Development 120, 3563-3569 (1994)
Printed in Great Britain © The Company of Biologists Limited 1994

Control of CNS midline transcription by asymmetric E-box-like elements: similarity to xenobiotic responsive regulation

Keith A. Wharton, Jr.[1,†], Robert G. Franks[1,†], Yumi Kasai[1,‡] and Stephen T. Crews[2,*]

[1]Molecular Biology Institute, Department of Biology, University of California, Los Angeles, CA 90024-1570, USA
[2]Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA

*Author for correspondence
†Authors contributed equally
‡Present address: Monsanto Corporation, 700 Chesterfield Parkway North, St. Louis, MO 63198, USA

SUMMARY

Central nervous system midline cells constitute a discrete group of *Drosophila* embryonic cells with numerous functional and developmental roles. Corresponding to their separate identity, the midline cells display patterns of gene expression distinct from the lateral central nervous system. A conserved 5 base pair sequence (ACGTG) was identified in central nervous system midline transcriptional enhancers of three genes. Germ-line transformation experiments indicate that this motif forms the core of an element required for central nervous system midline transcription. The central nervous system midline element is related to the mammalian xenobiotic response element, which regulates transcription of genes that metabolize aromatic hydrocarbons. These data suggest a model whereby related basic-helix-loop-helix-PAS proteins interact with asymmetric E-box-like target sequences to control these disparate processes.

Key words: CNS midline, *Drosophila*, enhancer, neurogenesis, *single-minded*, transcription, xenobiotic response

INTRODUCTION

Development of the nervous system in a bilaterally symmetric organism critically depends on cells that lie on its midline. Midline cells in various metazoans have been shown to be important as functional neurons and glia, as sources of signals for migrating commissural growth cones and for patterning adjacent tissues during embryogenesis (reviewed in Nambu et al., 1993). Consequently, these cells often display patterns of gene expression distinct from lateral neuroectoderm. Single gene mutants and lineage analysis in the fruit fly *Drosophila melanogaster* have begun to decipher the genesis and function of CNS midline cells (Nambu et al., 1993).

CNS midline gene expression can be temporally divided into three embryonic components: (1) initial mesectodermal (MEC) expression in the blastoderm and gastrula (stages 5-7), (2) midline precursor (MLP) expression (stages 8-11) and (3) later midline expression (stages 12-17), which is restricted to subsets of differentiated cells such as the midline glia (MLG) and neurons. MEC expression is controlled by dorsal/ventral patterning genes (Nambu et al., 1990; Kosman et al., 1991; Leptin, 1991; Rao et al., 1991), whereas MLP expression requires activity of the *single-minded (sim)* gene product, which encodes a basic-helix-loop-helix-PAS (bHLH-PAS) transcription factor (Nambu et al., 1990, 1991). Later midline expression is more complicated, genetically dependent on *sim* and other transcription factors (Nambu et al., 1993).

Previous experiments identified CNS midline enhancers from the *sim*, *Krüppel*, *slit* (*sli*), *Toll* (*Tl*) and *rhomboid* (also known as *veinlet*) genes (Nambu et al., 1990; Hoch et al., 1992; Ip et al., 1992b; Wharton and Crews, 1993). In this paper, we focus on three of these genes: *sim*, *sli* and *Tl*. Genetic analysis has shown that *sim* positively autoregulates its own expression in MLP cells beginning around stage 8 (Nambu et al., 1991). Germ-line transformation experiments demonstrated that a 3.7 kilobase (kb) fragment of the *sim* gene, fused to the *E. coli lacZ* gene, drives β-galactosidase expression in CNS midline cells in a pattern indistinguishable from the endogenous *sim* gene (Nambu et al., 1991). The *Tl* gene contains a 900 base pair (bp) fragment that drives CNS midline expression of β-galactosidase beginning at stage 9 when coupled to a *lacZ* enhancer-tester vector (Wharton and Crews, 1993). The *sli* gene has a 380 bp fragment that drives β-galactosidase expression later in development in MLG (Wharton and Crews, 1993).

Genetic evidence suggests that these enhancers are targets of the *sim* gene product (Nambu et al., 1990, 1991). The DNA-binding properties of SIM have not been demonstrated, but an educated guess can be entertained by comparison to related bHLH transcription factors. In many examples studied, the bHLH domain mediates dimerization and DNA binding to the consensus E-box (5′ CANNTG 3′) (Murre et al., 1989). The

adjacent SIM PAS domain is thought to mediate protein-protein interactions (Huang et al., 1993). The SIM bHLH and PAS domains closely resemble those of two subunits of the mammalian Aromatic Hydrocarbon Receptor Complex (AHRC) (Hoffman et al., 1991; Nambu et al., 1991; Burbach et al., 1992; Ema et al., 1992). The AHRC is a ligand-dependent transcription factor that induces expression of enzymes involved in drug and carcinogen metabolism (Whitlock, 1990). The AHRC interacts with the consensus xenobiotic response element (XRE) (5′ TNGCGTG 3′) present in multiple copies in responsive gene enhancers (Denison et al., 1988). The nuclear form of the AHRC is composed of two bHLH-PAS proteins, Aromatic Hydrocarbon Receptor (AHR) and AH Receptor Nuclear Translocator (ARNT), that bind to the XRE as a heterodimer (Reyes et al., 1992; Matsushita et al., 1993; Dolwick et al., 1993; Probst et al., 1993).

Sequence analysis of the *sim*, *sli* and *Tl* regulatory regions revealed a conserved sequence ACGTG that resembles the XRE. Mutation of these sequence elements followed by germline transformation revealed that this element is required for CNS midline transcription for all three genes. Multimerization of a 20 bp DNA fragment containing the ACGTG sequence is sufficient for MLP transcription. These experiments further reinforce the functional similarity between control of CNS midline transcription and aromatic hydrocarbon metabolism. They also suggest a model for how related bHLH-PAS domain proteins control transcription of these two processes.

