The EMBO Journal vol.8 no.10 pp.3095-3103, 1989

The *Drosophila hairy* protein acts in both segmentation and bristle patterning and shows homology to N-*myc*

Christine A.Rushlow[1], Aileen Hogan, Sheena M.Pinchin, Kathy M.Howe, Michael Lardelli and David Ish-Horowicz

ICRF Developmental Biology Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[1]Present address: 743A Fairchild, Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Communicated by D.Ish-Horowicz

The *Drosophila* segmentation gene, *hairy* (*h*), acts to regulate embryonic segmentation and bristle pattern. We present the DNA sequence of the *h* gene and of *h* cDNAs, thereby deducing the organization of the *h* transcripts. The *h* gene encodes a 337 amino acid protein that acts in both embryonic segmentation and adult bristle patterning. The *h* protein includes a domain that shows extensive similarity to a domain of the proto-oncogene N-*myc* that may be involved in DNA binding and/or protein dimerization. We discuss mechanisms of *h* action as a transcriptional regulator.

Key words: *hairy*/*Drosophila*/segmentation/*myc*/pattern formation

Introduction

The *Drosophila* pair-rule segmentation gene, *hairy* (*h*), is required during two distinct developmental steps: embryonic segmentation and the establishment of adult bristle pattern during larval/pupal stages (Nüsslein-Volhard and Wieschaus, 1980; Ingham *et al.*, 1985a,b). Its striped expression during blastoderm cellularization serves to establish the reiterated (metameric) pattern of parasegmental units that represent the basic embryonic body-plan (reviewed in Akam, 1987; Ingham, 1988). *h* is also required during larval/pupal development to suppress ectopic bristle production on various adult cuticular structures including the wing and the notum.

Genetic analysis has shown that individual *h* mutations affect segmentation and bristle patterning unequally (Ingham *et al.*, 1985b). Homozygous viable *h* alleles (including h^1 and h^2) have strong bristle phenotypes but show weak or no segmentation defects. Several lethal *h* alleles with strong segmentation phenotypes have only weak effects on bristle pattern. Genetic analysis alone cannot distinguish the basis for these observations: *h* could encode variant products that are differentially utilized during development. Alternatively, *h* mutations could affect regulatory regions that are required for one or other *h* function.

The *h* locus has been cloned and its transcription during early embryonic development has been analysed (Ish-Horowicz *et al.*, 1985; Ingham *et al.*, 1985a). *h* expression is first detected over all or most of the syncitial blastoderm. After the final (13th) syncitial cleavage, the pattern evolves into seven segment-wide stripes and an antero-dorsal patch in the head. *h* mutations cause loss of alternate metameres leading to embryos with half the wild-type number of segments. The deleted structures roughly correspond to sites of *h* expression, suggesting a localized requirement for *h* in establishing segmentation.

Analysis of segmentation gene expression in *h* mutant embryos indicates that *h* regulates the expression of other segmentation genes (Carroll and Scott, 1986, Howard and Ingham, 1986; Ish-Horowicz and Pinchin, 1987). Various lines of evidence indicate that *h* controls the expression of the pair-rule gene *fushi tarazu* (*ftz*). In *h* embryos, the *ftz* stripes are broader and fail to resolve completely, suggesting that *h* negatively regulates *ftz* expression. Strong evidence that this interaction is direct comes from experiments in which *h* is expressed ectopically from a heat shock promoter. Induction of an *hsp70-h* (*HSH*) fusion gene rapidly and irreversibly represses *ftz* transcription. In contrast, *h* patterning appears to be independent of *ftz* activity.

The *h* gene is also required at about the time of pupariation, when bristle primordia are being specified in the imaginal discs. Adult *h* mutant flies develop ectopic bristles (microchaetes) in several tissues including wings, scutellum and pleura, indicating that *h* acts to repress microchaete development (Ingham *et al.*, 1985b). *h* appears to be cell-autonomous—clones of *h* cells develop extra bristles, even in an otherwise h^+ background—again indicating a localized mechanism of action (Garcia-Bellido and Merriam, 1971; Ingham *et al.*, 1985b). Gene dosage experiments have suggested that *h* might act to repress *achaete* (*ac*), a gene within the achaete-scute complex (AS-C) that is required for microchaete development (Falk, 1963; Botas *et al.*, 1982; Moscoso del Prado and Garcia-Bellido, 1984).

In this paper, we report the DNA sequence from the chromosomal *h* locus and of several cDNAs from which we deduce a putative *h* protein sequence. We present evidence that both *h* functions are mediated by the same protein product, and support a role for *h* as a transcriptional regulator by describing protein homology with a region of the N-*myc* proto-oncogene. By P-element mediated gene-transfer, we demonstrate rescue of the *h* segmentation function with cloned *h* DNA. We discuss the timing and possible mechanisms of *h* function during bristle patterning.

Results

Rescue of h segmentation mutations with the cloned h gene

The cloning and characterization of the *h* locus has been previously described (Ish-Horowicz *et al.*, 1985). Figure 1 shows a restriction map of the *h* region spanning ~28 kb. The *h* gene encodes a major transcript (α_1) of ~2.3 kb, and a minor, slightly smaller transcript (α_2) which is much less abundant at blastoderm. The embryonic time-course and spatial distribution of the α_1 transcript is consistent with that expected for a pair-rule segmentation gene. Moreover, we

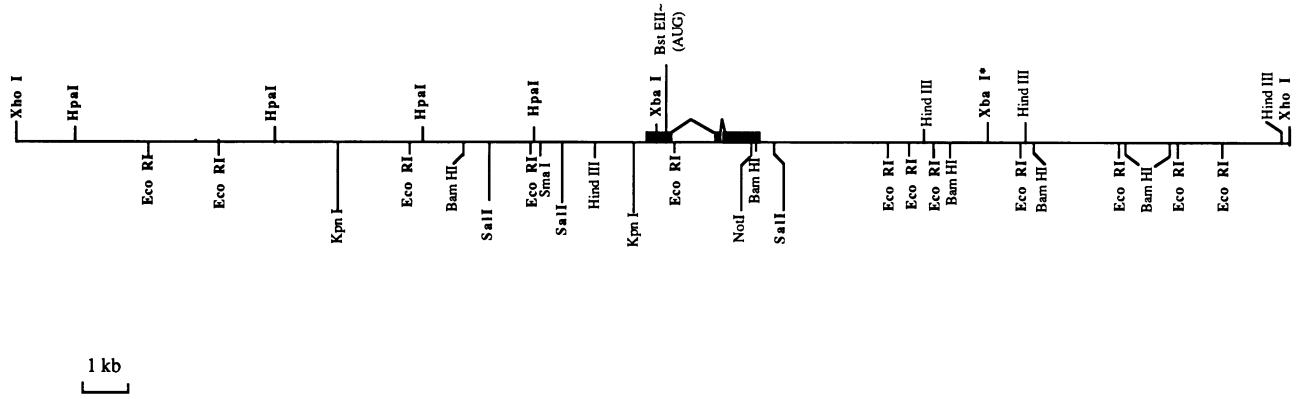


Fig. 1. A restriction map of the *h* region, the broad line indicating the *h* transcript structure. Distant sites for *Sal*I, *Sma*I, *Kpn*I, *Hpa*I and *Bst*EII have not been mapped. The downstream *Xba*I site used in later rescue experiments is asterisked.

have mapped several *h* mutations to the cloned region (Ish-Horowicz *et al.*, 1985). Nevertheless, there are now several examples of related gene-pairs that show similar expression patterns (Bopp *et al.*, 1986; Coleman *et al.*, 1987; Rushlow *et al.*, 1987). We therefore reintroduced the cloned *h* gene into flies by P-element transposition, and showed that it rescues *h* segmentation defects.

