RNA was prepared directly from embryoid bodies using the Fast-Track system (Invitrogen). First-strand cDNA was prepared from 5 μg of poly(A)+ RNA using SuperScript II reverse transcriptase (Gibco), and used for construction of a uni-directional cDNA library in the AZAP II vector (Stratagene). Unamplified phage were packaged and plated for library screening. Procedures for library screening, DNA sequencing, and northern blotting were as described (Sambrook et al., 1989).

For computer analysis of the Cryptic cDNA sequence and its putative protein product, we used the Lasergene software package (DNA/STAR, Madison, WI). Signal sequence cleavage was analyzed using a weight-matrix table (von Heijne, 1986), as implemented by the AnalyzeSignalase software program (authored by Ned Mantei). The PSORT analysis program (version 6.3), which predicts the cellular localization of proteins based on their sequence (Nakai and Kanehisa, 1992), was accessed via the World Wide Web. The PROSITE protein motifs database (release 13.0) (Bairoch and Bucher, 1994) was searched using MacPattern version 3.4 (Fuchs, 1994).

**In vitro translation and secretion analysis**

For in vitro translation analysis, we subcloned a full-length Cryptic cDNA into the pcDNA1/Amp vector (Invitrogen), followed by synthesis of capped transcripts from the T7 promoter using the Message Machine system (Ambion). In vitro translation was performed using rabbit reticulocyte lysates (Promega) in the absence or presence of canine pancreatic microsomal membranes (Boehringer), according to the manufacturers’ instructions. Translation products were labeled with [35S]cysteine (NEN), separated by SDS-PAGE, and visualized by autoradiography after fluorographic enhancement with Amplify reagent (Amersham). To assay glycosylation, we treated translation products with either EndoH or PNGase F (New England Biolabs). Control experiments were performed using a bacterial β-lactamase RNA (Promega) to confirm the signal sequence.
cleavage activity of the microsomal membrane preparation (data not shown).

To construct expression vectors for alkaline phosphatase fusion proteins, we used PCR to amplify the desired fragments of Cryptic, followed by cloning into ApTag-2/Amp (Y.-T. Yan and M. M. Shen, unpublished), which is derived from ApTag-2 (Cheng and Flanagan, 1994) by transfer of the alkaline phosphatase insert to the pCDNAI/Amp vector. To generate these Cryptic constructs, we synthesized the following oligonucleotides:

M87: (5′) AGCTGATGCTGCCCTTACGACCACACC (3′)
M88: (5′) AGCTAGATCTGAAGAGAATTTTCG (3′)
M90: (5′) AGCTGAAATTCGCCACCTGAGGGGACCAACC (3′)
M91: (5′) AGCTGAAATTCGCCACCTGACATCAGGCGCCTCAG (3′).

The M90 and M91 primers contain an EcoRI site and sequences corresponding to an optimal translation initiation site (Kozak, 1986), while the M87 and M88 primers contain an XbaI site, so that amplified fragments could be cloned into the EcoRI and XbaI sites of ApTag-2/Amp. Amplification using the primers M87 and M90 resulted in full-length Cryptic, M87 and M91 in deletion of the C-terminal hydrophobic domain, and M88 and M90 in deletion of the residues prior to the second translation initiation site.

For assays of fusion protein secretion, COS cells were transfected using Lipofectamine (Gibco), and conditioned medium was collected two days later for analysis of alkaline phosphatase activity. Alkaline phosphatase activity was determined by assaying the hydrolysis of p-nitrophenyl phosphate (Boehringer) in 96-well microplates, as previously described (Flanagan and Leder, 1990). For our experiments, we measured changes in absorbance (A) at 405 nm in an EL800 kinetic microplate reader (Bio-Tek Instruments), with the amount of AP activity defined as the maximum rate of change of A405/ml/hour, as calculated using Deltasoft 3 microplate analysis software (Bio-Metallics, Inc.).

Ribonuclease protection and whole-mount in situ hybridization assays

For embryonic RNA extraction, we dissected intact embryos free from decidual tissue and extraembryonic membranes at days 8.5 through 11.5 post coitum, when 0.5 days post coitum is defined as noon of the day of the vaginal plug. Total RNA was prepared from homogenized embryos, adult tissues, embryoid bodies, or tissue culture cells using standard methods (Chirgwin et al., 1979). Ribonuclease protection assays were carried out as described previously (Shen and Leder, 1992). To generate an anti-sense probe for Cryptic, a full-length cDNA with a short (20 residue) poly(A) tail was linearized at an EcoRI site and sequences (Cheng and Flanagan, 1992) (M. M. Shen, unpublished observations). We therefore

RESULTS

Strategy for the isolation of mesodermally expressed genes

To study gene expression during mesoderm formation, we have utilized the in vitro differentiation of embryonic stem (ES) cells into embryoid bodies as a model system. Previously, we have analyzed the role of LIF, a cytokine involved in maintaining ES cell identity (Smith et al., 1988; Williams et al., 1988). In the absence of LIF, aggregates of ES cells in suspension culture form embryoid bodies that can spontaneously differentiate into a wide range of embryonic cell types. The predominant cell types that arise in this system include the extraembryonic visceral and parietal endoderm, which are primitive endoderm (hypoblast) derivatives, and mesodermal cell types such as cardiac muscle, skeletal muscle, and blood, which arise from primitive ectoderm (epiblast) (Doetschman et al., 1985). In the presence of LIF, the monolayer differentiation of ES cells is completely inhibited (Smith et al., 1988; Williams et al., 1988). During embryoid body differentiation, however, we found that exogenously added LIF selectively inhibits the formation of primitive ectoderm, while permitting the differentiation of primitive endoderm (Shen and Leder, 1992).

