The EMBO Journal vol.7 no.7 pp.2175-2183, 1988

Sequence of the *twist* gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos

B.Thisse, C.Stoetzel, C.Gorostiza-Thisse and F.Perrin-Schmitt

Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Institut de Chimie Biologique, Faculté de Médecine, 11, rue Humann, 67085 Strasbourg Cédex, France

Communicated by C.Nüsslein-Volhard

The *twist* gene is involved in the establishment of germ layers in *Drosophila* embryos: *twist* homozygous mutant embryos fail to form the ventral furrow at gastrulation and lack mesoderm and all internal organs. We have determined the sequence of the *twist* gene, that contains 'CAX' repeats in its 5' moiety, and codes for a protein of 490 amino acids. We have raised *anti-twist* antibodies that were used to study the distribution of the *twist* protein in whole mounts and tissue sections of wild-type embryos. *Twist* protein appears to be a nuclear protein at all developmental stages. It is present over both poles and in the midventral region (endoderm and mesoderm anlagen) at cellular blastoderm stage; later in development, it is detected within the mesodermal layer until its differentiation into somatopleura and splanchnopleura in which some cells are still labelled by *anti-twist* antibodies.

Key words: Drosophila/gastrulation/germ layer establishment/protein localization/*twist* sequence

Introduction

At gastrulation in *Drosophila*, the cells of the mid ventral region of the embryo invaginate, forming a ventral furrow. The cells involved in this process have been shown to consist of most of the anterior midgut anlage (described as endoderm), the entire mesoderm anlage, the proctodeum and part of the posterior midgut anlagen (belonging to the endoderm) (Poulson, 1950; Campos-Ortega and Hartenstein, 1985). The primordia that derive from these anlagen give rise to most of the internal organs of the larva (Campos-Ortega and Hartenstein, 1985).

Embryos homozygous for *twist* become abnormal at gastrulation: they do not form the ventral furrow, and were described as entirely lacking mesoderm and all internal organs, developing as a hollow tube of hypoderm and dying at the end of embryogenesis (Simpson, 1983; Nüsslein-Volhard *et al.*, 1984). *Twist* is a zygotic gene whose function is essential only at gastrulation [~3 h of development (Thisse *et al.*, 1987a)]. It belongs to a class of genes involved in the process of germ layer formation. This class is composed of 11 genes acting early during oogenesis, also called the dorsal group (the maternal genes), and a number of zygotic genes acting early in embryogenesis (Anderson, 1987, for a review and references therein).

Among the latter, *snail* seems very similar to *twist* in that the phenotype of *snail* embryos resembles that of *twist* embryos (Nüsslein-Volhard *et al.*, 1984; and Simpson, 1983). Other zygotic genes involved in the formation of germ layers (such as *dpp-Hin*, *zen*, *tsg*, *tld* or *srw*) seem to act primarily on more dorsally located tissue (Jürgens *et al.*, 1984; Zusman and Wieschaus, 1985; Doyle *et al.*, 1986; Hoffman and Goodman,, 1987; Irish and Gelbart, 1987; St Johnston and Gelbart, 1987; Rushlow *et al.*, 1987b). Molecular cloning of the *twist* gene has been reported (Thisse *et al.*, 1987a). *In situ* hybridization to tissue sections has revealed that the *twist* transcripts are detected at the cellular blastoderm stage in the midventral cells that belong to the mesodermal anlage, whilst during gastrulation and germ band extension they accumulate in the mesodermal layer of the embryos (Thisse *et al.*, 1987b).

In the present study, we report the *twist* gene sequence and the deduced primary structure of its protein. We have raised antibodies against a fusion containing the major part of the *twist* coding region and have analysed the localization of the *twist* protein in wild-type embryos during development. We show that the *twist* protein seems to be a nuclear-bound protein, that is present not only in the mesodermal layer of the embryos until its differentiation, but also in cells forming the anterior and the posterior midgut primordia until their first post blastoderm mitoses.

Results

Molecular organization of the twist gene

Restriction maps of the genomic and cDNA *twist* clones, the polarity of this transcript unit as well as the existence of a small intron in the 3' part of the gene have been reported previously (Thisse *et al.*, 1987a) and see Figure 1A. The complete nucleotide sequence of the *twist* gene is presented in Figure 1B. The comparison of this sequence with the sequence of the *twist* cDNA clone reveals certain characteristics of the gene. The unique intron of 120 bp lies between +1383 to +1503, the exon/intron boundaries being in agreement with the consensus sequence (Breathnach and Chambon, 1981; Mount, 1982). There are two hexanucleotides present at positions +1773 and +1782 (brackets on Figure 1B) in the non-coding part of the genomic clone which were not found in the cDNA clone. This could be explained either by a polymorphism between the Canton S (cDNA) and Oregon R (genomic) strain libraries, or by a cloning artefact. Note also a striking repetition of CAX nucleotides within the 5' part of the gene (see below).

The 5' nucleotide of the cDNA falls within a gTCAGTT sequence that does resemble the consensus transcription start of *Drosophila*: ATCA($^{G}/_{T}$) T($^{T}/_{C}$) described by Hultmark *et al.* (1986); in fact, the best fit is with the *yolk protein 1* transcription start site whose nucleotide +1 is aligned on the fourth nucleotide of the consensus. Primer extension

