MOLECULAR AND CELLULAR BIOLOGY, June 1992, p. 2475–2483
0270-7306/92/062475-09$02.00/0

Vol. 12, No. 6

Point Mutations in the *Drosophila* hairy Gene Demonstrate In Vivo Requirements for Basic, Helix-Loop-Helix, and WRPW Domains

S. MARK WAINWRIGHT AND DAVID ISH-HOROWICZ*

Imperial Cancer Research Fund Developmental Biology Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, England

Received 4 November 1991/Accepted 25 February 1992

The *Drosophila* pair-rule gene, hairy (*h*), encodes a nuclear basic helix-loop-helix (bHLH) protein that regulates embryonic segmentation and adult bristle patterning. In both cases, the *h* protein behaves as a transcriptional repressor. In this study, we determined the molecular nature of 12 *h* alleles. One mutation maps within the HLH domain, consistent with *h* function requiring homodimerization or heterodimerization with other HLH proteins. A second mutation lies in the basic domain, suggesting that DNA binding is required for *h* activity. Several mutations show that the *h* C terminus, in particular the WRPW domain, is also required for *h* activity, perhaps by interacting with other proteins to mediate transcriptional repression. We show that the *h* protein in *Drosophila virilis* closely resembles that in *D. melanogaster* and includes completely conserved bHLH and WRPW domains.

A wide variety of developmental processes are regulated by the basic-helix-loop-helix (bHLH) class of transcription factors. bHLH proteins are characterized by two adjacent subdomains: the 13-amino-acid basic domain (b-domain) required for DNA binding, and the HLH domain comprising two amphipathic helices separated by a (presumed) nonhelical loop, functioning in protein dimerization (9, 19, 46, 50, 51, 71; Fig. 1).

At the molecular level, the best-characterized bHLH protein is MyoD, which controls myogenic cell lineage by activating transcription of muscle-specific genes, including *myoD* itself (20; reviewed in reference 72). In vitro evidence suggests that MyoD activates transcription as a heterodimer with a second bHLH protein, E12 (6, 51, 64). Although MyoD homodimers bind DNA, MyoD-E12 heterodimers bind target sites about 10-fold more strongly (6, 51). *myoD* mutations that prevent dimerization or DNA binding abolish the myogenic activity of *myoD* (19, 46).

In vivo evidence for interactions between bHLH proteins comes from genetic analysis of neural cell fate and sex determination in *Drosophila* spp. Development of the central and peripheral nervous systems is regulated by proneural bHLH proteins whose expression defines potential neural cell fate (11, 61; reviewed in reference 27). Genetic interactions between the daughterless (*da*) bHLH gene and the achaete-scute complex (AS-C) that encodes a family of bHLH proneural proteins (*T3*, *T4*, *T5*, and *T8*) are consistent with *da*/AS-C heterodimers acting as transcriptional activators of neural cell fates (13). Indeed, *da* forms heterodimers with AS-C *T3* that bind DNA in vitro and that activate transcription in yeast cells (10, 51). AS-C and *da* heterodimers are also involved in regulating *Drosophila* sex determination (reviewed in reference 29). Female development in *Drosophila* spp. depends on the X/A ratio and requires maternal *da* activity and zygotic expression of X-linked sisterless (*sis*) genes (15–17). Genetic analysis shows that the *sis-b* product is encoded by AS-C *T4* and acts with *da* to promote female development (23, 54, 66).

We previously reported that the *Drosophila* hairy (*h*) segmentation gene encodes a bHLH protein (62). Metameric (segmental) pattern in the *Drosophila* embryo is established via a cascade of transcriptional regulation that subdivides the embryo into successively more precise anteroposterior spatial domains (reviewed in references 33 and 34). *h* is a member of the pair-rule class of genes whose striped patterns of expression mark the prospective reiterated organization of the embryo (52). All pair-rule genes that have been characterized at the molecular level appear to encode transcriptional regulators.

h's role in segmentation is to regulate the expression of other pair-rule genes. The complex cross-regulatory network of segmentation genes makes it difficult to distinguish their direct and indirect interactions, but several lines of evidence argue that *h* acts directly as a transcriptional repressor of the fushi tarazu (*ftz*) gene. *ftz* stripes are broadened in *h* mutant embryos, as would be expected if *h* inhibits interstripe *ftz* expression (12, 32). Also, ectopic *h* expression at the late blastoderm stage leads to rapid and irreversible inhibition of *ftz* transcription (37).

h is also required for patterning adult sensory bristles (microchaetae), in which it also behaves as a negative regulator. *h* mutant flies are characterized by ectopic bristles in several regions, including the head, thorax, and wings (35). Genetic interactions between *h* and achaete (*ac*) (the AS-C *T5* transcript) are consistent with *h*'s inhibiting bristle development by antagonizing the action of the *ac* protein (8, 25, 49, 60, 62). Ectopic microchaetae in *h* mutant flies are suppressed by a pulse of *h* expression during early pupariation, the time when such sensory organs appear to be determined (62). Surprisingly, ectopic *h* appears not to suppress wild-type microchaetae, despite the fact that these also depend on *ac* activity.

The effects of premature *h* expression on *Drosophila* sex determination support *h*'s interacting with AS-C proteins (54). Misexpression of *h* during the early blastoderm stage causes female-specific lethality by antagonizing X-linked *sis* genes, including AS-C *T4*.

bHLH proteins could repress transcription in several ways. Two HLH transcriptional repressors completely lack the adjacent basic domain (Fig. 1) and appear to act by complexing bHLH activators as non-DNA-binding het-

* Corresponding author.

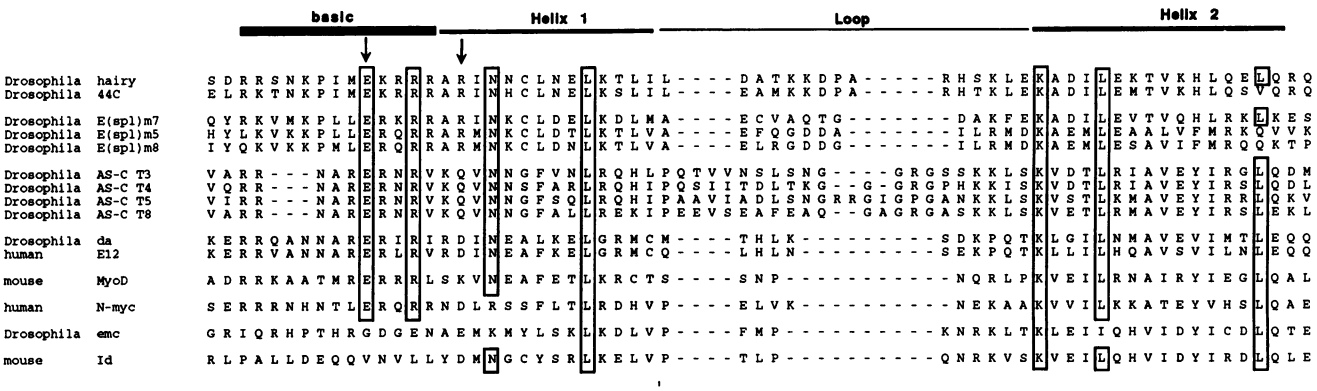


FIG. 1. Protein sequences of representative bHLH and HLH proteins. Demarcations between bHLH domains are as in reference 4. The boxed amino acids show positions of amino acid conservation. The two arrows show the amino acids mutated in h^{5H} and h^{14C} (see text). References for protein sequences: *h* (62); 44C (68); *E(spl)* (42); AS-C (1, 69); *emc* (22, 26); *da* (14); Id (4); E12 (50); MyoD (20); N-Myc (44).

erodimers. Id prevents *myoD* from activating myogenesis and muscle-specific transcription (4). In vitro, it heterodimerizes with MyoD and inhibits DNA binding to muscle-specific target sites. The other gene, extramacrochaetae (*emc*), encodes a suppressor of *Drosophila* bristle development that, like *h*, may act by antagonizing AS-C activity (18, 22, 26, 63, 67).