These observations prompted us to develop an in vitro differential screening approach for the identification of mesoderm-specific genes (Fig. 1). Thus, genes that are differentially expressed in untreated versus LIF-treated embryoid bodies should include markers for primitive ectoderm and early mesoderm. In order to identify such genes, we employed the differential display RT-PCR technique (Li and Pardee, 1992; Liang et al., 1993), a protocol that is capable of detecting rare transcripts, and that generates ‘fingerprints’ of mRNA populations. Since we were interested in identifying early mesodermally-expressed genes, we decided to analyze time points prior to or concurrent with the appearance of the earliest markers of terminally differentiated mesoderm. Our prior experience with this culture system indicated that early markers of terminal mesoderm differentiation, such as cardiac α-actin and β-globin, could first be detected at day 5 of differentiation, and were expressed in abundance at day 6 (Shen and Leder, 1992) (M. M. Shen, unpublished observations). We therefore

Fig. 1. Strategy for the isolation of mesoderm-specific genes. Embryoid bodies were formed by plating ES cells in suspension culture in the absence or the continual presence of 1,000 U/ml LIF (Shen and Leder, 1992). Under these conditions, LIF-treated embryoid bodies are inhibited from differentiating into mesodermal cell types, but are capable of forming extraembryonic endoderm. Mesoderm-specific genes will therefore be expressed in untreated embryoid bodies, but not LIF-treated embryoid bodies.
focused our analysis on days 4 and 6 of embryoid body differentiation in vitro.

Using this screening approach, we identified several bands on differential display gels that differed in intensity between RNA samples from embryoid bodies cultured in the absence versus the presence of LIF (Fig. 2). These candidate bands were excised for DNA elution, followed by PCR re-amplification. To confirm their differential expression, these re-amplified fragments were labeled and used as probes for northern blots of poly(A)^+ RNA from embryoid bodies. Of nine re-amplified fragments analyzed, four were differentially expressed at significant levels. The remaining five fragments were either not differentially expressed, or did not produce a detectable signal on northern blots (data not shown); similar outcomes have been observed previously by others (Liang et al., 1993). The four positive re-amplified fragments were cloned, and individual clones were then re-screened for differential expression on northern blots (Fig. 3).

One of these four cloned fragments (33-24) was expressed at higher levels in the presence of LIF. Database searching with this 288 bp fragment using the BLAST alignment program (Altschul et al., 1990) revealed that it contained a partial open reading frame with a 33 amino acid identity to the C terminus of the rat SCG10 gene. The SCG10 gene was originally identified as a specific neuronal marker (Anderson and Axel, 1985), and encodes a membrane-associated protein that is a member of the stathmin family (Stein et al., 1988; Schubart et al., 1989). The previously described expression of SCG10 in mouse morulae and blastocysts (Pampfer et al., 1992) is consistent with its expression in undifferentiated ES cells (Fig. 3). Thus, detection of the SCG10 gene correlated with our expectations for the differential display strategy.

We cloned three fragments from differential display gels that were expressed at higher levels in embryoid bodies cultured in the absence of LIF (Fig. 3). All three of these cloned fragments corresponded to novel genes. Two of these differentially expressed clones (32-1 and 35-3) were expressed at lower levels in LIF-treated embryoid bodies. In contrast, the third clone (34-25) was of particular interest because it hybridized to a single 1.0 kb transcript that was completely absent in RNA from LIF-treated embryoid bodies (Fig. 3). In subsequent analyses, as described below, we have focused on cloning and sequencing complete cDNAs for this gene, and evaluating its expression pattern in vivo.

**Isolation and sequence analysis of Cryptic cDNA clones**

By screening a primary cDNA library made from embryoid bodies at day 5 of differentiation, we isolated nine cDNAs corresponding to the 34-25 fragment. Of these nine independent cDNA isolates, four are apparently full-length, since they possess precisely the same 5'-end and have 3'-poly(A) tails. These putative full-length clones are 1039 bp in length, of which the 3'-terminal 407 bp corresponds to the complete 34-25 fragment. This cDNA sequence contains a single long open reading frame encoding a putative 202 amino acid protein (Fig. 4A). As described below, this predicted protein product is most similar to the Cripto gene product, and consequently we have named our novel gene **Cryptic**.

Examination of the predicted CRYPTIC product using the PROSITE database of protein motifs (Bairoch and Bucher,
identified an EGF-like motif located at residues 98-128. When we searched the translated GenBank database using the BLAST alignment program (Altschul et al., 1990), all of the high-scoring matches to the putative CRYPTIC protein produced alignments with its EGF-like motif. The only significant alignments that extended beyond the EGF-like motif occurred with the translated products of the murine and human Cripto genes, and the Xenopus FRL-1 gene (Fig. 5A).

