Crescent, a novel chick gene encoding a Frizzled-like Cysteine-Rich Domain, is expressed in anterior regions during early embryogenesis

PETER L. PFEFFER[1]*, EDDY M. DE ROBERTIS[2] and JUAN-CARLOS IZPISUA-BELMONTE[1]

[1]*The Salk Institute, La Jolla, CA and* [2]*Howard Hughes Medical Institute, Department of Biological Chemistry, University of California, Los Angeles, CA, USA*

ABSTRACT   We describe the isolation of a novel chicken gene that we have termed *crescent*, based on the most distinctive stage of its highly dynamic expression pattern during early embryogenesis. *Crescent* encodes a protein that in its N-terminal half shows the characteristic invariant 9 cysteine residues of the cysteine-rich domain (CRD) found in the Frizzled family of proteins, in Smoothened and in Collagen XVIII. The CRD of several Frizzled proteins have recently been shown to bind to Wg. Unlike Frizzled proteins, *crescent* does not contain a transmembrane domain and thus can not function as a receptor. *Crescent* expression is first found at stage XII (E-G&K) in the center of the area *pellucida*. On primitive streak formation, expression is detected in the entire anterior half of the area *pellucida* in the hypoblast layer. At maximal streak extension, *crescent* transcripts are localized primarily to the germinal crescent, where the primordial germ cells reside. During head process and head fold stages, *crescent* labels the anteriormost endodermal cells which will give rise to prospective foregut. With the commencement of somitogenesis, *crescent* expression rapidly wanes.

KEY WORDS: *Crescent, Frizzled, Cysteine Rich Domain, hypoblast, primordial germ cells, endoderm, early chick development*

Introduction

During a search for organizer-specific genes we fortuitously isolated a clone that is expressed specifically in an anterior arc at the late gastrula stage. This anterior arc has been termed the germinal crescent (Swift, 1914) on account of the localization of the chick's primordial germ cells to this area. In chicken, germ cells originate in the epiblast (Eyal-Giladi *et al.*, 1981), translocate into the hypoblast and migrate or are carried forward to the germinal crescent by the hypoblast layer. The hypoblast is composed of primary and secondary hypoblast cells. The primary hypoblast forms by inward movement of cells (polyingression) from the epiblast, an epithelial sheet composing the area *pellucida*. The islands of primary hypoblast cells are joined by secondary hypoblast cells emerging from the posterior margin (Koller's sickle) of the area *pellucida* to form a complete layer one-cell thick which is termed the hypoblast. Before the hypoblast sheet has reached completion, cells of the junctional endoblast emerge from the posterior margin and displace the hypoblast layer in an anterior direction. Simultaneously gastrulation commences with the formation of the primitive streak at Koller's sickle. As the primitive streak extends, cells from the anterior tip (Hensen's node) insert themselves into the center of the hypoblast sheet. These cells which will

form the endoderm displace the hypoblast to the periphery of the area *pellucida*. The hypoblast and primordial germ cells embedded in it thus land up in the anterior-lateral arc of the germinal crescent by the end of gastrulation (see Stern, 1990 and Schoenwolf, 1991). The primordial germ cells subsequently, at head fold stages, enter the forming blood vessels and migrate passively until at stages 20-24 they leave the circulatory system to populate the gonads (Ukeshima *et al.*, 1991). The hypoblast in contrast, will form extraembryonic structures. We address the issue of the distribution of a novel gene, *crescent*, in relation to the hypoblast, primordial germ cells and embryonic endoderm.

The protein encoded by *crescent* contains a cysteine-rich domain (CRD), a recently proposed protein motif that was found when comparing the α-chain of a Collagen XVIII with the *Drosophila* Frizzled protein and its vertebrate homologs, rat Fz-1 and Fz-2 (Rehn and Pihlajaniemi, 1995). Frizzled-like proteins have now been found in numerous vertebrate and invertebrate phyla and comprise a large family (Wang *et al.*, 1996). The prototype gene,

---

***Present address for reprints:** Institute of Molecular Pathology, Dr. Bohr Gasse 7, A-1030 Vienna, Austria. FAX: 1.7987153. e-mail: PFEFFER@AIMP.UNA.AC.AT

A

CLONE 5.2

```
AGGCTATAAA AGGCGTTCTG CAGGCTGGAG CGGCTCAGTC CGGGTGTGTG CTGCGAGCCC   60
         *  K   A   F   C    R   L   E    R   L   S   P    G   V   C    C   E   P

ACTCCGAGCA CCATGAGGGC TGCGAGCACG GAGAAGGCCT CCCGCAGACG GGGAGCCCCG  120
 T   P   S   T   (M)  R   A    A   S   T    E   K   A   S    R   R   R    G   A   P

TGGCTGCCAG GCCTGCTGCT GCGCCTGCTG CTGTGGGGCT GCGCGGGGGG GCCGGCCAGC  180
 W   L   P   G    L   L   L    R   L   L    L   W   G   C    A   G   G    R   A   S

TACCTGCGGA GGTCCTCCAG CTGCACCGCC ATCCCGCGCA GCATGGCCCT GTGCTATGAC  240
 Y   L   R   R    S   S   S    C   T   A    I   P   R   S    M   A   L    C   Y   D

ATCGGTTACT CGGAGATGAG GATCCCCAAC CTGCTGGAGC ACGAGACCAT GCCGGAGGTG  300
 I   G   Y   S    E   M   R    I   P   N    L   L   E   H    E   T   M    P   E   V

ATCCAGCAGT CCTCCAGCTG GCTGCCCTTG CTGGCCAGGG AGTGCCATCC CGACGCCAGG  360
 I   Q   Q   S    S   S   W    L   P   L    L   A   R   E    C   H   P    D   A   R

ATTTTCCTCT GCTCCCTCTT CGCGCCCATC TGCCTGGACA GGCTCATCTA CCCCTGCCGC  420
 I   F   L   C    S   L   F    A   P   I    C   L   D   R    L   I   Y    P   C   R

AGCCTGTGCG AGGCCGTCAA GAGGAGCTGT GCGCCCGTCA TGGCTTGTTA CGGCTATCCC  480
 S   L   C   E    A   V   K    R   S   C    A   P   V   M    A   C   Y    G   Y   P

TGGCCCGAAA TCCTCAACTG CAACAAGTTC CCTGCCGGACC ACGAGCTGTG CATCGCTGCC  540
 W   P   E   I    L   N   C    N   K   F    P   A   D   H    E   L   C    I   A   A

GTCTCCACGG ATGAAAGTTC CTCCAGCAGG AGAATGCCTC GAGCCAGCTG CAAGGACTGC  600
 V   S   T   D    E   S   S    S   R   R    M   P   R    A   S   C    K   D   C

GAACTGGAGG AGGCCAGCAC CGCCAGGGAG ATCCTCGACA ACCTCTGCGC CAACGATTTC  660
 E   L   E   E    A   S   T    A   R   E    I   L   D   N    L   C   A    N   D   F

ACAGTGAAAA TCCGAATCCT CAGGAAGAAC ACGACCACAA CCGTCTCCGA CTTCGACCTG  720
 T   V   K   I    R   I   L    R   K   N    T   T   T   T    V   S   D    F   D   L

GACCCCTCCA GGGTGGAGGT GCTGAAGCCA GGCCCGCTCC TCAGAACCGA AATCCCAGGC  780
 D   P   S   R    V   E   V    L   K   P    G   P   L   L    R   T   E    I   P   G

AGGCTGCAGC AGTGGCTGGA CATAGACGCT ACCTGCGCCC ACAACATCAT GAGAGGCACT  840
 R   L   Q   Q    W   L   D    I   D   A    T   C   A   H    N   I   M    R   G   T

CACGCGGGGG TCTTCGTCAT CAGCGGAGAA GTGCGGAGCG ACAAAGTGGT GGTGAATAAG  900
 H   A   G   V    F   V   I    S   G   E    V   R   S   D    K   V   V    V   N   K

GCCTACGCCT GGCAGAAGAA GAACAGGAAC CTGCACCAGG CCGTGCGGAG ATGGAAACAT  960
 A   Y   A   W    Q   K   K    N   R   N    L   H   Q   A    V   R   R    W   K   H

CACCGCTGCC CTGAACAGGC CGGCCAGGAAG GTCTGAGACC TCAGTCCTAA CGCTAAGAAA  1020
 H   R   C   P    E   Q   A    G   R   K    V   *
                  #
CGGCTCCCCT CCCCACGGCT AGAAAACTCC TTCCCCTGGC AGTTGGTGTT TCGGAGATGT  1080
ACGTTCAGTA TTCTTAGGTG AATCAAAGCA GTCATAGTCA GCAGGTAGGG CCTCACTACA  1140
GACCCCCGAC ATCACTAAAA CCCTTTTCCC GAGACAGAGC TTAGGGATGC AGACTTAACA  1200
TATTCCTCTG GCAAGGTCCT GCAAACCCAG GACTTACAGC GGATGGCTTT GTGCAGCACC  1260
ACTGACCTGA GACCTGCAGG CGAGCACAGA CCAGAAGAGA CCTTTAATAA CACGGGCATC  1320
TGTACAGGTG AAAGGAAACC GCCCAGCTGA GGATTTCTGC TGAGGATAAC TGGGAAGGAC  1380
TGTGCTGTAA CTGAAGGACAT TCCGTTTGGG TCAGAAAGTT CTGGTCATTT GGAAATACCG  1440
AAGATACCCC CAGAGCCTCC AGGGAGGTGG CAACGACACT GAATTCCCTG CGGATCGCTC  1500
TCACCCAAAC ACAACACAGA CTGCAGTCGC TCTCAGAGGT GAATACAGCC ACAGCTCTAC  1560
ACGAGCAGAA AGCAGAAAAC AGCAAATCTC CAAGTGAGAA AATGAGTTGG ATTTCTACTG  1620
CCTAACAAAG CACTGTCCGT CCAAAGTTTT TCTACACTGG ATTAAACTTG AGAACGGAAG  1680
AAAAATGACA GAGAAAATAA AAAAAACCCA CACAGCCTTG GGGATCAATA CGTTCTTCCC  1740
TGTGCCCAAA GCCTGCAGCC TCAGCTGCGG AGAGAACCGC CGGCTCTGAT GGAGCACACAT  1800
TGCCTCCCTG CGCTCATAAA AGAACACCGA CAGCCCAGCC AAGACCGTCT CTTCAGATGC  1860
GATGTGGGCT GAGCTCAGCA CTCACGGCAC TGGCACTGCA ATTCCGCCAC GCTCCGACAT  1920
TTGTGCAGCA ATTACTCCCT ATTTCCCTGA GGCTCAGAAG AAAAACACGG AGAGATTCCC  1980
TGTTTTTATT CCTCCCCGCT GCCTCGTATC ACAAACTGCC TCCCGACCCG CAGGGCCTGC  2040
TGCCAGGTGG TCCCCGCGAC GGCAGCATCC TCTGAACAGA CTGCGTTTTG AAAACACAAC  2100
AAGTAAAAGC TCTGATGGTT GAAAA                                         2125
```