In contrast, *h* retains a substantially conserved b-domain, suggesting that *h* heterodimers (or homodimers) are capable of site-specific DNA binding and raising the possibility that *h* might repress transcription differently from Id and *emc*. This idea is supported by structural similarities between *h* and other bHLH proteins encoded by the Enhancer-of-split [*E(spl)*] complex of neurogenic genes (42). The *h* and *E(spl)* b-domains include a characteristic proline residue at position 6 (Fig. 1) and are terminated by a specific C-terminal Trp-Arg-Pro-Trp tetrapeptide (WRPW). Also, both behave as antagonists of neural development, *E(spl)* being required for lateral inhibition between cells (43), a process that restricts neural cell fate to individual cells within clusters of proneural cells (reviewed in reference 2).

There is no direct assay for *h* DNA-binding activity. Although putative target genes have been identified genetically, specific DNA sites through which *h* acts are currently unknown. In this study, we investigated structural requirements for in vivo *h* activity by analyzing mutant *h* proteins. We have mapped protein lesions in 12 *h* mutations and demonstrated three important domains required for *h* function: the b-domain, the HLH domain, and the C-terminal WRPW domain. We also show that the *h* bHLH and WRPW domains are completely conserved in the distantly related *Drosophila* species, *D. virilis*.

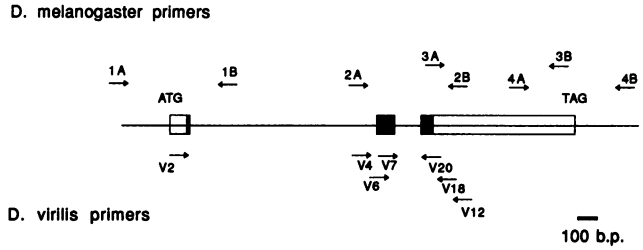
MATERIALS AND METHODS

Generation of novel *h* mutations. Four hundred males were fed overnight on 25 mM ethyl methanesulfonate in 1% sucrose (28) and mated to h^2/h^2 virgin females. A total of 9,000 adult progeny were scored for ectopic bristles on the wing blade, from which three *h* mutations were recovered and designated h^{w4}, h^{w15}, and h^{w24}.

PCR cloning of *D. melanogaster h* DNA and DNA sequencing. For each allele, 100 24-h-old homozygous *h* embryos were identified and selected by their pair-rule phenotypes. DNA was prepared by sodium dodecyl sulfate extraction and potassium acetate precipitation (3). One-tenth of the DNA was used in a 25-μl polymerase chain reaction (PCR) reaction (94°C for 1 min; 58°C for 2 min; 72°C for 2 min; 45 cycles), using the primer pairs shown in Fig. 2 (24). PCR products were digested with *Hin*dIII and *Sal*I, gel purified, and cloned into Bluescript KS+ (Stratagene) for sequencing. Standard dideoxy sequencing methods were used (U.S. Biochemical Sequenase II kit) on single-stranded templates (Stratagene). Each mutation was confirmed by sequencing clones derived from two independent PCR reactions.

Cloning of *h* from *D. virilis*. Genomic DNA and adult RNA were prepared from adult flies by conventional methods (3). First-strand cDNA was generated by random priming and amplified by using degenerate primers (v4-v18 and v6-v18; Fig. 2) at an annealing temperature of 55°C (40). Subsequent amplifications (v2-v20 and v7-v12) also made use of *D. virilis*-specific primers. PCR products were treated with T4 polymerase, kinase treated, gel purified, and cloned into Bluescript KS+ (Stratagene) for sequencing.

Genomic *D. virilis* clones were identified from *D. virilis* EMBL3 libraries generously provided by P. Schedl and R.



1A TCGGAAGCTTAATCGCACTGCAGCCAGAAC
1B GAATGTCGACGGTTCGCTTCAGTTAATCCAC
2A CAATAAGCTTGAAGACCCATGCAGCTTCCT
2B GGGAGTCGACCTAACCTCGTTCACACAGTCG
3A GCAGAAGCTTGCCGATCCCAAGATTGTGAAC
3B GACCGTCGACTTGATGTCCATGGGCTCGTAG
4A AACCAAGCTTTTGGTTAGCATGCCCCAGCGTACA
4B GTGCGTCTGATGAAGGTCCTTAGCC

V2 CCIGCICAGCTIAAGGASAC
V4 TCIAACAAGCCSATIATGGA
V6 GCICGIATIAACAASTGSST
V7 GACGTAATGATCTCTAGAGTC
V12 GAIACSTCSTTIACSCASTC
V18 CAGGTGCTTIACIGTCTTSTC
V20 TCTCGAGCTTGGAATGACGC

FIG. 2. PCR primers for amplifying *h* mutations and cloning *h* from *D. virilis*. Dark shading indicates the bHLH domain, split between three exons. Primers v7 and v20 are specific for *D. virilis h*. I = inosine; S = G and C.

TABLE 1. Characterization of ethyl methanesulfonate-induced alleles of hairy and comparison with h^{i22a}

Allele	Reference or discoverer	Parental chromosome	Mutation	*h* protein expression	Cuticle phenotype[b] (h^n/h^{i22})	*ftz* protein expression
5H	38	rucuca	E40V (GAG→GTG)	Stripes 2–7 reduced[c]	+++/++/+	Stripes fused
14H	38	*st e*	R46C (CGT→TGT)	Normal[c]	+++	Stripes fused
9K	38	rucuca	P336L (CCC→CTC)	Stripes 2–7 reduced[c]	++	Not determined
12C	38	*st e*	P336L (CCC→CTC)	Normal[c]	+	Stripes broadened
C1	30	*Ki* p^p Aldox	R43[d] (CGA→TGA)	Undetectable[e]	+++	Stripes fused
7H	38	rucuca	L72[d] (TTG→TAG)	Undetectable[c,e]	+++	Stripes fused[f]
IL79	38	Wild type	C114[d] (TGT→TGA)	Normal[g,e]	+++	Stripes fused
w15	This study	*st*	Q150[d] (CAG→TAG)	Normal[e]	+++	Stripes fused
IK93	38	Wild type	K206[d] (AAG→TAA)	Normal[c]	+++/++/+	Stripes broadened
11D	38	rucuca	Q218[d] (CAG→TAG)	Stripes 2–7 reduced[c]	+	Stripes broadened
w4	This study	*st*	Q309[d] (CAG→TAG)	Normal	+++	Stripes fused
8K	38	*st e*	Q300 onwards deleted (see text)	Delayed separation of stripes 3 and 4[c]	++/+	Stripes broadened
w24	This study	*st*	None found	Normal	+	Stripes broadened
i22	J. Belote	*Ki* roe p^p	h protein deletion[b]	None[c]	+++	Stripes fused