MATERIALS AND METHODS

sim gene deletion constructs

Deletion constructs of the 3.7 kb *sim* early promoter region were generated using an Erase-a-base kit (Promega). The 5′ deletion fragments of 2.8, 2.1, 1.5 and 0.9 kb were cloned into Casper-AUG-β-gal (Thummel et al., 1988), and the 5′-ends sequenced to localize the deletion site.

Isolation of *Drosophila virilis sim* gene

The *Drosophila melanogaster* 2.8 kb fragment containing the *sim* early promoter is sufficient for MEC and MLP expression (see Results), and the corresponding DNA was desired from *Drosophila virilis*. The *Drosophila virilis sim* homolog was cloned from a genomic library (kindly provided by John Tamkun) using a *sim* bHLH-PAS cDNA probe. Two overlapping genomic clones were isolated and restriction mapped. The *Drosophila virilis* region corresponding to the *Drosophila melanogaster* 2.8 kb early promoter region was sequenced and aligned to the *Drosophila melanogaster* sequence using the University of Wisconsin GCG program COMPARE (Devereux et al., 1984). This analysis identified 13 conserved sequence blocks, 20 to 60 bp long, of greater than 90% identity interspersed by nonconserved sequence. The order of the conserved sequence blocks was preserved between the two species, but spacing of the blocks varied.

DNA sequence analysis

The *sim* 2.8 kb, *sli* 1.0 HV and *Tl* 1.4 RR (Wharton and Crews, 1993) restriction fragments that include the 380 bp *sli* and 900 bp *Tl* sequences were bidirectionally deleted using an Erase-a-base kit and sequenced by the dideoxynucleotide chain termination method. These sequences were examined for the presence of elements related to the XRE (TNGCGTG). Because of SIM's arginine residue at basic region position 13, which specifies a CG dinucleotide pair in E-box response elements (Dang et al., 1992; Ferre-D'Amare et al., 1993), sequences containing CG dinucleotide pairs in the *sim*, *sli* and *Tl* enhancers were aligned using University of Wisconsin GCG program FIND (Devereux et al., 1984). These sequences were then examined for the presence of asymmetric E-box-like motifs.

Site-directed mutagenesis

Oligonucleotide site-directed mutagenesis (Amersham) was performed on the 900 bp *Tl* 950 Rd fragment (*Tl950* refers to the name of the construct published in Wharton and Crews, 1993, whereas the size of the fragment from sequence analysis is 900 bp). Sites were either mutated individually or all four sites were altered together. Each NACGTG motif was mutated to the *Bam*HI restriction site GGATCC. The mutant clones residing in Bluescript II KS were cut with *Not*I and *Kpn*I, sequenced to verify the mutated sequence and cloned into the *lacZ* enhancer-tester vector C4PLZ (Wharton and Crews, 1993). Three independent transformants of a construct that harbored a specific mutation in *Tl* site 2 failed to produce any tissue-specific staining. Since a deletion of *Tl* sites 2 and 3 shows normal salivary gland placode staining and low levels of midline staining (Wharton, 1992), the *Tl* site 2 mutation may represent an artifact associated with the mutagenesis, cloning or transformation. Site-directed mutagenesis of *sli* fragment 380 dNc (Wharton and Crews, 1993) replaced its single AAACGTG motif with the *Hin*dIII restriction site TAAGCTT. The mutant fragment was sequenced and cloned into *Not*I-*Kpn*I cut C4PLZ. Site-directed mutations in the 2.8 kb *sim* fragment replaced each NACGTG motif with a *Bam*HI site. The mutant construct was sequenced and cloned into Casper-AUG-βgal (Thummel et al., 1988).

Multimerization of CME

Two 24-mers that duplicate opposite strands of the 20 bp encompassing *Tl* site 4 were synthesized. The coding strand oligonucleotide was 5′ ctagAAATTTGTACGTGCCACAGA 3′ and the complementary strand was 5′ ctagTCTGTGGCACGTACAAATTT 3′ (non *Tl* sequence that introduces an *Xba*I half-site is in small case; ACGTG motif is underlined). These oligos were annealed and multimerized with T4

```
sim 1        ACTTCAGACGTGCATGTTG
                   |||||
sim 2        TTAAGATACGTGACCTAAG
                   |||||
sim 3        TGTATTTACGTGCGAATTC
                   |||||
sim 4        TTTTCGAACGTGATTTTGG
                   |||||
sim 5        AGCCAAAACGTGTGCCATG
                   |||||
Tl 1         CCTGAGTACGTGTTAATTT
                  | |||||
Tl 2         GAGTCGAACGTGTGTGATA
                  | |||||
Tl 3         AGTTTGTACGTGTGTGTGA
                  | |||||
Tl 4         AATTTGTACGTGCCACAGA
                   |||||
sli          AACTCAAACGTGCCGCAGA

Consensus          GTACGTG
                   AA
```

Fig. 1. Sequence alignment of ACGTG motifs. The sequence ACGTG was found 10 times in the three CNS midline enhancers. The *sim* gene contains five motifs within a 3.7 kb fragment, and *sim* motifs 2-5 are identical between *Drosophila melanogaster* and *Drosophila virilis*. Four ACGTG motifs were found within the 900 bp *Tl* fragment. The *sli* 380 bp fragment contains a single motif. At each nucleotide position, nine out of ten sequences yield the extended consensus (G/A)(T/A)ACGTG.

DNA ligase. Multimers of 3 and 4 were visualized with UV light, isolated on a 15% polyacrylamide gel and purified by standard techniques. Multimer fragments were cloned into *Xba*I-cut Bluescript II KS$^-$ and sequenced. Clones containing 3 and 4 copies of *Tl* site 4 were cloned into *Not*I-*Kpn*I cut C4PLZ followed by germline transformation.

