Cell, Vol. 67, 1157–1167, December 20, 1991, Copyright © 1991 by Cell Press

The Drosophila *single-minded* Gene Encodes a Helix-Loop-Helix Protein That Acts as a Master Regulator of CNS Midline Development

John R. Nambu, Josephine O. Lewis, Keith A. Wharton, Jr., and Stephen T. Crews
Molecular Biology Institute
Department of Biology
University of California
Los Angeles, California 90024

Summary

Development of the Drosophila CNS midline cells is dependent upon the function of the *single-minded* (*sim*) gene. Sequence analysis shows that *sim* is a member of the basic-helix-loop-helix class of transcription factors. Cell fate experiments establish that *sim* is required for early events in midline cell development, including a synchronized cell division, proper formation of nerve cell precursors, and positive autoregulation of its midline expression. Induction of ectopic *sim* protein under the control of the *hsp70* promoter shows that *sim* can direct cells of the lateral CNS to exhibit midline cell morphology and patterns of gene expression. We propose that *sim* functions as a master developmental regulator of the CNS midline lineage.

Introduction

Within the embryonic CNS of Drosophila there exists a specialized group of cells that lie along the midline of the ventral nerve cord and are characterized by a number of unique anatomical, developmental, and morphological properties (Thomas et al., 1984; Crews et al., 1988; Klämbt et al., 1991). These midline cells consist of both neurons and glia and define a discrete cellular substructure that plays an important role in the proper elaboration of the axon scaffold during CNS formation (Thomas et al., 1988; Jacobs and Goodman, 1989a, 1989b; Klämbt et al., 1991). In the blastoderm, the midline cells arise as two single cell wide stripes that occupy a unique position between cells that will give rise to the mesoderm and lateral neuroectoderm. As gastrulation proceeds, the two stripes of midline cells are brought together along the ventral furrow to form a single column. The midline cells constitute a discrete mitotic domain and undergo a synchronous postgastrulation cell division (Foe, 1989). During neurogenesis, the midline cells extend their nuclei into the nerve cell precursor layer, yet characteristically retain cytoplasmic connections with the underlying epidermis. They later differentiate into 25–30 neurons and glia per segment, which constitute the axis of symmetry for the ventral nerve cord.

The *single-minded* (*sim*) gene plays an important role in the development of the midline cells and formation of the CNS. Mutations in *sim* are embryonic lethals characterized by fusion of the longitudinal axon bundles and collapse of the axon scaffold (Thomas et al., 1988). The *sim* gene encodes a nuclear protein expressed in the midline cells throughout embryogenesis (Crews et al., 1988). Genetic and cell biological studies have indicated that *sim* is required for proper midline cell differentiation and gene expression (Nambu et al., 1990).

In this study, we show that *sim* is a member of the basic-helix-loop-helix (bHLH) family of transcription factors, possessing both putative DNA-binding and protein dimerization domains. The use of cell-marking techniques to follow the development of the midline cells in *sim* null mutant embryos indicates that *sim* functions early in the midline lineage. In *sim* null mutant backgrounds, the midline cells fail to undergo synchronous postgastrulation cell division, exhibit defects in the formation of nerve cell precursors, and do not properly autoregulate *sim* gene expression. Function of the *sim* gene is necessary for all aspects of midline cell development, including early events in the formation of midline nerve cell precursors as well as later events in their differentiation into mature neurons and glia.

The developmental consequences of the ectopic expression of *sim* protein were investigated using transformant fly strains in which *sim* transcription is driven by the *hsp70* promoter. Induction of ectopic *sim* during early embryogenesis results in the expression of midline cell properties, including cellular morphology and patterns of gene expression, in most or all cells of the lateral CNS. The ectopic expression of *sim* is thus sufficient to promote other cell types to take on midline identities. The data suggest that *sim* normally acts to direct a specific subset of the neuroectoderm to develop into the CNS midline.

Results

The *sim* Gene Encodes a Member of the bHLH Family of Transcription Factors

The previously described *sim* cDNA sequence contains an open reading frame (ORF) of 655 amino acids (Crews et al., 1988). As noted, it was not possible to define the amino terminus of the protein, since the ORF fully extends to the 5′ end of the cDNA clone. We have now isolated and characterized full-length *sim* cDNA clones, which indicate that the *sim* coding region contains an additional 18 amino acids at the amino-terminal end. This yields an ORF of 673 amino acids. Detailed sequence analysis, as initially noted by S. Kohtz, suggests that the amino terminus of the *sim* protein contains a bHLH sequence motif, characteristic of a large family of transcription factors (Murre et al., 1989). This family has both vertebrate and invertebrate members, including *myc*, *MyoD*, *E2A*, *daughterless*, *twist*, and genes of the *achaete-scute* complex (reviewed in Murre et al., 1989; Benezra et al., 1990). These transcription factors are characterized by the presence of two amphipathic α helices separated by a variable loop. This region directs the formation of protein dimers while an adjacent basic region mediates DNA binding (Murre et al., 1989; Davis et al., 1990).

Figure 1A illustrates sequence alignment of *sim* with representative members of the major subfamilies of bHLH proteins. The bHLH sequence of *sim* is most similar to