[a] All alleles are lethal in the homozygous condition. The rucuca chromosome contains the h^1 allele.
[b] From reference 36. +, Most embryos exhibit cuticle deletions less severe than a classic pair-rule phenotype; ++, most embryos exhibit a classic pair-rule phenotype; +++, most embryos exhibit cuticle deletions more severe than a classic pair-rule phenotype.
[c] From reference 31.
[d] Nonsense mutation.
[e] In these cases, embryos were stained with a mixture of *h* monoclonal antibodies recognizing N-terminal *h* epitopes (see Materials and Methods).
[f] From reference 12.
[g] This study.

Blackman. A total of 50,000 plaques were screened by using nick-translated, subcloned PCR fragment v4-v18 (Fig. 2). DNA was prepared from four positive clones, and a 6.5-kb *Eco*RI fragment hybridizing with the v4-v18 probe was subcloned into Bluescript KS+ for sequencing.

Embryo techniques. Embryos for staining and cuticle preparations were collected and handled according to standard methods (75). Immunohistochemistry with polyclonal mouse anti-*h* (55) or rabbit anti-*ftz* (45) was performed by using alkaline phosphatase-coupled secondary antibodies (Jackson Immunoresearch) and standard methods. In later experiments, we used a cocktail of three mouse monoclonal antibodies, 24/1, 30/3, and 32/4, generated against bacterially expressed *h* protein (56). These antibodies detect *h* epitopes near the N terminus of the protein, as shown by their individual abilities to immunoprecipitate in vitro-translated hybrid proteins retaining *h* amino acids 1 to 43 (24/1 and 30/3) and 1 to 88 (all three) (55).

In situ hybridization was performed with DIG-labeled *h* probes as previously described (65) except that labeled RNA probes were used.

In each staining experiment, 75% of embryos are wild type and provide internal controls for staining intensity. Normal staining implies the inability to distinguish wild-type and mutant staining patterns. Embryos were dehydrated in ethanol and mounted in methacrylate (JB-4; Polysciences).

Nucleotide sequence accession number. The GenBank accession number for the *D. virilis h* cDNA sequence reported in this paper is M87885.

RESULTS

***h* mutations demonstrate three protein domains required for *h* activity.** We analyzed 13 ethyl methanesulfonate-induced *h* mutations, most of which were recovered in the original screens for segmentation mutations (38, 52; Table 1). We generated a further three alleles by screening for noncomplementation of the recessive *h* bristle patterning requirement (Materials and Methods). All mutations analyzed are recessive and, except for the very weak allele h^{w24}, result from lesions in the protein-coding region. We could not identify an altered coding sequence in h^{w24} which may represent a regulatory mutation.

Conventional gene cloning from heterozygous mutant stocks presents the considerable problem of distinguishing between mutant and wild-type genes. We avoided this difficulty by making pure mutant DNA from embryos that were manually selected according to their cuticular phenotype. We used PCR to amplify the mutant genes and determined their DNA sequences by standard means (Materials and Methods; results are summarized in Table 1 and Fig. 3).

Four mutations that eliminate *h* function are due to missense mutations. h^{14H} is due to a C-to-T base change that alters Arg-46 to Cys in the first amphipathic helix of the HLH domain. In other bHLH proteins, this position in helix

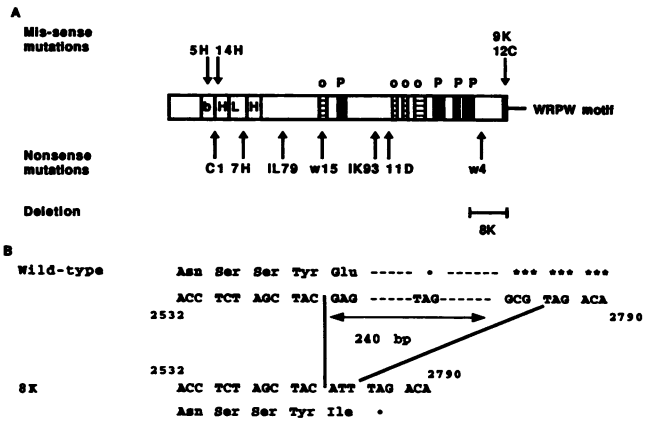


FIG. 3. (A) Diagram of the *h* gene, indicating the sites of *h* mutations. o, *opa* repeats (74); P, potential PEST sequences (59). (B) DNA sequence of the h^{8K} deletion, indicating its extent and altered protein sequence. The TAG at position 2785 is not the wild-type terminator but lies in the *h* 3′ untranslated region.

1 is almost invariably occupied by a hydrophilic amino acid that is usually charged (Fig. 1 and similar figure in reference 4). h^{14H} expresses normal levels of nuclear protein (31), so protein stability is not affected, and the lack of *h* activity must be due to disruption of the HLH motif. h^{14H} lacks both bristle and segmentation activity (35), indicating that both *h* functions require the dimerization domain.

h^{5H}, the other N-terminal missense mutant, is due to the mutation of a completely conserved amino acid in the basic domain. An A-to-T base transversion alters Glu-40 to Val. h^{5H} lacks segmentation and bristle-suppressing activity (35), indicating that the basic domain is essential for *h* function. h^{5H} was induced on an h^{1} chromosome in which a gypsy transposon was inserted into the upstream *h* promoter, reducing expression of stripes 2 to 7 (31). Such reduced expression may contribute to the h^{5H} phenotype but cannot be its major cause because the majority of h^{1} embryos retain sufficient expression for normal segmentation (35).

The other two missense mutations, h^{9K} and h^{12C}, reveal a precise requirement for the C terminus of the *h* protein. They include identical base substitutions (C to T) that alter the penultimate amino acid Pro-336 to Leu. The two mutations are independent alleles, having been induced in different parental chromosomes and retaining characteristic parental genetic markers (Table 1). They also include different third-base substitutions that do not affect the protein sequence (not shown). h^{12C} embryos suffer extensive pattern deletions, although the phenotype is weaker than that of the alternate segment deletions of prototypic pair-rule embryos (35; Table 1; Fig. 4B). h^{9K} embryos suffer more severe cuticular deletions (Table 1), presumably because the mutation was induced on an h^{1} chromosome that reduces *h* expression.