The *h* transcribed region lies within a 28 kb *Xho*I fragment (see Figure 1) that includes ~14 kb of DNA upstream of the α_1 promoter. This upstream region includes the breakpoints of several *h* mutations that affect the patterns of *h* expression (Ish-Horowicz *et al.*, 1985; Howard *et al.*, 1988; K.L.Hooper, S.M.Parkhurst and D.Ish-Horowicz, submitted). We generated a library of *Xho*I fragments cloned into a modified cosmid vector encoding G418-resistance that packages fragments between 25 and 38 kb (see Materials and methods).

From 2000 such cosmids, we recovered one *h* cosmid (*cosh*) that includes the entire 28 kb fragment. This was used to transform flies to G418-resistance and three transformed lines were recovered, two mapping to chromosome 2 and one to the X chromosome. Southern analysis showed that each results from chromosomal integration of an apparently unrearranged copy of the *cosh* insert (data not shown).

We showed that *cosh* includes all sequences necessary for embryonic function by rescuing the *h* segmentation phenotype. Two of the three lines are able to rescue *h* segmentation defects and give rise to viable adults bearing two lethal *h* chromosomes. The *cosh* insertion in the other line may be associated with a position effect or with a rearrangement not apparent on a Southern blot. Rescued flies eclose and show normal segmentation, but all show a severe *h* bristle phenotype (Figure 2). The phenotype is more extreme than that of h^1 flies and comparable with that associated with h^{R47}, a rearrangement broken ~15 kb upstream of the α_1 promoter that also shows a strong bristle phenotype but retains full segmentation *h* activity (Howard *et al.*, 1988). Thus, sequences upstream of −15 kb are necessary for the late *h* function.

We have shown that 5 kb of 3′ sequence suffice for *h* function by introducing a construct lacking further downstream sequences. Flies containing a 23 kb insert running from the upstream *Xho*I site to the downstream *Xba*I site (Figure 1) also rescue the segmentation phenotype of *h* mutant embryos (data not shown).

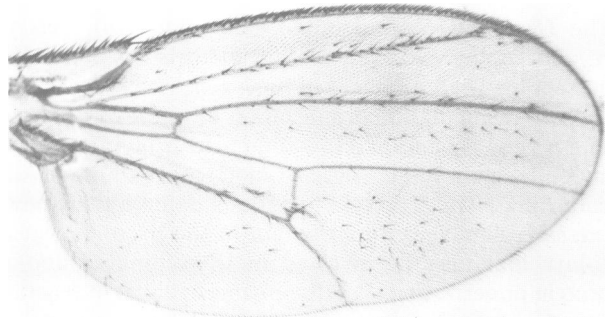


Fig. 2. Ectopic bristle on the wing of a *cosh*;h^{5H07}/h^{IL79K} fly containing an insert at 58F.

h DNA sequence

The DNA sequence of a 10 kb region surrounding the transcribed *h* region was determined by dideoxynucleotide sequencing (see Materials and methods). Several *h* cDNAs were isolated, using probe fragments from this region to screen a phage λgt10 cDNA library prepared from 0−3 h wild-type embryos (Poole *et al.*, 1985). These clones presumably derive from the α_1 transcript which represents at least 90% of the embryonic *h* RNA (Ish-Horowicz *et al.*, 1985). Their DNA sequences were also determined and compared with the genomic sequence in order to define transcript structure.

Figure 3 shows the nucleotide sequence of ~10 kb of genomic DNA which includes the cDNA plus the regions that immediately flank the 5′ and 3′ ends. In addition to the sequence data, results of primer extension and S1 nuclease analyses (see below) indicate that the *h* transcriptional unit is 2.3 kb in length. One of the cDNA clones (D2E) terminates with a stretch of adenosines that presumably reflects the polyadenylated mRNA. Three polymorphisms were found between the cDNA and genomic sequences, only one of which lies in the protein-coding region (see below and legend to Figure 3).

Comparison of the genomic and cDNA sequences reveals that the *h* transcription unit is interrupted by two introns (Figures 1 and 3) with conventional donor and acceptor splice sites (Mount, 1984). The first ATG codon is located at position 492 downstream from the transcript start site

(Figure 3; see below for numbering system) and begins an open translational reading frame of 337 amino acid residues. However, a second ATG codon lying a further nine amino acids downstream could encode a slightly smaller protein; in each case, the reading frame is the same. Analysis of the sequences that flank the ATG codons reveals that the second ATG (CAACATG) fits the *Drosophila* consensus sequence (C/A AA A/C ATG (Cavener, 1987) better than does the first (CGAAATG). Eight-three per cent of *Drosophila* mRNAs have A in the second position, and 13% have G. The first ATG in the *h* mRNA, which lies in a weaker context, could be frequently bypassed, causing translation of a shorter *h* protein (see Kozak, 1986). However, results using the *HSH* fusion gene argue that the full-length product is likely to be the active protein (Ish-Horowicz and Pinchin, 1987 and below).

The *h* mRNA has an unusually long 5′-untranslated leader of ~500 bases. Such leaders often have ATG codons upstream of the open reading frame that cause false starts and therefore inefficient translation of the proper coding region (Hunt, 1985). There are no such ATGs in the *h* untranslated leader; however, it may include sequences that reduce translational efficiency by increasing secondary structure (Stanton *et al.*, 1986).

The open translational reading frame is followed by an untranslated trailer of 833 bases. About 25 bases upstream of the polyadenylation site lies the consensus signal AATAAA (reviewed in Birnstiel *et al.*, 1985). A CA dinucleotide lies 23 residues downstream of the AATAAA, representing the processing site at which adenosine residues are added to the transcript (Moore and Sharp, 1984). It is followed by a GT-rich region and oligo-T stretches. We have not examined the site of polyadenylation further.

Within ~6 kb upstream of the transcription start site should lie sequences required for *h* expression in the antero-dorsal region ('stripe' 0) and stripe 1 (Howard *et al.*, 1988). Such expression must lie directly or indirectly under the control of the *bicoid* (*bcd*) gene that regulates anterior patterning (Frohnhöfer and Nüsslein-Volhard, 1986; Driever and Nüsslein-Volhard, 1988). Indeed, we have studied *h* expression in mutations of maternal and zygotic segmentation genes that might be expected to affect *h* pattern. *bcd* was the only gene tested that eliminated *h* stripe 1 (K.L.Hooper *et al.*, submitted). We have searched *h* upstream sequences for the *bcd* binding sites shown to be important in regulating *hunchback* expression (Driever and Nüsslein-Volhard, 1989). TCTAATCC at positions −4438 to −4430 shows 8/9 matches to the TCTAATCCC consensus. This region lies outside the domain required for stripe 0 expression, but within a region that plays a role in controlling stripe 1 (Howard *et al.*, 1988; K.L.Hooper *et al.*, submitted; M.Pankratz, personal communication; G.Riddihough, personal communication). The presence of a putative *bcd* binding site suggests that *bcd* may act directly in defining the stripe 1 domain. We have not searched for other consensus homeobox binding sites as they are too degenerate to indicate whether they might be functional (e.g. see Odenwald *et al.*, 1989).