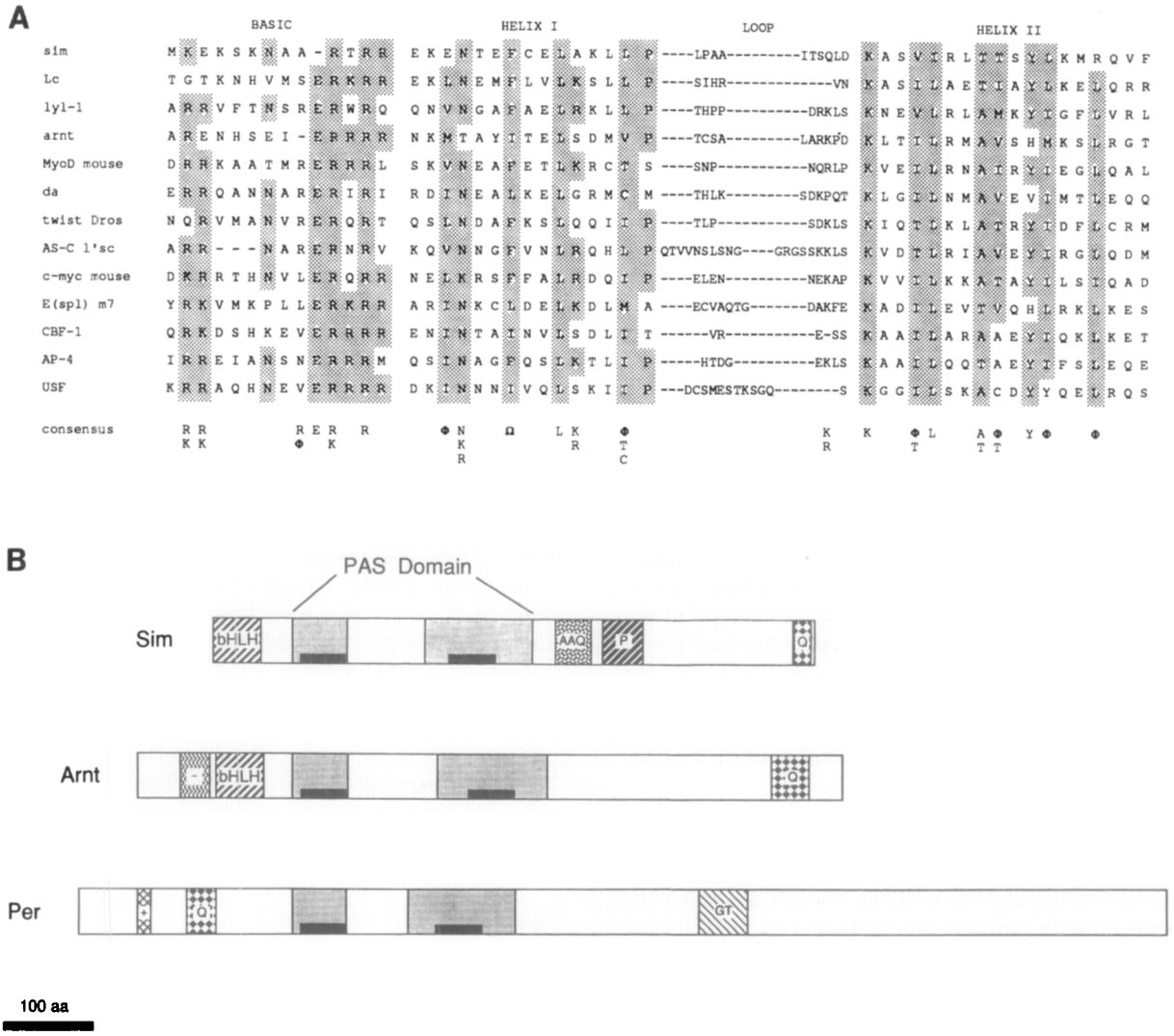


Figure 1. Comparison of Several bHLH Domains and of the PAS Domain Proteins

(A) Alignment of the amino-terminal region of the *sim* protein with a representative group of vertebrate and invertebrate bHLH domains. The basic region, helix I, loop, and helix II are denoted. Amino acid residues shared between proteins are indicated by shaded boxes. A bHLH consensus sequence from Blackwood and Eisenman (1991) is shown below: Φ = leucine, isoleucine, valine, methionine; Ω = phenylalanine, leucine, isoleucine, tyrosine.

(B) Schematic comparison of the structures of the PAS domain proteins (*sim*, *per*, and *arnt*). Discrete structural motifs are noted, including the bHLH domain, PAS domain, alanine-alanine-glutamine repeats (AAQ), proline-rich region (P), glutamine-rich region (Q), negatively charged region (–), positively charged region (+), and glycine-threonine repeats (GT). Two 51 amino acid direct repeats present in the PAS domain are denoted by two solid black bars. Light gray indicates regions of sequence similarity between the PAS domain proteins (Hoffman et al., 1991).

the vertebrate *lyl-1* (Mellentin et al., 1989) and maize *Lc* (Ludwig et al., 1989) proteins. Residues predicted to be along the face of amphipathic helices are generally hydrophobic and similar to other HLH proteins. Sequence analysis using a hydrophobic correlation method (Sweet and Eisenberg, 1983) indicates that predicted amphipathic helices of *sim* can be folded similarly to those of other HLH proteins. These findings strongly suggest that *sim* is a DNA-binding transcriptional regulator.

The *sim* protein also contains an extended region that is found on the Drosophila *period* (*per*) and vertebrate *arnt* proteins (Figure 1B). This conserved region, which we refer to as the PAS domain (*per*, *arnt*, and *sim*), spans approximately 270 amino acids in *sim* and contains two direct repeats of 51 amino acids (Crews et al., 1988; Hoffman et al., 1991). *per* is a nuclear protein that may act as a transcriptional regulator to control biological rhythms (Siwicki et al., 1988; Hardin et al., 1990). *per* does not appear to possess a bHLH region or other DNA-binding domain (M. Rosbash, personal communication). *arnt* is a subunit of the aromatic hydrocarbon receptor complex, which serves as a DNA-binding transcriptional activator of xenobiotic response genes (Whitlock, 1990; Hoffman et al., 1991). Interestingly, *arnt* is also a bHLH protein (Hoffman et al., 1991). The carboxy-terminal half of the *sim* protein contains a series of alanine-alanine-glutamine repeats, a

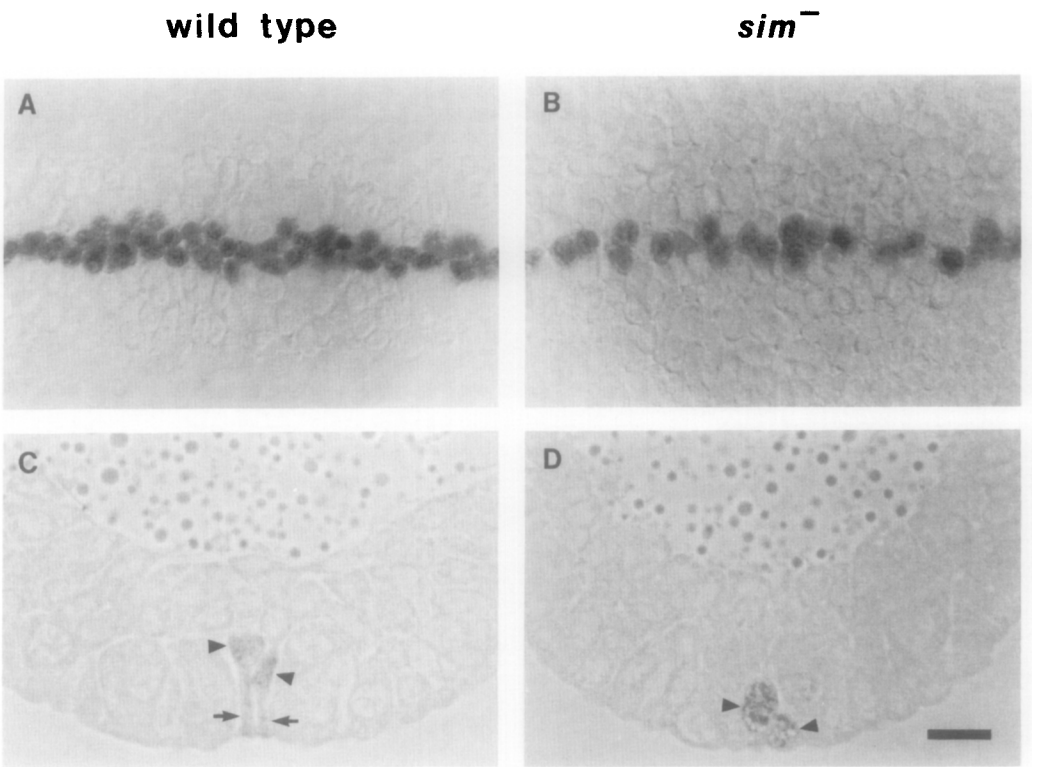


Figure 2. Defects in Early Midline Cell Development Observed in *sim* Mutant Embryos

Midline cell development was followed in wild-type (A and C) and *sim* mutant (B and D) embryos using the P[3.7*sim/lacZ*] marker. Expression of *lacZ* was detected via immunohistochemistry using a monoclonal antibody against β-galactosidase. These embryos were examined as whole mounts (A and B) or via 3 μm thick tissue cross sections (C and D). Bar, 16 μm (A and B); 12.8 μm (C and D).

(A) Ventral view of a stage 8/9 wild-type germband extended embryo. Anterior is to the left. By this stage a postgastrulation synchronous cell division has occurred.

(B) Ventral view of a stage 9 sim^{H9} mutant embryo. Note that there are approximately half as many midline cells present as in wild type, indicating that the cells did not divide synchronously.

(C) Cross section through a stage 10 wild-type embryo carrying P[3.7*sim/lacZ*] and stained with anti-β-galactosidase. Dorsal is up. During nerve cell precursor formation the midline cells extend nuclei (arrowheads) into the neuroblast layer. These cells maintain cytoplasmic extensions (arrows) to the epidermis.