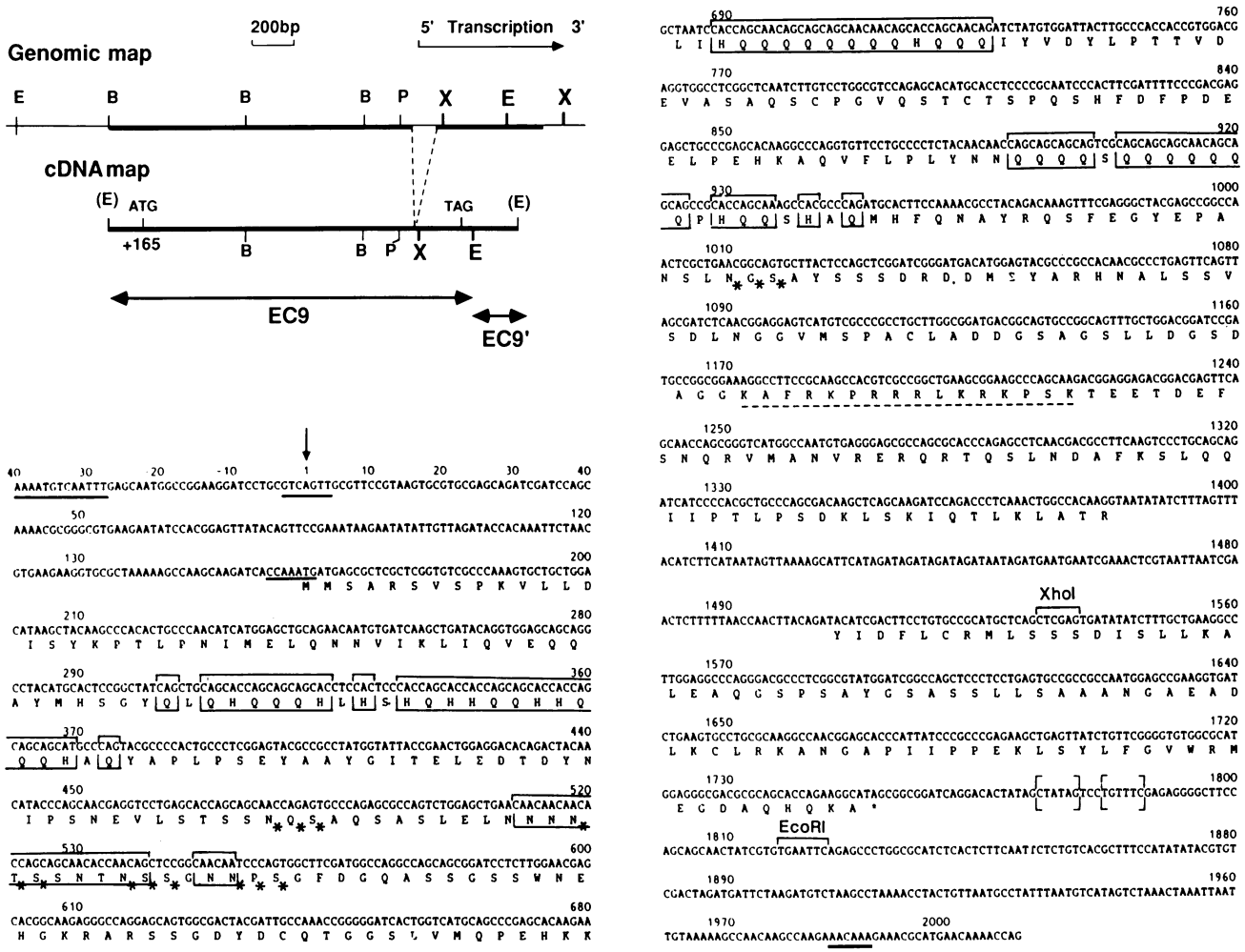


Fig. 1. (**A**) Schematic maps of the *twist*-genomic and cDNA *Eco*RI fragments. EC9 and EC9′ of the *twist* cDNA are aligned with respect to the cDNA map. B:*Bam*HI; E: *Eco*RI; P: *Pst*I; X: *Xho*I. The thick line on the genomic map corresponds to the exon sequences. The initiation (ATG) and termination (TAG) codons are figured on the cDNA. (**B**) Nucleotides sequence of the *twist* gene and deduced amino acid sequence of the encoded protein. The arrow indicates the major transcription start +1. The A/T rich region between −40 and −28, the ATG initiation codon and the possible polyadenylation sites are underlined. The TAG stop codon lies at +1748. The four 'CAX' stretches in the 5′ portion are boxed. The five potential glycosylation sites (N−X−S) are indicated by an asterisk. The dashed line indicates a region rich in basic amino acids.

analysis and nuclease S1 mapping showed that this sequence corresponds probably to the 5′ end of the *twist* mRNA (see Figure 2). At the 3′ end of the cDNA, we do not find the consensus AATAAA signal of polyadenylation, but there is an AACAAA sequenced at +1984 to +1989, lying 22 bp upstream from the poly(A) tail, that might be used as a polyadenylation signal. Therefore, the mature *twist* mRNA is 1878 bp long excluding the poly(A) tail and the *twist* cDNA clone EC9 is full length.

In the genomic sequence, upstream of the 5′ initiation site for transcription, there is no obvious TATA box (Breathnach and Chambon, 1981), although there is a $^{A}/_{T}$ rich region between positions −40 to −28. The translation process may use the first ATG codon (at position +160) which lies within a CcAAATG sequence that fulfils the requirements of a translation start for a *Drosophila* gene as described by Cavener (1987), namely: ($^{C}/_{A}$)AA($^{A}/_{C}$)ATG. Moreover, this ATG is followed by a 1470 nt ORF that ends at position 1748 on the genomic sequence with a TAG codon, coding therefore for a 490 amino acid protein. This is the only long ORF detected. It is followed by a 3′ untranslated region of 240 nt. There is a striking repetition of the CAX nucleotide triplet in the 5′ moiety of the gene. These repeats are translated as four stretches of Gln/His, Asn/Ser or Thr, Gln and Gln. Such sequences were described as 'OPA boxes' (Wharton *et al.*, 1985) or 'M repeats' (McGinnis *et al.*, 1984) and were found in other *Drosophila* genes, such as *Notch* (Wharton *et al.*, 1985), *Antennapedia* (McGinnis *et al.*, 1984), *Deformed* (Regulski *et al.*, 1985) *engrailed* (Kassis *et al.*, 1986) and recently *dorsal* (Steward, 1987).

Besides this structure, a region very rich in Arg or Lys (10 basic amino acids amongst 16) is observed (nt 1172−1217, dashed underlined region on Figure 1B, as well as five possible glycosylation sites (Asn−X−Ser) in the 5′ moiety of the gene (asterices on Figure 1B). A putative cAMP dependent phoshorylation site (described as basic−basic−X−Ser; Krebs and Beavo, 1979) lies within the Arg/Lys rich dashed underlined region. The *twist* translation product is mostly hydrophilic, and has a calculated pH_i of 6.15. Computer searches with the primary structure of the *twist* gene (GenBank, EMBL) and of the deduced *twist* protein (NBRF) did not reveal any homology with other genes or proteins, besides homologies with the 'OPA boxes' or 'M repeats' that we have already mentioned.

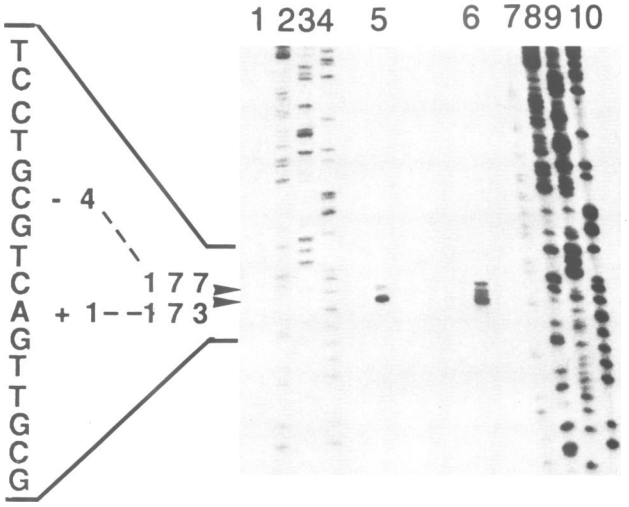


Fig. 2. Transcription start mapping. The single-stranded probe synthesized for S1 nuclease mapping, as well as the 17mer oligonucleotide used in primer extension analysis, were hybridized to poly(A)$^+$ RNA from 2 to 4 h embryos (~10-fold probe excess versus *twist* RNA molecules). **Lanes 1–4** and **7–10** are sequencing reactions primed with the 17mer. The polymerization reaction was done with unlabelled triphosphate nucleotides. **Lane 5**: primer extension reaction, the major product migrates as a 173 nt fragment, the minor one as a +177 nt fragment. **Lane 6**: defining the transcription initiation site by S1 nuclease mapping. A discrete number of fragments are protected (incomplete digestions), the major one migrates as a fragment of 173 nt.