Fig. 1. Sequence, ORF, transcript size and genomic characterization of crescent. (A) *Nucleotide sequence and ORF of the longest cDNA isolated (clone 5.2) and upstream genomic sequence to indicate the absence of further start codons between the putative translational start at position 73 (circled) and the stop codon at position 7. The presence of introns is indicated by short vertical arrows. A possible poly-A adenylation signal 429 bp from the end of the cDNA is shown in bold. The CRD is underlined. A possible amidation site is indicated by a star.*

Drosophila frizzled, is of developmental interest as it has been shown to be necessary for establishing the polarity of cuticular hairs and bristles (Gubb and Garcia-Bellido, 1982; Adler *et al.*, 1987). Mitotic clone analysis (Vinson and Adler, 1987) indicated that Frizzled has dual functions. It is required for the proximo-distal transmission of an intercellular polarity signal and it is required for cells to respond to such a signal. As *frizzled* encodes a putative transmembrane protein belonging to the superfamily of G-coupled receptors, it was proposed that the extracellular domain functions in the transmission, while the cytoplasmic domain in the interpretation of polarity information (Vinson *et al.*, 1989). The Fz-motif is located in the extracellular domain, which suggested that it may be involved in the binding of a ligand. This is indeed the case as was shown for a second *Drosophila frizzled* gene (D*fz*-2) which has been shown to bind the signaling molecule Wg via the CRD (Bahnot *et al.*, 1996). Intriguingly, yet another CRD domain was found in a developmentally critical protein, namely *Drosophila* Smoothened, a transmembrane protein that cooperates with the receptor for Hedgehog (Alcedo *et al.*, 1996). The relationship of Crescent to these other cysteine-rich domain-containing proteins will be presented.

Results

Genomic structure of crescent

During a differential screen for organizer-specific genes, we isolated clone 5.1 from a stage 4 (Hamburger and Hamilton staging, 1951; H&H) chicken Hensen's Node cDNA library. Sequence analysis revealed that the 1500 bp isolate was not full length. A RACE approach yielded 250 bp of additional 5' sequence but a potential start site was only encountered on a second exhaustive high stringency screen of the original library. The resulting clone 5.2 of 2075 bp encompasses a single open reading frame (ORF) of 921 bp (Fig. 1A). This ORF falls within the 95% confidence limit of third-base positional constraints as measured by the GCG program Testcode (Fig. 1B; Fickett, 1982). The ORF is followed by 1130 bp of untranslatable 3' trailer containing a possible polyadenylation signal 520 bp from the 3' end of the message.

Translation of this ORF from the first methionine gives a hypothetical protein of 307 amino acids. Northern blot analyses (Fig. 1C) show two transcripts of 3 and 2.4 kbp. Both transcripts are larger than the cDNA clone 5.2 of 2.1 kbp, yet this clone is likely to contain the full coding sequence, since sequencing of a genomic clone isolated from a chicken cosmid library revealed a stop codon 45 bp upstream of the 5' end of the cDNA. This assumes, of course, that no splice acceptor sites are located in these 45 bp of genomic sequence. We have determined the size and position of two introns by restriction mapping and sequencing. They are located in the ORF at positions 401 and 664 (Fig. 1A) and are 900 and 500 bp respectively in size.

Crescent is a single copy gene in the chicken as shown by high stringency hybridization of clone 5 DNA to restriction endonuclease digested genomic DNA (Fig. 1D).
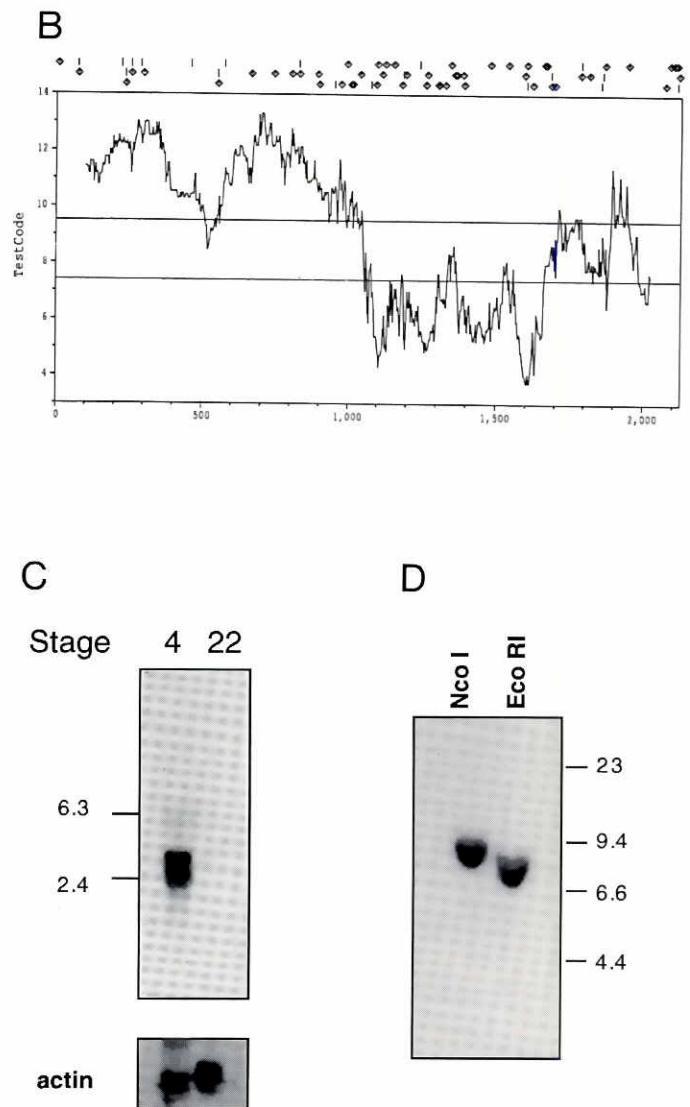
Features of crescent

Sequence databank searches revealed similarity at both the DNA and protein level between *crescent* and a collection

of genes related to the *Drosophila* segment polarity gene *frizzled (fz)* (Vinson *et al.,* 1989), namely the recently described *Drosophila fz2* (Bahnot *et al.,* 1996), rat *fz-1* and *fz-2* (Chan *et al.,* 1992), human *fz-2* and *fz-5* (Zhao *et al.,* 1995; Wang *et al.,* 1996), mouse *fz-3,4,6,7,8* and *C. elegans Cfz1* (Waterston *et al.,* 1992; Wang *et al.,* 1996).