**Fig. 4.** Sequence analysis of the Cryptic cDNA. (A) Complete sequence of the full-length Cryptic cDNA, with predicted translation product for initiation at the first methionine (GenBank accession number U57720). The positions of sequence motifs are marked as follows: the second initiator methionine (thick underline); the predicted signal sequence cleavage site (triangle); a potential N-glycosylation site (circled); the EGF-like motif (thick gray bar); and the polyadenylation signal (overline). There are also two overlapping matches to the consensus sequence for GPI cleavage and attachment (double underline), which consists of a tri-peptide with small amino acids (Ser, Ala, Gly, Asp, Asn, or Cys) in the first and third positions, located 7-14 amino acids upstream of a carboxy-terminal hydrophobic domain (Gerber et al., 1992; Coyne et al., 1993). One of the Cryptic cDNA clones has an additional 6 bp extension (GCCCTC) at its 3' end, prior to the poly(A) tail. (B) Hydropathy plot of the predicted CRYPTIC protein, produced using the Kyte and Doolittle algorithm (Kyte and Doolittle, 1982), with a window size of 9, as implemented by the Protean program of the LASERGENE software package (DNASTAR, Madison, WI). Positive scores indicate increasing hydrophobicity. (C) Plot of scores for signal sequence cleavage, as predicted by the von Heijne algorithm (von Heijne, 1986). A score over 3.5 is indicative of potential cleavage. In the case of the putative CRYPTIC protein, initiation at the second methionine is predicted to result in signal sequence cleavage after the serine at position 35 (score 3.91), as marked by the arrow. Note that cleavage after the leucine at position 32 is unlikely, since signal sequence cleavage after a leucine is rare in eukaryotic proteins (von Heijne, 1985).
The predicted CRYPTIC protein displays striking sequence similarities with the putative murine and human CRIPTO and frog FRL-1 proteins (Fig. 5A), with 29% amino acid identity to murine CRIPTO and 23% identity to FRL-1. Most of these sequence similarities are found in the single EGF-like motif and a second conserved cysteine-rich region that has not been found in any other known proteins (Fig. 5B). Furthermore, these genes share a predicted overall alignment that is comprised of a potential N-terminal leader sequence, the EGF-like motif, the second cysteine-rich motif, and a C-terminal hydrophobic domain; in addition, CRIPTO and FRL-1 share a short region of similarity upstream of the EGF-like motif, which is not found in CRYPTIC (Fig. 5B). However, the distant relationship of FRL-1 to both Cryptic and Cripto suggests that

**Fig. 5.** Sequence comparison of the putative Cryptic, Cripto, and FRL-1 protein products. (A) Alignment of the putative murine CRYPTIC, murine CRIPTO, human CRIPTO, and Xenopus FRL-1 proteins. Amino acids that are identical among at least three of the four proteins are boxed. The sequence alignment was produced using the default parameters of the Clustal algorithm in the MegAlign program of the LASERGENE software package (DNASTAR, Madison, WI). The sequence of murine CRIPTO is from Doono et al. (1993), human CRIPTO is from Ciccodicola et al. (1991), and frog FRL-1 is from Kinoshita et al. (1995). (B) Schematic organization of the putative Cryptic, Cripto, and FRL-1 encoded proteins. Shown are the positions of the EGF-like motif and a second, unrelated cysteine-rich motif (CFC domain), with the approximate positions of the conserved cysteines (vertical lines). Also depicted is the percentage amino acid identity in these two conserved regions between Cripto and Cripto, and between Cripto and FRL-1; there is no significant amino acid conservation in the remainder of these proteins. The positions of the putative signal sequence (white box) and C-terminal hydrophobic region (gray box) are also shown. For Cripto, the sequence signal is shown in two parts, corresponding to the two potential translation initiation sites. (C) Alignment of EGF-like motifs from the putative products of Cryptic, Cripto, FRL-1 and several EGF-related growth factors. Note the conservation of the six cysteines and two glycines (boxed), as well as an aromatic amino acid (shaded box) found two residues before the second glycine, but not of the arginine residue (shaded) that has been implicated in high-affinity erbB receptor binding (Groenen et al., 1994). In addition, the consensus residues implicated in Ca2+ binding by many EGF-like domains (Handford et al., 1991) are absent from CRYPTIC, Cripto, and FRL-1. The sequences and alignment of murine EGF, human TGF-α, human amphiregulin (AR), and human heparin-binding EGF (HB-EGF) are from Groenen et al. (1994).
A novel EGF-related gene

