ELSEVIER SCIENCE IRELAND

Mechanisms of Development 45 (1994) 269–277

Transcriptional activation domains of the single-minded bHLH protein are required for CNS midline cell development

Robert G. Franks, Stephen T. Crews *

Molecular Biology Institute and Department of Biology, University of California, Los Angeles, Los Angeles, CA 90024, USA

(Received 22 November 1993; accepted 2 December 1993)

Abstract

The *single-minded* gene functions as a master developmental regulator within the midline cell lineage of the embryonic central nervous system of *Drosophila melanogaster*. Genetic experiments suggest that Single-minded can function as a transcriptional activator. Regions of the Single-minded protein were fused to the DNA binding domain of the mammalian transcription factor Sp1 and shown to activate transcription from a reporter gene linked to Sp1 binding sites. Three independent activation domains were identified in the carboxy terminal region of Single-minded that include areas rich in serine, threonine, glutamine and proline residues. Germ line transformation experiments indicate that the carboxy terminal activation domains, the PAS dimerization domain, and the putative DNA binding basic domain of Single-minded are required for expression of CNS midline genes in vivo. These results define in vivo a functional activation domain within Single-minded and suggest a model in which Single-minded activates transcription through a direct interaction with promoter elements of CNS midline genes.

Key words: Basic-helix-loop-helix; *Drosophila*; PAS domain; *single-minded*; Transcriptional activator

1. Introduction

The *single-minded* (*sim*) gene functions as a genetic switch which instructs a specific subset of neuroectodermal cells to develop into central nervous system (CNS) midline cells (Thomas et al., 1988; Nambu et al., 1990, 1991). Genetic analysis reveals that *sim* is required for all known developmental steps of the CNS midline cell lineage including the formation of CNS midline precursor cells and their subsequent differentiation into mature neurons and glia. Additional experiments indicate that *sim* is required for the transcription of all genes with distinct expression in the CNS midline cells. Consistent with its critical role in CNS midline cell development, ectopic expression of Sim can transform lateral neuroectodermal cells into CNS midline cells.

The *sim* gene encodes a nuclear protein that has a structure consistent with a direct role in transcriptional regulation (Nambu et al., 1991). Sequence analysis reveals the presence of a basic-helix-loop-helix (bHLH) motif at the amino terminus. Functional studies on other bHLH proteins have shown that the basic region is involved in DNA binding and the HLH region mediates dimerization (Murre et al., 1989; Davis et al., 1990; Voronova and Baltimore, 1990). Most often bHLH proteins function in vivo as heterodimers (Lassar et al., 1991; Amati et al., 1992). The bHLH region of Sim is followed by a second domain, the Per-Arnt-Sim (PAS) homology domain (Crews et al., 1988), that is also found adjacent to the bHLH domains of both subunits of the mammalian aryl hydrocarbon receptor complex (AHRC) (Hoffman et al., 1991; Burbach et al. 1992; Ema et al., 1992). Sim and the AHRC proteins are members of a distinct family of transcriptional regulators that contain both bHLH and PAS domains. Biochemical experiments reveal that the PAS domain mediates protein dimerization (Huang et al., 1993). The carboxy terminal third of Sim contains regions with a high percentage of hydrophilic amino acids including serine, threonine, and glutamine residues as

* Corresponding author. Present address: Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA. Phone: (919) 962-4380. Fax: (919) 962-3155. E-mail: steve crews@.unc.edu.

SSDI 0925-4773(93)E0078-M

well as a proline rich region. In other proteins, regions rich in these amino acids act as transcriptional activation domains (Courey and Tjian, 1988; Mitchell and Tjian, 1989).

The genetic and protein sequence data suggest that *sim* encodes a DNA binding transcriptional activator that is required for gene expression in the CNS midline cell lineage. In this paper, we have utilized cell culture and in vivo assays to provide evidence that the Sim protein contains functional domains consistent with its proposed role as a transcriptional activator. The results indicate that there are at least three independent domains within the carboxy terminal portion of Sim that are sufficient for transcriptional activation when attached to a heterologous DNA binding domain. Germline transformation experiments further show that this combined region is required for midline transcription in vivo. These results verify in vivo the activation domains identified in the *Drosophila* cell culture system. Additional germline transformation experiments indicate that the basic domain and PAS domain are required for CNS midline transcription. These experiments support a model in which Sim determines the fate of the CNS midline cell lineage by directly interacting with DNA control elements of CNS midline genes and activating transcription from their promoters.