Whereas the Frizzled-related proteins show identity over nearly their entire length, it is only the N-terminal half of Crescent that bears similarity to the most N-terminal regions of members of the Frizzled family (see Fig. 1A underlined region and Fig. 2A). Interestingly, this is exactly the region that is related to a domain found in one N-terminal variant of the mouse α1-chain of Collagen XVIII. This domain has been coined the Fz-motif (Rehn and Pihlajaniemi, 1995) or simply the CRD (cysteine-rich domain; Bhanot *et al.,* 1996). Such a CRD is also found in the amino terminal end of Smoothened (Smo), a *Drosophila* transmembrane protein that interacts with the Hedgehog receptor (Alcedo *et al.,* 1996). Crosswise sequence comparisons reveal that the Frizzled-related proteins are more closely related to each other in the CRD than to Crescent, Smoothened or Collagen XVIII (Table 1; Fig. 2A and 2B). The Frizzled-related proteins can be grouped into 3 subfamilies: (i) mammalian Fz-1,2,5,7,8 to which *Drosophila* Frizzled and Fz-2 belong – note that this subfamily can be further split into two groups – the Dfz-2, Hfz-5, Mfz-8 group (~90% identical to each other; 76% to the second group) and the mammalian Fz-1,2,7 group (>97% identity), Dfz-1 is more closely related to the second group (77% as opposed to ~70%), (ii) mouse Fz-4 together with the nematode Fz-1, and (iii) mouse Fz-3,6 containing the most divergent Fz-motif among the Frizzled-related proteins (see Table 1). Crescent is about equally closely related (~60% a.a. similarity) to the first two subfamilies and shows 10% less similarity to the Fz-3,6 subfamily. Crescent, Smoothened and Collagen XVIII fall into three separate groups. It should be noted that the spacing of the first two cysteines in the CRD is increased by two residues only in these three proteins.

An alignment of Crescent, the Frizzled proteins, Smoothened and Collagen XVIII over the cysteine-rich domain of 103 to 109 amino acids reveals the complete conservation of 9 cysteine residues (Fig. 2B, Smo has only eight of the cysteines). Also, the position of eight α-helix breaking prolines in this domain is nearly fully conserved. It is thus likely that this region adopts a distinctive secondary structure probably stabilized by intrachain cysteine disulfide bonds. A profile search (Gribskov, 1989) using the Fz-motif revealed a faint similarity to the Wnt family, which is characterized by 9 invariant cysteines in a 120 amino acid domain (Profilesearch z-score of 4,65 compare with 28,6 for Frizzled).
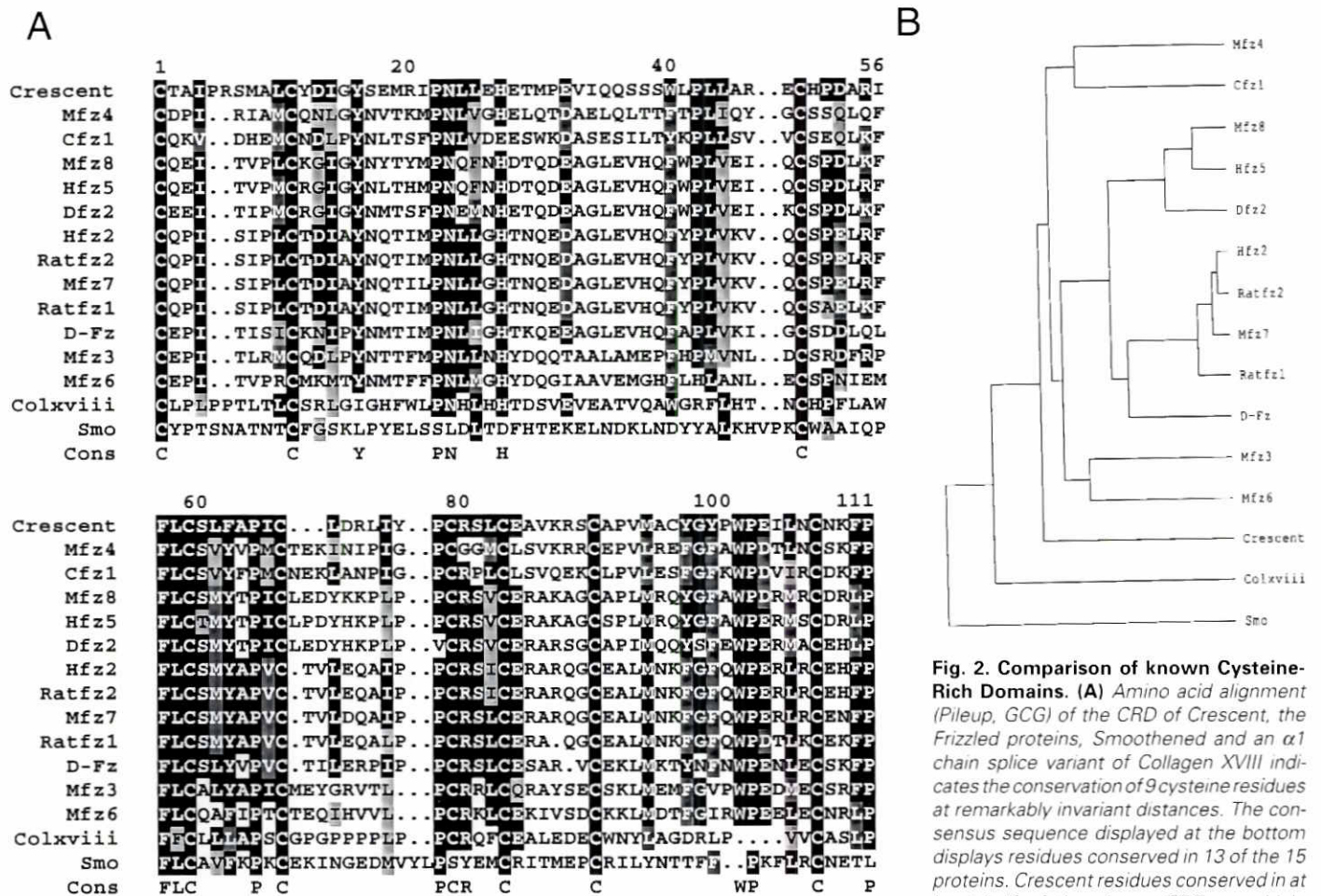
In the *Drosophila*, *C. elegans* and vertebrate Frizzled proteins as well as in Smoothened, the N-terminal CRD has been postulated to lie extracellularly and is followed by a highly conserved transmembrane domain containing seven hydrophobic stretches, each of sufficient length to be membrane-spanning (Vinson *et al.,*

1989; Chan *et al.,* 1992). In contrast, *crescent* does not encode a potentially membrane-spanning domain. Instead, Crescent terminates with a highly basic region and bears at its C-terminus a potential amidation site at the ultimate alanine.

