Structure-function analysis of the EGF-CFC family member Cripto identifies residues essential for nodal signalling

Gabriella Minchiotti1,‡, Giuseppe Manco2, Silvia Parisi1, Carmine T. Lago1, Frederic Rosa3,* and M. Graziella Persico1,*

1International Institute of Genetics and Biophysics, CNR, Via G. Marconi 12, 80125 Naples, Italy
2Institute of Protein Biochemistry and Enzymology, CNR, Via G. Marconi 12, 80125 Naples, Italy
3Ecole Normale Superieure, 46 rue d’Ulm, 75230 Paris CEDEX 05, France
*These two authors contributed equally to this work
‡Author for correspondence (e-mail: minchiot@iigb.na.cnr.it)

Accepted 21 August 2001

SUMMARY

Cripto is the founding member of the family of EGF-CFC genes, a class of extracellular factors essential for early vertebrate development. In this study we show that injection of Cripto recombinant protein in mid to late zebrafish Maternal-Zygotic one-eyed pinhead (MZoep) blastulae was able to fully rescue the mutant phenotype, thus providing the first direct evidence that Cripto activity can be added extracellularly to recover oep-encoded function in zebrafish early embryos. Moreover, 15 point mutations and two deletion mutants were generated to assess in vivo their functional relevance by comparing the ability of cripto wild-type and mutant RNAs to rescue the zebrafish MZoep mutant. From this study we concluded that the EGF-CFC domain is sufficient for Cripto biological activity and identified ten point mutations with a functional defective phenotype, two of which, located in the EGF-like domain, correspond to loss-of-function mutations. Finally, we have developed a three-dimensional structural model of Cripto protein and used it as a guide to predict amino acid residues potentially implicated in protein-protein interaction.

Key words: Cripto, nodal signalling, zebrafish, one-eyed pinhead, structure-function

INTRODUCTION

The process of embryonic induction in which one population of cells influences the developmental fate of another population plays an essential role in establishing the body plan of all multicellular organisms. These inductive events as well as subsequent patterning events, rely upon cell-cell interactions mediated by extracellular molecules such as the members of the EGF-CFC family. Cripto is the founding member of this gene family that includes human, mouse and chicken cripto, Xenopus FRL1, mouse and human cryptic, and zebrafish one-eyed pinhead (oep) (Ciccodicola et al., 1989; Dono et al., 1993; Kinoshita et al., 1995; Shen et al., 1997; Zhang et al., 1998; Bamford et al., 2000; Colas and Schoenwolf, 2000). All the proteins of this family contain a signal sequence, a characteristic EGF-like domain, a second cysteine-rich region called the CFC domain, and a hydrophobic C terminus. Both Oep and Cripto proteins are membrane-bound (Zhang et al., 1998; Minchiotti et al., 2000). Furthermore, Cripto is associated with the cytoplasmic membrane through a GPI-anchor, thus suggesting that the anchorage to the membrane by a GPI-linkage could be important in determining the cell-cell interactions required during vertebrate embryo development (Minchiotti et al., 2000).

Comparison of the phenotypes of zebrafish embryos lacking both the maternal and zygotic expression of oep (MZoep embryos) and of mouse cripto mutant has revealed surprising similarities, both mutants displaying defective germ layer and trunk formation. Furthermore, the position and the direction of the anterior-posterior axis appear abnormal in both mutants (Ding et al., 1998; Gritsman et al., 1999). The injection of cripto mRNA is able to fully rescue MZoep mutants (Zhang et al., 1998), thus supporting the idea that despite potential differences in morphogenetic movements, vertebrate germ layer formation and positioning of the anterior-posterior axis are controlled by conserved mechanisms involving the EGF-CFC proteins (Zhang et al., 1998; Gritsman et al., 1999).

The developmental roles of cripto, cryptic and oep genes have been analysed in detail (Ding et al., 1998; Gritsman et al., 1999; Xu et al., 1999; Yan et al., 1999) but the signal transduction pathways activated by these molecules remain elusive. Although originally described as an EGF-related factor (Ciccodicola et al., 1989), Cripto is unable to bind any of the four known members of the ErbB receptor family (Salomon et al., 1989). This lack of binding is consistent with the absence, in the Cripto protein, of some amino acids essential for high affinity binding to EGFRs, including the entire A loop between the first two cysteines (Salomon et al., 1999). In vitro studies on mammary epithelial cells have suggested that Cripto acts as a non-autonomous signal leading to Ras/raf/Mek/MAPK
activation (Bianco et al., 1999; Salomon et al., 1999). However, strong genetic evidence has indicated that Oep and Cripto do not act as an instructive signal in MAP signalling but render cells competent to respond to an instructive signal such as Nodal or another TGFβ ligand (Gritsman et al., 1999; Shen and Schier, 2000). Moreover, chimeric embryos of wild-type and Cripto+/- ES cells develop normally and both cell types are present in adult tissues, suggesting that wild-type cells can rescue the mutant cell phenotype (Xu et al., 1999). These results can either indicate a juxtacrine activity of the GPI-anchored Cripto or that the rescue could be due to a soluble form released from the membrane of wild-type cells.