(D) Cross section through a stage 10 sim^{H9} mutant embryo carrying P[3.7*sim/lacZ*] and stained with anti-β-galactosidase. The mutant midline cells (arrowheads) do not extend nuclei into the neuroblast precursor layer and do not possess cytoplasmic extensions. Cells appear rounded and are located along the epidermal cell layer.

proline-rich region, and a glutamine-rich region. In other regulatory proteins, proline-rich and glutamine-rich regions have been identified as potential transcriptional activation domains (Courey and Tjian, 1988; Mermod et al., 1989; Mitchell and Tjian, 1989).

The *sim* Gene Is Required for the Earliest Developmental Events of the CNS Midline Lineage

Since *sim* protein is present in the midline cells by gastrulation, we were interested in determining whether *sim* functions in the formation of midline nerve cell precursors during neurogenesis. In a previous study we utilized a P[7.8*sim/lacZ*] construct to determine that in *sim* mutant embryos, the midline cells do not properly differentiate into mature neurons and glia and fail to take their appropriate positions within the developing CNS (Nambu et al., 1990). In contrast, the midline nerve cell precursors appeared to form normally. However, since this P element construct contains a substantial amount of the *sim* coding region, including the bHLH region and PAS domain, fused to *lacZ*, it was noted that the fusion protein could provide a partial rescue of *sim* function.

We have here addressed the issue of *sim* function during nerve cell precursor formation through the use of a new P[*sim/lacZ*] transformant strain that allows us to follow the midline cells in a *sim* genetic null background. The construct, P[3.7*sim/lacZ*], contains 3.7 kb of *sim* 5′ flanking DNA and 18 amino acids of coding sequence fused in-frame to *lacZ* (Y. Kasai and S. T. Crews, unpublished data) and yields β-galactosidase expression in a pattern similar to the endogenous *sim* protein.

Two different *sim* mutant alleles were used in these studies, sim^{H9} *and* $sim^{B13.4}$. Both are likely to be null alleles, as they do not express any detectable *sim* protein. Hybridizations in situ do indicate that *sim* RNA is present in these alleles. Midline cell fate in wild-type and mutant embryos was assessed by following the P[3.7*sim/lacZ*] marker. These experiments indicate that in both sim^{H9} *and* $sim^{B13.4}$

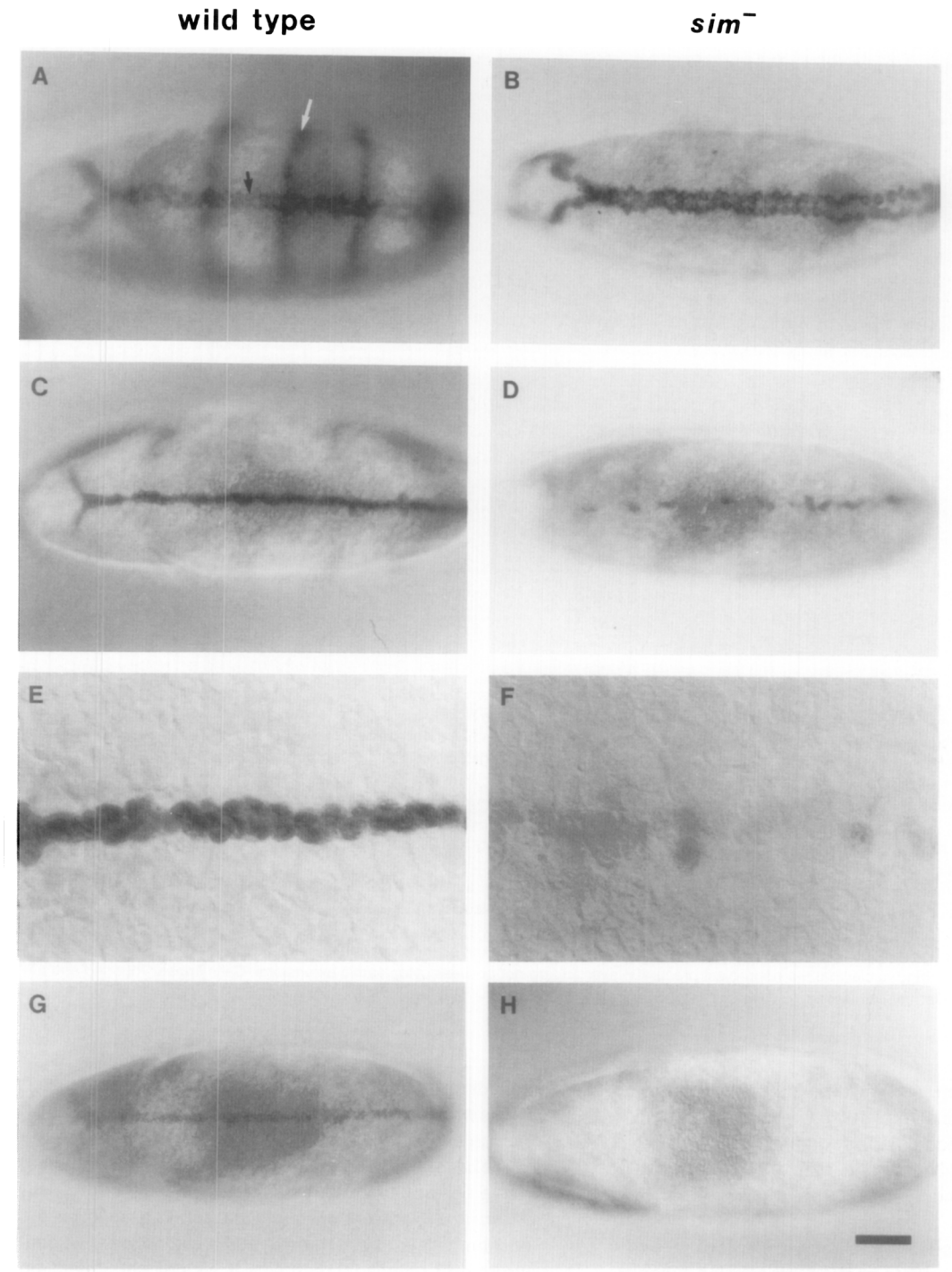


Figure 3. Autoregulation of *sim* Gene Expression

Expression from the *sim* early and *sim* late promoters was compared in both wild-type (A, C, E, and G) and *sim* mutant (B, D, F, and H) embryos. All views are ventral with anterior to the left. Bar, 64 μm (A–D, G, and H); 12.8 μm (E and F).

(A) Whole-mount in situ hybridization using a *sim* cDNA probe to detect RNA derived from the *sim* early promoter in a stage 7 wild-type embryo. The dark arrow indicates the position of midline cells. The light arrow indicates expression of P[*ftz/lacZ*] in pair rule stripes from a TM3 balancer chromosome that has been detected with a β-galactosidase gene probe. Gastrulation is nearly complete, and two rows of cells that strongly express *sim* have just come together at ventral midline.

(B) A stage 7 sim^{H9} mutant embryo. The mutant midline cells continue to express *sim* RNA at high levels and still appear morphologically normal.

(C) Whole-mount in situ hybridization of the *sim* cDNA probe to a stage 9 wild-type embryo, showing strong expression of *sim* RNA in midline nerve cell precursors.

(D) A stage 9 $sim^{B13.4}$ mutant embryo showing greatly reduced levels of *sim* RNA on a per-cell basis. There are also fewer cells expressing *sim* transcripts than in wild-type embryos.

embryos, the midline cells appear normal as they migrate to the ventral midline at gastrulation. Thereafter, as detected by inspection of midline cell nuclei, stage 9 *sim* mutant embryos (all stages are from Campos-Ortega and Hartenstein, 1985) have only half as many midline cells as in comparable wild-type embryos (compare Figures 2A and 2B). Thus, in *sim* mutants the midline cells fail to undergo a characteristic synchronized cell division.