The *crescent* product is likely to be located extracellularly, even though it lacks a conventional signal sequence. The highly charged

Fig. 1. Sequence, ORF, transcript size and genomic characterization of *crescent*. (B) *ORF determination by the GCG program Testcode. Values above the first horizontal line fall within the 95% confidence limit of third-base positional constraints thus representing potential coding regions, values below the lower line were non-coding and values between the lines indeterminate. Above the graph start and stop codons are indicated by vertical lines and diamonds respectively, for each reading frame.* (C) *Northern blot of stage 4 and 22 total RNA hybridized to a full length crescent probe and reprobed with actin (below). Expression is only seen at stage 4 with two main transcripts of 3 and 2.4 kbp.* (D) *Southern blot of chick genomic DNA digested with NC and EcoRI respectively, and probed with a full length crescent probe at high stringency. Only one band is seen in both cases, indicating that crescent is a single copy gene.*

Fig. 2. Comparison of known Cysteine-Rich Domains. (A) *Amino acid alignment (Pileup, GCG) of the CRD of Crescent, the Frizzled proteins, Smoothened and an α1 chain splice variant of Collagen XVIII indicates the conservation of 9 cysteine residues at remarkably invariant distances. The consensus sequence displayed at the bottom displays residues conserved in 13 of the 15 proteins. Crescent residues conserved in at least half of the shown CRDs are high-lighted in black (identical), dark grey (highly conservation) or light grey (weak conservation). The amino acid similarity between members is depicted graphically in (B) (using the GCG program Pileup) and numerically in Table 1. References for the respective sequences are listed in the text (section Features of crescent).*

N-terminal 13 amino acids of Crescent followed by a highly hydrophobic stretch of 19 amino acids (Fig. 1A) may be an example of a signal-anchor sequence of type II membrane bound proteins. Such proteins are anchored to the membrane via their N-terminus.

Expression pattern during embryogenesis

Early expression of crescent in anterior hypoblast

The early pattern of *crescent* expression parallels the distribution and migration of the anterior part of the hypoblast as well as that of primordial germ cells. The earliest component of the hypoblast (the primary hypoblast) consists of islands of cells which are probably derived by ingression from the epiblast at stages X to XII (Eyal-Giladi and Kochav, 1976). *Crescent* expression is first detectable in 9 hour embryos (E-G&K stage XII) as a faint patch toward the center of the area *pellucida* (Fig. 3A). At this stage a sheet of cells extending from the posterior limit of the area *pellucida* to about halfway across the embryo can be seen. This sheet is the forming hypoblast layer arising by the anterior migration of secondary hypoblast cells originating at the posterior margin of the area *pellucida*.

During gastrulation, marked by the appearance and extension of the primitive streak (H&H stages 2+ to 4) the hypoblast layer is displaced in an anterior direction by the junctional endoblast layer originating at the posterior margin (Vakaet, 1970). Furthermore, first prospective endodermal and then (from stage 3c onward) prospective mesodermal cells ingress at the anterior end of the continuously lengthening primitive streak and insert into the hypoblast layer. The net result of these events is that the hypoblast cells are squeezed into the anterior periphery of the area *pellucida*. This retraction of the hypoblast layer is reflected by *crescent* expression. At stage 3c when the primitive streak has extended to the middle of the area *pellucida*, *crescent* staining is readily visible in the anterior half of this area (Fig. 3B). *In situ* hybridization of transverse sections through this anterior half of the embryo clearly shows that it is the underlying cell layer, the hypoblast, that is labeled whereas the epiblast shows no expression (Fig. 3E and 3F). Slightly later, at stage 3d, when the primitive streak has nearly reached its maximal extension, *crescent* expression is seen predominantly in a crescent shaped region around the anterior limit of the area *pellucida* (Fig. 3C and 3D). No staining is seen using a sense probe (not shown).

TABLE 1

AMINO ACID SIMILARITY OF THE CRD OF 15 PROTEINS CALCULATED BY THE GCG PROGRAM DISTANCES

Crescent	Mfz4	Cfz1	Mfz8	Hfz5	Dfz2	Hfz2	Rfz2	Mfz7	Rfz1	D-Fz	Mfz3	Mfz6	Col	Smo			
1.00	0.61	0.57	0.63	0.61	0.59	0.62	0.62	0.62	0.61	0.55	0.50	0.50	0.47	0.35	Crescent		
	1.00	0.65	0.65	0.64	0.62	0.63	0.63	0.64	0.65	0.71	0.59	0.57	0.41	0.32	Mfz4	1	
		1.00	0.60	0.58	0.62	0.60	0.60	0.61	0.61	0.61	0.55	0.54	0.40	0.39	Cfz1		
			1.00	0.93	0.89	0.77	0.77	0.77	0.76	0.71	0.66	0.62	0.46	0.35	Mfz8		F
				1.00	0.88	0.75	0.75	0.75	0.74	0.71	0.64	0.62	0.46	0.34	Hfz5		R
					1.00	0.75	0.75	0.75	0.72	0.68	0.63	0.63	0.43	0.34	Dfz2		I
						1.00	1.00	0.99	0.97	0.77	0.64	0.62	0.48	0.38	Hfz2	2	Z
							1.00	0.99	0.97	0.77	0.64	0.62	0.48	0.38	Rfz2		Z
								1.00	0.97	0.77	0.64	0.63	0.48	0.39	Mfz7		L
									1.00	0.77	0.66	0.62	0.47	0.40	Rfz1		E
										1.00	0.67	0.64	0.44	0.39	D-Fz		D
											1.00	0.67	0.44	0.32	Mfz3	3	
												1.00	0.42	0.37	Mfz6		
													1.00	0.31	ColXVIII		
														1.00	Smo		

The hypoblast is destined to form extraembryonic endoblast in the yolk sac stalk but also contains the primordial germ cells. These germ cells are believed to originate from the central zone of the area *pellucida* in the epiblast (Eyal-Giladi *et al.*, 1981; Ginsburg and Eyal-Giladi, 1986, 1987). They drop into the advancing secondary hypoblast layer and are carried or actively migrate in an anterior direction until they reach the germinal crescent. The question thus arises as to whether *crescent* is a marker of primordial germ cells. Before stage 4 this is difficult to answer as no marker is known for the germ cells. However, from stage 4 onwards, primordial germ cells can be readily distinguished by alkaline phosphatase staining (described for mouse cells by Buehr and McLaren, 1993; for chick cells by Pain *et al.*, 1996). Hence whole-mount double labeling experiments were performed, initially staining for alkaline phosphatase activity using the Fast Red-α-naphthol reaction which yields an orange-reddish product, followed by *in situ* hybridization with a *crescent* RNA probe which results in a purple stain. The results of Figure 4 vividly demonstrate that *crescent* indeed labels the regions showing endogenous alkaline phosphatase activity at stage 4 (compare the germinal crescent of panels A and B) and stage 6 (panels C and D). However, whereas the primordial germ cells remain in the germinal crescent and are reported to do so until stage 8 (4 somite), when they actively migrate into the developing blood vessels (Fujimoto, 1975), *crescent* expression starts to be seen in the head process and at stage 6 in the forming head fold whilst waning in the germinal crescent (see below). Furthermore, at these stages there are only about 150-250 primordial germ cells per embryo (Fujimoto *et al.*, 1975). Yet at stage 3c (Fig. 3F) it is the entire hypoblast and not only part of it that expresses *crescent*. This suggests that the concordant pattern of primordial germ cells may only reflect the fact that these are trapped by the advancing hypoblast. What is certain, however, is that *crescent* is an early marker for the anterior hypoblast.

*Crescent is expressed in the anteriormost endoderm*

The second wave of *crescent* expression begins at the end of gastrulation (H&H stage 4). The primitive streak has reached its maximal extension and now begins its regression, leaving in its wake the notochord. Simultaneously the head process appears anterior to the maximal primitive streak extension. *Crescent* expression is seen accumulating anterior to Hensen's node in the head process primordium (Figs. 4B and 5A) and at stage 5 not only marks the head process but also a fan-shaped area surrounding it (Fig. 5B). This includes the region of the presumptive prechordal plate. Hensen's node itself is devoid of *crescent* expression. As the head folds develop (H&H stage 6) expression is seen to be contained within the headfold pocket. The expression domain narrows down to the midline caudally, thus creating a triangular shape (Fig. 5C). Expression stays strong until the two somite stage (stage 7+, Fig. 5D) and rapidly wanes thereafter. Cross sections through the head fold region of a stage 6 embryo indicates that only the ventral-most cell layer is stained (Fig. 5E). This layer represents the endoderm which will give rise to the foregut.