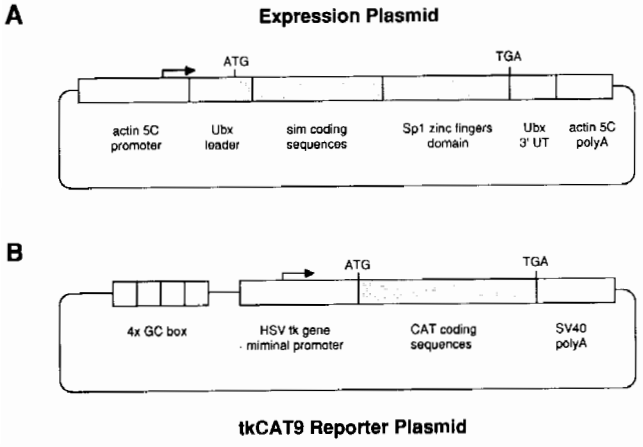


Fig. 1. Expression and reporter plasmids. (A) Diagram of the expression vector used to express the Sim/Sp1 fusion proteins. Transcripts are expressed from the actin 5C promoter and the arrow indicates the transcriptional initiation site. All fusion proteins begin with 8 amino acids from the *Ultrabithorax Ib* cDNA (Beachy et al., 1985). Shaded areas represent coding regions. Translational initiation and stop codons are shown as ATG and TGA, respectively. Sim coding sequences are followed by the zinc finger domain of Sp1 and by the 3′ untranslated region from the *Ultrabithorax Ib* cDNA (Ubx 3′ UT) and a polyadenylation signal from the *actin 5C* gene (actin 5C polyA). (B) Diagram of the tkCAT reporter plasmid. Four Sp1 binding GC boxes are located upstream of the herpes simplex virus thymidine kinase (HSVtk) gene promoter, which drives expression of the bacterial reporter gene encoding chloramphenicol acetyltransferase (CAT). Shaded areas represent coding regions and the arrow indicates the site of transcriptional initiation. The simian virus 40 (SV40) sequences contain the small t intron and polyadenylation signals. Figures are not drawn to scale.

2. Results

2.1. Identification of single-minded transcriptional activation sequences

There are several experimental systems including yeast, *Drosophila*, and mammalian cell culture that are available for identifying protein transcriptional activation domains. (Ptashne, 1988; Courey and Tjian, 1988). Since Sim is a *Drosophila* protein, we reasoned that a *Drosophila* cell culture system was most likely to detect relevant activation sequences. The Human Sp1 DNA binding protein can activate transcription in *Drosophila* Schneider line 2 (SL2) cells upon binding a reporter construct containing multiple Sp1 binding sites (Courey and Tjian, 1988). In our experiments, the activation domains of Sp1 were replaced with Sim sequences, and the ability of these fusion proteins to activate transcription from a reporter plasmid was determined.

The Sim/Sp1 fusion proteins were expressed using the vector pPacSp1 which utilizes the *Drosophila* actin 5C gene promoter. Restriction fragments of a *sim* cDNA clone were inserted into *Bam*HI sites of pPacSp1. This produced fusion proteins in which Sim sequences were fused to the carboxy terminal 168 amino acids of Sp1. The resulting Sim/Sp1 fusion proteins contain the zinc finger DNA binding domain of Sp1, but lack transcriptional activation domains of the Sp1 protein (Fig. 1).

The Sim/Sp1 expression plasmids were cotransfected into SL2 cells along with the reporter plasmid tkCAT9 (Courey et al., 1989). This reporter contains four Sp1-binding GC boxes linked to a basal promoter from the herpes simplex virus thymidine kinase (HSVtk) gene driving expression of chloramphenicol acetyl transferase (CAT) (Fig. 1). As an internal control for transfection efficiency, cotransfections also included a *copia* LTR-*lacZ* plasmid (Mount and Rubin, 1985) which drives a constitutive low level of β-galactosidase protein from the *Drosophila copia* gene LTR promoter. At 48 h after transfection, β-galactosidase activity levels were measured from cell lysates with a colorimetric assay. CAT activities were measured using a dual phase diffusion based assay (Neumann et al., 1987) and the results normalized to β-galactosidase activity.

The expression plasmid pPacsim N673/Sp1 contains the entire Sim coding sequence (673 amino acids) fused to the Sp1 DNA binding domain. When pPacsim N673/Sp1 is cotransfected with the reporter tkCAT9, there is an eighteen fold stimulation of CAT activity

relative to cells transfected with tkCAT9 alone (Fig. 2). Plasmid pPacSp1-168C, which contains only the DNA binding domain, failed to significantly stimulate CAT activity (1.7 fold) relative to tkCAT9 alone. This result demonstrates that the full length Sim protein contains sequences that can function as activation domains in SL2 cells.

Deletional analysis of the Sim protein was used to map the sequences that mediate transcriptional activation (Fig. 2). Carboxy terminal portions of Sim were able to dramatically stimulate CAT activity when fused to the Sp1 DNA binding domain. Amino acids 462–545 of Sim, expressed from the plasmid pPacsim 462–545/Sp1, raised CAT activity levels 160 fold versus tkCAT9 alone. This represents a potent stimulation that is stronger than that seen with full length Sp1 protein, which raises CAT levels 58 fold under these conditions. Similarly, the carboxy terminal 129 amino acids of Sim, expressed from the plasmid pPacsim 129C/Sp1, stimulated CAT activity 150 fold. Further constructs subdivided this region to yield three additional fragments that significantly enhanced CAT activity. Plasmids pPacsim 545–632/Sp1, pPacsim 41C/Sp1, and pPacsim 25C/Sp1 raised CAT activity levels 100, 140, and 26 fold respectively. The stimulation of CAT activity was entirely dependent on the presence of Sp1 DNA binding sites within the promoter of the reporter gene, since none of the expression plasmids stimulated CAT activity from the −37tkCAT control plasmid, which contains just the HSVtk basal promoter and the CAT reporter gene. Amino terminal portions of Sim expressed from the expression plasmid pPacsim N462/Sp1 failed to stimulate CAT activity when co-transfected with the tkCAT9 reporter.

These experiments indicate that the carboxy terminal portion of Sim contains multiple sequences that can independently function as activation domains. The Sim activation domains map to regions rich in serine, threonine, glutamine and proline residues (Fig. 2). Sequences rich in these amino acids have been shown to function as activation domains in other transcription factors (Courey and Tjian, 1988; Mitchell and Tjian, 1989). The bHLH and PAS domains do not function as activation domains in this assay. However, it remains possible that these sequences possess an activation function not revealed in these assays.