In addition to a decreased number of midline cells in *sim* mutant embryos, morphological defects in these cells become apparent during nerve cell precursor formation. In contrast to wild type, the midline cell nuclei in *sim* mutants do not migrate into the nerve cell precursor layer; rather, they remain along the ventral epidermis (Figures 2C and 2D). Additionally, the mutant midline cells retain the rounded shape of blastoderm neuroectodermal cells; they do not take on the polarized morphology of developing midline nerve cells.

Autoregulation of *sim* Gene Expression

Because *sim* is expressed throughout midline development and is required for normal midline gene expression, we investigated whether *sim* has an autoregulatory function in controlling its own expression. Analysis of *sim* autoregulation was pursued by assessing levels of either *sim* RNA or P[3.7*sim/lacZ*] expression in sim^{H9} and $sim^{B13.4}$ mutant embryos. These studies, along with ectopic expression experiments described in a later section, provide evidence that *sim* positively autoregulates its expression in the midline cells.

The *sim* gene has two distinct promoters that have different, though overlapping, developmental profiles (Crews et al., 1988; Y. Kasai, S. T. Crews, and J. B. Thomas, unpublished data). As monitored via in situ hybridization to *sim* RNA, the *sim* early promoter is first activated in the cellular blastoderm and continues to be expressed through germband extension (Figures 3A and 3C). In *sim* mutant embryos, normal levels of *sim* RNA are detected through stage 7 (Figure 3B) but then rapidly diminish, and by stage 9 *sim* transcripts are significantly decreased on a per-cell basis (Figure 3D). Thus, *sim* is required for maintaining proper levels of its own transcription. This conclusion is supported by the finding that expression of the P[3.7*sim/lacZ*] marker is also greatly reduced from wild-type levels in individual cells of mutant embryos (Figures 3E and 3F). Expression from the *sim* late promoter was monitored by in situ hybridizations using a probe specific to the late transcript. In wild-type embryos (Figure 3G), hybridization is detected starting at stage 9. In *sim* mutant embryos (Figure 3H), there is no detectable hybridization, indicating that *sim* is required for activation of *sim* late promoter transcription.

Ectopic Expression of *sim* Results in the Transformation of Lateral Cells into CNS Midline Cells

As *sim* acts early in embryogenesis and is required for midline cell development, we asked whether ectopic *sim* protein might be able to direct the expression of midline properties in non-midline cells. This issue was addressed through the use of P[*hsp70/sim*] transformant fly strains in which the ectopic expression of *sim* protein is driven by the *hsp70* promoter. Staining with a *sim* antiserum indicates that, upon heat induction, high levels of *sim* protein are induced throughout the P[*hsp70/sim*] embryo.

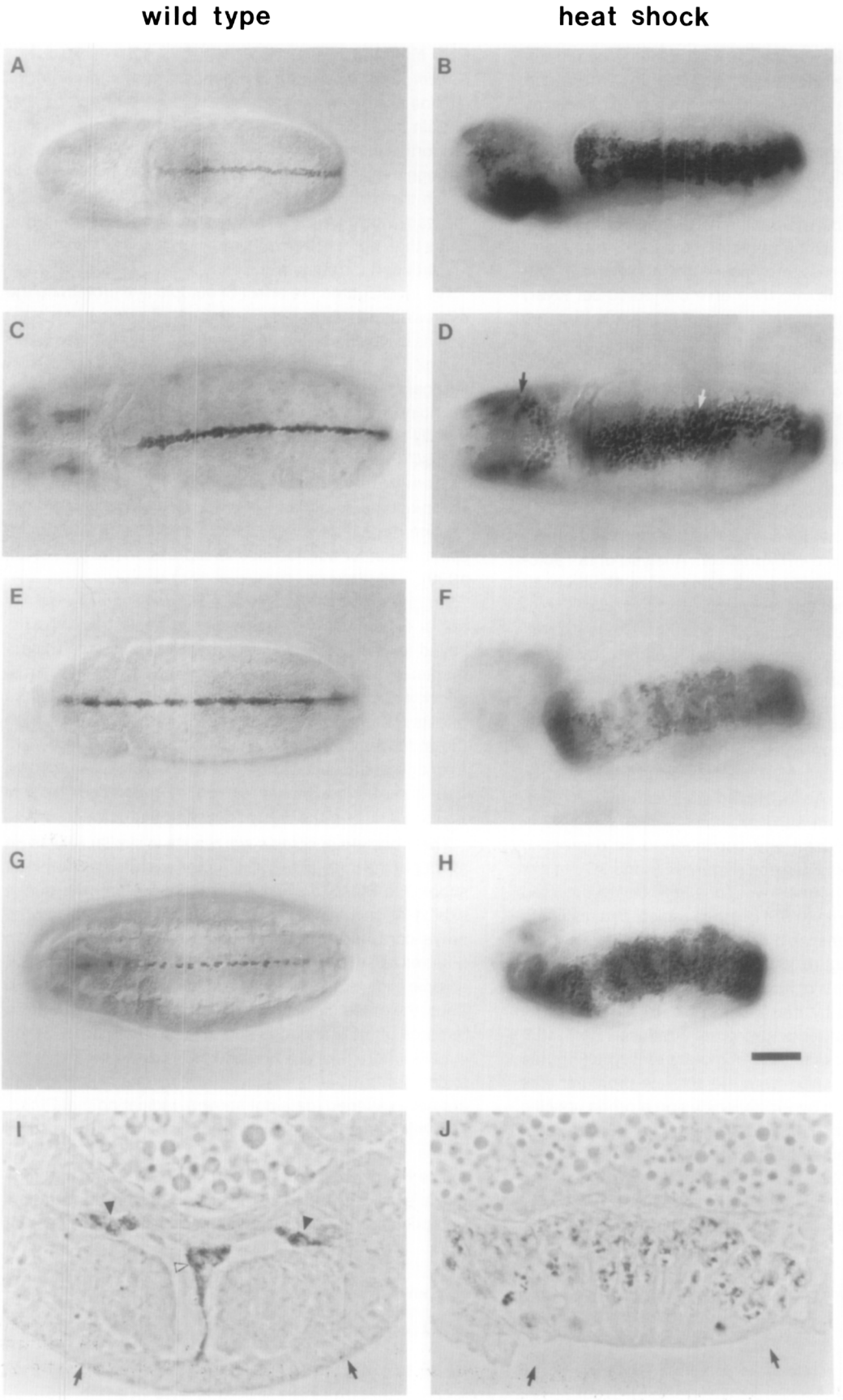
We investigated the effect of ectopic *sim* protein on the expression of midline genes through the use of reporter strains in which midline gene regulatory regions are fused to *lacZ*. Four genes that are prominently expressed in midline cells were analyzed (Figures 4A, 4C, 4E, and 4G), including *sim*, *rhomboid* (*rho*), *center divider* (*cdi*—a gene initially identified by enhancer trap insertions into 91F), and *slit* (*sli*). P[3.7*sim/lacZ*], P[*rho/lacZ*], and P[*cdi/lacZ*] 242 are expressed in the midline cells throughout most of embryogenesis, while P[1.0*sli/lacZ*] is a marker for later midline gene expression. All four of these genes require *sim* function for normal midline expression (Nambu et al., 1990).

Embryos from P[*hsp70/sim*] strains bearing *lacZ* reporter constructs were subjected to heat shock and assayed for the expression of β-galactosidase (see Experimental Procedures). For these experiments, the heat shock was delivered during early embryogenesis. All the effects of heat shock described below are dependent upon the induction of ectopic *sim* protein in embryos 0–4 hr after egg laying and are not observed in wild-type embryos heat shocked in a similar fashion, or in non-shocked embryos containing the P[*hsp70/sim*] construct.