**Fig. 6.** In vitro translation and secretion of CRYPTIC protein. (A) In vitro translation analysis performed with a full-length Cryptic cDNA. Translation products were labeled with [35S]cysteine and resolved on a 15% SDS-polyacrylamide gel; the position of molecular mass markers at 33.3 kDa, 28.6 kDa, and 19.4 kDa are indicated at left. The indicated lanes show translation products obtained in the absence (lane 1) or presence (lanes 2 through 9) of Cryptic RNA transcripts (‘template’), and in the absence (lanes 2 and 9) or presence of increasing concentrations of microsomal membranes (‘microsomes’), which were present at 0.5, 1.0, 1.5, or 2.0 units per reaction (lanes 3 through 8). For analysis of translation products by deglycosylation, reaction products obtained in the presence of 2 units of microsomes were treated with Endo H (lane 7) or PNGase F (lane 8). An equivalent amount of translation product was run in each lane. (B) Schematic structure of fusion proteins of CRYPTIC with alkaline phosphatase (AP), and levels of activity obtained in transfected COS cell supernatants. Five constructs were assayed for alkaline phosphatase activity: (1) fusion of full-length CRYPTIC to alkaline phosphatase; (2) same as construct 1, except for deletion of the C-terminal hydrophobic region; (4) the parental APtag-2/Amp vector as a negative control; and (5) the APtag-4 expression vector, which contains alkaline phosphatase with an intact signal sequence. Following parallel transfections of these constructs into COS cells, culture supernatants were examined for alkaline phosphatase activity: (1) fusion of full-length CRYPTIC to alkaline phosphatase; (2) same as construct 1, except for deletion of the C-terminal hydrophobic region; (3) same as construct 1, except for deletion of the C-terminal hydrophobic region; (4) the parental APtag-2/Amp vector as a negative control; and (5) the APtag-4 expression vector, which contains alkaline phosphatase with an intact signal sequence.

**Fig. 7.** Ribonuclease protection analysis of Cryptic expression. Each hybridization contained an anti-sense riboprobe for Cryptic, and an anti-sense riboprobe for ribosomal protein L32 (rpL32) as an internal standard. In this experiment, 15 µg of total RNA was used, with an autoradiographic exposure time of 7 days for Cryptic and 12 hours for ribosomal protein L32. RNA analyzed was from undifferentiated F9 cells (F9) or from F9 monolayers treated with retinoic acid for six (d6) or eight (d8) days; from undifferentiated ES cell monolayers (ES) or from ES embryoid bodies cultured in the absence of LIF for three, seven, or eleven days; from mouse embryos at 8.5, 9.5, 10.5, or 11.5 days post coitum; or from adult mouse brain (Br), small intestine (In), kidney (Ki), liver (Li), salivary gland (SG), spleen (Sp), testis (Te), or uterus (Ut). As a negative control, 50 µg of yeast tRNA was used. For the Cryptic riboprobe, faint protected bands are visible in the lanes corresponding to undifferentiated ES cells and to day 8.5 mouse embryos.

**FRL-1** is not an ortholog of either murine gene. Therefore, based on these extensive sequence similarities, we propose that these three genes comprise a new gene family, which we name ‘CFC’ (for Crypt, Frl-1, and Cryptic), and that their unique cysteine-rich domain be termed the ‘CFC motif’.

As noted previously for CRIPTO (Dono et al., 1993; Brandt et al., 1994), the variant EGF-like motif found in CFC family members is highly unusual. All EGF-like motifs contain six cysteines, which form three disulphide bonds in the case of EGF (Savage et al., 1973; Kohda and Inagaki, 1992; Montelione et al., 1992). In CRYPTIC, CRIPTO, and FRL-1, the first two cysteines of the EGF-like motif are adjacent, thereby eliminating the ‘A-loop’ that is normally found between these residues in other EGF-like repeats (Fig. 5C). In addition, the spacing between the third and fourth cysteines is reduced relative to other EGF-like repeats, resulting in a smaller ‘B-loop’. Despite these unusual sequence features of the EGF-like motif, however, several key residues that are found in all EGF-like repeats are conserved in CFC members (Fig. 5C).

**Secretion of the CRYPTIC protein product**

Examination of the Cryptic cDNA sequence revealed the presence of two in-frame candidate translation initiation sites (Fig. 4A). Interestingly, the 5’ proximal AUG at bp 256 occurs in a relatively poor context for translation initiation, since it lacks both a purine at the −3 position and a G at position +4 (Kozak, 1986). On the other hand, the second candidate initiation site possesses a purine at the −3 position, and would be expected to represent a strong initiation site (Kozak, 1986). Initiation at either of these AUG sequences would produce a protein product containing a hydrophobic region near the N terminus, which might correspond to a signal sequence (Fig. 4B). As predicted by the algorithm of von Heijne (1986) and by the PSORT protein sequence analysis program (Nakai and Kanehisa, 1992), this candidate signal sequence could potentially be cleaved after the serine at position 35 (or 22 amino acids following the second methionine) (Fig. 4C).

To determine whether these predictions were valid, we performed in vitro translation analysis utilizing a full-length Cryptic cDNA clone (Fig. 6A). We detected two translation products that migrated near the positions expected for translation initiation at the first and second AUG codons (Mr, 21,800 and 20,400, respectively) (Fig. 6A, lanes 2 and 9). In the
presence of increasing amounts of microsomes, which permit membrane translocation and signal sequence cleavage (Blobel and Dobberstein, 1975), these two bands were replaced by a single band migrating near 23 kDa, as well as faint bands at 21 kDa and 20 kDa (Fig. 6A, lanes 3 to 6). Because this result was consistent with signal sequence cleavage followed by glycosylation, we treated the translation products obtained in the presence of microsomes with the deglycosylating enzymes EndoH or PNGase F. After treatment with either of these enzymes, we observed a major band at 20 kDa, together with fainter bands at 23 kDa and 22 kDa, which presumably correspond to translation products with uncleaved signal sequences (Fig. 6A, lanes 7 and 8). These data are consistent with signal sequence cleavage of both translation products, resulting in a cleaved product of 20 kDa, followed by glycosylation to produce a mature protein migrating at approximately 23 kDa. Since glycosylation occurs in the microsomal lumen, these results indicate that the CRYPTIC protein product can undergo membrane translocation and signal sequence cleavage in vitro.