Structure of the 5′ end

Northern analysis using truncated upstream probes (not shown) suggests that the two *h*-specific transcripts α_1 and α_2 (Ish-Horowicz *et al.*, 1985) differ in the extent of their 5′ ends. Reverse transcription/primer-extension and S1 nuclease protection analyses were used to map the 5′ ends of these RNAs using polyadenylated RNA from 2−4 h embryos, first instar larvae and adults. Reverse transcription of a synthetic 24-base primer, complementary to the region 177−201 bases upstream of the initiation ATG (Figure 3), yields bands of 119 and 315 bases, corresponding to the two *h* transcripts (Figure 4). The ratio of the two extended products correlated well with the ratio of α_1 and α_2 transcripts in the Northern analysis. This places the transcription start points of α_1 at a T (= position +1; Figure 3); the α_2 transcript starts at +197. The difference accounts for the size difference between α_1 and α_2 estimated from the Northern analysis. These sites of transcription initiation are also consistent with nuclease protection of *h* transcripts (not shown) and *in vitro* transcription of *h* DNA (J.Topol, personal communication). Using probes that retain differing degrees of 5′ sequence, we can show that an upstream probe that extends to the start of the open translational reading frame (*Bst*EII site) detects both transcripts, whilst a shorter probe (316 bp shorter, *Pst*I site) detects only α_1 (not shown).

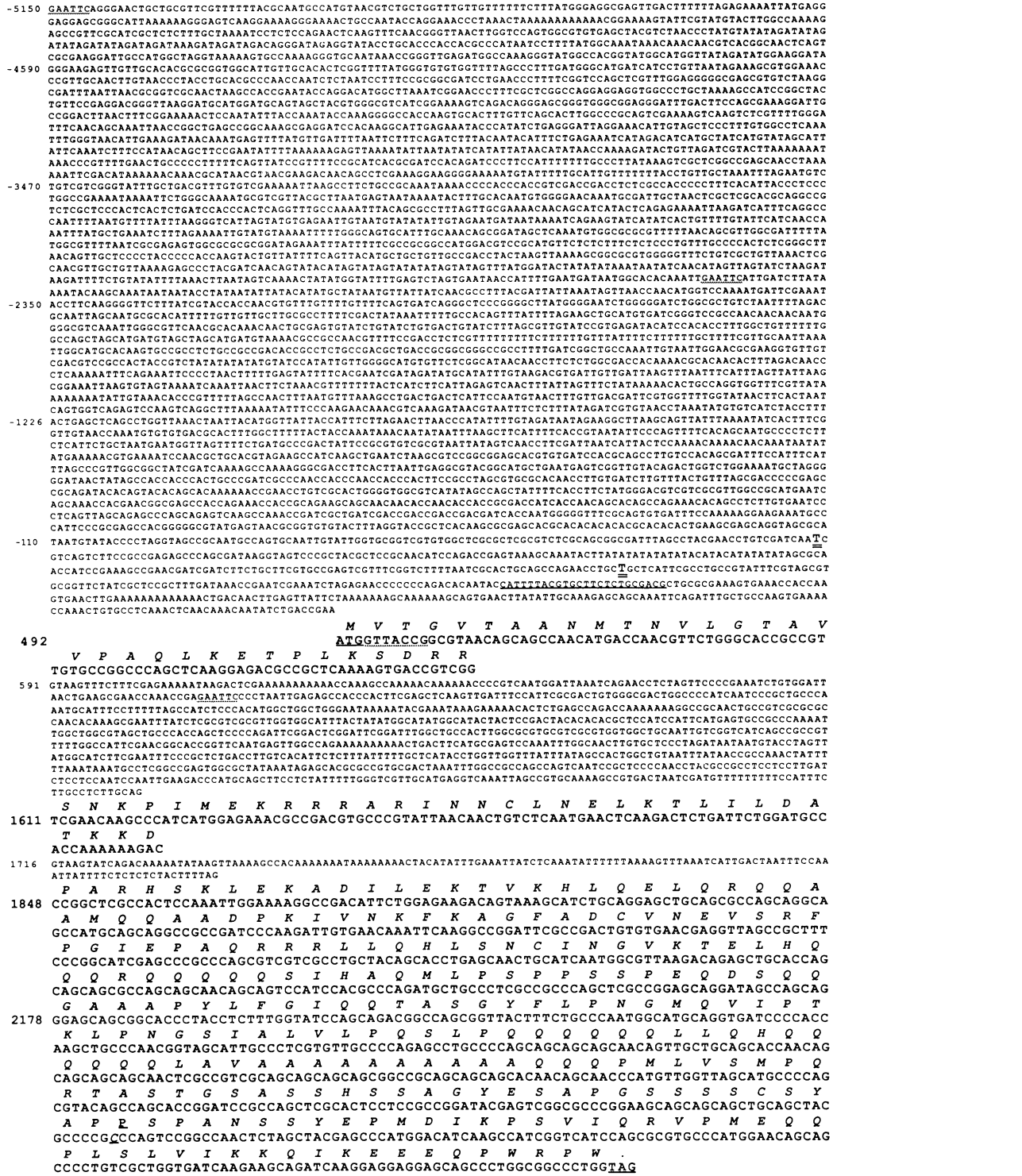
It is not clear which nearby sequences are involved in the initiation of transcription. Neither α_1 nor α_2 have putative TATA sequences ~30 bp upstream of their start sites. Nevertheless, each *h* transcript appears to start at a unique site, unlike some other TATA-less genes such as *engrailed* (Dierks *et al.*, 1983; Soeller *et al.*, 1988). The *h* transcripts also lack the upstream GC-rich regions that are found in the promoters of many housekeeping genes lacking TATA sequences (Dynan, 1986). There are two CAAT sequences upstream of α_1 at positions −76 and −87 but homology to the consensus GGCCAATCT sequence is poor (Benoist *et al.*, 1980).

The h protein

The amino acid sequence of the *h* open reading frame is shown in Figure 5A. One of the sequence polymorphisms between chromosomal and cDNA leads to an altered amino acid: a T to C change changes amino acid 292 from a Pro (codon CCC; chromosomal sequence) to a Ser (TCC; cDNA) (Figures 3 and 5A). It seems unlikely that this difference is functionally significant as this region of the *h* protein is already Pro/Ser rich. *h* appears to lack the DNA-binding motifs present in other segmentation genes that are thought to regulate gene expression, such as the homeo domain or the metal-binding finger domains (Laughon and Scott, 1984; McGinnis *et al.*, 1984; Vincent *et al.*, 1985; Rosenberg *et al.*, 1986; Tautz *et al.*, 1987).

We and others have made antibodies to the *h* protein based on the above sequence. This shows that the *h* protein is indeed localized in the nucleus, as expected for a transcriptional regulator (Carroll *et al.*, 1988; K.L.Hooper *et al.*, submitted). The signals for nuclear entry may lie in the N-terminal half which includes 14% Lys + Arg, clusters of which are implicated in nuclear localization (reviewed in Dingwall and Laskey, 1986). There are three such regions near the N terminus of the *h* protein that could specify nuclear entry.

h includes five stretches of repeated CAG-like sequences (*opa*; M-repeat) that are common to many other developmentally regulated genes (McGinnis *et al.*, 1984; Wharton *et al.*, 1985; Frigerio *et al.*, 1986). The *h* protein has four



short Gln stretches, one longer stretch of Ala (19 residues), and a Ser-rich sequence. The functional significance of such repeats is unclear though they may contribute to protein flexibility or instability.

As might be expected for a nuclear transcriptional regulator, the *h* protein shows no sign of the hydrophobic domain characteristic of trans-membrane or secreted proteins (Perlman and Halvorson, 1983), nor are there obvious nucleotide-binding or phosphorylation sites (Walker *et al.*, 1982; McCormick *et al.*, 1985). *h* contains three putative glycosylation sites at Asn positions 9, 209 and 296 (Marshall, 1972), although the presence of the necessary sequence does not guarantee that glycosylation occurs.