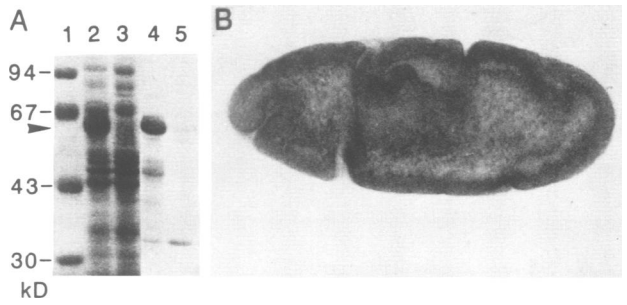


Fig. 3. Detection of the *βgal–twist* fusion protein and anlaysis of insert specific antibodies. (**A**) Crude extracts were prepared from JM109 bacteria containing the *pUC92-twist* vector grown in the presence (**lane 2**) or absence (**lane 3**) of IPTG, separated on 8% polyacrylamide–SDS gel, and stained with Coomassie Blue. Standard mol. wt markers were run in parallel (**lane 1**: phosphorylase B 94 kd, albumin 67 kd, ovalbumin 43 kd and carbonic anhydrase 30 kd). The induced *βgal–twist* fusion protein of ~62 kd is visible in lane 2 (arrow head). Replica gel was transferred to a nitrocellulose filter which was incubated with partially purified rabbit serum containing antibodies directed against the *βgal–twist* fusion protein (see Materials and methods). Goat anti-rabbit peroxidase coupled IgG were used as a secondary antibody to visualize by chemical coloration the presence of the *anti-twist* antibodies. **Lane 4** contains proteins extracted from IPTG induced cultures, and **lane 5** from cultures grown without IPTG. One major band is apparent in lane 4, corresponding to the *βgal–twist* fusion protein and some degradation products, while no label is detected in the control lane 5, showing the specificity of the *anti-twist* antibodies with regards to protein epitopes. (**B**) Test of the specificitiy of the antisera using a null *twist* allele. Embryos from the *b pr cn* twi^{ey50}/*CyO* strain were collected at 2 to 6 h of development, and processed as described in Materials and methods for whole mount immunocytology experiments. twi^{ey50} is a null allele (deletion of 1.8 kb within the 5′ coding part of the gene) (Thisse *et al.*, 1987a). About 25% of the observed embryos, with *twist* phenotype (Figure 3B), did not show any label over background.

Production of anti-twist antibodies

To produce a *βgal–twist* fusion protein we used the pUC bacterial expression vector (Hanna *et al.*, 1984) in which most of the *twist* cDNA sequences were inserted (see Figure 1A). Rabbits injected with a *βgal–twist* hybrid protein produced polyclonal antisera from which were eliminated the antibodies directed against the *Escherichia coli* protein epitopes (see Materials and methods and Figure 3A). Such antisera fail to label *twist* homozygous embryos (Figure 3B). In the following description we will call them '*anti-twist*' antibodies.

Localization of the twist protein in developing wild-type embryos

Although the embryonic development is a continuous process, an artificial subdivision into a series of different stages is useful to provide a temporal reference framework for embryonic events. We will refer to the stages described by Campos-Ortega and Hartenstein (1985).

During the first four stages following fertilization, the zygotic nuclei divide 13 times, giving rise to a syncytial blastoderm. At stage 5 (2 h 10 to 2 h 50), cellularization occurs, and soon after that the gastrulation process begins.

By using *in situ* hybridization to tissue sections, we have previously shown that the *twist* transcripts appear at stage 5 in the midventral cells that will invaginate during gastrulation to form the ventral furrow. Later, at germ band extension, those cells give rise to the mesodermal layer of the embryo, and we observed the *twist* transcripts accumulate in those mesodermal cells until stage 10 (end of germ band extension) (Thisse *et al.*, 1987b).

Anti-twist antibodies were used to reveal the cellular localization of the *twist* native protein in whole mounts, as well as in tissue sections of wild-type embryos. The presence of the *anti-twist* antibodies was revealed by histochemical coloration using a biotinyl goat antirabbit IgG that fixed avidin-peroxidase as a secondary antibody. We chose a peroxidase reaction to reveal the protein localization by histochemical colorations of the immunocytological preparations as it has the advantage over fluorescence of being permanent, and stable enough to allow the preparation of tissue sections even after the immunological reaction and coloration processes.

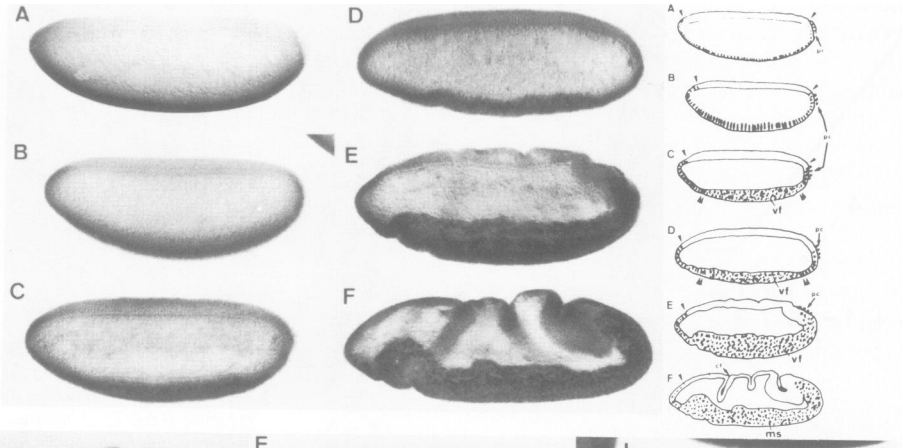
Blastoderm stages

We first detect the *twist* protein, albeit weakly, in early stage 5 embryos (~2 h 15), at the time where the blastoderm nuclei are still spherical, at the beginning of cellularization. The *twist* protein accumulates throughout the midventral cells of the embryo, extending around both anterior and posterior poles to label some dorsal cells (Figure 4–1A). During the formation of the cell walls, the nuclei elongate considerably and the level of labelling increases within the cells. However the number of cells that express *twist* remains constant and the intensity of the label is similar in all cells (Figure 4–1B).

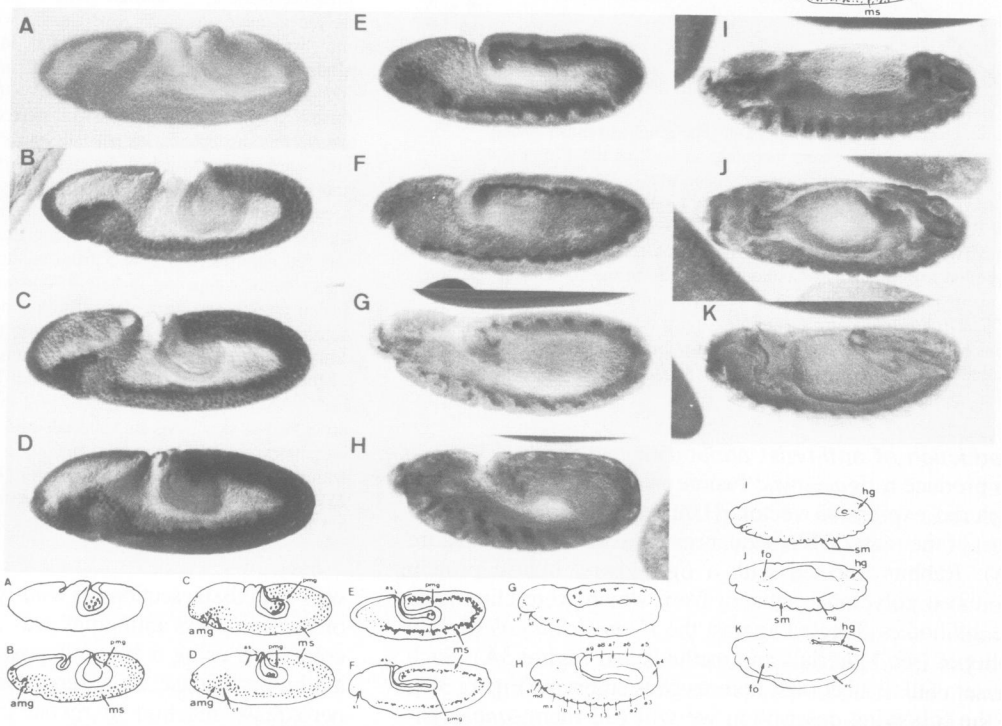
At this stage, as well as later on, the *twist* protein seems to be located within the nuclei (see in particular the tissue sections of (Figure 5C and D and Figure 6A).