In this study we provide the first direct evidence that Cripto activity can be provided extracellularly by injecting recombinant Cripto protein into the blastoderm of mid to late MZoep blastulae to recover oep-deficient function in zebrafish early embryos. Moreover, injection of cripto mutant RNAs into MZoep embryos provided a means of identifying both regions and residues within the Cripto protein required for its activity. Finally, as the three-dimensional (3D) structure of Cripto is not known at present, a programme of fold recognition was used to search for proteins distantly related to Cripto in order to obtain a model for the analysis of known and possible mutations in their potential 3D context.

MATERIALS AND METHODS

Plasmids and mutants

The cripto-His- (sequence from nucleotide –5 to +156 of the cripto cDNA) expressing vector was obtained by PCR, using the complete cripto cDNA as template and the appropriate oligonucleotides (Minchiotti et al., 2000). The amplified fragment was cloned into the EcoRI site of pcDNA3-His vector. The AgeI/HincII fragment from pMT/BiP/V5-HisB vector (Invitrogen), containing both the 6XHis epitope and the SV40 late polyadenylation signal, was cloned into the EcoRI site of pcDNA3 to obtain pcDNA3-His. When necessary, restriction sites were blunt-ended using Klenow polymerase. All the mutant derivatives of cripto (both deletions and amino acid substitutions) were generated by PCR mutagenesis using appropriate oligonucleotides (Ho et al., 1989). In all cases the amplified fragments were sequenced in both directions by the dideoxynucleotide method (Hattori and Sakaki, 1986).

Synthesis of RNA and rescue assay

Sense-capped RNA was synthesised using the mMESSAGE mMACHINE system (Ambion) and T7 RNA polymerase after XhoI digestion of cripto derivatives in both pCDNA3 and in pcDNA3-His vectors. In vitro synthesised RNA was microinjected in 1-4 cell stage oepm134 (Zhang et al., 1998) mutant embryos. Embryos were phenotypically analysed and documented 24-36 hours after fertilisation. Genotyping was not required as 100% of the progeny were MZoep mutants.

Cell cultures and western blot

Growth conditions of human embryonal kidney cells 293 and 293T (DuBridge et al., 1987), cell transfection, protein labelling and western blot analyses were performed as previously described (Minchiotti et al., 2000). The anti-Ha monoclonal antibody (No. Sc-7392, Santa Cruz Biotechnology) was used following the manufacturer’s instructions.

Cripto recombinant proteins

The soluble Cripto-Fc protein was produced using the plg-Tail expression system (No. MBK-006-5, R&D) and purified from conditioned medium of 293T cells following the manufacturer’s instructions. Anti-Cripto-Fc antibodies were used to show that Cripto-Fc exhibited the immunological determinants of Cripto (data not shown).

Cripto-His protein was purified from conditioned medium of stably transfected 293 cells using the QIaexpress protein purification system (Quiagen). Purified proteins were dialysed against 50 mM sodium phosphate buffer, pH 8.0. Protein concentration was quantified by BioRad assay (BioRad Laboratories) and aliquot samples were stored at –80°C.

NH2-terminal amino acid sequence analysis

NH2-terminal amino acid sequence analysis was performed after protein separation by SDS-PAGE and electrotransfer onto PVDF membrane ProBlott (Applied Biosystems), as described by Matsudaira (Matsudaira, 1987). The band of interest was cut out and subjected to automated Edman degradation by an Applied Biosystems gas-phase sequencer (model 477 A) equipped with an on-line 120 A phenylthiolyhydration analyser, following the manufacturer’s instructions.

Injection of recombinant protein

Both Cripto-His and Cripto-Fc proteins were diluted in 1 mg/ml BSA in Ringer’s medium. As tracer, 0.5% 10K dextran tetramethylrhodamin (Molecular Probes) was coinjected.

Remote homology detection and molecular modelling

To search for proteins distantly related to Cripto we made use of fold recognition methods (Rost, 1995), powerful means to identify remote homologues with sequence identity around the so-called “twilight zone” (Sander and Schneider, 1991; Manco et al., 1997; Caputo et al., 2000). The methods used were: the “TOPITS” programme (Rost, 1995) (http://www.embl-heidelberg.de/predictprotein/predictprotein.html); the GOR+ Sspred programme (Fischer and Eisenberg, 1996) (http://fold doe-mbi ucla ed u) and the H3P2 Scan programme (Rice and Eisenberg, 1997) (http://fold doe-mbi ucla ed u).