Discussion

Crescent as part of a family sharing the CRD-motif

We report on the cloning of a novel gene expressed from blastula to early neurula stages of the chick embryo. The predicted polypeptide encoded by *crescent* shares identity with the large family of Frizzled-related proteins with members reported in the phyla Vertebrata, Arthropoda, Nematoda and Echinodermata (for the latter only a partial PCR sequence of the transmembrane domain has been identified, Wang *et al.*, 1996). The area of similarity between Crescent and the Frizzled-family entails a 103-109 amino acid domain, characterized by the nearly invariant spacing of 9 cysteines. This cysteine-rich domain is also found in the Hedgehog receptor Smoothened and in an N-terminal variant of the α1-chain of type XVIII collagen. The presence of a CRD in a purely structural protein such as Collagen is puzzling. The Collagens, characterized by the presence of at least one triple helix sequence of a repeated Gly-X-Y motif, do though often also contain non-collagenous domains, several of which contain portions with
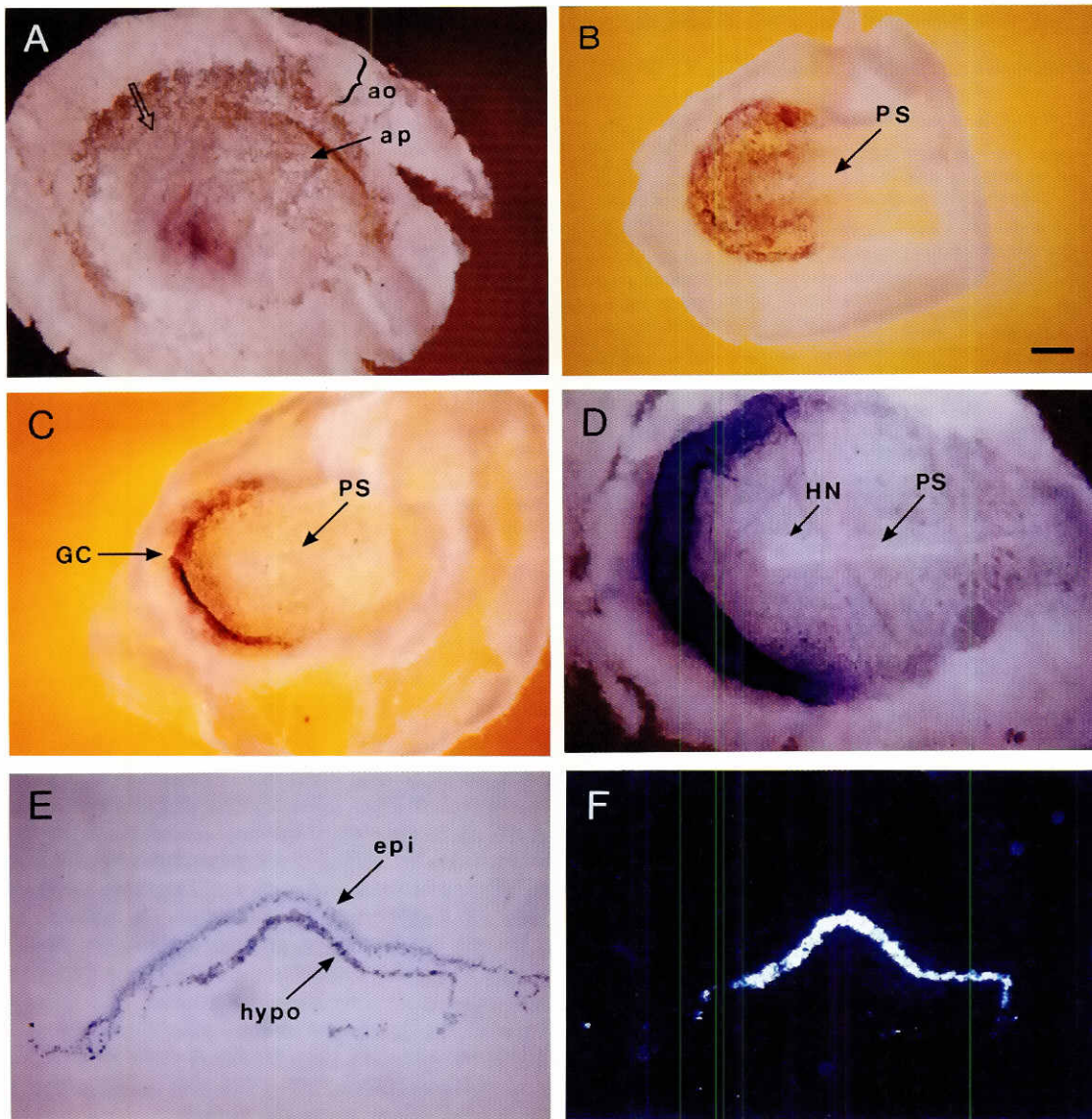
Fig. 3. Early *crescent* expression from the center of the area *pellucida* to the germinal crescent. (A) Whole-mount in situ hybridization of a 9 hour embryo (E-G&K stage XII) showing crescent expression in the center of the area pellucida (ap). Anterior is to the top and the edge of the advancing secondary hypoblast is indicated by an open arrow; ao, area opaqua; in the subsequent illustrations, anterior is always to the left. (B) Primitive streak (PS) stage (stage 3c) embryo showing expression in the anterior half of the area pellucida. (C and D) At stages 3d to 4 crescent expression has coalesced into an anterior arc, termed the germinal crescent; HN, Hensen's node. (E and F) Bright field (E) and dark field (F) images of a transverse section at the anterior-third level of a stage 3d embryo subjected to in situ by section hybridization. Only the lower hypoblast layer (hypo) is labeled. Epi, epiblast. Bar: A, 400 µm; B,C, 500 µm; D, 300 µm; E,F, 170 µm.

similarity to sequences found in proteins such as fibronectin, von Willebrand factor, thrombospondin (Bork, 1992) and chordin (Sasai et al., 1994).

The *Drosophila frizzled* and *fz-2* and vertebrate *fz-1* to *-8* genes code for membrane spanning proteins with an extracellular N-terminal region which contains the Fz-motif and a C-terminal two-thirds portion containing seven putative membrane-spanning domains. The C-terminal part resembles the superfamily of G-protein-coupled receptors. Mutations in *frizzled* affect the orientation of hairs and bristles in the *Drosophila* cuticle, suggesting that the Frizzled protein is required for the establishment of cell polarity in the epidermis (Adler et al., 1987). Genetic mosaic studies suggested that Frizzled has two mutably separate functions in establishing hair polarity. It is required for the intercellular transmission of a polarity signal, a process that is expected to involve the extracellular domain. The CRD encompasses most of the extracellular domain and could thus either function as a tethered ligand necessary to relay the polarity signal to neighboring cells or be

involved in the binding of a ligand which activates a signaling pathway leading to intercellular transmission of a signal. Secondly, Frizzled is also required for cells to respond to a polarity signal, probably via the cytoplasmic domain (Vinson et al., 1989). Recently, it was shown that the *dishevelled* gene lies downstream of *frizzled* in the intracellular response pathway (Krasnow et al., 1995).