It is clear at stage 5 that the *twist* protein is present around the anterior pole even in cells described as ectodermal and endodermal (which will later form the anterior midgut

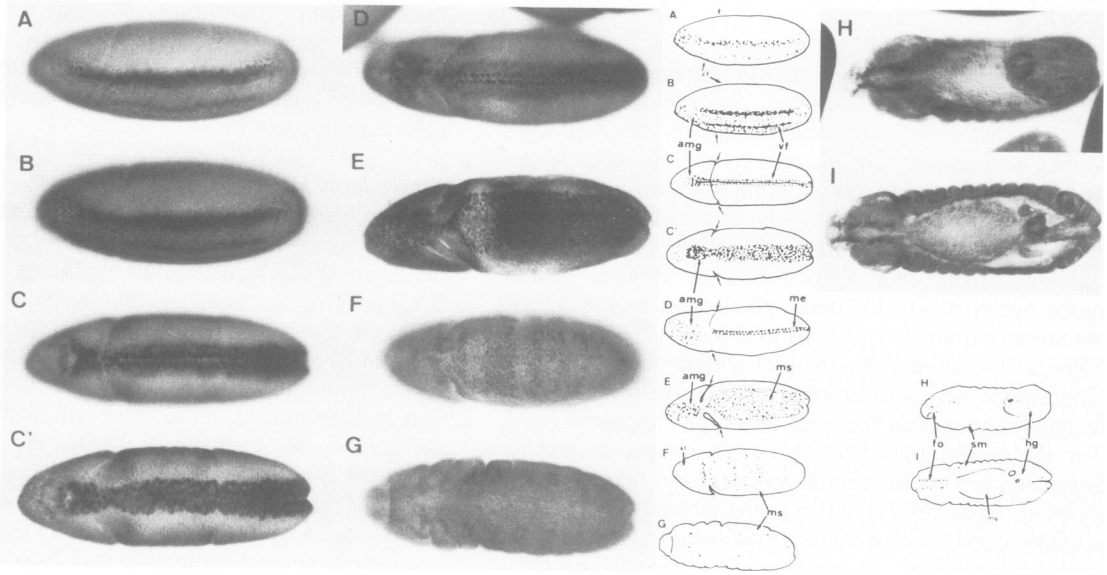
4-1


A
D
B
E
C
F
vf
ms

4-2


A
E
I
B
F
J
C
G
K
D
H
hg
fo
sm
amg
ms

4-3


A
D
H
B
E
I
C
F
C'
G
amg
vf
me
ms
fo
sm
hg

primordium), in continuity at the ventral side to the mesoderm primordium, and further to caudal part in continuity to the cells that belong to the proctodeum and posterior midgut anlagen (endoderm) (see Figure 4−1A−C). The *twist* protein at the posterior pole is present in the cells forming the posterior plate, that are situated immediately beneath the pole cells (the presumptive germline). No label is found in the pole cells (Figure 4−1D).

Gastrulation process

Stage 6 (2 h 50 to 3 h) is described as a short stage corresponding to early gastrulation, where mesodermal and endodermal primordia invaginate, and where cells of the posterior pole begin to migrate dorsally towards the anterior pole, forming a flat dorsal plate to which the pole cells adhere. The cephalic folds also appear at this stage.

The *twist* protein can be followed during these different steps within the ventral furrow of presumptive mesodermal cells and the endodermal cells at both the anterior and posterior part of the embryo and in the ectodermal cells of the anterior midgut primordium (Figure 4−1D−F for lateral views, Figure 4−3A−D for ventral views of whole embryos and Figure 5A for a parasagital section).

Until the middle of stage 6 when the cells participating in the ventral furrow invaginate, the intensity of the label seems similar in all cells. At the transition to stage 7 (Figure 4−1E and F), in whole mount embryos, cells that have invaginated appear to be less intensely stained than those of the anterior pole that have not yet invaginated. Such differences however are only observed with whole embryos, as sections at that stage do not show such differences in labelling (Figure 5A).

On transverse sections of labelled embryos, all the mesodermal cells forming the ventral furrow express the *twist* protein (we counted ~12 cells) as well as four mesectodermal cells connecting the ventral furrow to the ectoderm (see Figure 5C). Ventral side observation of embryos in whole mounts shows that only two mesectodermal cells at the midventral line appear to be labelled at the surface of the embryo (see in particular Figure 4−3C and D) although by varying the focus, the inner mesodermal cells also appear to be clearly labelled (Figure 4−3C′).

As gastrulation proceeds, the cell plate that carries the pole cells reaches a dorsal position (at the beginning of stage 7), and the *twist* protein label is barely visible in these cells (Figure 4−1E). As the cell plate sinks to form the aminoproctodeal invagination (late stage 7), no more *twist* protein is detected in those cells forming its bottom part, the cells which are going to form the posterior midgut rudiment. At the anterior pole of the embryo, however, the anterior midgut rudiment is beginning to form, and is still labelled with the *anti-twist* antibodies; in the mesodermal layer the *twist* protein is still abundant (Figure 4−1F for a lateral view of a whole mount embryo).

Germ band elongation

At the end of gastrulation, marked by the end of the formation of the mesoderm and endoderm layers, germ band elongation begins (stage 8—3 h 10 to 3 h 40).

The germ band consists of the mesodermal and ectodermal sheets of cells laying between the cephalic fold and the posterior plate. This germ band begins to elongate along the anterior−posterior axis, at stage 8, and moves around the posterior pole of the embryo along the dorsal surface, preceded by the proctodeal invagination.