To examine whether CRYPTIC protein can be secreted in vivo, we expressed tagged fusion proteins in transfected COS cells, and assayed for secretion of the tagged CRYPTOIC proteins into culture supernatants. For this purpose, we fused CRYPTIC at its C terminus to secreted placental alkaline phosphatase (Fig. 6B, construct 1), using a modified version of the AProtag-2 expression vector (Cheng and Flanagan, 1994). To examine whether levels of secretion were affected by the choice of translation initiation site or the absence of the C-terminal hydrophobic region, we also generated constructs in which the first AUG site or the C-terminal hydrophobic region were deleted (Fig. 6B, constructs 2 and 3). When we assayed culture supernatants from COS cells transfected with these constructs, we observed comparable levels of alkaline phosphatase activity (Fig. 6B). Note that these levels were approximately 20- to 40-fold lower than those obtained by transfection of a similar expression vector containing alkaline phosphatase with its intact signal sequence, which should direct efficient secretion (Fig. 6B, construct 5). These observations suggest either that the CRYPTIC signal sequence is relatively inefficient at directing secretion, and/or that the CRYPTIC protein frequently remains associated with the cell surface or extracellular matrix.

Notably, the C-terminal hydrophobic region of CRYPTIC might correspond to a transmembrane domain, resulting in a very short cytoplasmic tail, or might undergo processing for glycosyl-phosphatidylinositol (GPI) linkage. Indeed, two overlapping matches to the consensus sequence for GPI cleavage and attachment are located starting at amino acids 172 and 173 (Fig. 4A). However, comparable levels of secretion were observed with or without the C terminus (Fig. 6B, constructs 1 and 3), arguing against a transmembrane form of the CRYPTIC protein. In addition, preliminary experiments with constructs in which CRYPTIC is fused at its N terminus with alkaline phosphatase, rather than at the C terminus, show that similar levels of secretion are observed whether or not the C-terminal hydrophobic domain is present (H. Wang and M. M. Shen, unpublished), which also argues against GPI-linkage of CRYPTIC protein. Thus, the role of the C-terminal hydrophobic domain of CRYPTIC is presently unclear.

![Fig. 8. Whole-mount in situ hybridization analysis of Cryptic expression. Bright-field illumination of embryos from days 6.5 and 7.5 of gestation, corresponding to the early streak to the late head-fold stages of Downs and Davies (1993). Bars, 0.2 mm. (A) Lateral view of three early to late-streak stage embryos, showing patchy expression in the mesodermal wings and localized expression at the anterior end of the primitive streak (arrow). (B) Lateral view of three late-streak to early head-fold stage embryos, anterior to the left. Expression is found in the head process (arrow) and medial region of the egg cylinder. The staining in the head process is more intense at its caudal end, where the node is forming. (C) Lateral view of four early-to late-head-fold stage embryos, with the most advanced embryo on the left. Note the progressive narrowing of the lateral staining (arrow) with developmental age. (D) Ventral (bottom) view of late head-fold stage embryo, showing staining that fills the node (arrow) and weaker staining in the notochordal plate proceeding rostrally to the left. The lateral staining extends completely around the embryo, and is excluded from the paraxial mesoderm. (E) Rostral view of an early head-fold stage embryo, displaying expression in the node (arrowhead), in the axial midline, and in the anterior lateral plate mesoderm (arrow) at the base of the head folds. (F) Lateral view of a late head-fold stage embryo, anterior to the left, showing staining in the node (arrowhead), notochordal plate (arrow), and the lateral mesoderm. (G) Caudal view of an two-somite stage embryo, displaying faint and restricted staining laterally, converging at the base of the allantois. Low levels of staining are also present in scattered cells within the allantois. Staining within the node at this stage is only found in its central region (arrow).]
Temporal and spatial localization of Cryptic expression during gastrulation

To provide insights into potential roles for Cryptic during development, we have investigated its expression in embryonic and adult mouse tissues. As a first step, we used a ribonuclease protection assay to examine total RNA from mid-gestation embryos, and found very low levels of expression at day 8.5 post coitum, but no expression at later stages (Fig. 7). We also surveyed several adult tissues, and failed to detect any evidence of Cryptic expression in the tissues examined. Thus, unlike Cripto, which is expressed in several adult tissues, such as brain, heart, lung, and spleen (Dono et al., 1993), Cryptic expression was not observed in adult brain, intestine, kidney, liver, salivary gland, spleen, testis or uterus (Fig. 7). In addition, because Cryptic expression was up-regulated during ES cell differentiation, we investigated whether its expression could be induced by retinoic acid treatment of F9 embryonal carcinoma cells. We found that Cryptic is not expressed in undifferentiated F9 cells, and is weakly up-regulated during late stages of retinoic acid-induced differentiation (Fig. 7). Again, this contrasts with the high-level expression of Cripto in undifferentiated F9 cells and its rapid down-regulation after retinoic acid treatment (Dono et al., 1993).