The carboxy-half of the protein is rich in Pro (28%) and Ser (30%), Gln and Ala as mentioned above, and probably folds into a globular structure. The sulphur content of the protein is low; there are few Met residues (3%) and even fewer Cys residues (1%).

The h protein includes a putative DNA-binding/dimerization motif found in N-myc and other transcription factors

We compared the *h* protein sequence to the published databases, looking for structural similarities to previously characterized proteins. Although most search programs failed to uncover any similarities, the FASTA program (Pearson and Lippman, 1988) indicated that *h* has a structural homology to the N-*myc* proto-oncogene. Figure 5B shows that the 66 amino acid domain of *h* from positions 30 to 95 is 26% identical to a region of N-*myc*. A 'sequence jumbling' test indicates that the degree of homology is likely to be significant (4.95 standard deviations from the average) (see Doolittle, 1986).

This domain of N-*myc* is homologous to several other *Drosophila* proteins, including AS−C transcripts, *daughterless* (*da*), *twist* (*twi*) and the *Enhancer of split complex* (Villares and Cabrera, 1987; Alonso and Cabrera, 1988; Caudy *et al.*, 1988; Klämbt *et al.*, 1989; Murre *et al.*, 1989). It has recently been proposed that this domain functions in DNA binding and protein dimerization by forming two amphipathic helices (Murre *et al.*, 1989). The *h* protein fulfils the structural criteria for a similar domain. Figure 5B shows that there are only four departures from the conserved amino acids defined by Murre *et al.* (1989). Lys-41 represents a conservative substitution of an otherwise conserved Arg. In addition, Thr-81 replaces a conserved Ala, and His-84 replaces a conserved Val/Tyr. Pro-37 in *h* replaces a conserved Asn/Thr; however, this lies outside the proposed helix-forming regions. There is a lack of proline in the putative helix-forming regions although *h* is otherwise very proline-rich, and a Pro occurs between the helices.

The amino acids in the proposed helix domains are organized so that hydrophobic amino acids would occupy one helix face and charged residues be on the other. Figure 5C shows the two putative *h* helices arranged as amphipathic wheels with two hydrophobic faces. This structural conservation supports *h* acting as a transcriptional regulator (see Discussion).

The domain of *h/myc* protein homology does not extend

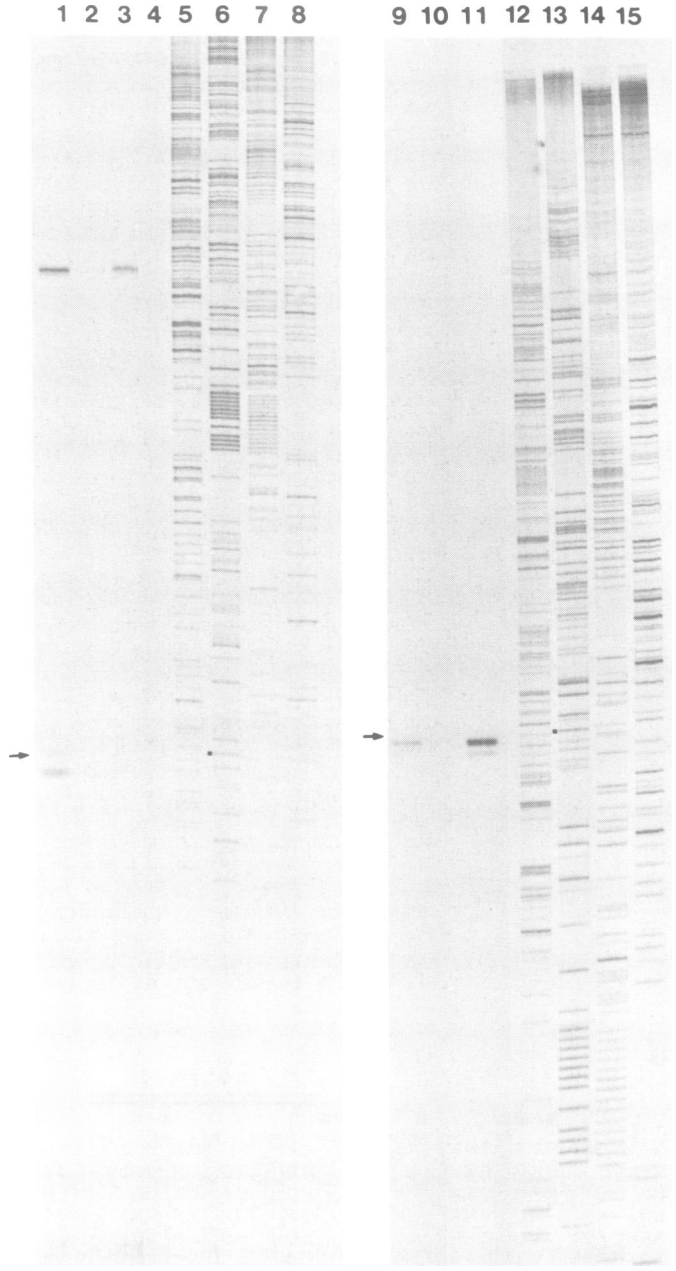


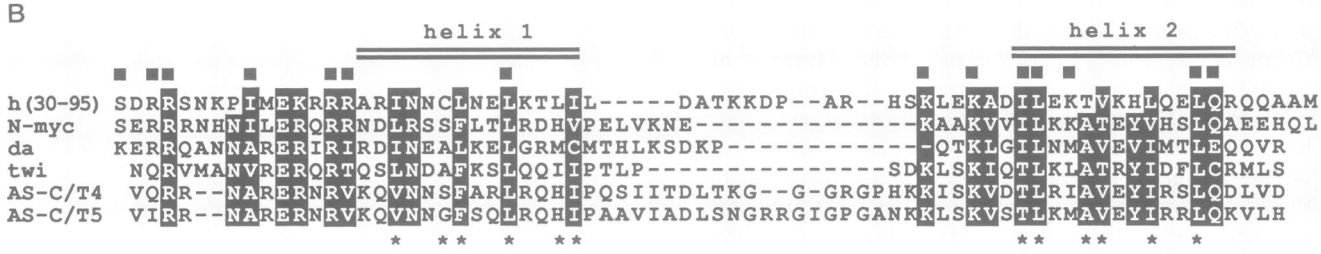
Fig. 4. Primer extension of *h* transcripts. Polyadenylated RNA (5 μg) was extended with reverse-transcriptase primed by a synthetic oligonucleotide (see text and Figure 3). Sizes of the elongated products were determined by comparison with sequence ladders (**lanes 5−8** and **12−15**). RNA samples are: **lanes 1** and **11**, adult RNA; **lanes 2** and **10**, first-instar larval RNA; **lanes 3** and **9**, 2−4 h embryonic RNA. The reduced signal in lanes 2 and 10 is due to partial degradation. The full-length extension products are 1 bp longer than the strongest bands due to premature termination caused by the methylated terminal base.

Fig. 3. Aligned chromosomal and cDNA sequences of the *h* region, indicating protein coding regions (single-letter amino acid code). The sites of transcription initiation are double-underlined, as is the site of 3′-polyadenylation. The initiator ATG and terminator TAG are single-underlined, as are the regions complementary to the 24 bp primer used for reverse-transcriptase mapping of the *h* 5′ ends, and the AATAAA polyadenylation signal. The *Eco*RI and *Bst*EII sites are dotted-underlined. The following (italicized-underlined) polymorphisms were found: (genomic first) C/T at base 2522 A/G at base 3010; T/C at base 3083.