During stage 8, all mesodermal cells perform two consecutive mitoses, and at the end of this stage, the cells of the anterior midgut primordium are reorganized. Embryos in the course of germ band elongation are shown in Figure 4−2A−D: Figure 4−2A being a control (incubation with preimmune serum). The *anti-twist* antibody label is visible in Figure 4−2B−D within the whole mesodermal layer of the germ band, as well as within the anterior midgut primordium. At the end of stage 8, the mesoderm begins to migrate anteriorly, giving rise to the cephalic mesoderm. The transient segmentation of the mesodermal layer during stage 9 is clearly visible (Figure 4−2E for a lateral view and 5B for a sagital section). At late stage 9 to early stage 10, the stomodeum had invaginated, and the cephalic mesoderm progresses towards the dorsal part of the

Fig. 4. Immunolocalization of the *twist* protein in whole mounts of wild-type embryos. The brown histochemical coloration that is obtained by peroxidase reaction is black on these pictures and labels the location of the *anti-twist* antibodies (see Materials and methods). (**1**) From beginning of cellular blastoderm to end of gastrulation. Embryos are oriented anterior to left, dorsal side up. The *twist* protein is detected around both poles and along the whole ventral side. (**A**) End stage 4−beginning stage 5; (**B−C**) stage 5 with elongated ventral nuclei; (**D**) stage 6: beginning of gastrulation, ventral nuclei begin to form the ventral furrow; (**E**) end stage 6, pole cell at dorsal position, *twist* protein is still present in the cell plate; (**F**) stage 7: pole cells are sinking, no *twist* protein detected within the amnioproctodeum invagination. On line drawings, the arrow heads point to the observed limits of the *twist* expression territory, and double arrow heads indicate the borders of the mesoderm anlage. Abbreviations: pc, pole cells; amg, anterior midgut anlage; pmg, posterior midgut anlage; cf, cephalic fold; vf, ventral furrow; ms, mesoderm layer. (**2**) From germ band extension to beginning of head involution stages. Embryos are oriented anterior to left, dorsal side up. (**A**) control embryo incubated with preimmune serum. (**B−K**) Embryos incubated with *anti-twist* antibodies. (**B−D**) Stage 8 embryos showing germ band elongation, with *twist* protein within the mesodermal layer, and at the anterior pole, within the growing anterior midgut primordium and still in some cells belonging to the ectodermal part of the anterior midgut primoridum. Transient segmentation of the mesodermal layer is obvious in a stage 9 embryo (**E**) where the mesodermal layer has progressed within the procephalic lobe towards its dorsal side. The stomodeum is formed at the end of stage 9 (**F**). No obvious *twist* protein is seen in the reorganized anterior midgut primordium. At stage 10, definitive segmentation is apparent within the mesodermal layer. At this stage the *twist* protein is only detected within the mesodermal layer. At stage 11, intersegmental furrows appear in the epidermis. *Twist* protein is still detected within some cells of each segment. (**I−K**) Stage 12−14 embryos. Differentiation of mesoderm into somatopleura and splanchnopleura has occurred, and the *twist* protein is still detected in some cells of each layer within cells giving rise to somatic musculature and along the fore and hindgut, among cells that will contribute to the visceral musculature. On line drawings, same abbreviations as in (1), plus: as, amnioserosa; st, stomodeum; cl, clypeolabrum; md, mandibular bud; mx, maxillary bud; lb, labial bud; t1−3, thoracic segments; a1−9, abdominal segments; Fo, foregut; hg, hindgut; vc, ventral chord; sm, somatopleura; som, somatic musculature. (**3**) Embryos observed from ventral (**A−G**) or from dorsal (**H**) and oblique (**I**) sides (anterior part ot left). (**A−C′**) formation of the ventral furrow and of the cephalic fold. (**C**) and (**C′**) are two views of the same embryo, with the focus on the surface (**C**) and on the presumptive mesoderm (**C′**). (**D**) Stage 8 embryos with obliterated ventral furrow; (**E**) stage 9; (**F**) stage 10; (**G**) stage 11; (**H** and **I**) stage 12.

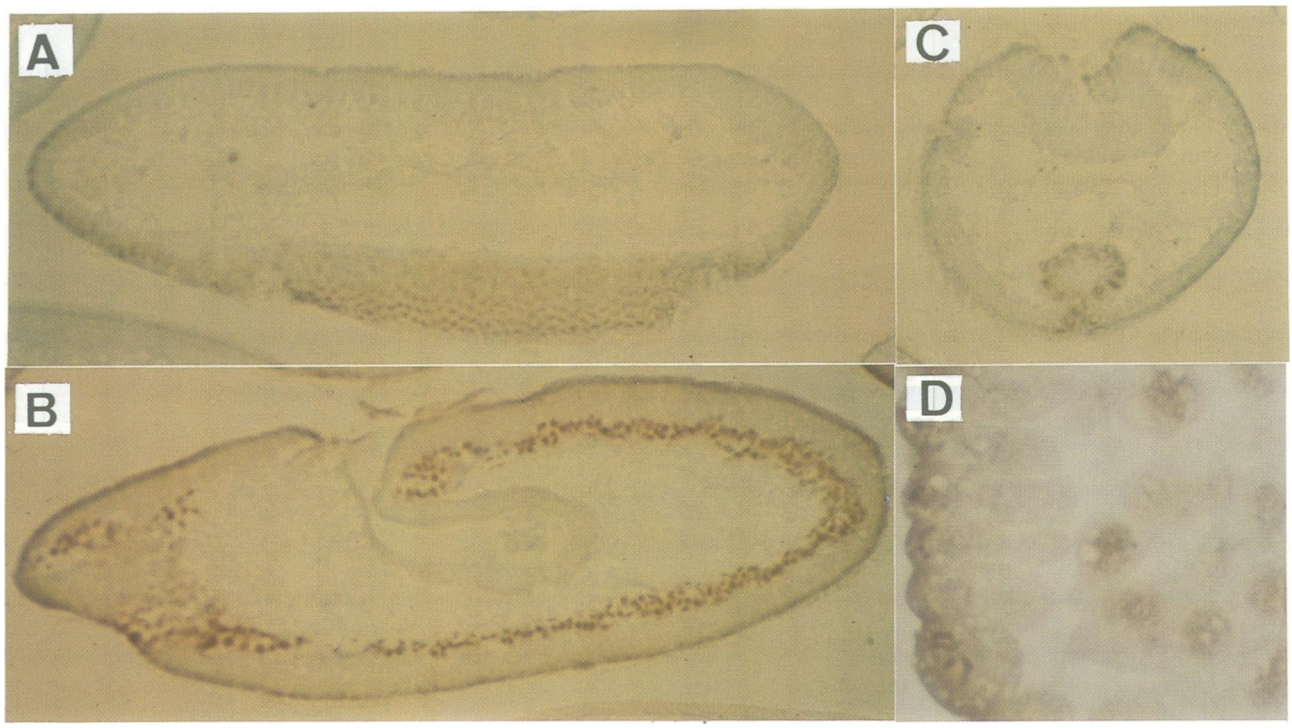


Fig. 5. (A–C) Paraffin sections (8 μm) of embryos (oil objective 17×). The localization of the *anti-twist* antibodies is revealed by the brown coloration developed by peroxidase reaction before sectioning. The blue counter stain of the cytoplasm is obtained by Giemsa treatment of the sections. The *twist* protein is visualized around poles and within the ventral furrow in (**A**) (parasagital section through a stage 6 embryo), whilst only the mesodermal layer is labelled in (**B**) (sagital section through a stage 9 embryo). The transverse section in (C) through the posterior part of a late stage 7 embryo (pole plate visible at the dorsal side) shows the presence of *twist* protein within both the mesodermal and the mesectodermal nuclei. (**D**) Thin section (1.5 μg) showing the nucleic localization of the *twist* protein within the mesoderm of procephalic lobe (oil objective 100 ×).

procephalic lobe (Figure 4–2F for a lateral view). At that time, only the mesodermal layer is labelled by the *anti-twist* antibodies.