The P[*4X950Tl4*] element was crossed into a *sim* null mutant (*sim*^{H9}) background and stained for β-galactosidase immunoreactivity (Nambu et al., 1990). The P[*4X950Tl4*] element was also crossed into a strain containing multiple copies of a *sim* cDNA fused to an *hsp70* heat-shock promoter (Nambu et al., 1991). Heat induction was similar to that described previously (Nambu et al., 1991; Franks and Crews, 1994): 2- to 4-hour-old embryos were treated for 1 hour at 37°C, allowed to recover for an additional 2-4 hours at 25°C and then stained for β-galactosidase immunoreactivity.

Germ-line transformation and detection of *lacZ* expression

P-element constructs were introduced into the *Drosophila* germ line by microinjection as described by Rubin and Spradling (1982) with *Pπ25.7wc* used as a source of transposase (Karess and Rubin, 1984). *w*^{1118} flies were used as the transformation strain, since all P elements used in this study contained the *white* mini-gene. Embryos were stained for β-galactosidase protein with a monoclonal antibody to β-galactosidase (Promega) (Nambu et al., 1991) or for *lacZ* transcripts with a *lacZ* antisense RNA probe (Tautz and Pfeifle, 1989; Kasai et al., 1992). At least three identically staining independent transformants for each construct were analyzed to control for position effects.

RESULTS

Conserved sequence motif in CNS midline enhancers

Given the functional relationship between SIM, the AHRC proteins and other HLH proteins, the midline enhancers from *sim*, *sli* and *Tl* were sequenced and searched for sequence motifs related to the XRE (TNGCGTG) and E-boxes (CANNTG). DNA corresponding to 2.8 kb of the *sim* gene was also sequenced in a related species *Drosophila virilis*, since conserved flanking sequences between homologous genes often identify regulatory elements (Bray and Hirsh, 1987). A candidate sequence motif, ACGTG, was found five times in *sim* within the 2.8 kb fragment, four times in the 900 bp *Tl*

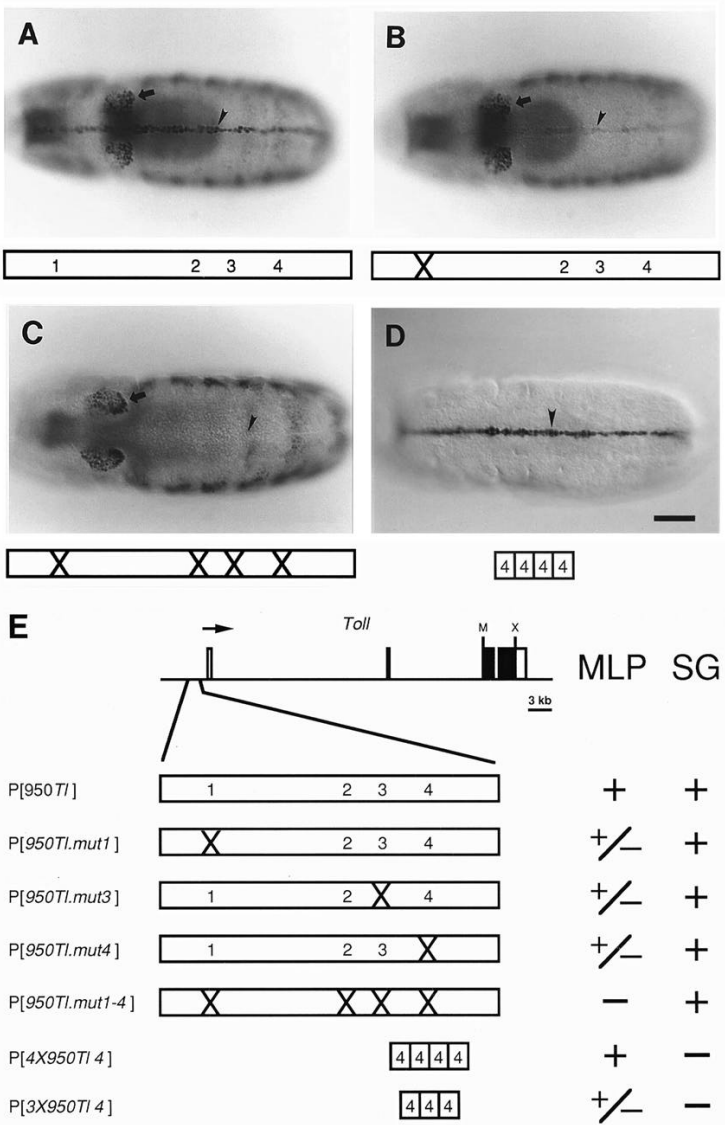

Fig. 2. *Tl* CMEs are necessary and sufficient for CNS midline expression. (A-D) Embryos containing P[*950Tl/lacZ*] constructs are stained with an antibody against β-galactosidase followed by HRP immunohistochemistry. Ventral surface is shown with anterior to the left. Scale bar, 50 μm. (A) Staining of P[*950Tl*] stage 10 embryo shows MLP (arrowhead) and salivary gland placode expression (arrow). (B) Staining of P[*950Tl.mut1*] stage 10 embryo, which has one of the four CMEs mutated, shows weak MLP expression (arrowhead) although salivary gland placode expression (arrow) is unaffected. Constructs P[*950Tl.mut3*] and P[*950Tl.mut4*] stain similarly. (C) Staining of P[*950Tl.mut1-4*] stage 10 embryo, which has four CMEs mutated, shows an absence of MLP expression (arrowhead), although salivary gland placode expression (arrow) is unaffected. (D) Staining of P[*4X950Tl4*] shows strong and specific MLP expression in a stage 10 embryo. This strain harbors four multimerized 20 bp fragments containing *Tl* site 4 driving *lacZ*. (E) Genomic map of the *Tl* gene (Wharton and Crews, 1993) showing location of the *950Tl* fragment and staining summary of each construct. Raised blocks represent exons with coding sequences filled and non-coding sequences unfilled. The location of the translational start and termination sites are marked with a 'M' and 'X', respectively. The *950Tl* fragment that confers CNS midline expression lies approximately 1.5 kb upstream from the start of transcription (arrow). Shown below are the different constructs analyzed in this paper. The first five constructs utilized the *950Tl* fragment fused to C4PLZ, which contains a weak P-element promoter fused to *lacZ*. This fragment has four CMEs, labeled 1-4. Specific mutations were generated at the sites marked with an 'X'. The bottom two constructs utilized multiple copies of a 20 bp fragment incorporating *Tl* site 4. P[*4X950Tl4*] had four copies tandemly linked to C4PLZ and P[*3X950Tl4*] had three copies. Each box with an enclosed '4' refers to a single copy of *Tl* site 4. Expression was monitored in CNS midline precursor cells (MLP) and the salivary gland placode (SG). High levels of expression are indicated by '+', weak levels by '+/–' and absence of expression by '–'.