Both the TOPITS and the H3P2 Scan programmes use the PHDsec method (Rost, 1997) for secondary structure predictions. Computer modelling was performed in preliminary analyses at the SWISS-MODEL web server (http://www.expasy.ch/swissmod), which employs the ProMod and Gromos 96 programmes (Guex and Peitsch, 1997; Guex et al., 1999) for comparative modelling and energy minimisation, respectively. Several 3D structures with the β-trefoil fold (see Results) were separately used for modelling and the best model was obtained with basic FGF (PDB code: 2bfh). The sequence alignment between Cripto and basic FGF, used for the modelling project was based on the output of the H3P2 Scan programme but was slightly modified in order to maximise the sequence identity, the correspondence in the type of residues (hydrophobic, basic, acid, polar, cysteines and tryptophans) and the threading energy, by exploiting the tools available under the 3.7b2 version of programme Swiss-PdbViewer (GlaxoWellcome Experimental Research, Geneva). Several models were automatically generated from slightly different alignments and the best one was subjected to further cycles of constrained energy minimisation to regularise the structure and geometrical parameters. The models were evaluated by the “What If” programme (Vriend, 1990) (http://biotech.ebi.ac.uk:8400/chk/whatif/index.html), the ERRAT programme (Colovos and Yeates, 1993) (http://www doe mbi ucla ed u), and by the tools implemented under the Swiss-PdbViewer programme. Multisequence alignment of the FGF family obtained from the Pfam database (Bateman et al., 1999) (http://pfam wustl edu/index.html) indicated the presence of a cysteine/glycine rich region in the middle of the protein. A manual alignment (not shown) of Cripto with these sequences that maximised amino acids identity and introduced only a small set of a few residue-insertion/deletions revealed that although cysteines
are not conserved among all sequences, they are frequently located at corresponding positions in Cripto. We were thus tempted to speculate on an evolutionary relationship among the EGF structure, the EGF-like module of Cripto and the middle part of the FGF molecules. To find more support for this hypothesis we exploited the Combinatorial Extension Programme (Shindyalov and Bourne, 1998) (http://cl.sdsc.edu/ce.html) to enquire about the structural similarity of EGF (PDBTM code: Iegf) and the possible Cripto distant relatives belonging to the cytokines superfamily (see Results). This programme can detect similarities between proteins not reported previously by other methods (Shindyalov and Bourne, 1998). Strikingly, the EGF structure can be superimposed on roughly the same cysteine-rich region mentioned above, with a root-mean square deviation (r.m.s.d.) lower than 5Å, in EGF as well as in other members of the superfamily (data not shown). In the light of the above findings, we attempted to model Cripto by using, as a reference, the 3D structure of the basic FGF (PDB code:2bfh). The r.m.s.d. of the backbone atoms (C, C-alfa and N atoms) for the two superimposed structures was 0.91Å on a total of 468 corresponding atoms. Finally, the paper of Lohmeyer et al. (Lohmeyer et al., 1997) demonstrated the absence of free cysteines in the refolded EGF-Cripto module. Because the cysteines potentially involved in the formation of the correct disulphide bonds pattern (Cys66-Cys73; Cys67-Cys79; Cys81-Cys90) were not too far away from each other in the model, constraints were applied in the model building procedure to include the arrangement of the correct disulphide bonds. The model was constructed with the Modeler module (Sali et al., 1995) within the software package INSIGHT II (Biosym/MSI, San Diego, CA). The obtained model was opportunely minimised using the Discover 3 module within INSIGHT II. The final result was comparable to the first model obtained with SWISS-MODEL Server in particular regarding the common refinement constraint values, indicating that if there are some errors these depend almost exclusively on the sequence alignment adopted, which could be inaccurate in some places owing to the low sequence identity between the reference and the test protein. The Ramachandran plot indicates that most residues have phi and psi angles in the core and allowed regions suggesting that the model has correct backbone conformations as found in well-refined structures. Most bond lengths, bond angles and torsion angles were in the range of values expected for a naturally folded protein.

**RESULTS**

**Recombinant Cripto protein provided in the extracellular space rescues MZoep mutants**

Expression of a potentially secreted form of zebrafish Oep can rescue the phenotype of MZoep mutants (Gritsman et al., 1999). However, in these experiments, the secretion of the protein in the extracellular space was not directly demonstrated. It was also not clear whether Oep or Cripto could act as a monomer or had to be released in the extracellular space as a homodimer or heterodimer. To directly address these issues, we produced and purified, from the culture medium of mammalian cells, two forms of soluble Cripto in which the hydrophobic C terminus was replaced by either a 6xHis epitope (Cripto-His) or the human IgG Fc domain (Cripto-Fc) (Fig. 1A,B; see Materials and Methods). Cripto-His protein behaved as a monomer when analysed by gel permeation chromatography (data not shown). Cripto-Fc fusion protein instead can assemble as a dimer because of the spontaneous dimerization of the Fc domain. As shown in Fig. 1B, protein A Sepharose immunoprecipitated, from the supernatant of Cripto-Fc transfected cells, a major protein species of the size expected for the Cripto-Fc fusion, when analysed under reducing conditions, and of the size of dimeric Cripto-Fc fusion under non-reducing conditions (for details see Materials and Methods).