2.2. *Single-minded activation domains are required for CNS midline gene expression in vivo*

The experiments described above demonstrate that Sim protein sequences can function as activation do-

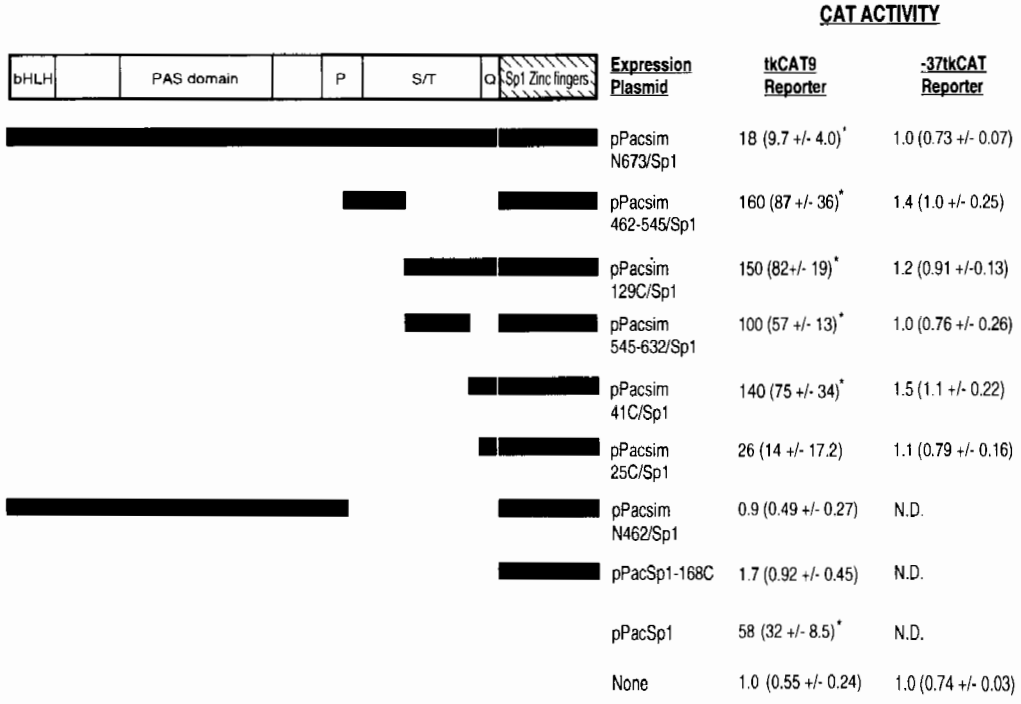
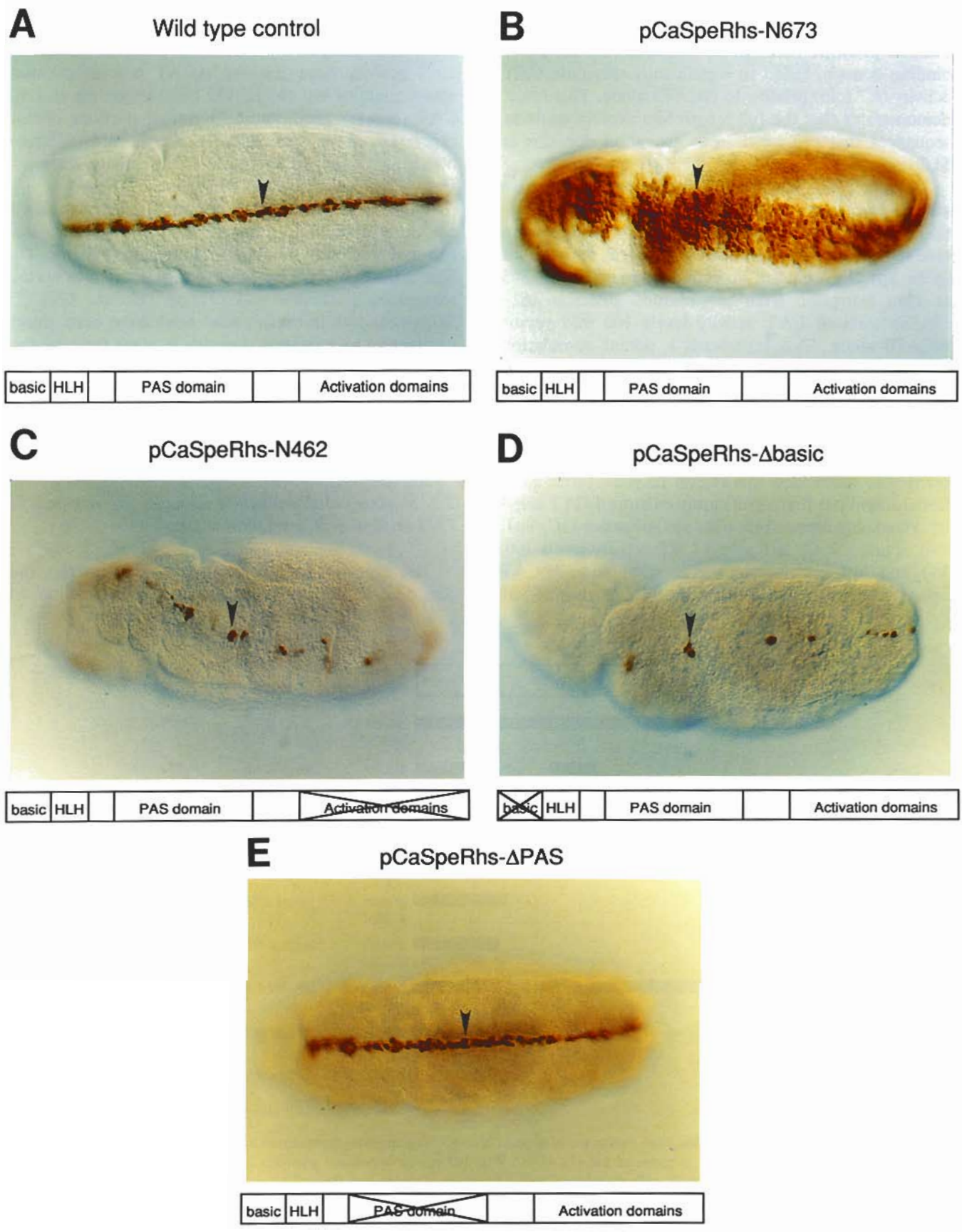