fragment (referred to as *950Tl*; Wharton and Crews, 1993) and once in the 380 bp *sli* fragment (Fig. 1). Four of five sites in *sim* were conserved between *Drosophila melanogaster* and *Drosophila virilis*.

Germ-line transformation shows requirement of conserved sequence element for CNS midline transcription

Germ-line transformation was utilized to test whether these motifs were required for CNS midline transcription. Specific mutations were introduced into the ACGTG sequences in DNA fragments of *sim*, *sli* and *Tl* that drive CNS midline transcription. Mutagenized fragments were fused to P element promoter-fusion (Thummel et al., 1988) or enhancer-tester vectors (Wharton and Crews, 1993), and introduced into germ-line DNA by microinjection. Embryos collected from homozygous transformed strains were stained using anti-β-galactosidase antibody or by in situ hybridization with a *lacZ* probe.

Tl

The P[*950Tl*] construct drives β-galactosidase expression in the CNS midline cells during embryonic stages 9-13 and is also expressed in the salivary gland placode, epidermis and gut (Wharton and Crews, 1993) (Fig. 2A,E). Specific mutagenesis of a single site, either 1, 3 or 4, shows a significant decrease in CNS midline expression (Fig. 2B,E). When all four sites are mutated, CNS midline expression is absent (Fig. 2C,E). Expression in other tissues, such as the salivary gland placode, is unaffected in all constructs providing an internal control. This experiment demonstrates that ACGTG motifs contribute to *Tl* expression within MLP cells.

sli

The 380 bp *sli* fragment drives β-galactosidase expression in MLG from stages 11-17 (Wharton and Crews, 1993) (Fig. 3A,C). This construct is also expressed in cells along the midline of the frontal sac. Mutation of the single ACGTG motif in *sli* eliminates MLG expression although expression in the frontal sac is unaltered (Fig. 3B,C).

sim

Previous experiments demonstrated that a 3.7 kb fragment of *sim* containing the early promoter (P_E) drives both blastoderm MEC and *sim*-dependent MLP transcription (Nambu et al., 1991; Kasai et al., 1992). Progressive 5′ deletions of this fragment reveal that high levels of MLP expression can be driven by fragments 2.8, 2.1 and 1.5 kb upstream of exon 2, while MLP expression is strongly reduced in an 0.9 kb fragment (Fig. 4E). MEC expression is at wild-type levels only in the 3.7 and 2.8 kb fragments (Fig. 4E). The unmutated 2.8 kb *sim-lacZ* construct maintained high levels of CNS midline transcripts at stage 11 and later (Fig. 4A). In contrast, specific mutations in ACGTG motifs were introduced together into sites 1, 2, 3 and 4 of the 2.8 kb fragment, and MLP expression was completely abolished by stage 11 (Fig. 4B). As an internal control, MEC and early MLP expression (stages 5-9) were identical to the unmutated fragment (Fig. 4C,D). This resembles *sim* gene transcription in *sim* mutant embryos. These results suggest that *sim* autoregulation, but not initial *sim* transcription, requires the ACGTG motifs.

In summary, these experiments indicate that the ACGTG motif is required for CNS midline expression in both the MLP cells early in neurogenesis and later in the MLG.

Minimal CNS midline element is sufficient for CNS midline transcription

To test whether the ACGTG motif by itself was capable of directing midline expression, multimers of three or four 20 bp fragments incorporating *Tl* site 4 were cloned into an enhancer-tester *lacZ* vector. Upon germ-line transformation and anti-β-galactosidase staining, embryos harboring these constructs expressed β-galactosidase in CNS midline cells beginning at stage 9 (Fig. 2D,E). Indeed, midline expression from these constructs appeared strikingly similar to *sim* expression: initially in all MLP cells, and later becoming restricted to MLG and a subset of ventral neurons (data not shown) (Crews et al., 1988). Four copies of this element drove strong MLP expression,