Purified recombinant Cripto proteins were used to rescue the MZoep phenotype. Different doses of purified recombinant Cripto proteins were microinjected, together with 10K dextran-rhodamin as a lineage tracer (Fig. 1C), into the blastoderm of mid to late MZoep blastulae. To better define the activity of Cripto in this in vivo assay, we have arbitrarily chosen 5 grades of rescue of the MZoep mutant embryos based on RNA injection results (see Results below, Fig. 2) ranging from no rescue (grade 0) to almost wild-type phenotype (grade 4). MZoep mutant embryos display severe defects including a single eye (cyclopia), lack of the notochord, trunk somites,
pronephros and blood. Grade 1 embryos exhibited trunk somites but lacked the notochord; grade 2 embryos exhibited both trunk somites and a notochord but still had a single median eye, while grade 3 embryos showed, in addition, partial rescue of cyclopia. Grade 4 embryos and wild-type embryos were indistinguishable (Fig. 2). Both Cripto-His and Cripto-Fc (Table 1) proteins were able to rescue MZoep embryos to a wild-type phenotype. In contrast, a variant of Cripto-His carrying the mutation Glu91Ala, which attenuates both membrane-anchored (see below) and secreted Cripto (data not shown) functions, was non-functional (Table 1). Moreover, intracellular injection of recombinant Cripto protein into the yolk cell at the 1-8 cell stage did not rescue MZoep embryos (Table 1). We can, therefore, conclude that Cripto activity can be provided extracellularly as a soluble protein, demonstrating that neither membrane anchorage through the C terminus nor multimerization/binding with other partners during synthesis and secretion are absolutely required. Finally, since Cripto-His and Cripto-Fc proteins have comparable level of activity (Table 1) our data strongly suggest that Cripto homodimerization is not absolutely required. Although injected wild-type Cripto protein is sufficient to completely rescue the early defects of MZoep embryos, it appeared unable to complement the loss of oep activity at later stages, such as direction of heart looping. Using morphological criteria we have found that the heart forms in Cripto-injected MZoep mutants but the direction of heart looping is randomised with respect to the left-right axis of the embryo (data not shown), according to what has been previously described after injection of oep mRNA into MZoep (Yan et al., 1999). Similar results were obtained with cripto RNA injections (data not shown). Thus, similar to cripto RNA injections, Cripto function provided extracellularly, can rescue the early but not the late phenotype of MZoep embryos.

The EGF-CFC domain is sufficient for Cripto biological activity in MZoep rescue assay

Unexpectedly, the N-terminal sequence analysis of both purified Cripto-His and Cripto-Fc proteins showed the three sequences: SVGIQNS, VGIQNSK and NSKSLNK, with the first sequence being the major component, which differs from the expected GRDLA sequence (Dono et al., 1993)

![Fig. 2. Phenotypes of MZoep embryos rescued by mutant cripto RNA injections. Live embryos at 32 hours post-fertilization, lateral view anterior to the left. (A) MZoep mutant embryo, uninjected. (B) Grade 1: rescue of somites. (C) Grade 2: rescue of somites and notochord. (D) Grade 3: partial rescue of cyclopia. (E) Grade 4: full rescue. (F) Wild-type embryo. Note the notochord (arrowhead in C-E) and somites (arrow in B-D).](image)

**Table 1. Rescue of MZoep embryos by injection of recombinant Cripto protein**

<table>
<thead>
<tr>
<th>Site of injection (quantity)</th>
<th>Embryos scored</th>
<th>Grades*</th>
</tr>
</thead>
<tbody>
<tr>
<td>YSL (1-8 cell)</td>
<td></td>
<td>0 1 2 3 4</td>
</tr>
<tr>
<td>BSA 200 pg</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>Cripto-His WT 2 ng</td>
<td>12</td>
<td>100</td>
</tr>
<tr>
<td>Cripto-His WT 500 pg</td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>Cripto-His WT 125 pg</td>
<td>18</td>
<td>44 34 22</td>
</tr>
<tr>
<td>Cripto-His WT 15 pg</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>Glu91Ala His 500 pg</td>
<td>13</td>
<td>100</td>
</tr>
<tr>
<td>Glu91Ala His 125 pg</td>
<td>14</td>
<td>100</td>
</tr>
<tr>
<td>Cripto-Fc 4 ng</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Cripto-Fc 1 ng</td>
<td>15</td>
<td>7 93</td>
</tr>
<tr>
<td>Cripto-Fc 250 pg</td>
<td>11</td>
<td>54 38 18</td>
</tr>
<tr>
<td>Cripto-Fc 32 pg</td>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>Cripto-His WT 2 ng</td>
<td>7</td>
<td>100</td>
</tr>
</tbody>
</table>

*Phenotypic rescue is indicated as a percentage of total embryos scored.
corresponding to the predicted mature protein (Fig. 3A). This result suggested that mammalian cells produce, almost exclusively, a proteolytically processed form of Cripto lacking the N terminus of the protein (Fig. 3A).