Based on its expression pattern in differentiating ES embryoid bodies, we anticipated that Cryptic would be expressed in early mesoderm derivatives in vivo. We have therefore examined Cryptic expression by whole-mount in situ hybridization (Wilkinson, 1992; Rosen and Beddington, 1993) of post-implantation mouse embryos, beginning with the onset of gastrulation. In the mouse, gastrulation begins at day 6.5 post coitum with the formation of the primitive streak, which marks the region in which epiblast cells delaminate, migrate, and become allocated to the nascent mesoderm (reviewed by Theiler, 1989; Kaufman, 1992; Hogan et al., 1994). Mesodermal cells appear to acquire regional diversity based on their antero-posterior position in the streak (Tam and Beddington, 1987; Lawson et al., 1991; Smith et al., 1994), becoming allocated to the axial mesoderm, which forms the notochord and prechordal plate, to the paraxial mesoderm, which forms the somites, to the intermediate and lateral mesoderm, which form the developing kidneys and visceral mesenchyme, and to the extraembryonic mesoderm. By the late primitive streak stage, the characteristic structure of the node appears at the anterior end of the streak, morphologically resembling a dimpled region at the tip of the egg cylinder. At the early head-fold stage, the node consists of a relatively broad oval region that tapers anteriorly into the contiguous notochordal plate, the precursor of the notochord (Sulik et al., 1994).

Our in situ hybridization results indicate that Cryptic is specifically expressed in two distinct spatial domains during gastrulation. First, Cryptic is expressed in the presumptive axial mesoderm, beginning with localized staining at the anterior end of the primitive streak by the late-streak stage (Fig. 8A, arrowhead), and is later found in the head process (Fig. 8B, arrowhead). Secondly, Cryptic is initially expressed by scattered cells in the mesodermal wings emerging from the primitive streak (Fig. 8A), and later becomes confined to a lateral band around the egg cylinder (Fig. 8B). In early and late head-fold stage embryos, Cryptic is expressed in a striking
‘Easter egg’ pattern (Fig. 8C). Strong staining is found within the node, with weaker staining extending anteriorly along the notochordal plate (Fig. 8D-F); notably, staining is absent from the prechordal plate and paraxial mesoderm. In addition, two symmetrical bands of Cryptic expression extend around the embryo, corresponding to the presumptive intermediate and lateral mesoderm (Fig. 8D-F). These lateral bands of Cryptic expression join rostrally at the anterior lateral plate, in the region of the cardiac mesoderm (Fig. 7E), and meet caudally at the base of the allantois (Fig. 8G). By early somite stages, Cryptic staining becomes faint, but can still be found in the center of the regressing node and in the intermediate mesoderm adjacent to the nascent somites (Fig. 8G). By midday 8.5, staining has apparently disappeared: after this embryonic stage we have not detected Cryptic expression by whole-mount in situ hybridization, consistent with our ribonuclease protection results (Fig. 7).

To examine the expression pattern of Cryptic in greater detail, we sectioned early and late head-fold stage embryos that had been stained by whole-mount in situ hybridization (Fig. 9A). In frontal sections through the node of an early head-fold stage embryo, Cryptic staining is restricted to the ventral mesoderm layer of the node, and is not observed in the overlying dorsal layer that is continuous with the epiblast (Fig. 9B and E). However, in sections that are immediately rostral to the node, staining is found in the notochordal plate and in a narrow patch of the adjacent midline neuroectoderm, which corresponds to the prospective floor plate of the neural tube (Fig. 9C and F). In frontal sections through more rostral regions of the notochordal plate, this midline neuroectoderm staining becomes slightly broader in extent (Fig. 9D and G). This expression in the prospective floor plate is also observed in transverse sections through the head folds of a late head-fold stage embryo (Fig. 9H and I). In addition, these sections show that Cryptic is expressed by both the somatopleural and splanchnopleural layers of the lateral plate mesoderm, flanking the coelomic cavity (Fig. 9I). Finally, sagittal sections confirm the specific expression of Cryptic by the anterior lateral plate mesoderm (Fig. 9J). Thus, these cryosections confirm that Cryptic is expressed by the axial and lateral plate mesoderm, and additionally reveal expression in the prospective floor plate of the neural tube.

**DISCUSSION**

By employing a differential display approach to identify mesoderm-specific genes using the in vitro differentiation of ES embryoid bodies, we have cloned a novel gene, Cryptic, which encodes a secreted growth factor-like molecule. Expression of the Cryptic gene is temporally and spatially restricted during early mouse embryogenesis, and is localized during gastrulation to known embryonic signaling centers, namely the node, notochordal plate, and prospective floor plate. Furthermore, Cryptic and its relatives Cripto and FRL-1 define a new family of growth factor-like molecules that are expressed in early vertebrate embryogenesis, and which may signal through activation of FGF receptors. Taken together, these results suggest that Cryptic may represent a signaling molecule involved in pattern formation and/or differentiation during mouse gastrulation.