Eight *h* alleles encode C-terminal truncated proteins. Seven are caused by nonsense mutations and form a series of mutant proteins of different lengths (Fig. 3; Table 1). Four alleles, h^{C1}, h^{7H}, h^{IL79}, and h^{w15}, show strong cuticular phenotypes consistent with a lack of active *h* protein (35; Table 1). h^{C1} and h^{7H} encode the shortest two proteins (42 and 71 amino acids, respectively) and do not accumulate immunologically detectable levels of *h* protein (31; Table 1).

h^{IL79} and h^{w15} encode longer proteins (113 and 149 amino acids, respectively). We have previously reported that *h* protein is not detectable in h^{IL79} embryos (31), and we were unable to detect h^{IL79} and h^{w15} proteins by using a polyclonal mouse antibody (data not shown). However, this result must be the consequence of the mutant proteins lacking important C-terminal epitopes, because monoclonal antibodies directed against the N-terminal portion of *h* protein reveal essentially wild-type levels of protein (Materials and Methods). Thus, truncated *h* proteins longer than 112 amino acids appear to accumulate normally (31; Table 1).

h^{IK93} protein may retain slight activity, as mutant embryos have a predominantly pair-rule phenotype (Table 1). h^{11D} encodes a protein that is only 12 amino acids longer, yet it must retain significant activity since mutant embryos show a considerably weaker phenotype (Table 1; Fig. 4). Moreover, h^{11D} expression is reduced, having been induced on an h^{1} chromosome (31).

The most C-terminal deletion, h^{w4}, has an unexpectedly strong phenotype, causing more severe pattern disruptions than do alleles encoding shorter proteins, h^{IK93} and h^{11D}. This phenotype appears to be characteristic of the truncated protein, as there is no other mutation in the *h* coding region (see Materials and Methods), and antibody staining shows that h^{w4} protein is expressed at normal levels (Table 1). This departure from a monotonic phenotypic series suggests that *h* may interact with several different proteins.

A final deletion mutation confirms that *h* requires an intact C terminus. h^{8K} is a 238-bp deletion/3-bp insertion at the 3′ end of the *h* gene that removes 117 bp of C-terminal *h* coding region and 121 bp of adjacent 3′ untranslated sequences (36; Fig. 3). It encodes a truncated *h* protein lacking the last 38 amino acids but including one extra amino acid derived from translation of the 3-bp insertion (Fig. 3B). The h^{8K} cuticular phenotype indicates that the modified protein largely lacks patterning activity (35; Table 1).

We also analyzed *ftz* patterning in *h* mutant embryos and showed that it correlates with the *h* cuticular phenotype (Table 1). Weak *h* alleles cause marked broadening of *ftz* stripes, although every stripe remains distinct and separated from its neighbors by unstained cells (Fig. 4F). Strong *h* alleles show more extensive broadening leading to fusion of adjacent *ftz* stripes. The least severe phenotype results in fusion of stripes 4 and 5; more severely affected embryos show fusion of stripes 6 and 7. Occasional strong *h* embryos express *ftz* almost continuously from about 15 to 65% egg length (Fig. 4G). h^{w4} embryos show a severely affected pattern of *ftz* expression (not shown), consistent with its cuticular phenotype.

The *h* protein from *D. virilis* is very highly conserved. If the entire *h* bHLH domain is critical for *h* function, it should be highly conserved between different *Drosophila* species. We used PCR to clone the *h* gene from *D. virilis*, a species that diverged from *D. melanogaster* about 60 million years ago (5), sufficient for extensive sequence divergence of functionally irrelevant or labile protein and DNA domains (39).

We amplified *D. virilis h* sequences from genomic DNA by using several degenerate oligonucleotide primers homologous to regions flanking or including the bHLH domain (Materials and Methods; Fig. 2). The *h* bHLH domain in *D. melanogaster* is interrupted by two introns, so we also amplified *h* sequences from adult *D. virilis* cDNA, chosen because *h* transcripts are abundant in adult flies (36). Several primer pairs amplified suitably sized DNA bands that were subcloned and sequenced to confirm that they derive from the *D. virilis h* gene (Materials and Methods).

The cloned PCR fragments were used to identify genomic *h* clones from a phage genomic library which were then subcloned and sequenced (Materials and Methods). The DNA sequence of the entire *h* protein-coding region and the predicted amino acid sequence for this region are shown in Fig. 5A. The *h* gene includes two introns that are present at positions equivalent to those in *D. melanogaster*. There is extensive sequence divergence outside the protein-coding regions (Fig. 5B) and at third-base positions of many codons.

In contrast, the *D. virilis* and *D. melanogaster h* proteins are very similar, although several amino acid insertions render the former 41 amino acids longer (Fig. 5C). Otherwise, the two proteins are 89% identical.

The *D. virilis* bHLH domain is identical to that in *D. melanogaster*. The next 34 amino acids are also homologous, incorporating two conservative substitutions, and include a set of hydrophobic heptad repeats that might form coiled-coil domains of protein-protein interaction (48; Fig. 5C). A second set of conserved repeats lies a further 11 amino acids downstream. The C-terminal 23 amino acids, including the terminal WRPW tetrapeptide, are invariant except for a single amino acid insertion in the *D. virilis* protein. Also retained are several *opa* repeat sequences and Pro-Ser-rich sequences that might served as PEST destabilization domains (59, 74; Fig. 5C).