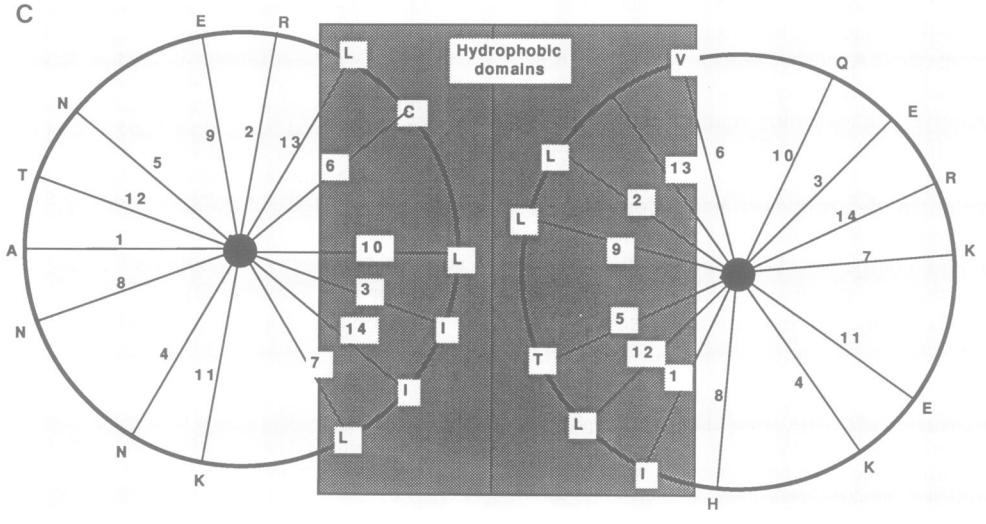
A

```
ATG GTT ACC GGC GTA ACA GCA GCC AAC ATG ACC AAC GTT CTG GGC ACC GCC GTT
MET Val Thr Gly Val Thr Ala Ala Asn MET Thr Asn Val Leu Gly Thr Ala Val

GTG CCG GCC CAG CTC AAG GAG ACG CCG CTC AAA AGT GAC CGT CGG TCG AAC AAG
Val Pro Ala Gln Leu Lys Glu Thr Pro Leu Lys Ser Asp Arg Arg Ser Asn Lys 36

CCC ATC ATG GAG AAA CGC CGA CGT GCC CGT ATT AAC AAC TGT CTC AAT GAA CTC
Pro Ile MET Glu Lys Arg Arg Arg Ala Arg Ile Asn Asn Cys Leu Asn Glu Leu

AAG ACT CTG ATT CTG GAT GCC ACC AAA AAA GAC CCG GCT CGC CAC TCC AAA TTG
Lys Thr Leu Ile Leu Asp Ala Thr Lys Lys Asp Pro Ala Arg His Ser Lys Leu 72

GAA AAG GCC GAC ATT CTG GAG AAG ACA GTA AAG CAT CTG CAG GAG CTG CAG CGC
Glu Lys Ala Asp Ile Leu Glu Lys Thr Val Lys His Leu Gln Glu Leu Gln Arg

CAG CAG GCA GCC ATG CAG CAG GCC GCC GAT CCC AAG ATT GTG AAC AAA TTC AAG
Gln Gln Ala Ala MET Gln Gln Ala Ala Asp Pro Lys Ile Val Asn Lys Phe Lys 108

GCC GGA TTC GCC GAC TGT GTG AAC GAG GTT AGC CGC TTT CCC GGC ATC GAG CCC
Ala Gly Phe Ala Asp Cys Val Asn Glu Val Ser Arg Phe Pro Gly Ile Glu Pro

GCC CAG CGT CGT CGC CTG CTA CAG CAC CTG AGC AAC TGC ATC AAT GGC GTT AAG
Ala Gln Arg Arg Arg Leu Leu Gln His Leu Ser Asn Cys Ile Asn Gly Val Lys 144

ACA GAG CTG CAC CAG CAG CAG CGC CAG CAG CAA CAG CAG TCC ATC CAC GCC CAG
Thr Glu Leu His Gln Gln Gln Arg Gln Gln Gln Gln Gln Ser Ile His Ala Gln

ATG CTG CCC TCG CCG CCC AGC TCG CCG GAG CAG GAT AGC CAG CAG GGA GCA GCG
MET Leu Pro Ser Pro Pro Ser Ser Pro Glu Gln Asp Ser Gln Gln Gly Ala Ala 180

GCA CCC TAC CTC TTT GGT ATC CAG CAG ACG GCC AGC GGT TAC TTT CTG CCC AAT
Ala Pro Tyr Leu Phe Gly Ile Gln Gln Thr Ala Ser Gly Tyr Phe Leu Pro Asn

GGC ATG CAG GTG ATC CCC ACC AAG CTG CCC AAC GGT AGC ATT GCC CTC GTG TTG
Gly MET Gln Val Ile Pro Thr Lys Leu Pro Asn Gly Ser Ile Ala Leu Val Leu 216

CCC CAG AGC CTG CCC CAG CAG CAG CAG CAA CAG TTG CTG CAG CAC CAA CAG CAG
Pro Gln Ser Leu Pro Gln Gln Gln Gln Gln Gln Leu Leu Gln His Gln Gln Gln

CAG CAG CAA CTC GCC GTC GCA GCA GCA GCA GCG GCC GCA GCA GCA GCA CAA CAG
Gln Gln Gln Leu Ala Val Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gln Gln 252

CAA CCC ATG TTG GTT AGC ATG CCC CAG CGT ACA GCC AGC ACC GGA TCC GCC AGC
Gln Pro MET Leu Val Ser MET Pro Gln Arg Thr Ala Ser Thr Gly Ser Ala Ser

TCG CAC TCC TCC GCC GGA TAC GAG TCG GCG CCC GGA AGC AGC AGC AGC TGC AGC
Ser His Ser Ser Ala Gly Tyr Glu Ser Ala Pro Gly Ser Ser Ser Ser Cys Ser 288

          C
TAC GCC CCG TCC AGT CCG GCC AAC TCT AGC TAC GAG CCC ATG GAC ATC AAG CCA
Tyr Ala Pro Ser Ser Pro Ala Asn Ser Ser Tyr Glu Pro MET Asp Ile Lys Pro
          Pro

TCG GTC ATC CAG CGC GTG CCC ATG GAA CAG CAG CCC CTG TCG CTG GTG ATC AAG
Ser Val Ile Gln Arg Val Pro MET Glu Gln Gln Pro Leu Ser Leu Val Ile Lys 324

AAG CAG ATC AAG GAG GAG GAG CAG CCC TGG CGG CCC TGG TAG
Lys Gln Ile Lys Glu Glu Glu Gln Pro Trp Arg Pro Trp  .  337
```

B



C



into the adjacent C-terminal domain that includes a series of repeated leucines in *myc* that play a role in protein-protein interactions (Landschulz *et al.*, 1988; Kouzarides and Ziff, 1988). Neither *h* nor other segmentation genes appear to have such a 'leucine zipper'.

An hsp70–h fusion rescues the h bristle phenotype

Although the DNA sequence provides no evidence for different *h* proteins, it does not completely exclude the possibility that the *h* segmentation and bristle functions require different proteins. We have previously shown that ectopic *h* induction at blastoderm disrupts segmentation (Ish-Horowicz and Pinchin, 1987). We therefore showed that the same protein can rescue the *h* bristle phenotype.