**A novel strategy for the isolation of genes expressed in early embryonic development**

Gene discovery approaches such as differential screening or gene-trapping are valuable because they can identify genes that have not previously been implicated in embryonic development. For this reason, a variety of differential screening approaches have been previously attempted in order to isolate candidate regulatory genes for mouse embryonic development. For example, differential screening and subtractive hybridization techniques have been employed to compare gene expression in pre-implantation embryos (Rothstein et al., 1992), in post-gastrulation embryos (Wilkinson et al., 1987), and in differentiating ES embryoid bodies (Poirier et al., 1991). However, relatively few candidate regulatory genes have been identified in such screens, perhaps because most regulatory genes are expressed at low abundances that are beyond the limit of detection of the screening approaches utilized. Moreover, a sensitive detection technique is necessary when comparing populations of mRNAs that are highly heterogeneous, such as those derived from intact embryos or from ES embryoid bodies.

To overcome such limitations, we have employed the differential display RT-PCR protocol, which can detect rare transcripts (Liang and Pardee, 1992). This protocol has recently been employed successfully to identify rare to moderately-expressed transcripts in a variety of systems, including the hematopoietic differentiation of ES embryoid bodies (Guimarães et al., 1995). In the case of Cryptic, it is clear from our findings that this gene is indeed expressed at low levels in differentiating embryoid bodies. Another advantage of the differential display approach is its ability to catalog the complete repertoire of differentially expressed genes, through the use of a sufficient number of random 5’ oligonucleotide primers. However, since we have only screened a small percentage of the oligonucleotides that would be necessary for a comprehensive analysis (Liang and Pardee, 1992), we anticipate that more many differentially expressed genes remain to be isolated using our screening strategy.

In our screening methodology, we compared ES embryoid bodies cultured in the absence versus the presence of LIF, resulting in two cell populations that differ primarily in the expression of mesoderm-specific genes (Shen and Leder, 1992). Since these two populations of embryoid bodies are initially identical, we decided to examine early time points of mesoderm differentiation, presumably enriching for candidate genes that are involved in mesoderm formation in vitro. As discussed below, Cryptic appears to represent a promising candidate for such a gene. Therefore, our results support the notion that this embryoid body screening strategy is an effective method for identifying candidate embryonic regulatory genes. Although our in vitro system can only represent a limited model for in vivo mesoderm differentiation, our experience thus far suggests that further investigation will be fruitful. Moreover, similar approaches can be utilized to isolate genes expressed during ES cell differentiation into other cell types, such as neuronal cells (Bain et al., 1995; Strübing et al., 1995).

**Potential roles for Cryptic during gastrulation**

The spatial and temporal localization of Cryptic expression during gastrulation is consistent with possible roles in
mesoderm and/or neuroectoderm differentiation and patterning. First, the expression of Cryptic in the axial and lateral mesoderm suggests a potential involvement in specifying the regional identity of newly formed mesoderm. Fate mapping studies have indicated that the regional identity of mesoderm correlates with the rostrocaudal origin of progenitor cells within the primitive streak (Tam and Beddington, 1987; Smith et al., 1994). Thus, at day 7.5 of embryogenesis, cells within the node primarily give rise to the notochord, but also may contribute to the prechordal plate, head mesenchyme, the definitive endoderm, and the floor plate of the neural tube (Lawson et al., 1991; Selleck and Stern, 1991; Beddington, 1994; Sulik et al., 1994). In addition, cells within the rostral third of the streak are fated to contribute primarily to somites in the paraxial mesoderm, the middle third primarily to the intermediate and lateral mesoderm, and the caudal third primarily to extra-embryonic mesoderm (Smith et al., 1994). However, few genes are known to be involved in the allocation of mesodermal progenitors among these regional subpopulations; notably, FGF receptor 1 has been implicated in this process (Deng et al., 1994; Yamaguchi et al., 1994). Since Cryptic does not appear to be expressed within the primitive streak, it is unlikely that it would be involved in the establishment of such regional differences, but might instead be involved in their maintenance.

Secondly, Cryptic may potentially be involved in signaling from the mesoderm to the overlying ectoderm. In this regard, it is noteworthy that Cryptic is one of the earliest known markers of the prospective floor plate, since its expression at the early head-fold stage in the midline neuroectoderm clearly precedes the expression of Sonic hedgehog (Echelard et al., 1993; Roelink et al., 1994), and may coincide with or precede the floor plate expression of the winged-helix transcription factor HNF-3β (Ang et al., 1993; Monaghan et al., 1993; Sasaki and Hogan, 1993), and of a murine homolog of the patched transmembrane protein (Goodrich et al., 1996). One possible explanation for the expression of Cryptic in the node, notochordal plate, and floor plate is that this pattern simply reflects the common origin of these tissues from cells located at the anterior end of the primitive streak (Lawson et al., 1991; Selleck and Stern, 1991; Beddington, 1994; Sulik et al., 1994). However, this explanation is inconsistent with the results of single-cell dye tracing experiments that indicate that floor plate cells are derived from the dorsal layer of the node (where Cryptic is not expressed), while notochordal plate cells are derived from the ventral layer (Beddington, 1994; Sulik et al., 1994). Alternatively, expression of Cryptic in the midline neuroectoderm may be dependent upon the homeo genetic induction of the floor plate by the notochordal plate (Placzek et al., 1993). These considerations also raise the possibility that Cryptic might itself be involved in neural tube patterning, which is particularly intriguing given the hypothesized role for an FGF-like activity in neural induction (Doniaich, 1995).