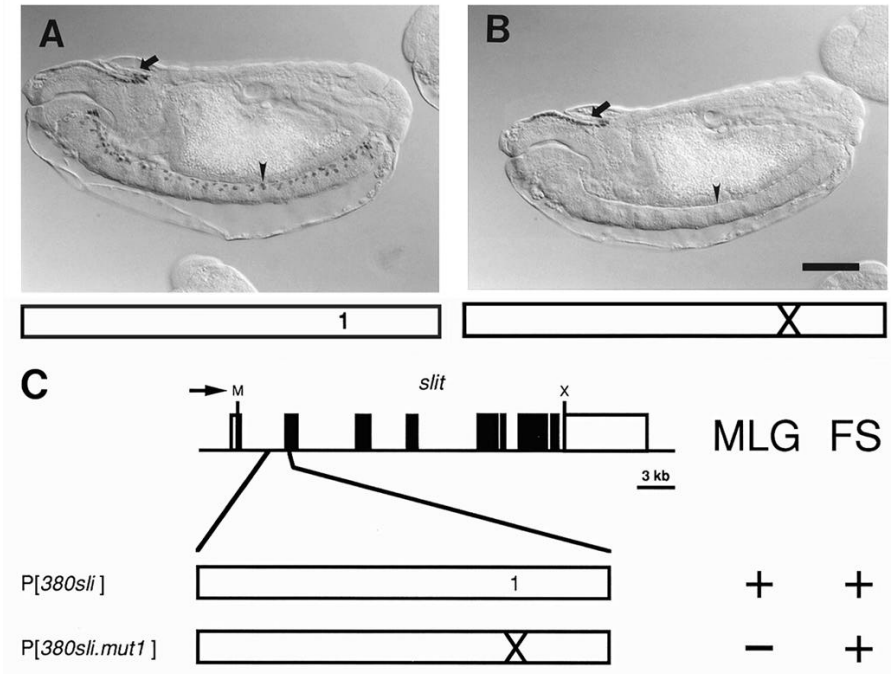

Fig. 3. Mutation of the single *sli* CME results in loss of MLG expression. (A,B) Embryos containing P[*sli*/*lacZ*] constructs are stained as in Fig. 2. Sagittal view is shown with anterior to the left. Scale bar, 50 μm. (A) Staining of P[*380sli*] stage 13 embryo shows strong late midline staining, including MLG (arrowhead) and midline frontal sac expression (arrow). This unmutated P element has an intact CME (labeled '1' in the box below). (B) Staining of P[*380sli.mut1*] stage 13 embryo, which has its CME mutated (marked with an 'X'), shows a complete absence of CNS midline expression (arrowhead) although frontal sac expression (arrow) remains strong. (C) Genomic map of the *sli* gene (Wharton and Crews, 1993) showing location of the 380 bp fragment downstream of the start site of transcription (arrow) and staining summary of each construct. Each fragment was fused to C4PLZ, transformed and assayed for midline glia (MLG) and frontal sac (FS) expression.

while three copies had weaker expression (Fig. 2E), suggesting a positive relationship between element dosage and expression level. This effect is similar to the dosage dependence of ACGTG copy number in the P[*950Tl*] mutagenesis experiments. Midline expression of P[*4X950Tl4*] was absent in a *sim*$^{-}$ background and expanded throughout the lateral neuroectoderm upon heat induction of a P[*hsp70/sim*] fly strain, demonstrating the dependence of its expression on *sim* gene function (data not shown). Similar results were previously obtained with midline enhancers of *sim*, *sli*, *Tl* (Nambu et al., 1990, 1991) and other CNS midline-expressed genes (Nambu et al., 1991; Muralidhar et al., 1993).

DISCUSSION

Transcriptional control of CNS midline development

These results define an ACGTG-containing CNS Midline Element (CME), which is required for *sim*-dependent CNS midline gene expression. For a gene to be activated in MLP, its regulatory region simply requires multiple CMEs and a promoter element. This is exemplified by the MLP expression of the *sim* and *Tl* genes. Each gene has 4-5 CMEs scattered over less than 1.5 kb of DNA and both genes show strong, comparable MLP expression. Mutations of individual CMEs within the *Tl* gene indicates that each CME contributes quantitatively to the overall level of expression. Comparison of the 1.5 kb *sim* fragment (5 CMEs) and the 0.9 kb *sim* fragment (2 CMEs) also suggests a correlation with CME copy number and midline expression. This relationship is further reinforced by the experiments in which either three or four CMEs were multimerized and fused to a heterologous promoter. Constructs with four CMEs reproducibly showed higher levels of expression than the constructs with three elements.

The *sli* gene is expressed in MLP and later in MLG, similar to the expression of *sim*. Previously, a 380 bp fragment was shown to be sufficient for MLG, but not MLP, expression. In this paper, we show that the 380 bp *sli* fragment contains a single CME that is required for MLG expression. Thus, a related control element is utilized for both MLP and MLG transcription. However, it is unknown whether the same transcription factors interact with the CME in MLP and MLG.