The induction of ectopic *sim* results in similar effects on all four of the midline reporter genes examined: After heat shock in a P[*hsp70/sim*] background, the reporter genes exhibit strong ectopic expression in the entire ventral nerve cord and in a broad band in the cephalic region (Figures 4B, 4D, 4F, and 4H). Other tissues did not exhibit ectopic midline gene transcription under these conditions. Thus, the presence of ectopic *sim* can alter the expression patterns of midline genes and activate their transcription in cells where they are normally not expressed. In similar experiments, the expression patterns of lateral CNS genes are not expanded (data not shown), indicating that P[*hsp70/sim*]-induced gene expression is specific to midline genes.

The heat shocked embryos exhibit severe morphological defects in the organization of the CNS and head region. In particular, there are significant effects on formation of

(E) A stage 11 wild-type embryo expressing the P[3.7*sim/lacZ*] reporter construct, which has been revealed by anti-β-galactosidase immunohistochemistry. Anterior is to the left. All midline cells continue to exhibit strong and uniform expression of the marker gene.
(F) A stage 10 sim^{H9} mutant embryo, indicating that by this stage the expression of *lacZ* from the P[3.7*sim/lacZ*] marker is greatly reduced in the absence of *sim* function.
(G) Whole-mount in situ hybridization to a stage 10 wild-type embryo, using a probe consisting of a 5.4 kb BamHI restriction fragment from a *sim* genomic DNA clone that is specific for the *sim* late promoter transcript. Note uniform expression of this transcript in all midline cells.
(H) A stage 10 sim^{H9} mutant embryo, indicating a lack of transcription from the *sim* late promoter.


wild type
heat shock
A
B
C
D
E
F
G
H
I
J

the CNS midline. Tissue sections through wild-type embryos stained with a *sim* antiserum illustrate the characteristic morphology of developing midline nerve cells as well as their position relative to the lateral CNS and ventral epidermis (Figure 4I). In embryos with ectopic *sim* expression, not only do all of the CNS cells express midline genes, but they also possess extended nuclei and long cytoplasmic projections (Figure 4J). This resembles the morphology of CNS midline cells and is distinct from the usual appearance of lateral neuroblasts and neurons. Furthermore, there is a lack of ventral epidermis in heat shocked embryos (Figure 4J), as well as a corresponding loss of medial denticle hairs in the ventral cuticle (data not shown). The ventral epidermis is derived from the same ectodermal precursor cells that give rise to the lateral CNS. In contrast, the CNS midline precursors do not give rise to epidermal cells. Thus, one interpretation of these results is that there is a transformation of ventral ectodermal cells into CNS midline cells, resulting in an overproduction of the CNS midline at the expense of both lateral CNS and ventral epidermis.

Discussion

The *sim* Gene Encodes a Midline Transcription Factor

Sequence analysis indicates that *sim* is a member of the bHLH family of transcription factors and contains both DNA-binding and protein dimerization motifs. This suggests that *sim* acts as a transcriptional regulator, a notion consistent with both the nuclear localization of *sim* and the requirement for midline cell gene expression. *sim* is thus likely to act by forming homodimers and/or heterodimers with other HLH proteins. It is not yet known what dimerization partner(s) *sim* interacts with. One requisite of any *sim* partner is that it be expressed in the midline cells. It might also be expected that mutations in such a partner would have significant effects on midline cell development and gene expression. Finally, an interpretation of the discrete spatial and temporal restrictions of the effects of ectopic *sim* expression is that a heterologous partner may be present in the ventral neuroectoderm and cephalic region during early embryogenesis. Along these lines, *daughterless* (Jimenez and Campos-Ortega, 1990) and genes of the *achaete-scute* complex (Cabrera et al., 1987; Romani et al., 1987) and *Enhancer-of-split* locus (Knust et al., 1987) meet some of these criteria and represent potential candidates for *sim* partners. A determination of whether any of these genes does in fact interact with *sim* will require both genetic and biochemical investigations. It is also possible that *sim* requires a novel type of partner, similar to the relationship between *myc* and *Max* (Blackwood and Eisenman, 1991) or *Myn* (Prendergast et al., 1991), which interact via bHLH and adjacent leucine zipper domains.

The PAS domain defines a protein family with at least three members, *sim*, *per*, and *arnt*, each of which is thought to have gene regulatory activities. An important aspect of *per* function appears to be in establishing a feedback loop to regulate its own expression in a circadian fashion (Hardin et al., 1990). *arnt* is a bHLH protein component of a heterologimeric receptor that regulates the transcription of genes involved in hydrocarbon metabolism (Whitlock, 1990; Hoffman et al., 1991). The spacing between the bHLH and PAS domains in *arnt* and *sim* is nearly identical; *sim* and *arnt* thus represent a novel bHLH subtype that may have distinct functional properties. A determination of the function of the PAS domain will be crucial for a better understanding of the activity of these three distinctive proteins.

sim Is a Master Developmental Regulator of the CNS Midline Lineage

We previously demonstrated that *sim* is required for the proper differentiation of the midline cells into mature neurons and glia (Nambu et al., 1990). These findings have now been extended to show that *sim* is also required for the earliest events in midline cell development, including a lineage-specific cell division that normally takes place shortly after *sim* protein has accumulated in midline cell nuclei. *sim* could act to regulate the expression or activity of genes required for this division, such as the *string* mitotic regulator (Edgar and O'Farrell, 1989). *sim* is also required for the proper formation of midline nerve cell precursors during neurogenesis. In *sim* mutant embryos the midline cell nuclei do not extend into the nerve cell precursor layer, and the cells fail to adopt their characteristic polarized morphology. This effect is similar to the defects in neuroblast formation observed in mutations of the proneural *achaete-scute* complex and *daughterless* genes (Jimenez and Campos-Ortega, 1990). Additionally, the autoregulation of *sim* gene expression is also similar to that of the

Figure 4. Ectopic Expression of *sim* Results in Transformation of Lateral Cells to the CNS Midline

(A–H) Embryos carrying two or more copies of the P[*hsp70/sim*] construct and one or two copies of midline-expressed P[*lacZ*] reporter constructs were subjected to heat shock and stained with anti-β-galactosidase. Dorsal (A–D) and ventral (E–H) views are presented. Anterior is to the left. Wild-type expression of P[3.7*sim/lacZ*] (A), P[*rho/lacZ*] (C), P[*cdi/lacZ*]242 (E), and P[1.0*sli/lacZ*] (G) are compared with expression in P[*hsp70/sim*] backgrounds (B, D, F, and H, respectively). In wild-type backgrounds, three genes yield expression in all midline cells of germband extended embryos (A, C, and E), while one gene exhibits later midline expression (G). In heat shocked embryos, note that all midline genes show a large domain of expression in most or all cells of the lateral CNS (D, white arrow). In addition, there is also ectopic expression of the genes in cells of the cephalic region (D, dark arrow). Bar, 64 μm.

(I and J) Cross section (3 μm) through a stage 12 wild-type embryo stained with a *sim* antiserum (I) and a heat shocked P[*hsp70/sim*] embryo carrying the P[7.8*sim/lacZ*] reporter construct and stained with anti-β-galactosidase (J). Dorsal is up. In the wild-type embryo (I), the midline cells have extended their nuclei (white arrowhead) into the dorsal surface of the nerve cord while maintaining cytoplasmic extensions to the epidermis (dark arrows). Arrowheads show non-midline cells above the CNS, also stained with the *sim* antiserum. In heat shocked embryos (J) there is an increase in the number of staining cells that exhibit midline-like morphology. There is a lack of unstained lateral CNS and of ventral epidermis in the expanded region of P[7.8*sim/lacZ*] expression (arrows). Bar, 12.8 μm.