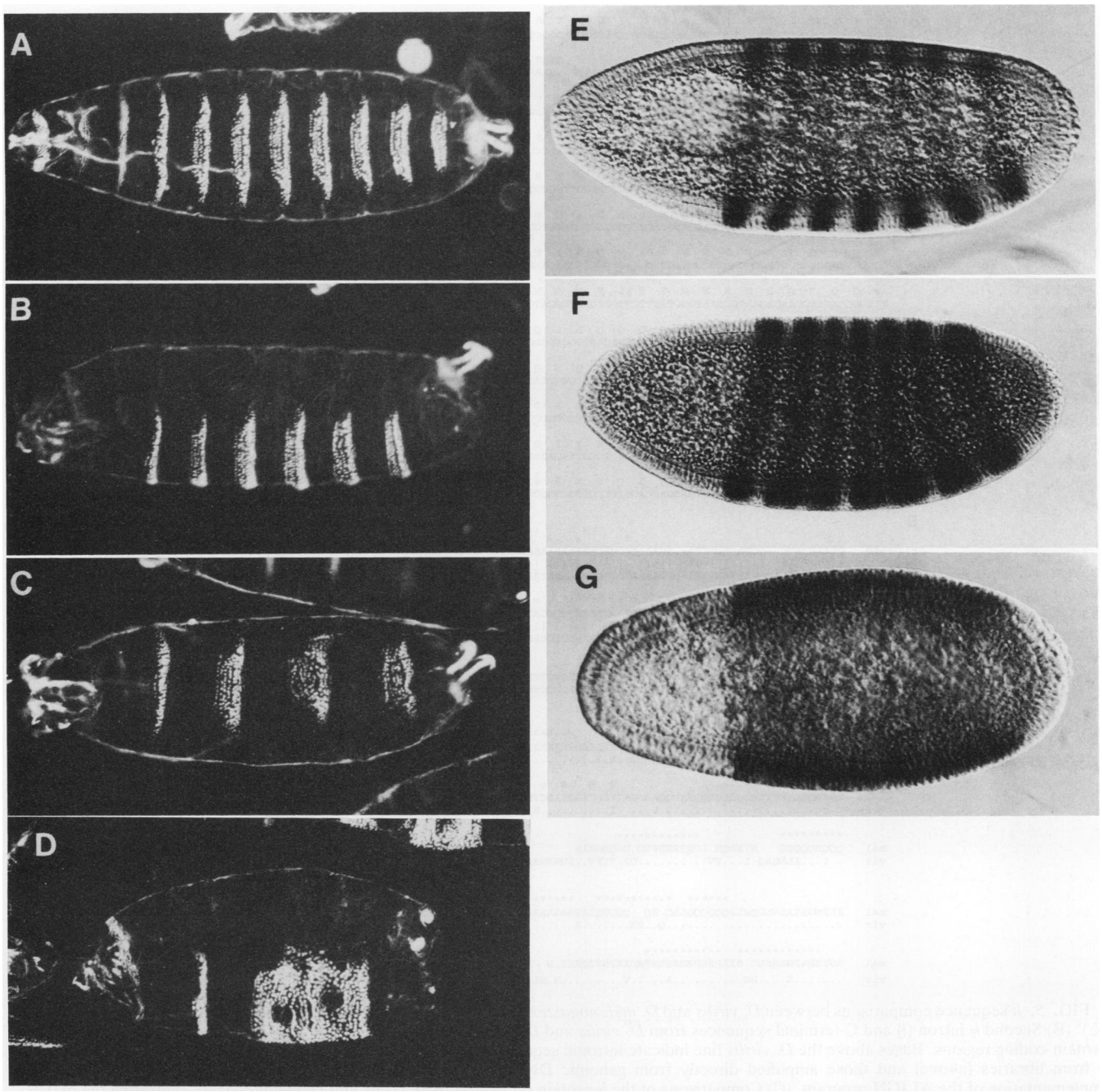


FIG. 4. Cuticle phenotype and anti-*ftz* staining of *h* mutant embryos. Cuticle preparations: (A) wild-type embryo; (B) h^{12C} embryos illustrating weaker than pair-rule phenotype (+ in Table 1); (C) pair-rule h^{1K93} embryo (++); (D) h^{C1} embryo illustrating more severe than pair-rule phenotype (+++). anti-*ftz* staining: (E) wild-type embryo; (F) h^{12C} (broadened stripes); (G) fused *ftz* stripes in h^{5H} embryo.

DISCUSSION

Structural requirements for *h* protein function. Our analysis of *h* mutant proteins demonstrates that the b-domain and the HLH are both required for *h* function. The h^{14H} mutation lies in the first HLH helix, substituting a hydrophilic for a hydrophobic amino acid. This amino acid should lie adjacent to the hydrophilic helix face (62) and could participate in electrostatic stabilization between helices such as occur in leucine zipper proteins (53). h^{5H} is caused by an altered b-domain, although reduced expression due to the upstream gypsy insertion may also contribute to its phenotype. h^{5H} results from substituting a valine for a critical glutamic acid that is absolutely conserved in all HLH b-domains, including *h*. Other mutations highlight a requirement for the *h* C terminus. Changing the penultimate amino acid inactivates *h* (h^{9K} and h^{12C}), as do all C-terminal truncations.

The deletion series maps structural requirements for protein stability and nuclear localization. h^{1K93} is the shortest protein that localizes unambiguously to the nucleus (31), showing that a nuclear localization signal must lie upstream

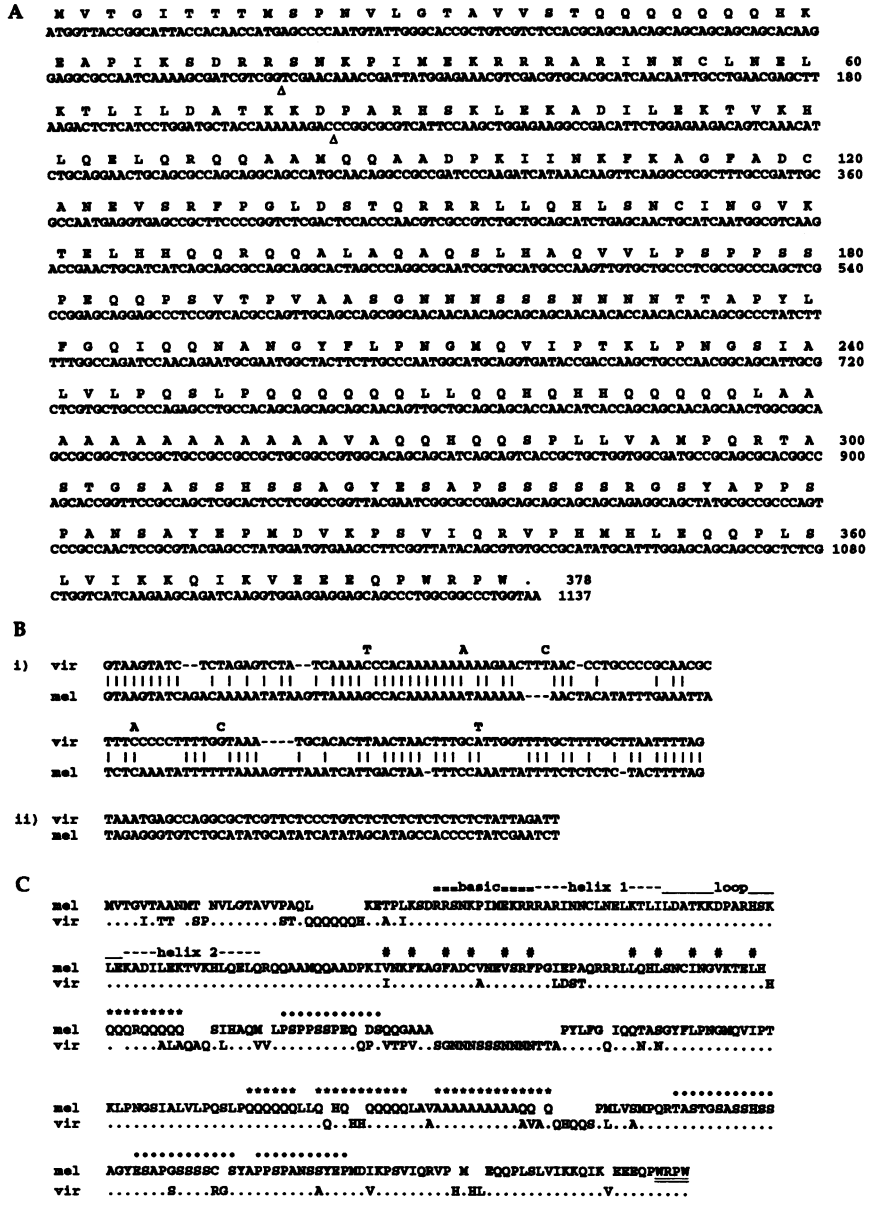


FIG. 5. *h* sequence comparisons between *D. virilis* and *D. melanogaster*. (A) *h* protein coding region of *D. virilis*, including intron positions (△). (B) Second *h* intron (i) and C-terminal sequences from *D. virilis* and *D. melanogaster* (ii), showing the sequence divergence outside the protein-coding regions. Bases above the *D. virilis* line indicate intronic sequence differences (two substitutions and an insertion) between the *h* from libraries (above) and those amplified directly from genomic DNA. Intron sequences were aligned by using the Intelligenetics implementation of the ALIGN program. (C) Comparisons of the *h* protein sequences of *D. virilis* and *D. melanogaster*, indicating the bHLH and WRPW (double-underlined) domains, the heptad arrays of hydrophobic amino acids (#), and *opa* (*) and PEST-like (●) motifs.