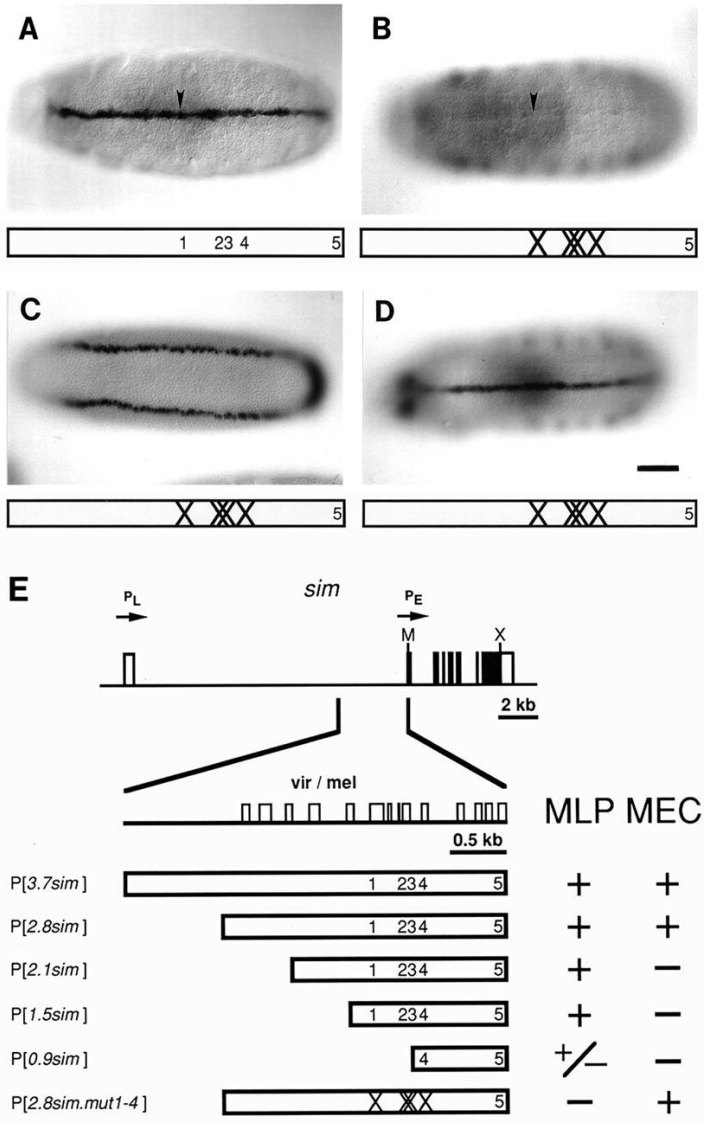

Fig. 4. Mutation of the *sim* CMEs results in loss of *sim* autoregulation. (A-D) Embryos containing P[*sim/lacZ*] constructs are hybridized with a labeled *lacZ* RNA probe followed by alkaline phosphatase histochemistry. Ventral surface is shown with anterior to the left. Scale bar, 50 μm. (A) Staining of P[*2.8sim*] stage 11 embryo shows MLP expression (arrowhead). This P element has all 5 CMEs intact. (B) Staining of P[*2.8sim.mut1-4*] stage 11 embryo, which has four of the CMEs mutated, shows absence of MLP expression (arrowhead). (C,D) Staining of P[*2.8sim.mut1-4*] stage 5 embryo (C) and stage 9 embryo (D) shows high levels of *lacZ* transcripts, indicating that the initial *sim*-independent transcription of the *sim* gene is unaffected. (E) Genomic map of the *sim* gene (Nambu et al., 1990) showing location of the different DNA fragments and expression summary of each construct. Location of the late promoter (P_L) and early promoter (P_E) are indicated with arrows. The 3.7 kb fragment that confers CNS midline expression lies upstream of exon 2 and contains P_E. Regions highly conserved between *Drosophila melanogaster* and *Drosophila virilis* within the 2.8 kb fragment are shown as open boxes. The location of the ACGTG motifs are indicated below and numbered 1-5; all but motif 1 are identical between *Drosophila melanogaster* and *Drosophila virilis*. Shown are the different constructs analyzed. The first five constructs are 5′ deletion fragments that contain P_E and are fused to Casper-AUG-β-gal. The P[*3.7sim*], P[*2.8sim*], P[*2.1sim*] and P[*1.5sim*] genes contained all 5 CMEs (numbered 1-5), whereas P[*0.9sim*] lacks the first three sites. The P[*2.8sim.mut1-4*] gene contains specific mutations in sites 1-4 (labeled with 'X's). Stage 11 embryos from each transformed line were analyzed by in situ hybridization with a *lacZ* probe for *sim*-dependent midline precursor expression 'MLP' and stage 5-7 embryos were analyzed for mesectoderm 'MEC' staining.

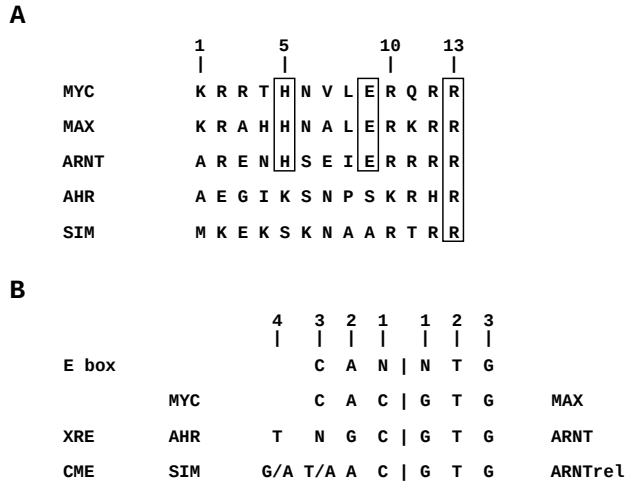

Fig. 5. Hypothetical model of bHLH-DNA interactions occurring at xenobiotic and CNS midline response elements. (A) Basic region amino acid sequences of human MAX (Blackwood and Eisenman, 1991), MYC, ARNT (Hoffman et al., 1991), murine AHR (Ema et al., 1992; Burbach et al., 1992) and *Drosophila* SIM (Nambu et al., 1990). Conserved amino acid residues that make sequence-specific DNA contacts in the MAX crystal structure are boxed (Ferre-D'Amare et al., 1993). (B) Nucleotide sequences of the E-box, MYC/MAX-binding site, XRE and CME with proposed binding proteins shown adjacent to each half site. One DNA strand is shown from 5′ to 3′. Nucleotide positions from the axis of symmetry are numbered above consensus. Axis of half-site is denoted by vertical dashed line. According to the MAX crystal structure, histidine 5 of MAX basic region makes contact with the G opposite C at position 3, glutamate 9 interacts with both the C and A at positions 3 and 2, and arginine 13 interacts with the G opposite C at position 1. AHR and SIM possess distinct amino acids at MAX positions 5 and 9 compared with MAX, MYC and ARNT. The existence of ARNTrel, a putative partner of SIM, is hypothesized from the sequence of the CME and SIM's relatedness to the AHRC proteins.