The *HSH* gene has the *h* cDNA coding region under the control of the *hsp70* heat shock promoter, fused at the first *h* ATG (Ish-Horowicz and Pinchin, 1987). As this translational start-site now lies downstream of IgG-derived consensus translational initiation sequences, *HSH* is likely to encode full-length *h* protein rather than the variant lacking nine N-terminal amino acids (see above). Exposure of *HSH;h²* larvae/pupae to a 1-h heat shock at 36°C leads to complete suppression of extra bristles on the h^2 L2 wing-vein (Table I). The rescue is due to ectopic *h* protein as the bristle phenotype is not rescued in heat shocked h^2 late larvae/early pupae (data not shown).

The efficiency of rescue depends critically on the developmental stage at heat shock. Initial experiments narrowed the temperature-sensitive period (TSP) down to day 1 of pupal development (see Materials and methods). Experiments with more accurately staged pupae shows that heat shock at 6–11 h post-pupariation leads to complete phenotypic rescue of the *h* phenotype in up to 80% of flies (Table I). The stronger h^1 allele is similarly rescued (Table I), again only in an *HSH* background. (The slight difference between the two TSPs is unlikely to be significant.) It is noticeable that *HSH*-induced ectopic *h* does not inhibit normal bristle development; neither macro- nor micro-chaete pattern appear altered in heat-shocked *HSH* flies.

Ectopic *h* may also regulate expression of the *ac* gene of the AS–C. *Hairy-wing* (*Hw*) mutations are dominant alleles at the AS–C locus in which extra bristles develop on the wing-veins. Hw^1 is associated with the insertion of a *gypsy* transposon near the *ac* region of AS–C that causes apparent derepression of the *ac* gene in certain tissues (Campuzano *et al.*, 1986). Ectopic bristle formation can be suppressed in heat shocked 6–9 h *HSH;Hw¹/+* pupae, indicating that the *ac* overexpression is sensitive to *h* (not shown). This suggests that *ac* might indeed be a direct target of *h* repression.

h transcription in late h mutations

In view of the pupal *HSH* TSP, we have re-examined post-embryonic *h* transcription in more detail (Figure 6). The α_1 transcript predominates at blastoderm and gastrulation. Thereafter, both *h* transcripts are expressed at equal levels except during the last 3rd larval instar, when the level of α_2 is 2–3 times that of α_1. This difference is only temporary as α_1 and α_2 become equal again in pupal RNA.

Table I. Pupal TSP for rescuing the *h* bristle phenotype

Time (h post-pupariation)	% complete rescue (total scored)	
	HSH;h¹	*HSH;h²*
5–6	0(67)	0(32)
6–7	39(43)	0(33)
7–8	52(29)	69(51)
8–9	40(10)	82(139)
9–10	0(15)	54(107
10–11	0(32)	12(69)
11–12	0(52)	0(62)

The genotypes used were: *HSH33; ru h¹ th st*, *HSH33; mwh jv h²* and controls lacking *HSH33*.

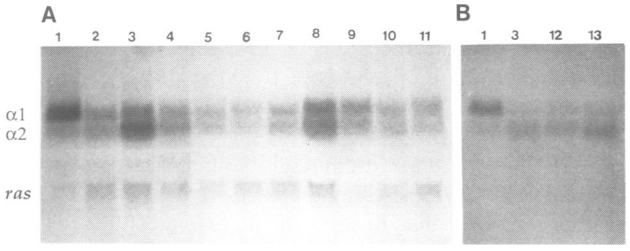


Fig. 6. Time-course of *h* transcription in wild-type, and pupal/larval transcription in *h* mutants. Levels of total RNA (~1 μg/lane), were visualized in (**A**) by rehybridizing the filter with a *ras* probe complementary to the 1.6-kb transcript of Dmras64B (Mozer *et al.*, 1985). (Note the h^2 chromosome carries a variant *ras* which does not produce a 1.6-kb transcript, but two transcripts, one of high and one of low mol. wt, not shown.) RNA levels in (**B**) were predetermined by visualizing ribosomal RNA by ethidium bromide staining. RNA samples were from wild-type individuals in lanes 1–8. **Lanes 1**, 2–4 h embryos; **2**, first larva instars; **3**, late third larval instars; **4**, pupae, day 1; **5**, pupae, day 2; **6**, pupae, day 3; **7**, pupae, day 4; **8**, adults; **9**, h^2 pupae, day 1; **10**, h^1 pupae, day 1; **11**, h^{R47}/h^{i22} pupae, day 1; **12**, h^1 late third larval instars; **13**, h^2 late third larval instars.

We have analysed late *h* transcription in several *h* mutations, but were unable to detect any quantitative or qualitative change that might relate to loss of late *h* function. Figure 6 shows that the levels of α_1 and α_2 transcription are unaltered in h^1 and h^2 larvae and pupae, as well as in RNA from h^{R47}/h^{i22} pupae. This suggests that the tissue specificity of *h* transcription is critical for late *h* function.

We have previously described a second class (β) of *h* transcripts that are present only in larval RNA (Ish-Horowicz *et al.*, 1985). However, their occurrence has proved irreproducible, and the hybridization signals are very susceptible to high-salt/RNAse treatment, unlike those from the α class that resist such digestion (K.R.Howard and C.A.Rushlow, unpublished observations). We now believe that the β signals are artefacts due to mismatch hybridization and the strength and sensitivity of RNA/RNA detection.

Discussion

In this paper, we have defined the structure of the *h* transcripts and of the *h* protein. This has allowed us to define and analyse the *h* protein sequence and show that the same

Fig. 5. (**A**) The *h* protein sequence. (**B**) Alignment of the *h* (30–95) domain sequence with homologous regions of *N-myc*, *da*, *twi*, and the T4 and T5 transcripts of AS–C. The black boxes indicate sequence homologies as described by Murre *et al.* (1989). The upper squares mark the sites of amino acid identity with N-*myc*; the lower asterisks show the hydrophobic amino acids within the amphipathic helices illustrated in Figure 5C. (**C**) The homologous *h* domain organized around helical 'wheels'.

protein acts in segmentation and bristle patterning. In particular, ectopic expression of the protein interferes with embryonic segmentation and rescues the *h* mutant bristle phenotype (Ish-Horowicz and Pinchin, 1987; this paper). Moreover, we and others have used the predicted sequence to make *h* fusion proteins and thereby generate antibodies that recognize authentic *h* protein in early embryos (Carroll *et al.*, 1988; K.L.Hooper *et al.*, submitted). The protein distribution corresponds to that expected from the pattern of *h* transcription and to the *h* cuticular phenotype.

Although *h* is clearly involved in regulating other segmentation genes, most particularly *ftz*, its mechanism of action is not yet clear. *h*'s nuclear localization suggests that it acts on *ftz* transcription, a view that is strongly supported by the structural homology to N-*myc*. Unfortunately, although *myc* proteins are implicated in transcriptional regulation, no clear mechanisms have yet emerged, nor has a specific recognition sequence for *myc* DNA binding yet been determined. However, a region including the conserved motif is required for DNA binding/dimerization in two proteins that regulate immunoglobulin gene transcription (Murre *et al.*, 1989). Although *h* acts to interfere with gene expression rather than to stimulate it, we expect that *h*'s role as a transcriptional regulator will involve this domain.