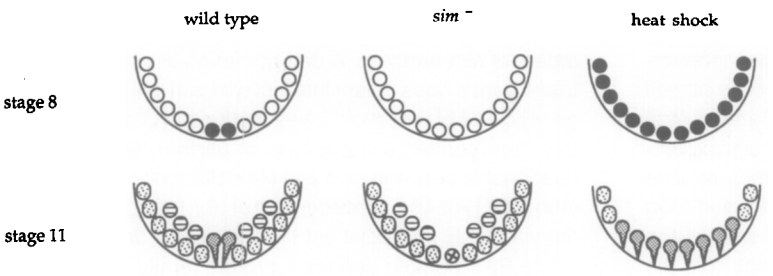


Figure 5. Function of *sim* as a Master Regulator of CNS Midline Development

Schematic representation of cross sections through ventral portions of wild-type, *sim* mutant, and heat shocked P[*hsp70/sim*] embryos at stages 8 and 11. In wild-type embryos, *sim* protein is expressed in a unique subset of cells (black) in the neuroectoderm and directs them to develop into the CNS midline. These cells (heavy stipple) take on distinctive morphologies and exhibit characteristic patterns of gene expression that differ from those of the lateral CNS (horizontal lines) and ventral epidermis (light stipple). In *sim* mutants, *sim* protein is not expressed in any cells of the neuroectoderm. In the absence of *sim* function (cross), the midline cells fail to undergo a synchronous postgastrulation cell division and do not properly form nerve cell precursors. The mutant midline cells do not exhibit normal morphology, nor do they express any downstream midline genes. Ultimately they fail to give rise to mature neurons and glia. In heat shocked P[*hsp70/sim*] embryos, *sim* protein is ectopically induced in all cells and directs most or all neuroectodermal cells to exhibit midline cell morphologies and patterns of gene expression. These embryos exhibit a corresponding defect in the formation of the ventral epidermis, while dorsolateral epidermis is still formed. *sim* is thus necessary for normal midline cell development, and ectopic *sim* expression is sufficient to transform the lateral CNS into CNS midline.

achaete-scute complex (Martinez and Modolell, 1991), as well as *MyoD* (Thayer et al., 1989; Weintraub et al., 1991), and is likely to be an important mechanism for maintaining the differentiation program of particular cell types.

The overexpression of *sim* protein in P[*hsp70/sim*] strains has important consequences for midline development. These experiments indicate that ectopic *sim* can induce other cell types to exhibit midline morphology and patterns of gene expression. Thus, the presence of *sim* protein early in embryogenesis can override the normal differentiation programs of cells in the lateral and cephalic regions and redirect them into the midline lineage. The spatial and temporal specificity of these effects implies that the responding cells must have a particular combination of gene products present in order to be sensitive to *sim*. This cellular transformation is reminiscent of the ability of *MyoD* and related genes to convert a variety of different cultured cell types to muscle cells (Davis et al., 1987; Choi et al., 1990; Weintraub et al., 1991) and of the effects of ectopic *achaete-scute* complex gene expression, in which ectopic sensory organs are formed (Alonso and Garcia-Bellido, 1986; Campuzano et al., 1986; Balcells et al., 1988, Rodriguez et al., 1990). As is also the case for *MyoD* and the *achaete-scute* complex, the transformation observed with ectopic *sim* expression is not ubiquitous, but is generally restricted to related cell types.

The results of cell fate experiments presented here and in previous studies indicate that *sim* is required for all developmental steps specific to the CNS midline lineage, including early events in the formation of nerve cell precursors as well as later steps in the differentiation of mature neurons and glia. Coupled with the cellular transformations observed with ectopic *sim* expression, these data suggest that *sim* acts as a master regulator of CNS midline cell development (Figure 5), which normally directs a discrete subset of the neuroectoderm to develop into the CNS midline.

The hierarchy of events underlying CNS midline development is proposed as follows: Maternal and zygotic genes that specify positional information along the dorsoventral axis activate *sim* at a unique position in the blastoderm. Through interactions with itself and/or other bHLH proteins, *sim* activates transcription of specific midline genes and in doing so initiates the developmental program of the CNS midline cell lineage. This program is maintained by a positive autoregulatory function of *sim* and is ultimately carried out by the functions of a large number of other midline-expressed genes, which all directly or indirectly require *sim* function.

Conserved Functions of Midline-Positioned Cells

The midline cells of the insect CNS constitute a distinct neuroanatomical structure whose development clearly differs from that of the rest of the nervous system. Although composed of functional neurons and glia, these cells have unique functions in the developing and mature CNS. It may be that, in the construction of a symmetric nervous system, there exist unique and conserved functions provided by specialized midline-positioned cells. In vertebrates, the floorplate cells lying at the base of the developing neural tube share general properties with the insect midline cells (Jessell et al., 1989; Klämbt et al., 1991). Both the midline and floorplate cells are thought to attract and interact with crossing commissural nerve cell growth cones, inducing changes in their direction of growth and expression of cell surface proteins (Bastiani et al., 1987; Dodd et al., 1988; Harrelson and Goodman, 1988; Tessier-Lavigne et al., 1988; Furley et al., 1990; Klämbt et al., 1991). Interestingly, genetic studies of zebrafish embryogenesis have identified a gene, *cyclops*, that is required for floorplate development (Hatta et al., 1991). *cyclops* mutants exhibit defective CNS formation, including fusion of the two lateral optic lobes. This phenotype is not unlike that of *sim* mutants and further suggests overlapping functions of midline and floorplate cells. Given the high degree of conservation exhibited by other bHLH proteins, including *twist* (Hopwood et al., 1989; Wolf et al., 1991), *achaete-scute* (Johnson et al., 1990), *daughterless* (Murre et al., 1989), and *MyoD* (Michelson et al., 1990), it will be of interest to determine whether a vertebrate homolog of *sim* functions in floorplate cell development.

Experimental Procedures

Drosophila Strains

The sim^{H9} and $sim^{B13.4}$ mutant alleles were obtained from the A. Chovnick laboratory and were maintained over a TM3 balancer chromosome containing a P[*ftz/lacZ*] construct, which directs the expression of β-galactosidase in strong pair rule stripes. This allowed for the unambiguous selection of homozygous mutant embryos by virtue of their lack of P[*ftz/lacZ*] expression.

Isolation and Characterization of *sim* cDNA Clones

sim cDNA clones were isolated from a 4–8 hr embryonic cDNA library kindly provided by N. Brown (Brown and Kafatos, 1988). The library was screened using restriction fragments from previously characterized *sim* cDNA and genomic DNA clones. Two clones contained sequences at their 5′ ends that extended well beyond those of previously described clones (Crews et al., 1988). These clones are estimated to be nearly full length and contain all of the *sim* protein coding region, as well as up to 511 bp of 5′ untranslated region. Sequence analysis of these two clones was performed using the dideoxy chain termination method (Sanger et al., 1977).

Midline Cell Markers

P[3.7sim/lacZ]

This strain was generated (Y. Kasai and S. T. Crews, unpublished data) by fusing a 3.7 kb BamHI–EcoRI restriction fragment of the *sim* gene, which contains the *sim* early promoter and 5′ regulatory elements, to *lacZ* in the P element vector, CaSpeR-β-gal (Thummel et al., 1988). The first 18 amino acids of the *sim* coding region are fused in-frame to *lacZ*. This strain exhibits β-galactosidase expression similar to native *sim* expression. Midline expression of β-galactosidase protein is first observed at stage 7 in all of the midline cells.

P[rho/lacZ]

This enhancer trap line was identified in our lab and contains a P element insertion into the *rho* locus (Nambu et al., 1990). The line exhibits expression patterns of β-galactosidase in a similar fashion to the *rho* gene (Bier et al., 1990). P[*rho/lacZ*] is expressed in all of the midline cells from stage 7 throughout embryogenesis.

P[cdi/lacZ]242

This enhancer trap line has a P element insertion into the *cdi* gene at 91F. *cdi* is expressed in the CNS midline cells, epidermis, and proventriculus (B. Matthews, Y. Hiromi, C. S. Goodman, and S. T. Crews, unpublished data); however, the P[*cdi/lacZ*]242 insertion shows only CNS midline expression (Nambu et al., 1990). Expression of this marker is observed in all of the midline cells from stage 9 and is maintained throughout embryogenesis.