Fig. 2. Transcriptional activation by Sim/Sp1 fusion proteins in SL2 cells. Diagrammatic representation of the *sim* cDNA coding sequences fused to the Sp1 zinc finger domain. Regions of the *sim* cDNA encoded by the expression plasmids are shown as filled black boxes. CAT activities were determined from lysates of cells transfected with expression plasmids and the Sp1 responsive tkCAT9 reporter or the control reporter plasmid −37tkCAT. CAT activities are shown as 'fold' stimulation versus basal activity from the reporter plasmid alone. Actual CAT activity levels in arbitrary CAT units ±standard deviations are shown in parentheses. The '*' indicates a statistically significant difference in level of CAT activity between lysates from cells containing the expression plasmid versus those containing a reporter plasmid alone. N.D. indicated that the value was not determined. Basic-helix-loop-helix domain (bHLH); Per, Arnt, Sim homology domain (PAS domain); proline, serine/threonine, glutamine rich domains (P, S/T, Q, respectively).


A
Wild type control
basic
HLH
PAS domain
Activation domains
B
pCaSpeRhs-N673
basic
HLH
PAS domain
Activation domains
C
pCaSpeRhs-N462
basic
HLH
PAS domain
Activation domains
D
pCaSpeRhs-Δbasic
basic
HLH
PAS domain
Activation domains
E
pCaSpeRhs-ΔPAS
basic
HLH
PAS domain
Activation domains

mains when fused to the Sp1 transcription factor in SL2 cells. Although activation domains have been mapped in many proteins using heterologous expression systems, the in vivo biological relevance of these domains has not often been addressed. The developing CNS midline cell lineage provides an excellent system to examine the in vivo function of these domains. The requirement of the Sim activation sequences for midline CNS gene expression in vivo was addressed using germline transformation. Deletion derivatives of Sim were created by placing restriction fragments of a *sim* cDNA clone in a P-element germline transformation vector under the control of the *hsp70* heat shock gene promoter. The constructs were introduced into germline DNA by microinjection and the embryos subjected to heat shock conditions. Previous results using a full length *sim* cDNA sequence showed that this treatment brought about expression of high levels of Sim in all cells (Nambu et al., 1991). This ectopic expression drives CNS midline gene expression throughout the entire CNS and procephalic region. Use of the ectopic Sim expression system provides a sensitive assay for the ability of Sim to activate CNS midline transcription in vivo.

A cDNA fragment encoding the amino terminal 462 amino acids of Sim, which lacks the activation domains identified in the cell culture assay, was cloned into the transformation vector pCaSpeRhs to create pCaSpeRhs-N462. Transgenic embryos carrying the pCaSpeRhs-N462 construct, when given a heat shock, ectopically express the truncated Sim protein at high levels in cell nuclei throughout the embryos as determined by staining with a Sim antisera (data not shown). In order to monitor the effect of ectopic Sim expression on CNS midline gene transcription, a chromosome containing a *P[3.7sim / lacZ]* reporter gene was introduced into the pCaSpeRhs-N462 strain. The *P[3.7sim / lacZ]* transgene contains the *sim* gene early promoter region fused to *lacZ* and β-galactosidase is expressed at high levels in the CNS midline cells (Fig. 3A; Nambu et al., 1991). Expression of full length Sim from one or two copies of pCaSpeRhs-N673 caused expansion of β-galactosidase expression into cells of the lateral CNS in 79% of the stage 10 through 11 embryos examined (Fig. 3B; Table 1). Ectopic expression of the truncated protein from the pCaSpeRhs-N462 construct failed to activate CNS midline gene expression in lateral CNS cells in all embryos examined (Fig. 3C; Table 1). Instead, ectopic expression of the pCaSpeRhs-N462 construct led to reduced expression from the *P[3.7sim / lacZ]* reporter. Expression of the *P[3.7sim / lacZ]* transgene was patchy and weaker on a per cell basis in 29% of the pCaSpeRhs-N462 embryos examined. Cell counts revealed that the number of cells expressing β-galactosidase was less than half the wild type number. Wild type stage 11 embryos (N = 14) contained an average of 115 (±12 s.d.) β-galactosidase positive cells while pCaSpeRhs-N462 stage 11 embryos that exhib-

Table 1
Phenotypes associated with the expression of the *hsp70 / sim* transgene constructs

hsp70 / sim construct	Percent expanded CNS midline	Percent patchy CNS midline	Percent normal CNS midline	
N673	79	0	21	$n = 47$
N462	0	29	71	$n = 98$
Δbasic	0	8	92	$n = 212$
ΔPAS	0	0	100	$n = 72$

Embryos carrying 1 or 2 copies of the indicated *hsp70 / sim* construct and 1 or 2 copies of the *P[3.7sim / lacZ]* transgene were heat shocked and stained with an anti-β-galactosidase antibody as described in Materials and methods. The percentage of stage 10 or stage 11 embryos displaying expanded, patchy, or normal CNS midline expression patterns are presented here for the four *hsp70 / sim* constructs. The number of embryos (*n*) counted for each *hsp70 / sim* construct is given in the right-most column.