Two lines of evidence suggest that *h* might act on *ftz* expression through interactions with other transcription factors. First, relatively little *h* is needed to inhibit *ftz* expression: the *HSH* gene is rather poorly induced at blastoderm and little or no *h* protein is directly detectable (Ish-Horowicz and Pinchin, 1987). Thus, low levels of *h* protein inhibit *ftz* expression, arguing that it does not act stoichiometrically in antagonizing wild-type expression. Second, the difficulty of demonstrating direct DNA binding of N-*myc*, suggests an analogy with the c-*fos* proto-oncogene (with which *myc* proteins show structural analogies—Vogt *et al.*, 1987; Kouzarides and Ziff, 1988), which only shows sequence-specific binding when complexed with the *jun/AP-1* protein (Halazonetis *et al.*, 1988; Kouzarides and Ziff, 1988; Nakabeppu *et al.*, 1988; Sassoni-Corsi *et al.*, 1988).

Presumably, the *h* protein also affects bristle patterning by regulating transcription, an obvious target being the *ac* locus (Botas *et al.*, 1982). The presence of dimerization motifs in both *h* and AS—C proteins raises the possibility that *h* interferes with *ac* action, although ectopic *h* appears not to interfere with the development of most *ac*-dependent bristles (see Results).

It is clear that the two *h* functions are differentially regulated. The failure of *cosh* to rescue the adult *h* phenotype indicates that sequences >14 kb upstream of the promoter are needed for pupal *h* expression. Indeed, the h^{R47} mutation is broken even further upstream of the promoter and also fails to complement adult *h* mutations (Howard *et al.*, 1988). Not all of this extensive upstream region need be required for pupal expression. The h^{1} mutation is due to the insertion of a *gypsy* transposon 5 kb upstream of the *h* promoter (Holmgren, 1984; Ish-Horowicz *et al.*, 1985). However, its suppression by the *trans*-acting *suppressor of Hairy wing* (*su-Hw*) locus suggests that the loss of *h* activity is due to transcriptional interference from the *gypsy* element (Parkhurst and Corces, 1986), not the inactivation of a pupal control element. The tissue-specificity of h^{1} would be related to that of *gypsy*. Transformation of suitable *h* DNA fragments into flies will define the *cis*-requirements for pupal expression more accurately.

The *HSH* TSP shows that *h* activity is required between 6 and 12 h post-pupariation for bristle patterning (Table I). This contrasts with the results using somatic recombination to generate mosaic clones of *h* cells which suggested that *h* is needed at least 14 h earlier, before 8 h pre-pupariation (Garcia-Bellido and Merriam, 1971). This discrepancy may reflect the chain of events that follow X-irradiation. Thus, the affected cells must first divide to yield *h* homozygous clones. Thereafter, the persistance ('perdurance' — Garcia-Bellido and Merriam, 1971) of *h* gene products introduces a second delay. Finally, if *h* indeed acts via *ac*, sufficient *ac* gene product must accumulate to direct cells towards bristle development.

Alternatively, the two experiments may measure different functional aspects of *h*. The clonal analysis estimates when epidermally committed cells no longer need h^{+} activity to maintain this decision. The *HSH* experiments measure when sensory committed cells can be reversed towards epidermal development by *h* expression. Imaginal microchaetes arise from individual precursor cells that divide to yield the bristle and three accessory cells. Such precursor cells are first detectable 14—18 h post-puparium when they commence division (J.Posakony and V.Hartenstein, personal communication). The *HSH* TSP suggests that precursor cell commitment may precede division by 4—8 h.

HSH-induced *h* is able to suppress the dominant Hw^{1} allele of the AS—C. This result is consistent with previous proposals that *h* acts as a repressor of the *ac* sub-function (Botas *et al.*, 1982; Moscoso del Prado and Garcia-Bellido, 1984). In this case, the upstream *gypsy* transposable element modulates the sensitivity of the *ac* promoter to *h* protein. As the Hw^{1} phenotype is suppressible by *su-Hw* it is likely to be due to transcriptional interference (Parkhurst and Corces, 1986). Alternatively, *h* overexpression could inhibit the action of ectopic *ac* expression.

Although *HSH*-induced ectopic *h* suppresses the *h* bristle phenotype, it fails to inhibit normal bristle development. Either insufficient *h* is made to affect normal cells, or normal bristle cells are insensitive to *h*. The latter explanation would suggest that only specific cells require *h* expression to regulate bristle patterning, and is consistent with our inability to visualize altered levels of *h* transcription in *h* mutant larvae and pupae (Figure 6). The tissue distributions of *h* and *ac* expression during late larval/early pupal development will define whether *h* expression is widespread, or restricted to specific *h*-requiring cells.

Materials and methods

DNA sequencing

The *h* gene was subcloned as *Eco*RI fragments into pEMBL vectors (Dente *et al.*, 1983). Both DNA strands were sequenced by dideoxy methods, using nested deletions made by unidirectional *Exo*III digestion (Henikoff, 1985). *h* cDNAs were isolated from the 0—3 h libraries of Poole *et al.* (1985), subcloned into pEMBL vectors and sequenced similarly.

RNA analysis

Northern blots were conducted using single-stranded RNA probes under stringent conditions (0.3 M NaCl, 65% formamide, 10% Dextran, 60°C) (Howard, 1986). Primer-extension using reverse transcriptase and nuclease protection were performed essentially as previously described (Berk and Sharp, 1978; McKnight and Kingsbury, 1982).

Constructing cosh

cosPneo (Steller and Pirrotta, 1985) has two sites for *SalI* and will not package fragments as small as 28 kb. The cosmid vector *cosA* that allows packaging of smaller DNA fragments was made by cloning the 4.5 kb *XhoI* fragment D of Ad5 DNA into the *SalI* site that does not lie between the P element ends, leaving one *SalI* cloning site. *Drosophila* DNA was cut with *XhoI*, enriched for the *h* fragment on sucrose gradient and packaged into cosmids using *SalI*-cut *cosA*. *cosh* was the only such cosmid to incorporate the *h XhoI* fragment, and was transformed into flies by selecting for resistance to G418 (Steller and Pirrotta, 1985).

Rescuing the h bristle phenotype by HSH

The coarse TSP was determined by exposing *HSH33;h* larvae/pupae from 24 h egg-lays to 3 h at 36°C and scoring for lack of ectopic bristles on wing-vein L2. *HSH33* has a X-chromosomal pHSH3 insertion (Ish-Horowicz and Pinchin, 1987). TSPs were further defined by collecting accurately timed cohorts of pupae. *HSH;Hw*^{1} flies were tested similarly. Only flies showing complete lack of bristles on the L2 wing-vein (the most sensitive site for ectopic bristle formation) were considered as showing rescue. Although the TSP for complete rescue is between 6 and 12 h post-pupariation, partial allevation of the *h* phenotype is seen over a broader period (not shown).

Acknowledgements

We should like to acknowledge Phil Ingham, Ken Howard and Mike Levine and our colleagues in the laboratory for their advice and criticism during the course of this work and Jim Posakony for communicating unpublished data on bristle cell development. We should also like to thank Susan Parkhurst for making some of the *cosh* transformant lines, and Ian Goldsmith for synthesizing DNA oligonucleotide primers. We particularly thank Jonathan Rothberg and Dave Hartley whose initial analysis of the *h* sequence with FASTA revealed the homology to N-*myc*.