We then went on to define the minimal Cripto region necessary to fulfil its activity. Comparison of EGF-CFC family member sequences reveals that the N- and C-terminal regions of different EGF-CFC proteins are apparently unrelated (Fig. 3A). Interestingly, even mouse and human cripto genes, whose sequence similarity rises up to 94% in the EGF-CFC region, showed a reduced sequence similarity in their N- and C-terminal regions (Dono et al., 1993). Similar results were described for mouse and human cryptic genes (Bamford et al., 2000). Moreover, the purified Cripto-His and Cripto-Fc proteins, both lacking the N terminus of the protein, are biologically active (see Results above). This observation strongly suggested that the EGF-CFC domain was sufficient to perform Cripto biological activity. To experimentally address this point we generated a cripto cDNA deletion derivative (EGF-CFC-His) lacking both the N terminus (i.e. spanning amino acids 24-52) and the hydrophobic C-terminal region required for membrane anchorage (Minchiotti et al., 2000) but exhibiting a C-terminal 6xHis epitope. The biological activities of the EGF-CFC-His and cripto-HisWT RNA were compared (Table 2) for their ability to rescue the MZoep phenotype when injected into the MZoep embryos at the 1-4 cell stage and phenotypic rescue was examined as previously described (see Results above and Fig. 2). Although high RNA doses were required when compared to Cripto-His, EGF-CFC-His was able to fully rescue MZoep embryos. These results demonstrated that the EGF-CFC domain is sufficient for Cripto activity and that, as suggested by the comparison of sequences, neither the N terminus nor the C terminus of Cripto are absolutely required for its activity.
Table 2. Rescue of MZoep embryos by injection of RNA coding for the cripto EGF-CFC region

<table>
<thead>
<tr>
<th>Injected RNAs</th>
<th>Quantity</th>
<th>Embryos scored</th>
<th>Grades*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CriptoHis</td>
<td>3 pg</td>
<td>8</td>
<td>62</td>
</tr>
<tr>
<td>CriptoWT</td>
<td>10 pg</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>EGFCFCHis</td>
<td>3 pg</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>EGFCFCHis</td>
<td>10 pg</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>EGFCFCHis</td>
<td>50 pg</td>
<td>16</td>
<td>50</td>
</tr>
<tr>
<td>EGFCFCHis</td>
<td>200 pg</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

*Phenotypic rescue is expressed as a percentage of total embryos scored.

Functional dissection of Cripto

To gain further insight into the functional importance of individual residues within Cripto, we created variants of the cripto cDNA by site-directed mutagenesis. We postulated that functionally important residues should be found within the most conserved residues of the EGF-CFC family. From the analysis of the multisequence alignment in the EGF-CFC family (Fig. 3A) 15 residues of the Cripto protein were mutagenised. Three main classes of amino acid residues were chosen: absolutely conserved (Pro52, Gly71, Phe78, Phe85, Arg88, Glu91, Arg95, His104, and Leu114), less conserved (Ser77, His92, Leu114, and Arg116) and not conserved (Asn63, and Glu97) in the EGF-CFC family. The Asn63 residue was chosen as a potential N-linked glycosylation site in Cripto (Minchiotti et al., 2000), while the Leu114 was conserved in all EGF-CFC members except Oep. Western blot analysis was performed on total lysates obtained from 293T cells transiently transfected with wild-type and mutant cDNAs (Fig. 3B) to ensure that for each mutant derivative the size was appropriate and the amount of protein was similar. Transfection efficiency was routinely monitored by cotransfecting cells with a vector expressing Jun-Ha fusion protein (Musti et al., 1997) (Fig. 3B). The expression level of Cripto mutant derivatives were comparable to wild-type protein except for Arg88Ala mutant, thus suggesting that Arg88Ala mutation could lead to a severe alteration (probably unfolding) of the protein structure that could explain protein aggregation and/or proteolytic degradation. The change of Asn63 to Ile resulted in an increased mobility of the mutant protein compared to the wild-type Cripto (Fig. 3B). The molecular mass of the mutant protein is consistent with that expected for the not-glycosylated protein (Minchiotti et al., 2000) (and data not shown) thus suggesting that the change Asn to Ile resulted in loss of glycosylation of Cripto.