Fig. 3. Ectopic expression of mutant sim proteins fails to transform lateral CNS and leads to dominant negative interference of CNS midline gene expression. Embryos carrying 1–2 copies of the *P[3.7sim / lacZ]* CNS midline reporter transgene and 1–2 copies of various *hsp70 / sim* expression constructs were subject to heat shock and stained with anti-β-galactosidase antibody. Ventral views of stage 10–11 embryos are shown with anterior to the left. The structure of the Sim protein and region deleted (X) in the corresponding transgene is indicated below each stained embryo. Schematic of the Sim protein shown includes the basic, HLH, PAS, and activation domains (not drawn to scale). (A) Wild type expression of the *P[3.7sim / lacZ]* reporter in a strain without an *hsp70 / sim* expression transgene is observed within the CNS midline cells (arrowhead). (B) Ectopic expression of full length Sim from the pCaSpeRhs-N673 construct transforms cells of the lateral CNS into midline CNS cells with a corresponding expansion of the *P[3.7sim / lacZ]* reporter (arrowhead). (C) Ectopic expression of the activation domain-deficient Sim protein from the pCaSpeRhs-N462 construct fails to expand the expression of the *P[3.7sim / lacZ]* transgene. Additionally, the expression of this reporter within the CNS midline cells is weaker on a per cell basis and is patchy (arrowhead) containing only half the number of β-galactosidase-positive staining cells per embryo as the control. The apparent disparity between the number of β-galactosidase expressing cells seen in (C) (approximately 25 cells) and the number reported in the text (on average 46 cells) is due to the existence of additional β-galactosidase expressing cells which are not visible because they are in a different focal plane or because they are lightly stained and not easily photographed. (D) Ectopic expression of the Sim basic domain mutant protein from the pCaSpeRhs-Δbasic construct also fails to expand the expression of the *P[3.7sim / lacZ]* midline reporter and leads to weak and patchy expression within the CNS midline cells (arrowhead). (E) Ectopic expression of the PAS domain deleted Sim protein from the pCaSpeRhs-ΔPAS construct fails to show an expansion of the *P[3.7sim / lacZ]* reporter gene. However, unlike pCaSpeRhs-N462 and pCaSpeRhs-Δbasic, the *P[3.7sim / lacZ]* reporter gene is expressed normally in the midline cells (arrowhead). Scale Bar equals 50 microns.

ited patchy CNS midline expression ($n = 21$) contained an average of 46 (± 17 s.d.) β-galactosidase positive cells. Previous work had shown that in *sim* mutant embryos a similar pattern of *P[3.7sim / lacZ]* expression was observed, due to the inability of Sim to positively autoregulate its own expression (Nambu et al., 1991), suggesting that induction of pCaSpeRhs-N462 leads to a dominant negative inhibition of the endogenous Sim protein. An interpretation consistent with these results proposes that the high level synthesis of the mutant Sim protein interferes with transcriptional activation through the formation of non-productive dimers or multimers. These results indicate that the carboxy terminal activation domains of Sim are required for CNS midline gene expression.

Mutations within the basic region of Sim which do not disrupt the HLH or PAS dimerization domains are also predicted to give rise to mutant proteins with dominant negative properties through the formation of non-productive complexes. To test this prediction and to demonstrate the in vivo importance of the basic domain, we ectopically expressed a mutant Sim protein from which the basic domain had been deleted. This mutant protein was expressed from the plasmid pCaSpeRhs-Δbasic, which encodes a full length Sim protein lacking 11 amino acids of the basic region. The Sim basic domain mutant protein, when ectopically expressed, was only weakly localized to cell nuclei (data not shown) confirming previous experiments that indicated the putative DNA binding basic domain acted as a nuclear localization signal (Nambu et al., 1990). Ectopic expression from pCaSpeRhs-Δbasic failed to ectopically activate CNS midline transcription and caused a weak, patchy midline expression of the *P[3.7sim / lacZ]* transgene consistent with its predicted dominant negative properties (Fig. 3D; Table 1).

Similar experiments were performed to test the function of the PAS domain in vivo. If this domain is involved in protein dimerization, it is predicted that ectopic expression will not expand CNS midline gene expression nor affect normal CNS midline transcription. We generated the construct pCaSpeRhs-ΔPAS that encodes a mutant Sim protein from which the PAS domain (amino acids 88–356 of Sim) had been deleted. Nuclear expression of this protein was verified by antibody staining (Fig. 3E). Ectopic expression from this construct failed to expand the expression of the *P[3.7sim / lacZ]* transgene demonstrating an in vivo function for this domain (Table 1). Additionally, expression of the ΔPAS protein failed to cause a dominant negative inhibition of the *P[3.7sim / lacZ]* reporter consistent with the requirement of this domain for dimerization. The failure of PAS domain mutant proteins to elicit a dominant negative phenotype suggests that the dominant negative phenotype seen with the pCaSpeRhs-Δbasic and pCaSpeRhs-N462 are not simply explained by non-specific dimerization interactions.

3. Discussion

Evidence is presented in this paper that the *Drosophila* Sim protein can function as a transcriptional activator. One set of experiments utilized a *Drosophila* cell culture system to express chimeric proteins in which the DNA binding domain of Sp1 was used to tether potential activation domains of Sim to an Sp1-responsive reporter plasmid. Three independent regions of Sim enhanced transcription from the Sp1-dependent promoter element. These domains mapped to carboxy terminal regions of Sim rich in serine, threonine, glutamine, and proline residues.