P[1.0sli/lacZ]

This line was constructed by fusing a 1.0 kb EcoRV–HindIII fragment of the *sli* gene to *lacZ* in the vector CPLZ. CPLZ is an enhancer-tester vector that uses the mini-*white* gene of CaSpeR (Pirrotta, 1988) and the P element promoter (up to −42 relative to the cap site) fused in-frame to *lacZ* to produce a transposase–β-galactosidase fusion protein. The *sli* gene is prominently expressed in the CNS midline, first in all of the precursors during germband extension and later in only the midline glia. The P[1.0*sli/lacZ*] construct yields highly restricted expression in only the midline glia beginning at stage 11 and continuing throughout embryogenesis.

Generation and Use of P[*hsp70/sim*] Strains

The P[*hsp70/sim*] construct was generated using a 2.7 kb DraI–NdeI *sim* cDNA fragment that contains all of the *sim* protein coding region, 167 bp of 5′ untranslated region, and 502 bp of 3′ untranslated region. KpnI linkers were added to this fragment, which was then cloned into the KpnI site of the pHT4 P element vector containing the *hsp70* promoter and 3′ trailer sequences (Schneuwly et al., 1987). This construct was coinjected with p 25.7wc (Karess and Rubin, 1984) into ry^{506} host flies (Rubin and Spradling, 1982). Five independent transformant fly lines with similar phenotypic characteristics were obtained.

In the heat shock experiments, embryos bearing two or more copies of P[*hsp70/sim*] and one or more copies of a midline gene marker were collected between 0 and 4 hr after egg laying. These embryos were dechorionated in 50% bleach, subjected to a heat pulse of 37°C for 1–1.5 hr, and then allowed to recover at 25°C for 2 hr or more. The embryos were subsequently prepared for immunocytochemical analysis as described below.

Immunohistochemistry and In Situ Hybridization

Immunohistochemical staining with anti-β-galactosidase (Promega) was performed as previously described (Nambu et al., 1990) except that some embryos were mounted in 80% Canada balsam, 20% methyl-salicylate (Lawrence et al., 1986). Whole-mount in situ hybridizations were carried out using digoxigenin-labeled probes (Boehringer Mannheim) as described by Tautz and Pfeifle (1989). The *sim* probe was a full-length cDNA, and the probe specific for the late promoter transcript consisted of a 5.4 kb BamHI genomic DNA restriction fragment that contains 5′ untranslated sequences not present in the early transcript. Stained embryos were viewed and photographed under Nomarski optics.

Tissue Sectioning

Embryos stained with anti-β-galactosidase or a *sim* antiserum were dehydrated through an ethanol series, which terminated in two 10 min incubations in 100% ethanol. Embryos were then transferred to acetone and left overnight in a staining dish with a 1:1 mixture of acetone and Epon-Araldite. Individual embryos were oriented in blocks of fresh Epon-Araldite and baked at 60°C for 12–15 hr. Sections (3 μm) were cut on a LKB Ultrotome, mounted with Permount, and photographed under phase–contrast optics.

Acknowledgments

The authors would particularly like to acknowledge the contribution of Steve Kohtz, who observed the similarity between *sim* and the bHLH family of proteins. We are indebted to Yumi Kasai for the P[3.7*sim/lacZ*] line and to Jonathan Rothberg and Spyros Artavanis-Tsakonas for the *slit* genomic DNA and for helpful advice. We would also like to thank David Eisenberg, Oliver Hankinson, Amelia Hartenstein, Emily Hoffman, Francesca Pignoni, and Michael Rosbash for useful advice and discussions. We appreciate the comments of Utpal Banerjee, Eddy DeRobertis, Judith Lengyel, and Larry Zipursky on the manuscript. The work of S. T. C. was supported by National Institutes of Health grant R01 HD25251. J. R. N. is supported by an NIH postdoctoral fellowship. J. O. L. and K. A. W. are supported by NIH predoctoral (GM-07185) and Medical Scientist Training Program (GM-08042) training grants. S. T. C. is a Lucille P. Markey Scholar, and this work was supported by a grant from the Lucille P. Markey Charitable Trust.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore hereby be marked "*advertisement*" in accordance with 18 USC Section 1734 solely to indicate this fact.

Received July 31, 1991; revised October 4, 1991.

References

Alonso, L. G., and Garcia-Bellido, A. (1986). Genetic analysis of *Hairy-wing* mutations. Roux's Arch. Dev. Biol. *195*, 259–264.

Balcells, L., Modolell, J., and Ruiz-Gomez, M. (1988). A unitary basis for different *Hairy-wing* mutations of *Drosophila melanogaster*. EMBO J. *7*, 3899–3906.

Bastiani, M. J., Harrelson, A. L., Snow, P. M., and Goodman, C. S. (1987). Expression of fasciclin I and II glycoproteins on subsets of axon pathways during neuronal development in the grasshopper. Cell *48*, 745–755.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990). The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell *61*, 49–59.

Bier, E., Jan, L. Y., and Jan, Y. N. (1990). *rhomboid*, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster*. Genes Dev. *3*, 190–203.

Blackwood, E. M., and Eisenman, R. N. (1991). *Max*: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with *myc*. Science *251*, 1211–1217.

Brown, N. M., and Kafatos, F. C. (1988). Functional cDNA libraries from *Drosophila* embryos. J. Mol. Biol. *203*, 425–437.

Cabrera, C. V., Martinez-Arias, A., and Bate, M. (1987). The expression of three members of the *achaete-scute* gene complex correlates with neuroblast segregation in Drosophila. Cell *50*, 425–433.

Campos-Ortega, J. A., and Hartenstein, V. (1985). The Embryonic Development of *Drosophila melanogaster* (New York: Springer-Verlag).

Campuzano, S., Balcells, L., Villares, R., Carramolino, L., Garcia-Alonso, L., and Modolell, J. (1986). Excess function *Hairy-wing* mutations caused by *gypsy* and *copia* insertions within structural genes of the *achaete-scute* locus of Drosophila. Cell *44*, 303–312.

Choi, J., Costa, M. L., Mermelstein, C. S., Chagas, C., Holtzer, S., and Holtzer, H. (1990). *MyoD* converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. Proc. Natl. Acad. Sci. USA *87*, 7988–7992.

Courey, A. J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell *55*, 887–898.

Crews, S. T., Thomas, J. B., and Goodman, C. S. (1988). The Drosophila *single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. Cell *52*, 143–151.

Crews, S., Franks, R., Hu, S., Matthews, B., and Nambu, J. (1991). The *Drosophila single-minded* gene and the molecular genetics of CNS midline development. J. Exp. Zool., in press.

Davis, R. L., Weintraub, H., and Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Davis, R. L., Cheng, P.-F., Lassar, A. B., and Weintraub, H. (1990). The MyoD DNA binding domain contains a recognition code for muscle-specific gene activation. Cell *60*, 733–746.

Dodd, J., Morton, S. B., Karagogeos, D., Yamamoto, M., and Jessell, T. M. (1988). Spatial regulation of axonal glycoprotein expression on subsets of embryonic spinal neurons. Neuron *1*, 105–116.

Edgar, B. A., and O'Farrell, P. H. (1989). Genetic control of cell division patterns in the Drosophila embryo. Cell *57*, 177–187.

Foe, V. E. (1989). Mitotic domains reveal early commitment of cells in *Drosophila* embryos. Development *107*, 1–22.

Furley, A. J., Morton, S. B., Manato, D., Karagogeos, D., Dodd, J., and Jessell, T. M. (1990). The axonal glycoprotein TAG-1 is an immunoglobulin superfamily member with neurite outgrowth–promoting activity. Cell *61*, 157–170.