The late stage 10 embryo shows the definitive segmentation pattern within the mesodermal layer (Figure 4–2G for lateral and 4–3F for ventral views). At the time of the appearance in the epidermis of the intersegmental furrows (beginning of stage 11) the number of cells with appreciable *twist* protein decreases, and only a few cells in each segment are labelled by *anti-twist* antibodies (Figure 4–2H for lateral and 4–3G for ventral views). The cells of the mesodermal layer perform two successive mitoses at the middle of stage 11, and then differentiate into somatopleura and splanchnopleura.

Germ band retraction and later stages

At 7 h 20, the germ band begins to retract, and the first regional differences of the mesodermal layer occur namely the separation of the somatopleura (the cells which will give rise to somatic musculature) from the splanchnopleura (the cells which will give rise to visceral musculature). Embryos at stage 12 (beginning of germband retraction) show *anti-twist* label both in some of the cells belonging to the somatopleura (some cells within each segment) as in cells of the splanchopleura (cells bordering the fore gut and the hind gut) (see Figure 4–2I and J for lateral views of whole mount embryos and Figure 6B–D for tissue sections).

A dorsal view from a stage 12 embryo in Figure 4–3I, as well as an oblic view in Figure 4–3H show both fore gut and hind gut labels. The splanchnopleura label is still visible in an embryo at about stage 14 (Figure 4–3K) which are the older embryos that were observed.

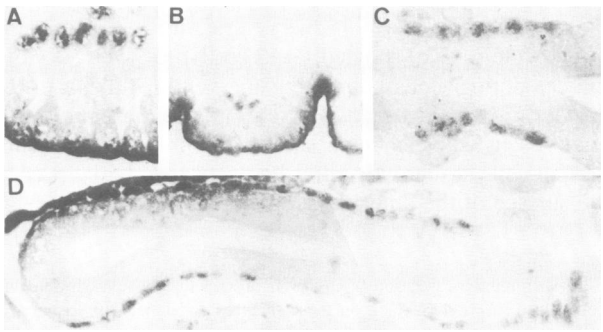


Fig. 6. Thin sections (1.5 μm) showing the mesodermal labelled layer of a stage 9 embryo (**A**), and in an about stage 12 embryo: the somatopleura label (**B**) and the fore gut (**C**) and hind gut (**D**) splanchnopleura labels.

Discussion

We have determined the nucleotide sequence of the *twist* gene. It is a simple transcription unit, with a unique 5′ end and one 120 bp intron near its 3′ end. The 1878 nt long *twist* mRNA is polyadenylated, and contains one large ORF generating a protein of 490 amino acids. The whole *twist* gene lies within the deficiency $Df(2R)twi^{S60}$ cloned during our chromosomal walk (Thisse *et al.*, 1987a), at position +3 to +5.

In the case of *twist*, searches for DNA sequence homologies or possible functions via tertiary structure did not reveal anything that would help in suggesting a function for the *twist* protein. Cell migrations during gastrulation involves cell surface receptors (see Hynes, 1987, for a review), and injection of R.G.D. peptides to *Drosophila* embryos mimics the *twist* phenotype (Naidet *et al.*, 1987); but nothing in the *twist* putative amino acid sequences shows any structural similarities with such molecules.

The *anti-twist* antibodies allows us to investigate the temporal and spatial distribution of the *twist* protein during embryogenesis, including its subcelluar localization.

For the *twist* gene, it appears that translation occurs just after the transcription process, as we detected the *twist* protein from stage 5 onwards, at about the same time when the *twist* RNA is detected by Northern blot analysis and *in situ* hybridizations to tissue sections (Thisse *et al.*, 1987b). The regions in the embryos that were shown to accumulate *twist* transcripts also accumulate the *twist* protein at about the same time. These observations suggest that there would be no major post-transcriptional regulation for the *twist* gene, in contrast to that observed in the case or *Krüppel* (Gaul *et al.*, 1987) or *Antennapedia* (Wirz *et al.*, 1986).

Our data show that the *twist* protein accumulates at early stages (from the beginning of cellular blastoderm to germ band extension) in all of cells of presumptive mesodermal layer (situated at the midventral part of the embryo), and also at both the anterior and posterior poles of the embryos.

How can we define the spatial territories of expression of the *twist* protein at early stages of development with respect to the embryonic layers?

In the cases of *dpp* (St Johnston and Gelbart, 1987), of the gene located at 95F possibly *crumbs* (Knust *et al.*, as well as for *zen* (Doyle *et al.*, 1986; Rushlow *et al.*, 1987b), transcript accumulations along the dorsal side and around both poles are clearly observed at beginning of cell formation. And in sagital sections, no RNA from these genes is detected within the mesodermal layer. For the *integrin PS2 α subunit* gene, the transcripts are detected at the same stage strictly within the mesodermal layer (Bogaert *et al.*, 1987). Therefore it is clear that the borders (at least of the mesodermal anlage) are already defined at that stage. And if the *twist* expression territory does not follow the strict subdivisions of the defined ectodermal, endodermal and mesodermal layers, it is not only the result of lack of strict delimitation of these anlagen at early stages. From compilation of fate maps of blastoderm stages (Hartenstein *et al.*, 1985) and of 6−8 min old gastrula (Technau and Campos-Ortega 1985), we would like to suggest that during germ layer determination, there exists a first subdivision into cells of the blastoderm that will invaginate during or after gastrulation, and cells which will stay at the surface of the embryos. *Twist* is expressed in the former group that is composed of cells of the anterior subdivision of the anterior midgut primordium (ectodermal cells which will form part of the stomodeum), cells of the posterior subdivision of the anterior midgut primordium (endodermal cells which form the tip of the ventral furrow at gastrulation), the whole mesodermal layer, the proctodeum and the posterior midgut primordium (endodermal cells).

However the *twist* protein function is apparently not necessary for cell invaginations. The observed mutant phenotype of the *twist* homozygous embryos (Simpson, 1983; Nüsselein-Volhard *et al.*, 1984) as well as ours is a defect in ventral furrow formation, although the cells forming the anterior midgut and the amnioproctodeal anlagen invaginate in the mutant embryo. Thus as regards invagination of cells, there are two independent mechanisms: cells of the anterior midgut and posterior midgut anlagen invaginate even in the absence of *twist* protein, while in contrast, the cells of the mesodermal layer do not invaginate in the absence of *twist* protein.

At the anterior pole of the embryo, it is clear that until the end of germ band extension *twist* protein is present in cells that belong to the anterior midgut anlage, which are considered to be of both endodermic as well as ectodermic origins, and that is disappears at about the time of the first post-blastoderm mitoses in these cells.

Within the anterior midgut primordium, the first post-blastoderm mitoses of the endodermal cells occurs at stage 8, the second one at stage 9. The *twist* protein is visible in the endodermal cells of the anterior midgut until the end of stage 8. At the posterior part of the embryo, the *twist* protein disappears when the pole plate sinks within the embryo, corresponding approximately to the period of the first mitoses in the posterior endodermal cells. However, in the mesodermal layer, the *twist* protein remains present after two mitoses (clearly until stage 12). Thus even if the *twist* protein expression territory does not follow strict ectoderm/endoderm/mesoderm limits, there are some clear cut layer subdivisions that are observed as different half-lives of the *twist* protein within the endodermal versus mesodermal layer, suggesting different controls in those cells.