Germline transformation experiments utilizing a truncated form of Sim, which does not contain the carboxy terminal activation domains, indicate that the regions identified in the cell culture assay are required for activation in vivo. The failure of the activation domain deficient protein, encoded by pCaSpeRhs-N462, to expand the expression of the *P[3.7sim / lacZ]* reporter gene suggests that these activation domains function to stimulate the expression of CNS midline gene expression during normal development of the CNS midline cells. This failure cannot be easily explained by low expression levels or instability of this mutant protein, since protein expression was detected using a Sim antisera and a likely dominant negative phenotype was observed.

The prevalence of the clusters of serine, threonine, glutamine and proline amino acids within the Sim activation domains matches the amino acid preferences of activation domains mapped within other transcriptional regulatory proteins (Courey and Tjian, 1988; Mitchell and Tjian, 1989). It is interesting to note that sequences rich in serine, threonine, and glutamine residues are found in a similar location in the closely related mammalian protein Ahr, the ligand binding subunit of the aryl hydrocarbon receptor complex, that functions as a ligand activated transcriptional activator (Burbach et al., 1992; Ema et al., 1992). It is thought that transcriptional activation domains stimulate transcription through interactions with proteins associated with the promoter (i.e. the basal transcriptional machinery and associated factors) and thus stabilize the formation of a transcriptionally active initiation complex (Ptashne, 1988; Hoey et al., 1993). Sim may belong to a class of transcriptional activators whose effects are mediated by the TATA-binding protein associated factor 110 (TAF110). This is suggested by the proximity in Sim of serine/threonine rich and glutamine rich activation domains in an organization reminiscent of the activation domain structure of Sp1. These domains of

Sp1 are required for transcriptional activation and multimerization, and have been shown to mediate a specific interaction with TAF110. (Courey and Tjian, 1988; Pascal and Tjian, 1991; Hoey et al., 1993).

The phenotype associated with the expression of the truncated Sim protein from pCaSpeRhs-N462 displayed weak, patchy *P[3.7sim / lacZ]* midline expression consistent with a dominant negative action of the truncated protein. A similar pattern of *P[3.7sim / lacZ]* expression was observed in *sim* mutant embryos, due to the inability of Sim to positively autoregulate its own expression (Nambu et al., 1991). Ectopic expression of another *sim* construct that was lacking the basic region and presumably the ability to bind DNA also showed a similar phenotype. It is proposed that the activation domain deficient protein and the basic domain mutant protein are able to form dimers with an endogenous bHLH partner molecule, but are unable to activate transcription. This leads to a reduction in the level of expression from the *P[3.7sim / lacZ]* reporter. The failure of these dimers to activate transcription could be a result of their inability to make contact with proteins associated with the promoter, to bind DNA, or to localize to cell nuclei. These results demonstrate an in vivo requirement for the carboxy-terminal activation domains and suggest that the in vivo function of the basic region of Sim is to mediate DNA binding. Expression of basic region mutants of other bHLH proteins (e.g. Myc, Myn, or MyoD) result in dominant negative inhibition of in vitro DNA binding as well as transcriptional activation in culture (Davis et al., 1990; Prendergast et al., 1991; Kretzner et al., 1992). Our germline transformation results showing that the *sim* basic region and activation domain deficient genes effectively induce a dominant negative phenotype suggest that a similar approach could be used to eliminate *sim* or bHLH gene function in vivo for organisms, such as *Xenopus,* in which conventional genetic analysis is difficult.

It is likely that the reduced Sim activity within the developing CNS midline cells exhibited by the basic region and activation domain deficient Sim strains fails to support other aspects of the CNS midline developmental program including the synchronous cell division of CNS midline precursors. This is supported by cell counts that indicate less than half the wild type number of β-galactosidase expressing cells are present on the CNS midline. Additionally, the β-galactosidase positive cells appear larger in size, a characteristic of the premitotic precursors. The reduction in β-galactosidase positive cells is likely due to the failure of the CNS midline cell precursors to divide as well as a failure by Sim to autoregulate its own expression, as also observed in *sim* mutant embryos.

The PAS domain, which mediates protein dimerization in vitro, was also required for Sim function. A mutant Sim protein, in which the PAS domain was deleted, was not able to support CNS midline transcription in vivo. However, the ΔPAS mutants did not lead to a dominant negative inhibition of the *P[3.7sim / lacZ]* reporter. This result is consistent with the proposed role of this domain as a mediator of dimerization in vivo and suggests the putative partner molecule(s) for Sim will have both bHLH and PAS domains.

The results presented here further define the role of *sim* as a genetic switch that controls CNS midline transcription. We suggest a model in which expression of Sim within neuroectodermal precursor cells allows the formation of a complex between Sim and unidentified bHLH/PAS proteins. This protein complex then binds to regulatory sequences of CNS midline genes and activates transcription. This is analogous to the ability of bHLH regulators such as MyoD, Achaete-Scute, Ahr, and Myc to directly activate transcription of target genes through interactions with bHLH or bHLH/PAS containing partner molecules (Davis et al., 1990; Cabrera and Alonso, 1991; Amati et al., 1992; Ema et al., 1992; Kretzner et al., 1992) In the case of Sim, it will be important to demonstrate its ability to bind DNA and to identify the partner molecule(s) with which it interacts.