The discovery that the second *Drosophila* Frizzled homolog, Fz-2, is a receptor for the signaling molecule Wg and that the CRD is sufficient to mediate this interaction, has verified the role of this domain as a ligand-binding moiety (Bahnot et al., 1996). The generality of this interaction is underscored by the observation that several of the vertebrate Fz receptors (namely Fz-4, 5 7 and 8) as well as Dfz can also interact with Wg (Bahnot et al., 1996). Interestingly, mouse Fz-3 and -6 do not bind Wg. These mouse Fz proteins fall into the third, most divergent, subfamily of Frizzled proteins as judged by a comparison of solely the CRDs (Table 1, Fig. 2). When one considers that there are around 20 different Wg
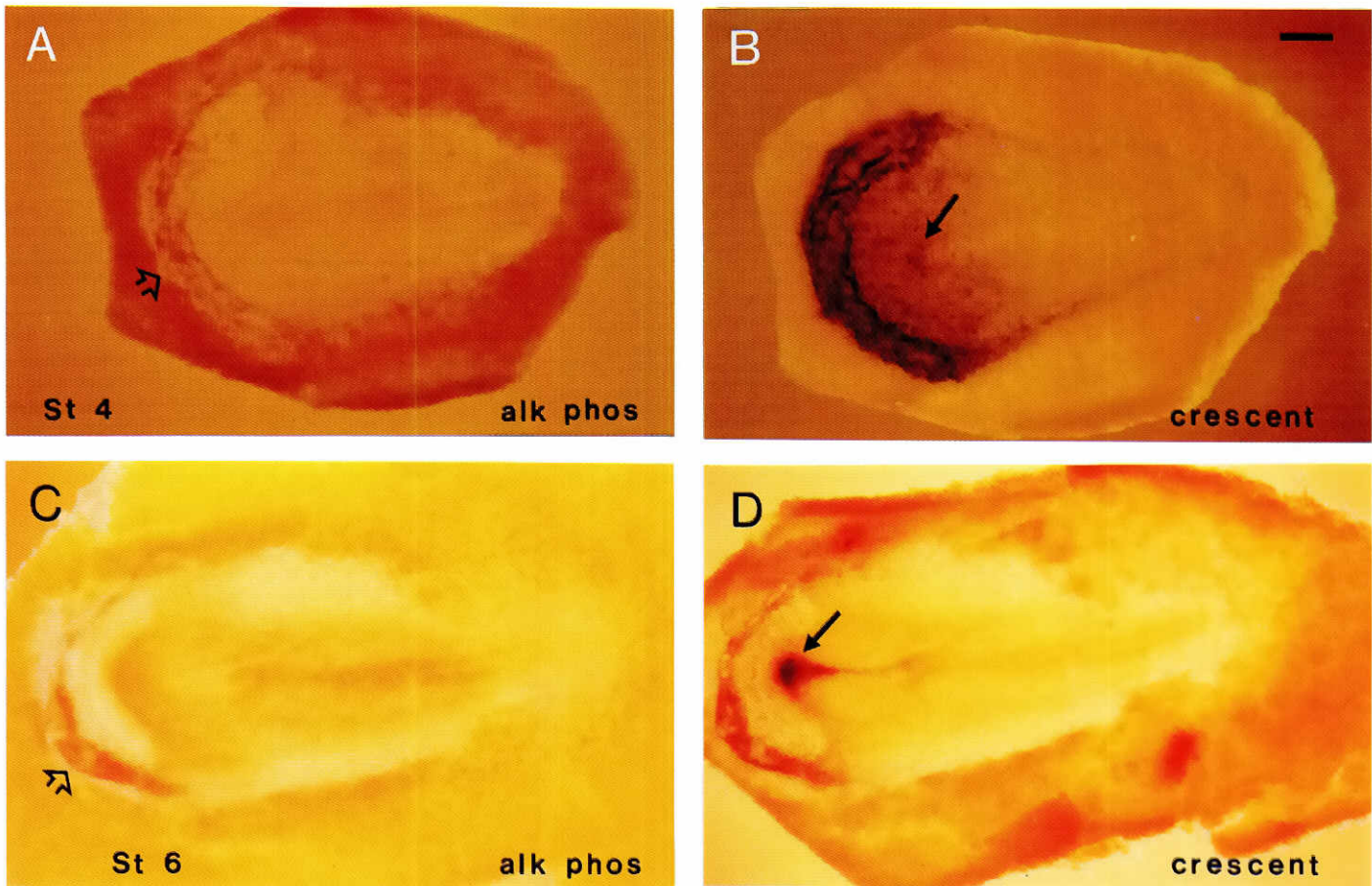
Fig. 4. *Crescent* is expressed in alkaline phosphatase positive cells. *Endogenous alkaline phosphatase staining* **(A,C)** *and subsequent whole-mount in situ hybridization to a full length crescent probe* **(B,D)** *of stage 4* **(A,B)** *and stage 6* **(C,D)** *chick embryos. Anterior is to the left. The open arrows indicate the germinal crescent population of alkaline positive cells which are also labeled by the crescent probe. The black arrow shows the emerging late phase of crescent expression just anterior to Hensen's node at stage 4 and in the head fold at stage 6. Bar: A,B, 430 μm; C,D, 500 μm.*

homologs in vertebrates (the Wnt proteins; Nusse and Varmus, 1992; Sidow, 1992) and at least 8 vertebrate Frizzled proteins, it is clear that the exact structure of the CRD will determine the array of ligands to which the particular receptor can bind. This provides a rationale for classifying these proteins solely on the basis of their CRDs.

The CRD of Smoothened is most highly diverged showing only some 35% identity to that of Crescent, the Frizzled family or to the Collagen XVIII variant. This is close to random were it not for the conserved Cysteines. This makes the conservation of the characteristic Cysteines intriguing and suggests an evolutionary very ancient structure where the location of the cysteines provides a stable molecular scaffold, probably via intramolecular cross-bridges, upon which a ligand-recognizing-domain or indeed a ligand-domain itself, can be modeled. The latter possibility is suggested by the observation that the Wnt family also shares a characteristic – though different – invariant spacing of 9 cysteines in a region of about 120 amino acids. It is possible that either the extracellular environment of these domains or a high binding specificity necessitates such a stable configuration.

It is likely that the *crescent* product is also located extracellularly although it does not have a *bona fide* signal sequence but rather a membrane anchor type sequence found in type II membrane proteins. This suggests that Crescent may be tethered to the cell surface, allowing it to interact with signaling molecules or act itself as a tethered ligand. The high divergence of the CRD would suggest that Crescent interacts with ligands/receptors other than those of Frizzled and Smoothened and may thus define a novel signaling pathway in the molecularly poorly understood genesis of the hypoblast and endomesoderm. In *Xenopus*, a secreted protein, *frzb*, containing a frizzled-like CRD has been isolated recently (L. Leyns, T. Bouwmeester, K. Sung, S. Piccolo and E.M.D.R., submitted).