Recent studies of *zen* expression in mutant embryos suggest that there exist separate regulatory clues at the poles (Rushlow *et al.*, 1987a). Work is in progress to test such a hypothesis for *twist* expression,

Another point is that the temperature sensitive period (TSP) of the *twist* gene has been determined to be around gastrulation (between 2 h 30 and 3 h 30) (Thisse *et al.*, 1987a). Therefore, *twist* gene function is vital at that time for the embryos, but not later. By using *anti-twist* antibodies we detect the *twist* protein until ~ 10 h of developmental. Is the *twist* protein that remains in the mesodermal cells later than the TSP only a persistence of a protein that has no further function?

Finally, we have observed that the *twist* protein appears to be localized in the nucleus even at every early stages, both with tissue sections of 8 μm (paraffin embedding) as well as with thin sections (1.5 μm). Preliminary results on ultra-thin sections observed with electron microscopy are also consistent with a nucleoplasmic localization (as opposed to membrane bound); we could not however determine whether or not the protein is attached to chromatin (work in progress). Such nuclear localization may also disfavour any hypothesis suggesting a direct role for the *twist* protein in cell adhesion or cell motility.

In this report, we have presented the whole structure of the *twist* gene, which appears to be a simple gene with one 5′ and one 3′ end and only one protein coding mRNA. However, the DNA sequence has not revealed the possible fucntion of the *twist* protein. Other than the fact that we know that the *twist* protein is vital for embryos at the gastrulation stage and that the *twist* protein is necessary for mesoderm formation, we cannot yet conclude its exact role in germ-layer establishment. However, by immunocytology, we

observed that the *twist* protein seems to be always localized in the nucleus, and that it is not only present in the mesoderm layer, but also in midgut primordia.

Two possible hypotheses could be suggested to speculate about the role of the *twist* protein. First, we know from Beer *et al.* (1987) that the cells of the midventral region of the embryos are committed to be mesodermal cells immediately after cellularization of the blastoderm, at about the same time as when the cells of the anterior and posterior midgut anlagen are committed to produce midgut cells (Technau and Campos-Ortega, 1986); and also, that cells of the anterior and posterior midgut anlagen are equivalent at early stages as they can be interchanged (Technau and Campos-Ortega, 1986). Therefore it is tempting to speculate that the *twist* protein might have a role in the commitment of cells to be 'internal cells' (midgut and mesodermal cells). Secondly, we know that the expression of *twist* is under the control of the *dorsal* gene (directly or indirectly) (Thisse *et al.*, 1987b); and preliminary results with *anti-dorsal* antibodies had shown that the *dorsal* protein is present broadly over the ventro-lateral part of the embryos, extending towards both poles (R.Steward, unpublished results). We could then suggest that the domain of *twist* expression follows the territories of *dorsal* expression along the anterior−posterior axis: the presence of the *twist* protein over part of ectodermal and endodermal cells in addition to its presence in the mesodermal cells would then result from the mechanism of its activation. At the poles, *twist*, *dpp*, *zen* as well as the *95F* gene are expressed at early stages. Among these, only *twist* is expressed within the mesodermal anlage; and in *twist* mutant embyros the only ventral furrow does not form (anterior and posterior midgut primordia invaginate). Therefore, the *twist* protein could act solely within the mesodermal primordium, and *twist* could be the (or one of the) moelcule(s) that are responsible for the commitment of the mesoderm.

In molecular terms for both hypotheses *twist* could regulate the expression of genes responsible for embryonic layer differentiation: in the mesodermal anlage, *twist* could at least control genes involved in cell adhesion and/or cell migration which enable ventral furrow formation at gastrulation.

Materials and methods

DNA sequencing

We used the two *Eco*RI fragments of the *twist* cDNA clone (EC9 and EC9′, see Figure 1A) as well as the 2.3 kb EcoRI genomic *twist* fragment (see Thisse *et al.*, 1987a), that were inserted into the *Eco*RI site of the pEMBL 18^+ vector (Dente *et al.*, 1983), and the dideoxy chain termination method (Sanger *et al.*, 1977) to determine the nucleotide sequence on both strands. Progressive overlapping clones deleted from one end were obtained by the technique of DNase I digestions in presence of Mn^{2+} (Lin *et al.*, 1985). A *Xho*I fragment of 0.6 kb, that overlaps the internal *Eco*RI site situated at +1819, has been isolated from the original genomic λ clone, subcloned and sequenced on both strands. The sequences bordering the internal *Eco*RI site were those that were determined by sequencing EC9 and EC9′ subclones.

Primer extension and S1 mapping

Primer extension analyses used two 17mer oligonucleotides (synthesized by F.Ruffenach in our laboratory). They are complementary to sequences +39 to +56 and +156 to +176 in Figure 1.

Single stranded 5′ end-labeled probes were synthesized on the pEMBL-EC9 recombinant, primed by these 17mer oligonucleotides, and synthesis extended to the *Eco*RI site at −442. Probes were hybridized to poly(A)$^+$ RNA prepared from 2−4 h embryos and S1 digested as described by Freund *et al.* (1986).

Construction of the expression plasmid

The *pUC-twist* expression plasmid was constructed as follows. The cDNA fragment (+366 to +1686), coding for 421 amino acids (see Figure 1A), was isolated from the EC9-D2 deletion clone (generated in pEMBL 18^+, for sequencing see above). This fragment was excised by *Hin*dIII (from the polylinker) and *Eco*RI and inserted at the *Eco*RI−*Hin*dIII sites in the polylinker of the pUC vectors (Hanna *et al.*, 1984). The *pUC-twist* recombinant DNA was used to transform the JM109 bacterial strain (Yanisch-Perron *et al.*, 1985) in order to produce a *βgal−twist* fusion protein. The *pUC92-twist* recombinant generates a *βgal−twist* protein. This resulting fusion protein contains (from N to C terminal): 8 amino acids from the *lac Z′ 5′ coding*, 10 amino acids from the pEBML polylinker, 421 amino acids from the *twist* coding sequence, 21 amino acids from the *twist*-non coding sequence (see below) and 100 amino acids from the *lac Z′ 3′ coding part*. Attempts to clone other fragments (i.e. from deletion clones) in pUC and pUR vectors (Rüther and Müller-Hill, 1983) resulted in unstable fusion proteins. The only stable construct contains an in-phase insert of the *twist* fragment at both the 5′ and 3′ ends of the *lac Z′* gene. It appears that the *twist* stop codon (UAG) is not recognized as a stop by bacteria of the JM series *sup E44* (Yanisch-Perron *et al.*, 1985), therefore the fusion protein also contains amino acids from the *twist* non-coding sequence.

Purification of the twist-fusion protein and of production antibodies

Bacterial cultures were grown in the presence or absence of IPTG. Crude extracts of proteins were prepared and separated by gel electrophoresis. The fusion protein band was identified as being specific to the bacterial strain grown in the presence of IPTG as inducer.

The band containing the *βgal−twist* protein was cut out from a preparative SDS−polyacrylamide gel, frozen at −80°C, ground to powder in a mortar with liquid nitrogen, and redissovled in PBS. After adding the same volume of complete Freund's adjuvant, ~200 μg of fusion protein was injected per rabbit. Rabbits were boosted at 4-week intervals with 100−150 μg fusion protein in incomplete Freund's adjuvant, and bled 8 days after the last boost.