References

Akam,M. (1987) *Development*, **101**, 1–22.
Alonso,M.C. and Cabrera,C.V. (1988) *EMBO J.*, **7**, 2585–2591.
Benoist,G., O'Hare,K., Breathnach,R. and Chambon,P. (1980) *Nucleic Acids Res.*, **8**, 127–142.
Berk,A. and Sharp,P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1274–1278.
Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
Bopp,D., Burri,M,. Baumgartner,S., Frigerio,G. and Noll,M. (1986) *Cell*, **47**, 1033–1040.
Botas,J., Moscoso del Prado,J. and Garcia-Bellido,A. (1982) *EMBO J.*, **1**, 307–310.
Campuzano,S., Barcells,L., Villares,R., Carramolino,L., Garcia-Alonso,L. and Modollel,J. (1986) *Cell*, **44**, 303–312.
Carroll,S.B. and Scott,M.P. (1986) *Cell*, **45**, 113–226.
Carroll,S.B,. Laughon,A. and Thalley,B.S. (1988) *Genes Dev.*, **2**, 883–890.
Caudy,M., Vässin,H., Brand,M., Tuma,R., Jan,L.Y. and Jan,Y.N. (1988) *Cell*, **55**, 1061–1067.
Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
Coleman,K.G., Poole,S.J., Weir,M.P., Soeller,W.C. and Kornberg,T. (1987) *Genes Dev.*, **1**, 19–28.
Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
Dierks,P., van Ooyen,A., Cochran,M., Dobkin,C., Reiser,J. and Weissman,C. (1983) *Cell*, **32**, 695–706.
Dingwall,C. and Laskey,R.A. (1986) *Annu. Rev. Cell Biol.*, **2**, 367–390.
Doolittle,R.F. (1986) *Of URFs and ORFs: A Primer on How to Analyze Derived Amino Acid Sequences.* University Sciences Books, Mill Valley, CA.
Driever,W. and Nüsslein-Volhard,C. (1988) *Cell*, **54**, 83–93.
Driever,W. and Nüsslein-Volhard,C. (1989) *Nature*, **337**, 138–143.
Dynan,W.S. (1986) *Trends Genet.*, **2**, 196–197.
Falk,R. (1963) *Am. Nat.*, **97**, 129–132.
Frigerio,G., Burri,M., Bopp,D., Baumgartner,S. and Noll,M. (1986) *Cell*, **47**, 735–746.
Frohnhöfer,H.G. and Nüsslein-Volhard,C. (1986) *Nature*, **324**, 120–125.
Garcia-Bellido,A. and Merriam,J. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 2222–2226.
Halazonetis,T.D., Georgeopolous,K., Greenberg,M. and Leder,P. (1988) *Cell*, **55**, 917–924.
Henikoff,S. (1985) *Gene*, **28**, 351–359.
Holmgren,R. (1984) *EMBO J.*, **3**, 569–573.
Howard,K.R. (1986) Ph.D. thesis, University of London.
Howard,K.R. and Ingham,P.W. (1986) *Cell*, **44**, 949–957.
Howard,K.R., Ingham,P.W. and Rushlow,C.A. (1988) *Genes Dev.*, **2**, 1037–1046.
Hunt,T. (1985) *Nature*, **316**, 580–581.
Ingham,P.W. (1988) *Nature*, **335**, 25–34.
Ingham,P.W., Howard,K.R. and Ish-Horowicz,D. (1985a) *Nature*, **318**, 439–445.
Ingham,P.W., Pinchin,S.M., Howard,K.R. and Ish-Horowicz,D. (1985b) *Genetics*, **111**, 463–486.
Ish-Horowicz,D. and Pinchin,S.M. (1987) *Cell*, **51**, 405–415.
Ish-Horowicz,D., Howard,K.R., Pinchin,S.M. and Ingham,P.W. (1985) *Cold Spring Harbor Symp. Quant. Biol.*, **50**, 135–144.
Klämbt,C., Knust,E., Tietze,K. and Campos-Ortega,J.A. (1989) *EMBO J.*, **8**, 203–210.
Kouzarides,T. and Ziff,E. (1988) *Nature*, **336**, 646–651.
Kozak,M. (1986) *Cell*, **44**, 283–292.
Landschulz,W.H,. Johnston,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759–1764.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.
Lees,A.D. and Waddington,C.H. (1942) *Proc. R. Soc. B*, **131**, 87–110.
McCormick,F., Clark,B.F.C., La Cour,T.F., Kjeldgaard,M., Norskov-Lauritsen,L. and Nyborg,J. (1985) *Science*, **230**, 78–82.
McGinnis,W. Levine,M.S,. Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) *Nature*, **308**, 428–433.
McKnight,S.L. and Kingsbury,R. (1982) *Science*, **217**, 316–324.
Marshall,R.D. (1972) *Annu. Rev. Biochem.*, **41**, 673–702.
Moore,C.L. and Sharp,P.A. (1984) *Cell*, **36**, 581–591.
Moscoso del Prado,J. and Garcia-Bellido,A. (1984) *Roux's Arch. Dev. Biol.*, **193**, 246–251.
Mount,S. (1984) *Nucleic Acids Res.*, **10**, 459–472.
Mozer,B., Marlor,R., Parkhurst,S. and Corces,V. (1985) *Mol. Cell. Biol.*, **5**, 885–889.
Murre,C., McCaw,P.S. and Baltimore,D. (1989) *Cell*, **56**, 777–783.
Nakabeppu,Y., Ryder,K. and Nathans,D. (1988) *Cell*, **55**, 907–915.
Nüsslein-Volhard,C. and Wieschaus,E. (1980) *Nature*, **287**, 795–801.
Odenwald,W.F., Garbern,J., Arnheiter,H., Tournier-Lasserve,E. and Lazzarini,R.A. (1989) *Genes Dev.*, **3**, 158–172.
Parkhurst,S.M .and Corces,V.G. (1986) *Mol. Cell. Biol.*, **6**, 47–53.
Pearson,W.R. and Lippman,D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
Perlman,D. and Halvorson,H.O. (1983) *J. Mol. Biol.*, **167**, 391–409.
Poole,S.J,. Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.
Rosenberg,U.B, Schröder,C,. Preiss,A., Kienlin,A., Cote,S., Riede,I. and Jäckle,H. (1986) *Nature*, **319**, 336–339.
Rushlow,C.A., Doyle,H., Hoey,T. and Levine,M. (1987) *Genes Dev.*, **1**, 1268–1279.
Sassoni-Corsi,P., Ransome,L.J,. Lamph,W.W. and Verma,I.M. (1988) *Nature*, **336**, 629–695.
Soeller,W.C., Poole,S.J. and Sharp,P. (1988) *Genes Dev.*, **2**, 68–81.
Stanton,L.W., Schwab,M. and Bishop,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1772–1777.
Steller,H. and Pirrotta,V. (1985) *EMBO J.*, **4**, 167–171.
Tautz,D., Lehmann,R., Schnürch,H., Schuh,R., Seifert,E., Kienlin,A., Jones,K. and Jäckle,H. (1987) *Nature*, **327**, 383–389.
Villares,R. and Cabrera,C.V. (1987) *Cell*, **50**, 415–424.
Vincent,A., Colot,H.V. and Rosbach,M. (1985) *J. Mol. Biol.*, **186**, 149–166.
Vogt,P.K., Bos,T.J. and Doolittle,R.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3316–3319.
Walker,J.E., Saraste,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.
Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55–62.

Received on April 7, 1989; revised on June 27, 1989

Note added in proof

The *h* chromosomal and cDNA sequences have been assigned the accession numbers X15904 and X15905 respectively in the EMBL database. The manuscript referred to throughout as 'Hooper, Parkhurst and Ish-Horowicz, submitted' is now in press: Hooper,K.L., Parkhurst,S.M. and Ish-Horowicz,D. (1989). Spatial control of hairy protein expression during embryogenesis. *Development*, in press.