of amino acid 205. Although h^{IL79} and h^{w15} express normal levels of *h* protein, we were unable to define their intracellular localization because the monoclonal antibodies give weak and imprecise staining that does not distinguish clearly between nucleus and cytoplasm.

h proteins of 112 amino acids (h^{IL79}) and longer are expressed at normal levels. Their failure to overaccumulate (relative to internal wild-type control embryos) indicates that the proteins are as unstable as wild-type *h* protein, which has a half-life of about 6 min (21). Analysis of *ftz* mutations encoding stabilized proteins indicate that degradation is an active process involving PEST-like destabilization signals (59) and that rapid turnover is important for pair-rule striping and patterning (41). *h* includes several such sequences (Fig. 5C). However, h^{IL79} protein lacks all obvious PEST-like sequences and is still very unstable, raising the possibility that its instability is PEST independent.

h^{C1} and h^{7H} fail to accumulate detectable levels of *h* protein, perhaps because they fail to enter the nucleus and are subject to rapid cytoplasmic degradation. In situ hybridization to h^{C1} and h^{7H} embryos showed no obvious effects on transcript expression (data not shown), indicating that transcript stability is not affected and confirming the lack of feedback regulation on *h* expression (31).

***h* activity requires the DNA-binding domain and interactions with other proteins.** The *h* HLH domain is required for *h* function, presumably mediating protein dimerization. We consider it unlikely that *h* acts via protein homodimers, although such a model cannot be completely excluded. We and others have not succeeded in identifying *h* protein homodimers or homomeric sequence-specific in vitro DNA-binding activity. In this respect, *h* might resemble the *myc* bHLH proteins for which efficient DNA binding requires heterodimerization with a partner protein (7, 57).

The effects of *h* on sex determination and bristle patterning argue that *h* acts with partner bHLH proteins, perhaps preventing them from activating transcription. One model for such transcriptional inhibition is that, like HLH proteins which lack b-domains, *h* "poisons" activator molecules by forming heterodimers that cannot bind DNA (4, 22, 26). The three-dimensional structure of a bHLH basic domain has not yet been determined, but it may resemble the dimeric DNA-binding basic domains of leucine zipper transcriptional regulator proteins which have been modeled as α helices that interact with the DNA major groove (70). The proline in the *h* b-domain might disrupt a helical conformation and inhibit DNA binding (4). Indeed, substituting a proline into the MyoD basic domain inactivates DNA binding and myogenic activity (19).

Our results argue for an alternative view, that *h* is an active regulator that forms DNA-binding complexes with other bHLH proteins. First, the extremely high degree of conservation between the *h* proteins of *D. melanogaster* and *D. virilis* indicates that their structures are under strong functional constraints. This is consistent with *h* protein playing an active role in transcriptional regulation, rather than its being a passive poison.

Also, the *h* b-domain retains almost all of the structural features that are conserved in other DNA-binding bHLH proteins, including the array of characteristic basic amino acids that appear to be required for DNA-binding activity and which are completely lacking in Id and *emc*. The b-domain is completely conserved in the *D. virilis h* protein, indicating that its structural features are necessary for *h* activity. Finally, mutation of the absolutely conserved glutamic acid in the *h* b-domain eliminates *h* activity in h^{5H} without affecting nuclear localization. These results indicate that the b-domain is required for *h* function, presumably for DNA binding, although we cannot exclude its having a changed role, e.g., in protein-protein interactions.

The b-domain proline is retained in the *D. virilis h* protein and is likely to be functionally important, despite any helical disruption that it might cause. This view is supported by the almost complete conservation of the *h* bHLH domain in a *Drosophila* bHLH protein (44C) expressed in the developing nervous system (68; Fig. 1). The *E(spl)* proteins also include a similarly positioned proline (42; Fig. 1), and the character of the amino acid at this position is important for the myogenic activity of myogenin (9) and MyoD (73). We note that the equivalent amino acid in most other bHLH proteins is asparagine, which can also terminate α helices (58), perhaps helping fit the DNA-binding domain into the major DNA groove (70).

The simplest model for *h* function is that heterodimers between *h* and a partner bHLH protein bind DNA and interfere with transcription, perhaps by occupying an otherwise activating site. For several reasons, we favor a more complex model in which repression involves interactions with other transcription factors. First, *h* activity requires the terminal WRPW domain, which is unlikely to play a role in bHLH heterodimerization and DNA binding. The only other proteins terminated by this tetrapeptide are 44C bH with the *h*-like bHLH domain and the neurogenic *E(spl)* complex of bHLH proteins that, like *h*, inhibit proneural development, possibly by antagonizing AS-C activity (42, 68). Assuming that the terminal WRPW is not involved in DNA binding or HLH dimerization, it may interfere with transcription by interacting with or inhibiting other protein factors, for example, adaptor or coactivator proteins (47). Second, the complete evolutionary conservation of the *h* bHLH domain argues for interactions with several proteins that constrain *h* sequence divergence. The loop domain is identical between *D. virilis* and *D. melanogaster*, suggesting the possibility that its role is not merely to tether together two amphipathic helices. It may interact with other protein factors or contribute to the specificity of HLH dimerization.

Finally, the properties of a hybrid MyoD-E12 protein suggest that the b-domain's role is not limited to determining DNA binding specificity. Substituting the b-domain from E12 or AS-C *T4* into MyoD gives hybrid proteins that still heterodimerize with E12 in vitro but which no longer activate myogenesis in vivo (19). This suggests that the E12 and *T4* b-domains are unable to interact with other factors required for muscle-specific transcription (73).

The next phase in defining *h*'s mechanism requires characterization of other proteins with which it interacts and their target DNA sites. Sex-specific effects of premature *h* expression provide an in vivo assay for *h* activity (54), but we currently lack in vitro assays for *h* DNA binding. Identification and characterization of *h* partner proteins should illuminate its interactions with DNA and with other components of the transcriptional machinery.