Crescent expression marks the anterior hypoblast and endoderm

Crescent expression during embryonic development was analyzed in stage 2 to 25 H&H (Hamburger and Hamilton, 1951; Schoenwolf, 1988) chick embryos. Expression is dynamic and can be described as being composed of three phases:
i) An early phase involving the anterior half of the gastrula (stage 3 to 4 H&H) hypoblast.
ii) A phase of intense expression in the germinal crescent during late gastrula to headfold stages (stage 3d to 6 H&H).
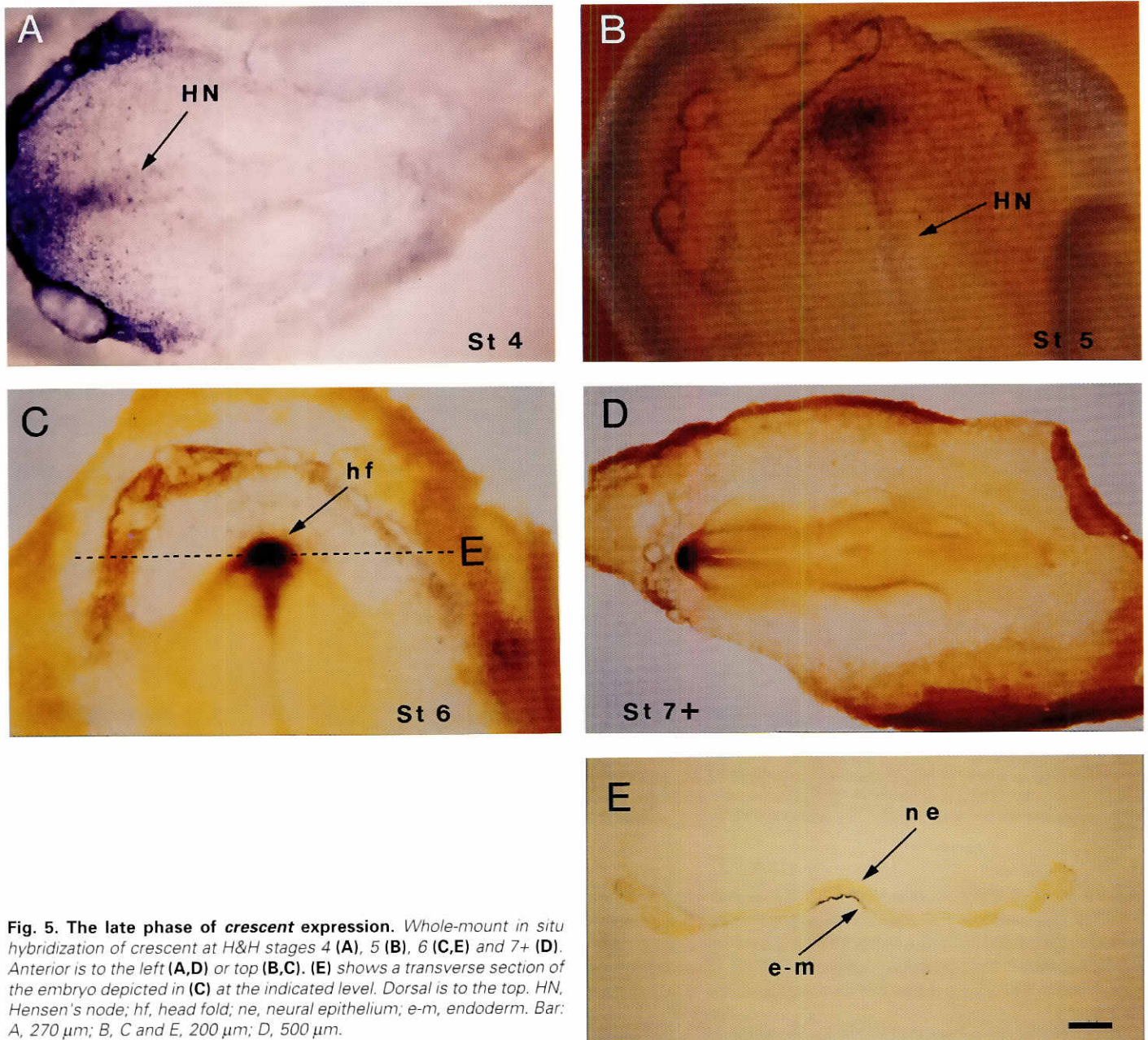
Fig. 5. The late phase of *crescent* expression. *Whole-mount in situ hybridization of crescent at H&H stages 4 (**A**), 5 (**B**), 6 (**C,E**) and 7+ (**D**). Anterior is to the left (**A,D**) or top (**B,C**). (**E**) shows a transverse section of the embryo depicted in (**C**) at the indicated level. Dorsal is to the top. HN, Hensen's node; hf, head fold; ne, neural epithelium; e-m, endoderm. Bar: A, 270 μm; B, C and E, 200 μm; D, 500 μm.*

iii) A second wave of expression in the endodermal layer of the head process (stage 5 to 7).

No expression could be detected at later stages by Northern blot analysis, whole-mount *in situ* hybridization or *in situ* hybridization by section.

The first two phases of expression resemble the pattern of migration of both the hypoblast and the primordial germ cells embedded in it. We have demonstrated the concordant pattern of germ cells and *crescent* RNA from stage 4 to 6 H&H, using alkaline phosphatase staining and *in situ* hybridization. Nonetheless, lacking single cell resolution, this does not prove that germ cells do express *crescent*. Indeed, *in situ* by section at stage 3c clearly shows the entire hypoblast to express *crescent*, whereas at this stage, at which germ cells cannot be detected by alkaline phos-

phatase staining, there exist only isolated pockets of primordial germ cells totalling less than 250 (Fujimoto *et al.*, 1975). Further, germ cells remain in the germinal crescent up to stage 8 H&H, when they actively migrate into the developing blood vessels (Fujimoto *et al.*, 1975; Ukeshima *et al.*, 1991), whereas *crescent* transcripts are no longer seen in this area after stage 6. This indicates that if germ cells express *crescent*, this is only transient. We favor the idea that *crescent* is an early marker of hypoblast cells, reflecting the progressive restriction of this cell layer into the anterior periphery due to the morphogenetic movements establishing the definitive germ layers.

In the third phase of expression commencing at stage 4 H&H (maximal primitive streak extension), *crescent* is expressed again in (1) anterior and (2) in the inner cell layers. *Crescent* transcripts

are detected in the head process primordium but not in Hensen's node. The fan-shaped area exhibiting *crescent* expression is constricted to the midline posteriorly and spreads laterally at the cranial end. Cross sections showed that only the endodermal layer is labeled. Thus by comparisons to fate maps and *goosecoid* expression (Izpisua-Belmonte *et al.,* 1993), *crescent* marks the prospective foregut.

Crescent expression is always seen anterior to the primitive streak. This expression resembles that of the recently described frog *cerberus* gene (Bouwmeester *et al.,* 1996). This gene codes for a secreted protein that is expressed in the anterior endoderm and is involved in the induction of head structures. However, *cerberus* in the frog is not expressed in the prechordal plate as revealed by comparison with frog *goosecoid,* a marker for this region. Comparison with chick *goosecoid* (see Fig. 2 of Izpisua-Belmonte *et al.,* 1993) reveals that *crescent* expression labels both the region underlying the presumptive prechordal plate and more anterior regions. Thus similar to frog *cerberus, crescent* is an early marker for the anterior endoderm but unlike *cerberus* and akin to *goosecoid,* it also marks the region of the prechordal plate, though (unlike *goosecoid*) only the deepest endodermal layer.

Materials and Methods

Cloning of crescent

A stage 4 (H&H) chicken Hensen's node library (kindly supplied by Claudio Stern) was screened under low stringency conditions (40% formamide; 5xSSC; 20 mM tris-HCl pH 7.6; 2x Denhardt's; 0.1% SDS at 37°C) with a probe derived by subtracting stage 21 tailbud RNA with trunk RNA in an effort to obtain genes expressed in the organizing centers of both Hensen's node and the tailbud. Strong positives of this directionally-cloned λZAP library were purified, excised to generate pBluescript clones, and transcribed with T3 RNA polymerase to generate digoxygenin-labeled RNA probes. These probes were used for whole-mount *in situ* hybridization of stage 4 and 21 embryos. Clone 5 surprisingly showed no expression at stage 21 but interestingly stained the germinal crescent at stage 4 and was thus analyzed further. To obtain the 5' end of the transcript, a) RACE was performed using a Clontech kit, b) the Hensen's node library was rescreened at high stringency and c) a chick cosmid library (Stratagene) was screened to obtain a cosmid containing the entire clone.

Sequencing was performed using the Sequenase kit (USB). Sequence data analysis was done using the GCG programs and the BLAST network.

RNA isolation and hybridization

Embryos were dissected at the relevant stages and immediately frozen in liquid nitrogen. RNA was isolated from the frozen samples by adding 10 volumes of STAT 60, a phenol/GITC based reagent (Tel-Test 'B', Inc.) and homogenizing using a polytron. Further steps were performed as described in the STAT 60 protocol. Northern blots were done using formaldehyde gels as described in Sambrook *et al.* (1989). The probe used was a EcoRI fragment covering the 5' two-thirds of clone 5.1.

In situ hybridization

Whole-mount *in situ* hybridization was carried out as described by Wilkinson (1993) but for staining BM purple (Boehringer Mannheim) was used instead of the classical BCIP/NBT. Embryos were fixed at 4°C in 4% paraformaldehyde/PBS for 30 min (stage 2 to 7) to 2 h (stage 25) and stained for 30 min to overnight depending on the intensity of the signal. The entire clone 5.1 was transcribed using digoxygenin-labeled rUTP and T7 RNA polymerase for sense and T3 RNA polymerase for antisense probes.

In situ by section hybridization was performed as described previously (Izpisua-Belmonte *et al.,* 1991). The probe was hydrolyzed to an average length of 300 nucleotides prior to use.

Alkaline phosphatase staining

Embryos were collected at appropriate stages and dehydrated in an ethanol series and left in absolute ethanol overnight. They were then rehydrated in stages and washed in three changes of distilled water of 10 min each. Embryos were transferred to freshly prepared stain as described by Buehr and McLaren, 1993 (25 mg Fast Red TR diazonium salt, 5 mg α-Naphthyl phosphate, 44.7 ml water, 150 μl 1 M MgCl$_2$ and 5 ml 4.5% Borax mixed in this order) and stained for 15 to 20 min. The reaction was stopped by three 10 minute washes in distilled water. Embryos were photographed and then subjected to whole-mount *in situ* hybridization for double labeling experiments.