4. Materials and methods

4.1. Plasmid constructs and Drosophila strains

The following plasmids have been previously described: pPacSp1, pPacSp1-168C, tkCAT9, -37tkCAT (Courey and Tjian, 1988; Courey et al. 1989), and *copia* LTR-*lacZ* (Mount and Rubin, 1985). The *P[3.7sim / lacZ]* fly line has also been previously described (Nambu et al., 1991). All expression vectors used for the cell culture experiments were made by cloning *sim* cDNA sequences into *Bam*HI-cut pPacSp1. The *Bam*HI digest removes sequences encoding the amino terminal portions of Sp1 and leaves the 168 carboxy terminal amino acids which encode the DNA binding zinc fingers. All portions of Sim were prepared from a cDNA clone NB-F1(Bgl) which is a mutated form of the cDNA clone NB-F1 (provided by J. Thomas). The NB-F1(Bgl) clone was created by oligonucleotide directed mutagenesis (Amersham; Sayers et al., 1988) that added *Bgl*II sites at the initiator methionine and at the termination codon of NB-F1. The plasmid pPacsim N673/Sp1 was made by directly cloning the 2 kb *Bgl*II fragment released from NB-F1(Bgl) into the *Bam*HI-cut pPacSp1. Restriction fragments used for the plasmids pPacsim N462/Sp1 and pPacsim 129C/Sp1 were isolated from a *Bgl*II/*Bam*HI double digest of NB-F1(Bgl). The pPacsim 545-632/Sp1 plasmid was

generated by truncating the 129C fragment at an internal Sau3A site. The plasmids pPacsim 462-545/Sp1, pPacsim 41C/Sp1, and pPacsim 25C/Sp1 were prepared from PCR amplification products of the NB-F1 cDNA using appropriate oligonucleotides.

4.2. DNA transfection and transient expression assay

Drosophila Schneider line 2 (SL2) cells (Schneider, 1972) were grown at 25°C in M3 media (Lindquist et al., 1982) supplemented with 2% fetal calf serum that was heat inactivated for 30 min at 60°C. Cells were plated at 8×10^6/100 mm plate and allowed to attach for 16–24 h before transfection. Cells were transfected by calcium phosphate precipitation according to Di Nocera and Dawid (1983) with the following modifications. Precipitations contained 125 m*M* $CaCl_2$, 140 m*M* NaCl, 25 m*M* Hepes, 0.75 m*M* sodium phosphate, pH 7.1 given as final concentrations. Typical transfections contained 5 μg reporter plasmid, 0.1 μg expression plasmid and 2 μg *copia* LTR-*lacZ* internal control plasmid. Sixteen to 24 h after transfection, cells were replenished with fresh media. Forty eight h after transfection, cells were rinsed in buffered saline then harvested with rubber policemen in 300 μl homogenizing buffer (0.25 *M* sucrose, 100 m*M* sodium phosphate buffer, pH 7.0). Cells were then lysed using three cycles of freeze thaw on dry ice. β-galactosidase activities were determined as described by Miller (1972). Typical reactions contained 150 μl of cell lysate and were carried-out at 37°C for 40 min. Chloramphenicol acetyl transferase (CAT) activities were determined as described by Neumann et al. (1987) with the modification of Eastman (1987). Typically, 50 ml of extract was heated to 65°C for 5 min and then added to the reaction mixture. CAT activities were read on a scintillation counter after 1 h and then normalized for β-galactosidase activity. For each experiment transfections were done in duplicate plates. For transfections with the tkCAT9 reporter, the average values reflect the results of 3 to 5 independent experiments. Transfections containing the −37tkCAT reporter represent the average value of two independent experiments. Statistical significance was determined using a students two sample *t*-test and gave a significance beyond $t = 0.005$.

4.3. Generation and use of P[hsp70/sim] mutant strains

In order to ectopically express mutant forms of Sim during embryogenesis, we utilized the P element transformation vector pCaSpeRhs (provided by C. Thummel). This vector contains the *Drosophila* heat shock 70 protein (*hsp70*) promoter, and was used to express altered Sim coding sequences during development. The pCaSpeRhs-N462 construct was created by cloning a 1.4 kb *Bgl*II/*Bam*HI fragment isolated from the *sim* cDNA clone NB-F1(Bgl) into a unique *Bgl*II site within the polylinker of the pCaSpeRhs vector. The pCaSpeRhs-N673 construct was prepared by subcloning a 2 kb *Bgl*II fragment excised from the NB-F1(Bgl) cDNA into the *Bgl*II site of pCaSpeRhs. This fragment contains the entire *sim* coding sequence. Three copies of a hemagglutinin epitope (provided by B. Futcher; Field et al., 1988) were placed in-frame at the extreme carboxy terminus of Sim within the pCaSpeR-N673 construct. Addition of this triple epitope to the full length Sim protein does not alter the in vivo activation properties of Sim as the epitope tagged version of Sim is able to transform lateral CNS to midline CNS with the same apparent efficiency as full length wild type Sim (data not shown). To prepare the Δbasic mutant we carried-out oligonucleotide directed mutagenesis on the NB-F1(Bgl) clone and removed the first 11 residues of the Sim basic domain. These residues directly follow the initiator methionine of Sim. The mutagenesis also altered 10 nucleotides of the 5′ untranslated sequence. The nucleotides immediately 5′ to the initiator methionine of *sim* were replaced with those from the *hsp70* gene that better fit the *Drosophila* translational initiation consensus sequence (Cavener, 1987). This did not alter the sequence of the protein product and was done in an attempt to achieve better translation of the message under heat shock conditions. The resulting clone NB-F1(Bgl-Δbasic) was digested with *Bgl*II and the 2 kb fragment cloned into the *Bgl*II site of pCaSpeRhs, producing the plasmid pCaSpeRhs-Δbasic. The hemagglutinin triple epitope was also added to the pCaSpeRhs-Δbasic construct. The ΔPAS deletion mutant was generated by oligonucleotide directed mutagenesis. A mutant oligonucleotide was designed to create a deletion of the coding sequences in the NB-F1(Bgl) c-DNA clone. This deletion removed amino acids 88–356 of Sim producing the plasmid NB-F1(ΔPAS). The NB-F1(ΔPAS) plasmid was cut with *Bgl*II and cloned into pCaSpeRhs to generate pCaSpeRhs-ΔPAS. The hemagglutinin triple epitope was also added to this construct. Transformed *Drosophila* lines were generated using the microinjection method of Rubin and Spradling (1982). The heat shock regimen involved collecting embryos 2–4 h old from stably transformed flies and subjecting them to a heat shock at 37°C for 1 h. The embryos were then allowed to recover for 3–4 h at 25°C before processing for immunohistochemistry. For the pCaSpeRhs-N462, pCaSpeRhs-Δbasic, and pCaSpeRhs-ΔPAS constructs, two to three independent transformed fly lines were examined and all gave similar results. Wild type embryos that were heat shocked or *hsp70/sim* constructs that did not receive a heat shock served as control embryos and failed to display the mutant phenotypes described (data not shown). Staining of transformed fly lines with an anti-