ACKNOWLEDGMENTS

We thank P. Schedl and R. Blackman for gifts of *D. virilis* genomic libraries and Sheena Pinchin for the anti-*h* mono- and polyclonal antibodies. We thank Susan Parkhurst and other members of the Developmental Genetics laboratory for their advice and encouragement and for criticisms of the manuscript. We also thank Ze'ev Paroush and Steve Hanes for criticisms of the manuscript.

REFERENCES

1. **Alonso, M. C., and C. V. Cabrera.** 1988. The *achaete-scute* gene complex of *Drosophila melanogaster* comprises four homologous genes. EMBO J. **7**:2585–2591.
2. **Artavanis-Tsakonas, S., and P. Simpson.** 1991. Choosing a cell fate: a view from the *Notch* locus. Trends Genet. Sci. **7**:403–407.
3. **Ashburner, M.** 1989. *Drosophila*: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell **61**:49–59.
5. **Beverley, S. M., and A. C. Wilson.** 1984. Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. J. Mol. Evol. **21**:1–13.
6. **Blackwell, T. K., and H. Weintraub.** 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science **250**:1104–1110.
7. **Blackwood, E. M., and R. N. Eisenman.** 1991. Max—a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with myc. Science **251**:1211–1217.
8. **Botas, J., J. Moscoso del Prado, and A. García-Bellido.** 1982. Gene dose titration analysis in the search for trans-regulatory genes in *Drosophila*. EMBO J. **1**:307–310.
9. **Brennan, T. J., T. Chakraborty, and E. N. Olson.** 1991. Mutagenesis of the *myogenin* basic region identifies an ancient

protein motif critical for activation of myogenesis. Proc. Natl. Acad. Sci. USA **88**:5675–5679.
10. **Cabrera, C. V., and M. C. Alonso.** 1991. Transcriptional activation by heterodimers of the *achaete-scute* and *daughterless* gene products of *Drosophila*. EMBO J. **10**:2965–2973.
11. **Cabrera, C. V., A. Martinez-Arías, and M. Bate.** 1987. The expression of three members of the *achaete-scute* gene complex correlates with neuroblast segregation in *Drosophila*. Cell **50**: 425–433.
12. **Carroll, S. B., and M. P. Scott.** 1986. Zygotically active genes that affect the spatial expression of the *fushi tarazu* segmentation gene during early *Drosophila* embryogenesis. Cell **45**:113–126.
13. **Caudy, M., E. H. Grell, C. Dambly-Chaudière, A. Ghysen, L. Y. Jan, and Y. N. Jan.** 1988. The maternal sex determination gene *daughterless* has zygotic activity necessary for the formation of peripheral neurons in *Drosophila*. Genes Dev. **2**:843–852.
14. **Caudy, M., H. Vässin, M. Brand, R. Tuma, L. Y. Jan, and Y. N. Jan.** 1988. *daughterless*, a Drosophila gene essential for both neurogenesis and sex determination, has sequence similarity to *myc* and the achaete-scute complex. Cell **55**:1061–1067.
15. **Cline, T. W.** 1976. A sex-specific, temperature-sensitive maternal effect of the daughterless mutation of Drosophila melanogaster. Genetics **84**:723–742.
16. **Cline, T. W.** 1986. A female-specific lesion in an X-linked positive regulator of the Drosophila sex determination gene, *Sex-lethal*. Genetics **113**:641–663.
17. **Cline, T. W.** 1988. Evidence that *sisterless-a* and *sisterless-b* are two of several discrete "numerator elements" of the X/A sex determination signal in Drosophila that switch *Sxl* between two alternative stable expression states. Genetics **119**:829–862.
18. **Cubas, P., J. F. de Celis, S. Campuzano, and J. Modolell.** 1991. Proneural clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disk. Genes Dev. **5**:996–1008.
19. **Davis, R. L., P.-F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The *MyoD* binding domain contains a recognition code for muscle-specific gene activation. Cell **60**:733–746.
20. **Davis, R. L., H. Weintraub, and A. B. Lassar.** 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51**:987–1000.
21. **Edgar, B. A., G. M. Odell, and G. Schubiger.** 1989. A genetic switch based on negative regulation sharpens stripes in *Drosophila* embryos. Dev. Genet. **10**:124–142.
22. **Ellis, H. M., D. R. Spann, and J. W. Posakony.** 1990. *extramacrochaete*, a negative regulator of sensory organ development in *Drosophila*, defines a new class of helix-loop-helix proteins. Cell **61**:27–38.
23. **Erikson, J. W., and T. W. Cline.** 1991. Molecular nature of the *Drosophila* sex determination signal and its link to neurogenesis. Science **251**:1071–1074.
24. **Erlich, H. A.** 1989. PCR technology. Stockton Press, New York.
25. **Falk, R.** 1963. A search for a gene control system in Drosophila. Am. Nat. **97**:129–132.
26. **Garrell, J., and J. Modolell.** 1990. The *Drosophila extramacrochaete* locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. Cell **61**:39–48.
27. **Ghysen, A., and C. Dambly-Chaudière.** 1989. Genesis of the *Drosophila* peripheral nervous system. Trends Genet. Sci. **5**:251–255.
28. **Grigliatti, T.** 1986. Mutagenesis, p. 39–58. *In* D. B. Roberts (ed.), *Drosophila*, a practical approach. IRL Press, Oxford.
29. **Hodgkin, J.** 1990. Sex determination compared in *Drosophila* and *Caenorhabditis*. Nature (London) **344**:721–728.
30. **Holmgren, R.** 1984. Cloning sequences from the *hairy* gene of *Drosophila*. EMBO J. **3**:569–573.
31. **Hooper, K. L., S. M. Parkhurst, and D. Ish-Horowicz.** 1989. Spatial control of *hairy* protein expression during embryogenesis. Development **107**:489–504.
32. **Howard, K., and P. Ingham.** 1986. Regulatory interactions between the segmentation genes *fushi tarazu*, *hairy* and *engrailed* in the *Drosophila* blastoderm. Cell **44**:949–957.
33. **Howard, K. R.** 1990. The blastoderm pattern. Semin. Cell Biol. **1**:161–172.
34. **Ingham, P. W.** 1988. The molecular genetics of embryonic pattern formation in *Drosophila*. Nature (London) **335**:25–34.
35. **Ingham, P. W., S. M. Pinchin, K. R. Howard, and D. Ish-Horowicz.** 1985. Genetic analysis of the *hairy* locus in *Drosophila melanogaster*. Genetics **111**:463–486.
36. **Ish-Horowicz, D., K. R. Howard, S. M. Pinchin, and P. W. Ingham.** 1985. Molecular and genetic analysis of the *hairy* locus in *Drosophila*. Cold Spring Harbor Symp. Quant. Biol. **50**:135–144.
37. **Ish-Horowicz, D., and S. M. Pinchin.** 1987. Pattern abnormalities induced by ectopic expression of the *Drosophila* gene *hairy* are associated with repression of *fushi tarazu* transcription. Cell **51**:405–415.
38. **Jürgens, G., E. Wieschaus, C. Nüsslein-Volhard, and H. Kluding.** 1984. Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. II. Zygotic loci on the third chromosome. Roux's Arch. Dev. Biol. **193**:283–295.
39. **Kassis, J., S. Poole, D. Wright, and P. O'Farrell.** 1986. Sequence conservation in the protein coding and intron regions of the *engrailed* transcription unit. EMBO J. **5**:3583–3589.
40. **Kawasaki, E. S.** 1990. Amplification of RNA, p. 21–27. *In* M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols. A guide to methods and applications. Academic Press, San Diego, Calif.
41. **Kellerman, K. A., D. M. Mattson, and I. Duncan.** 1990. Mutations affecting the stability of the *fushi tarazu* protein of *Drosophila*. Genes Dev. **4**:1936–1950.
42. **Klämbt, C., E. Knust, K. Tietze, and J. A. Campos-Ortega.** 1989. Closely related transcripts encoded by the neurogenic gene complex *Enhancer of split* of *Drosophila melanogaster*. EMBO J. **8**:203–210.
43. **Knust, E., K. A. Bremer, H. Vässin, A. Ziemer, U. Tepass, and J. A. Campos-Ortega.** 1987. The *Enhancer of split* locus and neurogenesis in *Drosophila melanogaster*. Dev. Biol. **122**:262–273.
44. **Kohl, N., E. Legouy, R. DePinto, P. Nisen, R. Smith, C. Gee, and F. Alt.** 1986. Human N-myc is closely related in organization and nucleotide sequence to c-myc. Nature (London) **319**:73–77.
45. **Krause, H. M., R. Klemenz, and W. J. Gehring.** 1988. Expression, modification and localisation of the *fushi tarazu* protein in *Drosophila* embryos. Genes Dev. **2**:1021–1036.
46. **Lassar, A. B., J. N. Buskin, D. Lockshon, R. L. Davis, S. Apone, S. D. Hauschka, and H. Weintraub.** 1989. MyoD is a sequence-specific DNA binding protein requiring a region of myc homology to bind to the muscle creatine kinase enhancer. Cell **58**:823–831.
47. **Lewin, B.** 1990. Commitment and activation at PolII promoters: a tail of protein-protein interactions. Cell **61**:1161–1164.
48. **McKeon, F. D., M. W. Kirschner, and D. Capon.** 1986. Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature (London) **319**:463–468.
49. **Moscoso del Prado, J., and A. García-Bellido.** 1984. Cell interactions in the generation of chaetae pattern in *Drosophila*. Roux's Arch. Dev. Biol. **193**:246–251.
50. **Murre, C., P. Schonleber McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerisation motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56**:777–783.
51. **Murre, C., P. Schonleber McCaw, H. Vässin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lassar, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell **58**:537–544.
52. **Nüsslein-Volhard, C., and E. Wieschaus.** 1980. Mutations affecting segment number and polarity in *Drosophila*. Nature (London) **287**:795–801.
53. **O'Shea, E. K., R. Rutkowski, W. F. Stafford, and P. S. Kim.** 1989. Prefential heterodimer formation by isolated leucine zippers from Fos and Jun. Science **245**:646–648.