Although, we have a identified one component of *sli* MLG expression, the complete array of elements required for *sli* MLP and MLG transcription remain unknown. The CME multimerization experiments revealed that multiple copies of the *Tl4* CME are able to drive transcription in MLPs. Since the 380 bp *sli* fragment has a only a single CME, additional CMEs residing elsewhere within the *sli* gene may be required for *sli* MLP transcription. Additional elements may also be necessary for *sli* MLG transcription, although it is possible that the single *sli* CME is sufficient for MLG expression. For instance, there may be sites for additional transcriptional repressors and activators that direct MLG transcription. Further mutational dissection of the *sli* 380 bp fragment and analysis of the *sli* CME will further resolve these issues.

bHLH-PAS transcription factor interactions

As shown in Fig. 1, sequence alignment of the ten ACGTG motifs indicate that additional nucleotides are favored, yielding a 7 bp consensus [5′ (G/A)(T/A)ACGTG 3′] whose 3′ proximal residues resemble both the XRE (TNGCGTG) and a half-site of the symmetric E-box motif (CANNTG). bHLH proteins usually recognize the E-box as heterodimers, each partner forming sequence-specific contacts within the major groove of one half-site (Blackwell and Weintraub, 1990; Ferre-D'Amare et al., 1993). Examination of the amino acid sequences of the AHRC bHLH-PAS subunits suggests a model for how an AHR/ARNT bHLH heterodimer recognizes the XRE that may be relevant in understanding the function of SIM and the CME.

ARNT possesses basic region amino acids identical to both MYC and MAX in positions important for sequence specific DNA-protein contacts: a histidine at basic region position 5, glutamate at 9 and arginine at 13 (Fig. 5A) (Ferre-D'Amare et al., 1993). AHR, by contrast, possesses a lysine and serine at positions 5 and 9, respectively, suggesting that AHR would not interact with an E-box half site. Thus, ARNT would recognize the E-box half site, GTG, and AHR would recognize the non-E-box half of the XRE, TNGC (Fig. 5B). Note that AHR and ARNT both possess arginine 13 that by mutational and crystallographic analysis (Dang et al., 1992; Halazonetis and Kandil, 1992; Ferre-D'Amare et al., 1993) specifies a CG dinucleotide pair adjacent to the axis of symmetry in the cognate binding site for bHLH dimers, and those sequences are also present at the center of the XRE core sequence.

This model also provides a hypothesis for recognition of the CME by SIM. In this scheme, the CME is composed of a GTG E-box half site and a diverged (G/A)(T/A)AC sequence. *sim* also encodes a bHLH-PAS domain protein that is more related in primary sequence and genomic structure to AHR than ARNT is to AHR, suggesting SIM and AHR diverged from a common ancestral gene (Burbach et al., 1992; Ema et al., 1992; Schmidt et al., 1993). SIM, like AHR, possesses amino acid residues at positions 5 and 9, different from MYC, MAX and ARNT, suggesting it is unlikely to bind a GTG half site. SIM also possesses an arginine at position 13, suggesting a CG pair adjacent to the axis of symmetry of its cognate binding site and that sequence is also present at the center of the CME. By analogy to the AHRC, we propose that SIM binds the (G/A)(T/A)AC half site of the CME and that a *Drosophila* ARNT-related protein forms a heterodimer with SIM to interact with the GTG half site of the CME (Fig. 5). However, it is worth noting that asymmetric E-box-like sequences can be recognized in vitro (and probably in vivo) by homodimers of the bHLH proteins E47 (Blackwell and Weintraub, 1990), TWIST (Ip et al., 1992a, b; Kasai and Crews, unpubl.), ENHANCER OF SPLIT (Tietze et al., 1992), HES-5 (Akazawa et al., 1992) and HAIRY (Van Doren and Posakony, personal communication). Thus, the occurrence of an asymmetric E-box-like sequence is not definitive evidence that the corresponding bHLH transcription factors act as heterodimers.

In summary, this paper provides evidence that a family of related bHLH transcription factors and cognate DNA sequence elements control CNS midline development and carcinogen metabolism. The CME is required for CNS midline transcription throughout *Drosophila* embryonic development. MLP expression requires multiple CMEs and a promoter element. The control of later CNS midline expression in differentiated glial cells can also involve the CME, but may include additional regulatory elements. The *sim* gene is genetically required for all CNS midline transcription and for the reasons described in this paper, it is likely that SIM acts directly through the CME to exert its control over CNS midline transcription and development. Future experiments will seek to identify ARNT-related or other bHLH-PAS dimerization partners for the SIM protein, biochemically analyze their predicted interactions with the CME

and further understand the multiplicity of control elements that restrict midline expression in differentiated neurons and glia.

S. T. C. would like to acknowledge the intellectual contributions of Eric Davidson in stimulating his interest in the control of gene expression during development. We would like to thank Al Courey, Oliver Hankinson and Song Hu for a critical reading of the manuscript, Jim Posakony for advice and the Crews laboratory for support throughout this project. K. A. W was supported by a NIH Medical Scientist Training Program (GM-08042). 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. Y. K. is a postdoctoral trainee supported by DHHS PHS National Institutional Research Service Award T 32 CA09056 and an NIH postdoctoral grant. This work was supported by grants from the Lucille P. Markey Charitable Trust, Jonsson Cancer Center Foundation (J911209) and the NIH (HD25251) to S. T. C.

REFERENCES

Akazawa, C., Sasai, Y., Nakanishi, S. and Kageyama, R. (1992). Molecular characterization of a rat negative regulator with a basic helix-loop-helix structure predominantly expressed in the developing nervous system. *J. Biol. Chem.* **267**, 21879-21885.

Blackwell, T. K. and Weintraub, H. (1990). Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250**, 1104-1110.

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.

Bray, S. J. and Hirsh, J. (1987). The *Drosophila virilis* dopa decarboxylase gene is developmentally regulated when integrated into *Drosophila melanogaster*. *EMBO J.* **5**, 2305-2311.

Burbach, K. M., Poland, A. and Bradfield, C. A. (1992). Cloning of the Ah-recptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc. Nat. Acad. Sci. USA* **89**, 8185-8189.