Hardin, P. E., Hall, J. C., and Rosbash, M. (1990). Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. Nature *343*, 536–540.

Harrelson, A. L., and Goodman, C. S. (1988). Growth cone guidance in insects: *fasciclin* II is a member of the immunoglobulin superfamily. Science *242*, 700–708.

Hatta, K., Kimmel, C. B., Ho, R. K., and Walker, C. (1991). The *cyclops* mutation blocks specification of the floor plate of the zebrafish central nervous system. Nature *350*, 339–341.

Hoffman, E. C., Reyes, H., Chu, F.-F., Sander, F., Conley, L. H., Brooks, B. A., and Hankinson, O. (1991). Cloning of a factor required for activity of the Ah (dioxin) receptor. Science *252*, 954–958.

Hopwood, N. D., Pluck, A., and Gurdon, J. B. (1989). A Xenopus mRNA related to Drosophila *twist* is expressed in response to induction in the mesoderm and the neural crest. Cell *59*, 893–903.

Jacobs, J. R., and Goodman, C. S. (1989a). Embryonic development of axon pathways in the *Drosophila* CNS. I. A glial scaffold appears before the first growth cones. J. Neurosci. *9*, 2402–2411.

Jacobs, J. R., and Goodman, C. S. (1989b). Embryonic development of axon pathways in the *Drosophila* CNS. II. Behavior of pioneer growth cones. J. Neurosci. *9*, 2412–2422.

Jessell, T. M., Bovolenta, P., Placzek, M., Tessier-Lavigne, M., and Dodd, J. (1989). Polarity and patterning in the neural tube: the origin and function of the floor plate. In Cellular Basis of Morphogenesis (Ciba Foundation Symposium 144), D. Evered and J. Marsh, eds. (Chichester, England: John Wiley & Sons), pp. 255–280.

Jimenez, F., and Campos-Ortega, J. A. (1990). Defective neuroblast commitment in mutants of the *achaete-scute* complex and adjacent genes of D. melanogaster. Neuron *5*, 81–89.

Johnson, J. E., Birren, S. J., and Anderson, D. J. (1990). Two rat homologues of *Drosophila achaete-scute* specifically expressed in neuronal precursors. Nature *346*, 858–861.

Karess, R. E., and Rubin, G. M. (1984). Analysis of P transposable element functions in Drosophila. Cell *38*, 135–146.

Klämbt, C., Jacobs, J. R., and Goodman, C. S. (1991). The midline of the Drosophila central nervous system: a model for the genetic analysis of cell fate, cell migration, and growth cone guidance. Cell *64*, 801–815.

Knust, E., Tietze, K., and Campos-Ortega, J. A. (1987). Molecular analysis of the neurogenic locus *Enhancer of split* of *Drosophila melanogaster*. EMBO J. *6*, 4113–4123.

Lawrence, P. A., Johnston, P., and Morata, G. (1986). Methods of marking cells. In *Drosophila*: A Practical Approach, D. B. Roberts, ed. (Oxford, England: IRL Press), pp. 229–242.

Ludwig, S. R., Habera, L. F., Dellaporta, S. L., and Wessler, S. R. (1989). *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the *myc*-homology region. Proc. Natl. Acad. Sci. USA *86*, 7092–7096.

Martinez, C., and Modolell, J. (1991). Cross-regulatory interactions between the proneural *achaete* and *scute* genes of *Drosophila*. Science *251*, 1485–1487.

Mellentin, J. D., Smith, S. D., and Cleary, M. L. (1989). *lyl-1*, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell *58*, 77–83.

Mermod, N., O'Neill, E. A., Kelly, T. J., and Tjian, R. (1989). The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. Cell *58*, 741–753.

Michelson, A. M., Abmayr, S. M., Bate, M., Martinez-Arias, A., and Maniatis, T. (1990). Expression of a *MyoD* family member prefigures muscle pattern in *Drosophila* embryos. Genes Dev. *4*, 2086–2097.

Mitchell, P. M., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Murre, C., McCaw, P. S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. Cell *56*, 777–783.

Nambu, J. R., Franks, R. G., Hu, S., and Crews, S. T. (1990). The *single-minded* gene of Drosophila is required for the expression of genes important for the development of CNS midline cells. Cell *63*, 63–75.

Pirrotta, V. (1988). Vectors for P-mediated transformation in *Drosophila*. In Vectors, A Survey of Molecular Cloning Vectors and Their Uses, R. L. Rodriguez and D. T. Denhardt, eds. (Boston: Butterworths), pp. 437–456.

Prendergast, G. C., Lawe, D., and Ziff, E. B. (1991). Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. Cell *65*, 395–407.

Rodriguez, I., Hernandez, R., Modolell, J., and Ruiz-Gomez, M. (1990). Competence to develop sensory organs is temporally and spatially regulated in *Drosophila* epidermal primordia. EMBO J. *9*, 3583–3592.

Romani, S., Campuzano, S., and Modolell, J. (1987). The *achaete-scute* complex is expressed in neurogenic regions of *Drosophila* embryos. EMBO J. *6*, 2085–2092.

Rubin, G. M., and Spradling, A. C. (1982). Genetic transformation of *Drosophila* with transposable element vectors. Science *218*, 348–353.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schneuwly, S., Klemenz, R., and Gehring, W. J. (1987). Redesigning the body plan of *Drosophila* by ectopic expression of the homeotic gene *Antennapedia*. Nature *325*, 816–818.

Siwicki, K. K., Eastman, C., Petersen, G., Rosbash, M., and Hall, J. C. (1988). Antibodies to the *period* gene product of Drosophila reveal

diverse tissue distribution and rhythmic changes in the visual system. Neuron *1*, 141–150.

Sweet, R. M., and Eisenberg, D. (1983). Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J. Mol. Biol. *171*, 479–488.

Tautz, D., and Pfeifle, C. (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 *98*, 81–85.

Tessier-Lavigne, M., Placzek, M., Lumsden, A. G. S., Dodd, J., and Jessell, T. M. (1988). Chemotropic guidance of developing axons in the mammalian central nervous system. Nature *336*, 775–778.

Thayer, M. J., Tapscott, S. J., Davis, R. L., Wright, W. E., Lassar, A. B., and Weintraub, H. (1989). Positive autoregulation of the myogenic determination gene MyoD1. Cell *58*, 241–248.

Thomas, J. B., Bastiani, M. J., Bate, C. M., and Goodman, C. S. (1984). From grasshopper to *Drosophila*: a common plan for neuronal development. Nature *310*, 203–207.

Thomas, J. B., Crews, S. T., and Goodman, C. S. (1988). Molecular genetics of the *single-minded* locus: a gene involved in the development of the Drosophila nervous system. Cell *52*, 133–141.

Thummel, C. S., Boulet, A. M., and Lipshitz, H. D. (1988). Vectors for *Drosophila* P-element mediated transformation and tissue culture transfection. Gene *74*, 445–456.

Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benezra, R., Blackwell, T. K., Turner, D., Rupp, R., Hollenberg, S., Zhuang, Y., and Lassar, A. (1991). The *myoD* gene family: nodal point during specification of the muscle cell lineage. Science *251*, 761–766.

Whitlock, J. P., Jr. (1990). Genetic and molecular aspects of 2,3,7,8-tetra-chlorodibenzo-p-dioxin action. Annu. Rev. Pharmacol. Toxicol. *30*, 251–277.

Wolf, C., Thisse, C., Stoetzel, C., Thisse, B., Gerlinger, P., and Perrin-Schmitt, F. (1991). The *M-twist* gene of *Mus* is expressed in subsets of mesodermal cells and is closely related to the *Xenopus X-twi* and the *Drosophila twist* genes. Dev. Biol. *143*, 363–373.