**Cryptic, Cripto and FRL-1 define a new family of growth factor-like molecules**

The extensive sequence similarities between the putative Cryptic, Cripto, and FRL-1 encoded proteins define a new family of EGF-related genes, which we have named the ‘CFC’ family. Among the identifying features of this family are the variant EGF-like motif, the novel cysteine-rich domain (the CFC motif), and the C-terminal hydrophobic region. These sequence similarities suggest that members of the CFC family may possess similar biochemical properties. Moreover, the available evidence indicates that CFC family members exhibit growth factor-like activities and are expressed in specific temporal and spatial patterns during early embryogenesis.

First, previous studies of the expression patterns of Cripto and FRL-1 are consistent with their involvement in early vertebrate embryogenesis. Thus, the Cripto gene was originally cloned from the human teratocarcinoma cell line NTERA2, and was shown to be down-regulated upon retinoic acid-induced differentiation (Ciccodicola et al., 1989). Expression of murine Cripto is found in early and late-streak stages of gastrulation in the epiblast and newly formed mesoderm of the primitive streak, and later in the outflow tract of the heart at day 8.5 (Dono et al., 1993; Johnson et al., 1994). By contrast, expression of FRL-1 in *Xenopus* is temporally restricted to the period of gastrulation, disappearing during early neurulation, but does not display any apparent spatial localization at the mRNA level (Kinoshita et al., 1995). Interestingly, we have found that there is relatively little overlap in the expression domains of Cripto and Cryptic during gastrulation (H. Wang and M. M. Shen, unpublished). These observations suggest that the members of the CFC family may perform non-redundant functions, and that their biological activities may be relatively unique.

Secondly, there are several lines of evidence that indicate that the Cripto protein possesses potent growth factor activity. Thus, conditioned medium from CHO cells transfected with a human Cripto expression construct can stimulate the proliferation of a variety of human breast carcinoma cell lines (Brandt et al., 1994). In addition, transfection of human Cripto can transform NIH/3T3 cells and NOG-8 mouse mammary epithelial cells (Ciccodicola et al., 1989). Moreover, several studies have documented Cripto overexpression in human breast, colorectal, gastric, and pancreatic carcinomas (Ciardiello et al., 1991; Kuniyasu et al., 1991; Saeki et al., 1992; Friess et al., 1994; Qi et al., 1994), suggesting that Cripto may involved in the autocrine or paracrine stimulation of tumor cell growth.

Finally, the available data suggest that members of the CFC family interact with receptors that are distinct from the four known members of the erbB receptor family, unlike other EGF-related growth factors. For example, the mitogenic effects of EGF are not competed by recombinant Cripto protein, suggesting that Cripto does not bind to the EGF receptor p170erbB-1 (Brandt et al., 1994). Moreover, the variant EGF-like motif of the CFC family lacks important residues that mediate high-affinity binding to erbB receptors, including the residues of the missing ‘A loop’ and the conserved arginine residue prior to the sixth cysteine (Groenen et al., 1994; Nagata et al., 1994). Consistent with these observations, the isolation of FRL-1 in a functional assay for FGF receptor autophosphorylation in yeast has suggested that it may represent a novel ligand for FGF receptors (Kinoshita et al., 1995). Notably, the mesoderm and neural-inducing activities of FRL-1 in *Xenopus* animal cap assays can be blocked by a truncated FGF receptor lacking the cytoplasmic tyrosine kinase domain (Kinoshita et al., 1995), which acts in a dominant-negative manner (Amaya et al., 1991, 1993). However, the activities of FRL-1 can also be blocked by a mutant truncated FGF receptor, which has no effect on FGF-mediated signaling (Byers et al., 1992; Amaya...
et al., 1993), indicating that the molecular mechanism of FRL-1 activation of FGF receptors is not identical to that employed by FGFs (Kinoshita et al., 1995). At present, the interaction of FRL-1 with FGF receptors is poorly understood, and there is no evidence that FRL-1 protein directly binds to FGF receptors. An alternative possibility is that FRL-1 indirectly activates the FGF receptor after binding to a novel receptor molecule, as has been found for activation of the EGF receptor by ligands for G-protein-coupled receptors (Daub et al., 1996). In either case, the sequence similarities of Cryptic and Crito to FRL-1 suggest that these two mammalian relatives may also utilize the FGF signal transduction pathway. If so, it will be essential to examine whether members of the CFC family may be responsible in vivo for some of the physiological effects that have been previously attributed to FGFs.

We thank Peng Liang for helpful suggestions on differential display; Randy Johnson for a gift of differential display primers; Yu-Ting Yan, Sandy Price, and Sheryl Corado for assistance with cDNA cloning and sequencing; John Flanagan for generously providing alkaline phosphatase expression vectors; and Isaac Edery for use of his cryostat. We are particularly grateful to Yu-Ting Yan for his help in designing the experiments in Figure 6, and to Cory Abate-Shen for encouragement. This work was supported in part by a grant to M.M.S. in her invaluable support and advice. We also thank Debra Bader, Rich his cryostat. We are particularly grateful to Y u-Ting Y an for his help in designing the experiments in Figure 6, and to Cory Abate-Shen for encoura...
...tissues in the mouse embryo during gastrulation and early organogenesis. 


(Accepted 24 October 1996)