Crews, S. T., Thomas, J. T. and Goodman, C. S. (1988). The Drosophila *single-minded* gene and the molecular genetics of CNS midline development. *Cell* **52**, 143-151.

Dang, C. V., Dolde, C., Gillison, M. L. and Kato, G. J. (1992). Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc. Natl Acad. Sci. USA* **89**, 599-602.

Denison, M. S., Fisher, J. M. and Whitlock, Jr., J. W. (1988). The DNA recognition site for the dioxin-Ah receptor complex: nucleotide sequence and functional analysis. *J. Biol. Chem.* **263**, 17221-17224.

Devereux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

Dolwick, K. M., Swanson, H. I. and Bradfield, C. A. (1993). *In vitro* analysis of Ah receptor domains involved in ligand-activated DNA recognition. *Proc. Nat. Acad. Sci. USA* **90**, 8566-8570.

Ema M., Sogawa, K., Watanabe, N., Chujoh, Y., Matsushita, N., Gotoh, O., Funae, Y. and Fujii-Kuriyama, Y. (1992). cDNA cloning and structure of mouse putative Ah receptor. *Biochem. Biophys. Res. Comm.* **184**, 246-253.

Ferre-D'Amare, A. R., Prendergast, G. C., Ziff, E. B. and Burley, S. K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363**, 38-45.

Franks, R. G. and Crews, S. T. (1994). Transcriptional activation domains of the single-minded bHLH protein are required for CNS midline cell development. *Mech. Dev.* **45**, 269-277.

Halazonetis, T. D. and Kandil, A. N. (1992). Predicted structural similarities of the DNA binding domains of c-myc and endonuclease EcoRI. *Science* **255**, 464-465.

Hoch, M., Schroder, C., Seifert, E. and Jäckle, H. (1992). *cis*-acting control elements for *Krüppel* expression in the *Drosophila* embryo. *EMBO J.* **9**, 2587-2595.

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.

Huang, Z. J., Edery, I. and Rosbash, M. (1993). PAS is a dimerization domain common to *Drosophila period* and several transcription factors. *Nature* **364**, 259-262.

Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K. and Levine, M. (1992a). *dorsal-twist* interactions establish *snail* expression in the presumptive mesoderm of the *Drosophila* embryo. *Genes Dev.* **6**, 1518-1530.

Ip, Y. T., Park, R. E., Kosman, D., Bier, E. and Levine, M. (1992b). The *dorsal* gradient morphogen regulates stripes of *rhomboid* expression in the presumptive neuroectoderm of the *Drosophila* embryo. *Genes Dev.* **6**, 1728-1739.

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

Kasai, Y., Nambu, J. R., Lieberman, P. M. and Crews, S. T. (1992). Dorsal-ventral patterning in *Drosophila*: DNA binding of snail protein to the single-mined gene. *Proc. Natl. Acad. Sci. USA* **89**, 3414-3418.

Kosman, D., Ip, Y. T., Levine, M. and Arora, K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. *Science* **254**, 118-122.

Leptin, M. (1991). *twist* and *snail* as positive and negative regulators during *Drosophila* mesoderm development. *Genes Dev.* **5**, 1568-1576.

Matsushita, N., Sogawa, K., Ema, M., Yoshida, A. and Fujii-Kuriyama, Y. (1993). A factor binding to the xenobiotic responsive element (XRE) of P-4501A1 gene consists of at least two helix-loop-helix proteins, Ah receptor and Arnt. *J. Biol. Chem.* **268**, 21002-21006.

Muralidhar, M. G., Callahan, C. A. and Thomas, J. B. (1993). *Single-minded* regulation of genes in the embryonic midline of the *Drosophila* central nervous system. *Mech. Dev.* **41**, 129-138.

Murre C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y. and Jan. Y. N., Cabrera, C., Buskin, J. N., Hauschka, S. D. and Lassar, A. B., Weintraub, H. and Baltimore, D. (1989). Interactions between heterologous helix loop helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537-544.

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.

Nambu, J. R., Lewis, J. O., Wharton, Jr. K. A. and Crews, S. T. (1991). The Drosophila *single-minded* gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67**, 1157-1167.

Nambu, J. R., Lewis, J. O. and Crews, S. T. (1993). The development and function of the *Drosophila* CNS midline cells. *Comp. Biochem. Physiol.* **104A**, 399-409.

Probst, M. R., Reisz-Porszasz, S., Agbunag, R. V., Ong, M. S. and Hankinson, O. (1993). Role of the aryl hydrocarbon receptor nuclear translocator protein in aryl hydrocarbon (dioxin) receptor action. *Mol. Pharmacol.* **44**, 511-518.

Rao, Y., Vaessin, H., Jan, L. Y. and Jan, Y. N. (1991). Neuroectoderm in *Drosophila* embryos is dependent on the mesoderm for positioning but not for formation. *Genes Dev.* **5**, 1577-1588.

Reyes, H., Reisz-Porszasz, S. and Hankinson, O. (1992). Identification of the Ah receptor nuclear translocator protein (ARNT) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193-1195.

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

Schmidt, J. V., Carver, L. A. and Bradfield, C. A. (1993). Molecular characterization of the murine Ahr gene. Organization, promoter analysis, and chromosomal assignment. *J. Biol. Chem.* **268**, 22203-22209.

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.

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.

Tietze, K., Oellers, N. and Knust, E. (1992). Enhancer of split^{D}, a dominant mutation of *Drosophila*, and its use in the study of functional domains of a helix-loop-helix protein. *Proc. Natl. Acad. Sci. USA* **89**, 6152-6156.

Wharton, Jr., K. A. (1992). *Ph.D. Thesis*, University of California at Los Angeles.

Wharton, Jr., K. A. and Crews, S. T. (1993). CNS midline enhancers of the *Drosophila slit* and *Toll* genes. *Mech. Dev.* **40**, 141-154.

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

(Accepted 24 August 1994)