The influence of different mutations (highlighted in Fig. 3A) on Cripto activity was tested by comparing wild-type and mutant RNAs for their ability to rescue the MZoep phenotype as previously described (see Results above and Fig. 2). The results of injections are presented in Table 3. Injections of wild-type cripto RNA rescued MZoep embryos in a dose-dependent manner, consistent with previous reports (Gritsman et al., 1999) (Fig. 2). Injections of the fifteen cripto variants defined three main categories of mutations. Mutations of residues Pro52, Phe85, His92, Arg95 and Glu97 did not alter Cripto activity, while residues Gly71 and Phe78 were absolutely critical for Cripto activity since injections of cripto variants Gly71Asn and Phe78Ala, even at high RNA doses, proved unable to rescue the MZoep phenotype (Table 3). Finally, mutations of residues Asn63, Ser77, Arg88, Glu91, His104, Leu114, Leu122 and Arg116 led to an intermediate phenotype.

Identification of a putative remote homologue and modelling

To elucidate whether mutated Cripto residues correspond to amino acids potentially involved in maintaining the structural integrity of the protein or to potentially exposed residues that might be involved in interaction with a partner/receptor essential for activity, a molecular model of Cripto was derived to be used as a guide model until the Cripto 3D structure is solved.

The only protein of known 3D structure closely related to human CRIPTO and present in the Protein Data Bank (PDB™) is the EGF-like cysteine/glycine-rich motif, which has been recently modelled onto the EGF structure by using the murine
Cripto variants in nodal signalling

EGF, the human transforming growth factor α (TGFα) and the EGF domain of human Factor-IX as references (Lohmeyer et al., 1997). However, although structurally related to EGF, Cripto is unable to bind any of the four known members of the ErbB receptor family (Salomon et al., 1999). One possibility is for Cripto to adopt an EGF-like fold allowing the interaction with members of a different class of receptors (Lohmeyer et al., 1997). A 3D fold recognition software was used to identify the best structural match for Cripto, in the PDBTM (see Materials and Methods). As expected, the Topits (Rost, 1995) and Gon+predss (Fischer and Eisenberg, 1996) prediction methods (available at UCLA-DOE Fold Recognition Server http://fold.doe-mbi.ucla.edu) were biased by the similarity between the EGF structures and the EGF-like motif present in Cripto (data not shown). We, therefore, tried the more relaxed algorithm H3P2 Scan (Rice and Eisenberg, 1997). In addition to a high similarity score (Z) with small-domain cysteine-rich proteins, we observed a lower but significant level of similarity between Cripto and proteins of the so-called β-trefoil fold structure. According to the Class-Architecture-Topology-Homology (CATH) classification (Orengo et al., 1997), these proteins belong to the class “mainly beta”, in agreement with the secondary structure prediction for Cripto (α=0%, β=14.2%, rest 85.8%) obtained with the PHDsec programme (Rost and Sander, 1993) which indicates the absence of α-helices.

Proteins identified as similar to Cripto in terms of secondary structure similarity and sharing the β-trefoil fold were acidic FGFs (PDB™ codes: 1barb;2afgc), interleukin-1beta (PDB™ code: 21lb) and basic FGF (PDB™ code: 2bfh). These proteins ranked in the top 20 proteins of the programme output we observed two other proteins with a β-trefoil fold: the interleukin 1 receptor antagonist protein (PDB™ code: 1lir1) and the plant cytotoxin B-chain of ricin (PDB™ code: 1abrb). Therefore, it appears that, besides a strong similarity to small domain proteins adopting an EGF-like fold, the whole Cripto protein shows distant structural similarities with proteins adopting a β-trefoil topology.

In the light of the above findings, we attempted to model Cripto by using as reference the 3D structure of the basic FGF (PDB™ code: 2bfh) (for details see Materials and Methods). The two superimposed traces for the model and the reference structure are shown in Fig. 4A.

The fold is that of a classical β-trefoil with two β-sheets made up of four and six β-strands, respectively. The model includes the three disulphide bridges (Cys66-Cys73; Cys67-Cys79; Cys81-Cys90) typical of the EGF structure (Fig. 4B; for details see Materials and Methods).

3D structural context of Cripto mutations

We indicated mutated residues on the 3D model and colour-coded them according to the capacity of the mutation to interfere with Cripto rescuing activity (Fig. 4C). Red residues (Pro52, Phe85, His92, Arg95 and Glu97) do not alter Cripto activity whereas blue residues (Gly71 and Phe78) are absolutely critical for Cripto. Finally, the remaining residues, showing an intermediate phenotype, are indicated in cyan (Asn63, Ser77, Arg88, Glu91, His104, Leu114, Leu122 and Arg116). Most mutations affecting activity are located on the same side of the molecule and are exposed (mean residue accessibility >30% except for Arg88 and Ser77 ranging 5-15%) thus suggesting that these residues might be involved in interactions with a partner/receptor essential for activity. Interestingly, the analysis of the model indicated that Asn63,
predicted to be a potential N-linked glycosylation site in Cripto (Minchiotti et al., 2000), was fully exposed and located in a β-turn (Fig. 4C), thereby being a good candidate for glycosylation. The change of Asn63 to Ile, resulting in loss of glycosylation of Cripto, confirmed this hypothesis.