Antibodies directed against bacterial protein epitopes were removed as described (Carrel *et al.*, 1969; Martinez-Hernandez *et al.*, 1975).

The specificity of the effluent was tested on Western blots of the *βgal−twist* hybrid protein, with proteins of the induced and non-induced *pUC92-twist* JM109 strain as control (see Figure 3A).

Immunocytology

Embryos were dechorinated, permeabilized and fixed as described by Mitchison and Sedat (1983) except that 4% paraformaldehyde was dissolved in PBS. After rehydration, embryos were preincubated in PBS, 5% NGS, 0.2% NP-40 for at least 30 min, and incubated with *anti-twist* antibodies overnight at 4°C (serum dilution 1:300). Three washes in PBS, 5% NGS, 0.2% NP-40 were followed by incubation with goat biotinyl antirabbit IgG from BRL at 1:200 in the same buffer (2 h at 20°C). After three successive washes, the last in PBS alone, embryos were incubated with ABC-kit reagents from Vectastein (45 min at 20°C). The peroxydase reaction (1 mg/ml DAB; 0.03% H_2O_2) was followed under a stereomicroscope and stopped with PBS. Embryos were then dehydrated, mounted in 1:2 Canada Balsam:methylsalicylate and observed.

For sections, after the peroxidase reaction, embryos were either directly processed for paraffin inclusion (8 μm sections) or osmium-acid treated (1%) before inclusion into Spurr resin (from Taab Laboratory Equip.) to enhance contrast (1.5 μm sections) (Roussel *et al.*, 1987).

Acknowledgements

We are grateful to P.Chambon for his continued interest in this project, and to C.Nüsslein-Volhard for critically reading the manuscript. We thank Timo J.Ylilomi for helpful suggestions in histochemistry, Guy Roussel for advice on thin-sectioning procedures and use of his material, G.Richards for comments on the manuscript, B.Boulay and C.Werlé for help with the pictures and Hélène Martina for typing the manuscript. This work was supported by grants from CNRS, INSERM and the Association pour la Recherche sur le Cancer.

References

Anderson,K.V. (1987) *Trends Genet. Sci.*, **3**, 91−97.

Beer,J., Technau,G.M. and Campos-Ortega,J.A. (1987) *Roux's Arch. Dev. Biol.*, **196**, 222−230.

Bogaert,T., Brown,N. and Wilcox,M. (1987) *Cell*, **51**, 929–940.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Carrel,S., Gerber,H. and Barandum,S. (1969) *Nature*, **221**, 385–386.
Campos-Ortega,J.A. and Hartenstein,V. (1985) *The Embryonic Development of D.melanogaster*. Springer, Berlin.
Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
Doyle,H.J., Harding,K., Hoey,T. and Levine,M. (1986) *Nature*, **323**, 76–79.
Freund,J.N., Vergis,W., Schedl,P. and Jarry,B.P. (1986) *J. Mol. Biol.*, **189**, 25–36.
Gaul,U., Seifert,E., Schuh,R. and Jäckle,H. (1987) *Cell*, **50**, 639–647.
Hanna,Z., Fregeau,C., Prefontaine,G. and Brousseau,R. (1984) *Gene*, **30**, 247–250.
Hartenstein,V., Technau,G.M., Campos-Ortega,J.A. (1985) *Roux's Arch. Dev. Biol.*, **194**, 213–216.
Hoffman,F.M. and Goodman,W. (1987) *Genes Devel.*, **1**, 615–625.
Hultmark,D., Klemenz,R. and Gehring,W.J. (1986) *Cell*, **44**, 429–438.
Hynes,R.O. (1987) *Cell*, **48**, 549–554.
Irish,V.F. and Gelbart,W.M. (1987) *Genes Devel.*, **1**, 868–897.
Jürgens,G., Wieschaus,E., Nüsslein-Volhard,C. and Kluding,H. (1984) *Roux's Arch,Dev. Biol.*, **193**, 283–295.
Kassis,J.A., Poole,S.J., Wright,D.K., O'Farell,P.H. (1986) *EMBO J.*, **5**, 3583–3589.
Knust,E., Dietrich,U., Tepass,U., Bremer,K.A., Weigel,D., Vässin,H. and Campos-Ortega,J.A. (1987) *EMBO J.*, **6**, 761–766.
Krebs,E.G. and Beavo,J.A. (1979) *Annu. Rev. Biochem.*, **48**, 923–959.
Lin,H.C., Lei,S. and Wilcox,G. (1985) *Anal. Biochem.*, **147**, 114–119.
Martinez-Hernandez,A., Merril,D.A. and Naughton,M.A. (1975) *J. Histochem. Cytochem.*, **23**, 146–148.
McGinnis,W., Levine,M., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) *Nature*, **308**, 478–483.
Mitchison,T.J. and Sedat,J. (1983) *J. Dev. Biol.*, **99**, 261–264.
Mount,S. (1982) *Nucleic Acids Res.*, **12**, 857–872.
Naidet,C., Sémériva,M., Yamada,K.M. and Thiery,J.P. (1987) *Nature*, **325**, 348–350.
Nüsslein-Volhard,C., Wieschaus,E. and Kluding,H. (1984) *Roux's Arch. Dev. Biol.*, **193**, 267–282.
Poulson,D.F. (1950) In Demerec,M. (ed.), *Biology of Drosophila*. Wiley, NY. pp. 168–274.
Regulski,M., Harding,K., Kostriken,R., Karch,F., Levine,M. and McGinnis,W. (1985) *Cell*, **43**, 71–80.
Roussel,G., Neskovic,N.M., Trifilieff,E., Artault,J.C. and Nussbaum,J.L. (1987) *J. Neurocytology*, **16**, 195–204.
Rushlow,C., Frasch,M., Doyle,H. and Levine,M. (1987a) *Nature*, **330**, 583–586.
Rushlow,C., Doyle,H., Hoey,T. and Levine,M. (1987b) *Genes Devel.*, **1**, 1268–1279.
Rüther,V. and Müller-Hill,G. (1983) *EMBO J.*, **2**, 1791–1794.
Sanger,F., Nühlen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci,. USA*, **74**, 5463–5463.
St Johnston,R.D. and Gelbart,W.M. (1987) *EMBO J.*, **9**, 2785–2791.
Simpson,P. (1983) *Genetics*, **105**, 615–632.
Steward,R. (1987) *Science*, **238**, 692–694.
Technau,G.M. and Campos-Ortega,J.A. (1986) *Roux's Arch. Dev. Biol.*, **194**, 196–212.
Thisse,B., El Messal,M. and Perrin-Schmitt,F. (1987a) *Nucleic Acids Res.*, **8**, 3439–3453.
Thisse,B., Stoetzel,C., El Messal,M. and Perrin-Schmitt,F. (1987b) *Genes Devel.*, **1**, 709–715.
Wharton,K.A., Johansen,K.M., Xu,T. and Artavanis-Tsakonas,S. (1985) *Cell*, **43**, 565–581.
Wirz,J., Fessler,L.I. and Gehring,W.J. (1986) *EMBO J.*, **12**, 3327–3334.
Yanisch-Perron,C., Wieira,J. and Messing,J. (1985) *Gene*, **33**,103–110.
Zusman,S.B. and Wieschaus,E.F. (1985) *Dev. Biol.*, **111**, 359–371.

Received on February 8, 1988; revised on April 25, 1988