54. **Parkhurst, S. M., D. Bopp, and D. Ish-Horowicz.** 1990. X:A ratio in *Drosophila* is transduced by helix-loop-helix proteins. Cell **63**:1179–1191.
55. **Pinchin, S. M.** Unpublished data.
56. **Pinchin, S. M., J. Gannon, D. Lane, and D. Ish-Horowicz.** Unpublished data.
57. **Prendergast, G. C., D. Lawe, and E. B. Ziff.** 1991. Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell **65**:395–407.
58. **Richardson, J., and D. Richardson.** 1988. Amino acid preferences for specific locations at the ends of alpha helices. Science **240**:1648–1652.
59. **Rogers, S., R. Wells, and M. Rechsteiner.** 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science **234**:364–368.
60. **Romani, S., S. Campuzano, E. R. Macagno, and J. Modolell.** 1989. Expression of *achaete* and *scute* genes in *Drosophila* imaginal discs and their function in sensory organ development. Genes Dev. **3**:997–1007.
61. **Romani, S., S. Campuzano, and J. Modolell.** 1987. The achaete-scute complex is expressed in neurogenic regions of Drosophila embryos. EMBO J. **6**:2085–2092.
62. **Rushlow, C. A., A. Hogan, S. M. Pinchin, K. R. Howe, M. T. Lardelli, and D. Ish-Horowicz.** 1989. The *Drosophila hairy* protein acts in both segmentation and bristle patterning and shows homology to *N-myc*. EMBO J. **8**:3095–3103.
63. **Skeath, J. B., and S. B. Carroll.** 1991. Regulation of *achaete-scute* gene-expression and sensory organ pattern-formation in the *Drosophila* wing. Genes Dev. **5**:984–995.
64. **Sun, X. H., and D. Baltimore.** 1991. An inhibitory domain of E12 transcription factor prevents DNA binding in E12 homodimers but not in E12 heterodimers. Cell **64**:459–470.
65. **Tautz, D., and C. Pfeifle.** 1989. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. Chromosoma (Berlin) **98**:81–85.
66. **Torres, M., and L. Sánchez.** 1989. The *scute* (T4) gene acts as a numerator element of the X:A signal that determines the state of activity of *Sex-lethal* in *Drosophila*. EMBO J. **8**:3079–3086.
67. **Van Doren, M., H. M. Ellis, and J. W. Posakony.** 1991. The *Drosophila extramacrochaetae* protein antagonizes sequence-specific DNA binding by *daughterless/achaete-scute* protein complex. Development **113**:245–255.
68. **Vässin, H., M. Caudy, E. Bier, L.-Y. Jan, and Y.-N. Jan.** 1990. Role of helix-loop-helix proteins in *Drosophila* neurogenesis. Cold Spring Harbor Symp. Quant. Biol. **60**:239–245.
69. **Villares, R., and C. V. Cabrera.** 1987. The *achaete-scute* gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to *myc*. Cell **50**:415–424.
70. **Vinson, C. R., P. B. Sigler, and S. L. McKnight.** 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science **246**:911–916.
71. **Voronova, A., and D. Baltimore.** 1990. Mutations that disrupt DNA-binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains. Proc. Natl. Acad. Sci. USA **87**:4722–4726.
72. **Weintraub, H., R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar.** 1991. The *myoD* gene family: nodal point during specification of the muscle cell lineage. Science **251**:761–766.
73. **Weintraub, H., V. J. Dwarki, I. Verma, R. Davis, S. Hollenberg, L. Snider, A. Lassar, and S. J. Tapscott.** 1991. Muscle-specific transcriptional activation by MyoD. Genes Dev. **5**:1377–1386.
74. **Wharton, K. A., B. Yedvobnick, V. G. Finnerty, and S. Artavanis-Tsakonas.** 1985. *opa*: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster*. Cell **40**:55–62.
75. **Wieschaus, E., and C. Nüsslein-Volhard.** 1986. Looking at embryos, p. 199–226. *In* D. B. Roberts (ed.), *Drosophila*, a practical approach. IRL Press, Oxford.