**DISCUSSION**

Recombinant secreted Cripto protein is active in rescuing the MZoep phenotype

The production of a recombinant and biologically active Cripto protein allowed us to answer two main questions of the biology of Cripto and, possibly, of all the members of the EGF-CFC family, which could not be experimentally investigated by means of RNA injection: (i) does Cripto require multimerization/binding with other partners during synthesis and secretion? (ii) does Cripto require to be presented in a clustered state, which mimics membrane attachment, to efficiently activate downstream signalling events?

The ability of recombinant Cripto protein to fully rescue the MZoep mutant when injected into the blastoderm of mutated embryos demonstrates first that Cripto is active as a single molecular component and does not require to be associated with a specific intracellular molecule during synthesis and secretion, and second that Cripto can act as an extracellular factor to enable proper nodal signalling propagation. In addition, the fact that Cripto activity can be provided extracellularly is in agreement with previous data obtained with chimeric mouse embryos established from the combination of wild-type and cripto−/− ES cells; cripto−/− wild-type chimeras develop normally and both wild-type and cripto−/− cells are present in adult tissues, demonstrating that cripto acts non-autonomously during development (Xu et al., 1999). These results could either indicate a juxtacrine activity of the GPI-anchored Cripto or that the rescue can be due to a soluble form released from the membrane of wild-type cells.

Cripto-His and Cripto-Fc proteins displayed the same level of activity, strongly suggesting that Cripto dimerization/clustering was not absolutely required to efficiently activate downstream signalling events. Furthermore, the secreted form of Cripto, although lacking its C-terminal domain, is still active, thus demonstrating that this domain of Cripto is not strictly required for the biological activity tested. Similarly to Cripto and Oep, GFRα, a GPI-anchored protein, which mediates glial cell line-derived neurotrophic factor (GDNF) signalling (Jing et al., 1996), has been shown to also act as a diffusible co-factor by binding to its ligands and associating with the transmembrane receptor (Airaksinen et al., 1999; Paratcha et al., 2001). Thus Cripto surface localisation may serve to restrict activity to specific microenvironments eventually guaranteeing a high mobility in the membrane as a result of GPI anchoring, whereas release may lead to a distal effect.

The EGF-CFC domain is sufficient for Cripto biological activity in MZoep rescue assay

The EGF-CFC domain is the highest homology region between EGF-CFC proteins. Here, we provide evidence that this domain is sufficient to perform Cripto biological activity. Expression of a cripto deletion derivative (EGF-CFC-His) was still able to rescue the MZoep phenotype. Consistent with this result, the biochemical characterisation of Cripto-His and Cripto-Fc proteins purified from the conditioned medium of 293 cells, unmasked the existence of a post-translational modification of Cripto and, possibly, of all the EGF-CFC proteins. Both recombinant Cripto proteins are expressed as proteolytically processed polypeptides resulting quite exclusively in the EGF-CFC domain. Limited endoproteolysis of polypeptide precursors is a general mechanism generating a diversity of biologically active peptides and proteins in all eukaryotic phyla (Jones et al., 1996; Seidah and Chretien, 1997; Piccolo et al., 1999) and may be important to control/activate Cripto activity.

Although the EGF-CFC region is active, it appears less efficient than the wild-type Cripto isoform since higher RNA doses were required to fully rescue MZoep mutants. Such a lower efficiency suggests that the N-terminal domain of Cripto could be important to modulate its activity. The functional importance of the N-terminal domain may differ between different EGF-CFC members. Interestingly, a loss-of-function mutation (Arg78Trp) of the human cryptic gene CFC1, located in the N-terminal region of the protein (i.e., just upstream of the EGF domain; see Fig. 3A), has been recently described as associated with human left-right laterality defects (Bamford et al., 2000). This mutated form was still able to fully rescue the MZoep mutant and, moreover, had no dominant negative effect (Bamford et al., 2000). This most intriguing result supported the hypothesis that the N terminus region of the protein was not essential for the role of EGF-CFC genes in nodal signalling but could confer functional specificity to each member of the family, thus suggesting that, in addition to their role in nodal signalling, the EGF-CFC proteins might have extra functions whose specificity is mediated by the N terminus region of the protein (Bamford et al., 2000). Accordingly, mutation of Proline52 (Pro52Ala), the only absolutely conserved residue in the N-terminal region of the EGF-CFC proteins, did not alter the capacity of Cripto to rescue MZoep mutants (see Results).

Functional dissection of Cripto

From the mutational dissection of the Cripto protein it emerged that ten out of the 15 point mutations analysed affected the biological activity of Cripto when tested in the rescue assay of MZoep mutant phenotype. These residues are located both in the EGF and in the CFC domains thus suggesting that, in the current assay, both regions are required for Cripto biological activity. Quite surprisingly, we did not find a good correlation between the high degree of conservation of a residue in the EGF-CFC family and essential functions. For instance, conserved amino acids (Pro52, Phe85 and Arg95) did not strictly correspond to functional residues in the MZoep assay. Conversely, Asn63, Ser77 and Leu114, which correspond to amino acid residues less conserved in the EGF-CFC, severely affected the biological activity of Cripto in the same assay. Criteria other than conservation might be useful to identify functionally important residues and with this idea in mind, a molecular model, such as the one we propose, could be a useful tool.