References

ADLER, P.N., CHARLTON, J. and VINSON, C. (1987). Allelic variation at the frizzled locus of *Drosophila. Dev. Genet. 8:* 99-119.

ALCEDO, J., AYZENZON, M., VON OHLEN, T., NOLL, M. and HOOPER, J.E. (1996). The *Drosophila* smoothened gene encodes a seven-pass membrane protein, a putative receptor for the hedgehog signal. *Cell 86:* 221-232.

BAHNOT, P., BRINK, M., SAMOS, C., HSIEH, J-C., WANG, Y., MACKE, J.P., ANDREW, D., NATHANS, J. and NUSSE, R. (1996). A new member of the frizzled family from *Drosophila* functions as a wingless receptor. *Nature 382:* 225-230.

BORK, P. (1992). The modular architecture of vertebrate collagens. *FEBS Lett. 307:* 49-54.

BOUWMEESTER, T., SUNG-HYUN, K., SASAI, Y., LU, B. and DE ROBERTIS, E. M. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature 382:* 595-601.

BUEHR, M. and McLAREN, A. (1993). Isolation and culture of primordial germ cells. *Methods Enzymol. 225:* 58-77.

CHAN, S.D.H., KARPF, D.B., FOWLKES, M.E., HOOKS, M., BRADLEY, M.S., VUONG, V., BAMBINO, T., LIU, M. Y., ARNAUD, C.D., STREWLER, G.J. and NISSENSON, R.A. (1992). Two homologs of the *Drosophila* polarity gene *frizzled* are widely expressed in mammalian tissues. *J. Biol. Chem. 267:* 25202-25207.

EYAL-GILADI, H. and KOCHAV, S. (1976). From cleavage to primitive streak formation: a complementary normal table and a new look at the first stages of the development of the chick. *Dev. Biol. 49:* 321-337.

EYAL-GILADI, H., GINSBURG, M. and FARBAROV, A. (1981). Avian primordial germ cells are of epiblastic origin. *J. Embryol. Exp. Morphol. 65:* 139-147.

FICKETT, J. (1982). Recognition of protein coding regions in DNA sequences. *Nucleic Acids Res. 10:* 5303-5318.

FUJIMOTO, T., UKESHIMA, A. and KIYOFUJI, R. (1975). The origin, migration and morphology of the primordial germ cells in the chick embryo. *Anat. Rec. 185:* 139-154.

GINSBURG, M. and EYAL-GILADI, H. (1986). Temporal and spatial aspects of the gradual migration of primordial germ cells from the epiblast into the germinal crescent in the avian embryo. *J. Embryol. Exp. Morphol. 95:* 53-71.

GINSBURG, M. and EYAL-GILADI, H. (1987). Primordial germ cells of the young chick blastoderm originate from the central zone of the area pellucida irrespective of the embryo-forming process. *Development 101:* 209-219.

GRIBSKOV, M. (1989). Profile analysis. *Methods Enzymol. 183:* 146-159.

GUBB, D. and GARCIA-BELLIDO, A. (1982). A genetic analysis of the determination of the cuticular polarity during development in *Drosophila* melanogaster. *J. Embryol. Exp. Morphol. 68:* 37-57.

HAMBURGER, V. and HAMILTON, H. (1951). A series of normal stages in the development of the chick. *J. Morphol. 88:* 49-92; see also reprint in *Dev. Dynamics 195:* 231-272, 1992.

IZPISUA-BELMONTE, J-C., DE ROBERTIS, E.M., STOREY, K. and STERN, C.D. (1993). The homeobox gene goosecoid and the origin of organizer cells in the early chick blastoderm. *Cell 74:* 645-659.

IZPISUA-BELMONTE, J-C., TICKLE, C., DOLLE, P., WOLPERT, L. and DUBOULE, D. (1991). Expression of the homeobox Hox-4 genes and the specification of position in chick wing development. *Nature 350:* 381-385.

KRASNOW, R.E., WONG, L.L. and ADLER, P.N. (1995). Dishevelled is a component of the frizzled signaling pathway in *Drosophila. Development 121:* 4095-4102.

NUSSE, R. and VARMUS, H.E. (1992). Wnt genes. *Cell 69:* 1073-1087.

PAIN, B., CLARK, M.E., SHEN, M., NAKAZAWA, H., SAKURAI, M., SAMARUT, J. and ETCHES, R.J. (1996). Long-term *in vitro* culture and characterization of avian embryonic stem cells with multiple morphogenetic potentialities. *Development 122:* 2339-2348.

REHN, M. and PIHLAJANIEMI, T. (1995). Identification of three N-terminal ends of type XVIII collagen chains and tissue-specific differences in the expression of the corresponding transcripts. *J. Biol. Chem. 270:* 4705-4711.

SAMBROOK, J., FRITSCH, E.F. and MANIATIS, T. (1989). *Molecular Cloning.* Cold Spring Harbor Laboratory, Cold Spring Harbor (NY).

SASAI, Y., LU, B., STEINBEISSER, H., GEISSERT, D., GONT, L.K. and DE ROBERTIS. E.M. (1994). *Xenopus* chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell 79:* 779-790.

SCHOENWOLF, G.C. (1988). Microsurgical analyses of avian neurulation: Separation of medial and lateral tissues. *J. Comp. Neurol. 276:* 498-507.

SCHOENWOLF, G.C. (1991). Cell movements in the epiblast during gastrulation and neurulation in avian embryos. In *Gastrulation* (Ed. R. Keller). Plenum Press, New York, pp. 1-27.

SIDOW, A. (1992). Diversification of the Wnt gene family on the ancestral lineage of vertebrates. *Proc. Natl. Acad. Sci. USA 89:* 5098-5102.

STERN, C. (1990). The marginal zone and its contribution to the hypoblast and primitive streak of the chick embryo. *Development 109:* 667-682.

SWIFT, C.H. (1914). Origin and early history of the primordial germ cells in the chick. *Am. J. Anat. 15:* 483-516.

UKESHIMA, A., YOSHINAGA, K. and FUJIMOTO, T.W. (1991). Scanning and transmission electron microscopic observation of chick primordial germ cells with special reference to the extravasation in their migration course. *J. Electron. Microsc. 40:* 124-128.

VAKAET, L. (1970). Cinephotomicrographic investigations of gastrulation in the chick blastoderm. *Arch. Biol. 81:* 387-426.

VINSON, C.R. and ADLER, P.N. (1987). Directional non-cell autonomy and the transmission of polarity information by the frizzled gene of *Drosophila. Nature 329:* 549-551.

VINSON, C.R., CONOVER, S. and ADLER, P.L. (1989). A *Drosophila* tissue polarity locus encodes a protein containing seven potential transmembrane domains. *Nature 338:* 263-264.

WANG, Y., MACKE, J.P., ABELLA, B.S., ANDREASSON, K., WORLEY, P., GILBERT, D.J., COPELAND, N.G., JENKINS, N.A. and NATHANS, J. (1996). A large family of putative transmembrane receptors homologous to the product of the *Drosophila* tissue polarity gene frizzled. *J. Biol. Chem. 271:* 4468-4476.

WATERSTON, R., MARTIN, C., CRAXTON, M., HUYNH, C., COULSON, A., HILLIER, L., DURBIN, R., GREEN, P., SHOWNKEEN, R., HALLORAN, N., METZSTEIN, M., HAWKINS, T., WILSON, R., BERKS, M., DU, Z., THOMAS, K., THIERRY-MIEG, J. and SULSTON, J. (1992). A survey of expressed genes in *Caenorhabditis elegans. Nature Genet. 1:* 114-123.

WILKINSON, D.G. (1993). *Essential Developmental Biology* (Eds. C. Stern, C. and P. Holland). Oxford University Press, Oxford.

ZHAO, Z., LEE, C.C., BALDINI, A. and CASKEY, C.T. (1995). A human homologue of the *Drosophila* polarity gene frizzled has been identified and mapped to 17q21.1. *Genomics 27:* 370-373.