β-galactosidase monoclonal antibody (Promega) was performed as previously described (Nambu et al., 1990). All embryonic developmental stages discussed are those defined by Campos-Ortega and Hartenstein (1985).

5. Acknowledgements

The authors thank Albert Courey, Josephine Lewis and Yumi Kasai for comments on the manuscript. We thank Arnold Berk and Albert Courey for helpful discussions and advice, Michael Rosbash for unpublished information concerning PAS domains, and Albert MacKrell for assistance with the *Drosophila* tissue culture experiments. R.G.F. is a predoctoral trainee in the UCLA Interdepartmental Program for Neuroscience and is supported in part by US Public Health Service National Research Service Award GM 07104 and by a Dr. Ursula Mandel Scholarship. This work was supported by grants from the Lucille P. Markey Charitable Trust and the NIH (HD25251).

6. References

Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I. and Land, H. (1992) Nature 359, 423–426.

Beachy, P.A., Helfand, S.L. and Hogness, D.S. (1985) Nature 313, 545–551.

Burbach, K.M., Poland, A. and Bradfield, C.A. (1992) Proc. Natl. Acad. Sci. USA 89, 8185–8189.

Cabrera, C.V. and Alonso M.C. (1991) EMBO J. 10, 2965–2973.

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

Cavener, D.R. (1987) Nucleic Acids Res. 15, 1353–1361.

Courey, A.J. and Tjian, R. (1988) Cell 55, 887–898.

Courey, A.J., Holtzman, D.A., Jackson, S.P. and Tjian, R. (1989) Cell 59, 827–836.

Crews, S.T., Thomas, J.B. and Goodman, C.S. (1988) Cell 52, 143–151.

Davis, R.L., Cheng, P.F., Lassar, A.B. and Weintraub, H. (1990) Cell 60, 733–746.

Di Nocera, P.P. and Dawid, I.B. (1983) Proc. Natl. Acad. Sci. USA 80, 7095–7098.

Eastman, A. (1987) BioTechniques 5, 730–732.

Ema, M., Sogawa, K., Watanabe, N., Chujoh, Y., Matsushita, N., Gotoh, O., Funae, Y. and Fujii-Kuriyama, Y. (1992) Biochem. Biophys. Res. Commun. 184, 246–253

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I.A., Lerner, R.A. and Wigler, M. (1988) Mol. Cell. Biol. 8, 2159–2165.

Hoey, T., Weinzierl, R.O. Gill, G., Chen, J., Dynlacht, B.D. and Tjian, R. (1993) Cell 72, 247–260.

Hoffman, E.C., Reyes, H., Chu, F.F., Sander, F., Conley, L.H., Brooks, B.A. and Hankinson, O. (1991) Science 252, 954–958.

Huang, Z.J., Edery, I. and Rosbash, M. (1993) Nature 364, 259–262.

Kretzner, L., Blackwood, E.M. and Eisenman, R.N. (1992) Nature 359, 426–429.

Lassar, A.B., Davis, R.L., Wright, W.E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintruab, H. (1991) Cell 66, 305–315.

Lindquist, S.L., Sonoda, S., Cox, T. and Slusser, K. (1982) Dros. Inf. Serv. 58, 163.

Miller, J.H. (1972) In Experiments in Molecular Genetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mitchell, P.M. and Tjian, R. (1989) Science 245, 371–378.

Mount, S. and Rubin, G.M. (1985) Mol. Cell. Biol. 5, 1630–1638.

Murre, C., Schonleber McCaw, P., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., Weintraub, H., Baltimore, D. (1989) Cell 58, 537–544.

Nambu, J.R., Franks, R.G., Hu, S. and Crews, S.T. (1990) Cell 63, 63–75.

Nambu, J.R., Lewis, J.L., Wharton, K.A. and Crews, S.T. (1991) Cell 67, 1157–1167.

Neumann, J.R., Morency, C.A. and Russian, K.O. (1987) BioTechniques 5, 444–447.

Pascal, E. and Tjian, R. (1991) Genes Dev. 5, 1646–1656.

Prendergast, G.C., Lawe, D. and Ziff, E.B. (1991) Cell 65, 395–407.

Ptashne, M. (1988) Nature 335, 683–689.

Rubin, G.M. and Spradling, A.C. (1982) Science 218, 348–353.

Sayers, J.R., Schmidt, W. and Eckstein, F. (1988) Nucl. Acids Res. 16, 791–802.

Schnieder, I. (1972) J. Embryol. Exp. Morphol. 27, 353–365.

Thomas, J.B., Crews, S.T. and Goodman, C.S. (1988) Cell 52, 133–141

Voronova, A. and Baltimore, D. (1990) Proc. Natl. Acad. Sci. 87, 4722–4726.