Positioning of Cripto mutated residues on the 3D model revealed that most of the mutations affecting activity are located on the same side of the molecule, defining an exposed area of approximately 800-1000Å². This is a relatively large
area, which suggests that Cripto could be engaged in functional interactions with several proteins. Interestingly, Gly71 and Phe78 residues, which appeared critical and absolutely required for Cripto activity, are fully exposed in the model. Phe78 has been previously hypothesised to be involved in Cripto receptor binding because it was found to be part of a surface hydrophobic patch and because mutations of the equivalent residue in human EGF reduced binding affinity (Lohmeyer et al., 1997). Thus, the severe effect of Phe78Ala mutant on Cripto activity suggests that the small domain adopting an EGF-like fold inside the Cripto protein may share functional similarities with the EGF molecule. Moreover, according to the 3D model, Gly71, the second absolutely essential residue, is also located in this hydrophobic patch, thus suggesting that Gly71 and Phe78 may identify a functional domain potentially involved in the interaction with a receptor molecule/complex.

The positioning onto the model of residue Arg88, whose mutation both reduced the expression level of the Cripto Arg88Ala protein and severely impaired its activity, suggested that this residue is required for the structural integrity of the protein and offered an explanation even for the phenotypic effect of Arg112Cys mutation recently reported for human Cryptic (Bamford et al., 2000). Cryptic Arg112 residue, corresponding to Arg88 in Cripto, has been recently shown to be associated with human laterality defects when mutated to Cys (Bamford et al., 2000). Interestingly, Arg112Cys was shown to be completely inactive in rescuing the MZoep mutant phenotype (Bamford et al., 2000). The differences in the rescuing ability of the Cripto Arg88Ala and the Cryptic Arg112Cys could be due to an interaction of Cys112 with adjacent cysteines potentially involved in disulphur bond(s), thus leading to a more severe alteration of the protein structure (unfolding) that could eventually explain the prominent clustering of staining observed when the Cryptic Arg112Cys protein was revealed on the surface of mammalian cells (Bamford et al., 2000).

**Cripto belongs to the β-trefoil superfamily**

We searched for Cripto remote homologues to be used as a guide model until the 3D structure of the protein is solved. Since the classical protein sequence search failed, we used a novel procedure based on sequence information to predict the secondary structure of a protein. Cripto appeared to have remote homologues among proteins that adopt a classical β-trefoil fold and in particular the Fibroblast Growth Factors (FGFs). This result is most intriguing since a functional similarity of EGF-CFC proteins to FGFs was suggested by the identification of the Cripto-related factor FRL1 as a potential activator of the FGF receptor in Xenopus (Kinoshita et al., 1995). Our results, obtained with a fold recognition programme, support the recent proposal of a common ancestor for all β-trefoil proteins (Ponting and Russell, 2000). By using Cripto as a probe we obtained similarity to several members of this newly recognised super-family thus reinforcing the significance of our fold assignment. Therefore, Cripto appears to be a member of the β-trefoil proteins and may be structurally related to the FGFs. On this basis we modelled Cripto on the 3D FGF structure and the result obtained is consistent with the parameters expected from a naturally folded molecule. One prediction from this model is that the region encompassing Val51-Pro52 is located in a β-turn, between β-strands Arg44-Phe50 and Ser53-Lys60, and is thus exposed. Regions submitted to proteolytic processing should be exposed. Quite strikingly, region Val51-Pro52 is a major site of proteolysis observed on the Cripto recombinant protein released from mammalian 293 cells. Similarly, Asn63 is exposed in the 3D model and thus appears to be a good candidate as a glycosylation site. The validity of the 3D model has been confirmed here by the biochemical analysis of the functional Cripto protein.

While this study has helped to define the structure-function relationships between Cripto and nodal signalling, future studies based on these findings will provide new insight in the molecular basis of signal transduction controlled by EGF-CFC proteins. Furthermore, the generation of mouse models bearing Cripto variants could provide additional information on the role of *cripto* during development and/or adult life.

We thank Mrs. M. Terracciano for technical assistance and Miss A. Secondulfo for correcting and typing the manuscript. G. M. was awarded an EMBO short term fellowship. G. M. would like to thank Dr F. Rosa’s group at the ENS, Paris, for advice, helpful discussions and comments and Dr C. Thierry for her hospitality. This work was supported by grants from the Associazione Italiana Ricerca sul Cancro (AIRC) to M. G. P. and ARC to F. R.

**REFERENCES**


Gritsans, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W. S. and

-Cripto variants in nodal signalling 4509


