Identification of Motifs That Are Conserved in 12 Drosophila Species and Regulate Midline Glia vs. Neuron Expression

Patricia Estes,1 Eric Fulkerson and Yi Zhang

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614
Manuscript received August 16, 2007
Accepted for publication December 15, 2007

ABSTRACT

Functional complexity of the central nervous system (CNS) is reflected by the large number and diversity of genes expressed in its many different cell types. Understanding the control of gene expression within cells of the CNS will help reveal how various neurons and glia develop and function. Midline cells of Drosophila differentiate into glial cells and several types of neurons and also serve as a signaling center for surrounding tissues. Here, we examine regulation of the midline gene, wrapper, required for both neuron–glia interactions and viability of midline glia. We identify a region upstream of wrapper required for midline expression that is highly conserved (87%) between 12 Drosophila species. Site-directed mutagenesis identifies four motifs necessary for midline glial expression: (1) a Single-minded/Tango binding site, (2) a motif resembling a pointed binding site, (3) a motif resembling a Sox binding site, and (4) a novel motif. An additional highly conserved 27 bp are required to restrict expression to midline glia and exclude it from midline neurons. These results suggest short, highly conserved genomic sequences flanking Drosophila midline genes are indicative of functional regulatory regions and that small changes within these sequences can alter the expression pattern of a gene.

In metazoan organisms, the central nervous system (CNS) is a complicated communication system characterized by diverse cells that make many intricate connections with a variety of cell types. To generate cellular diversity within the CNS, genes that control the specification, development, and function of cells must be tightly regulated in both space and time. Understanding the “regulatory code,” or how regulatory sequences flanking genes appropriately direct their expression, remains a major challenge to biologists within many fields, including molecular, cellular, developmental, evolutionary, and systems biology, as well as bioinformatics.

Regulatory regions of genes contain binding sites for transcription factors that activate or repress transcription. Such binding sites consist of DNA sequence motifs of between 4 and 20 bp, and oftentimes a particular motif is repeated several times within the regulatory regions of genes. New targets for certain transcription factors have been identified by searching the genome for shared motifs, particularly repeated motifs, in close proximity to one another (Rajewsky et al. 2002; Freeman et al. 2003) and to other binding sites for transcription factors in the same developmental pathway (Schroeder et al. 2004). Many false positives are identified in these studies, meaning that any putative regulatory region identified in silico must be confirmed in vivo. However, the success rate can be improved by including evolutionary comparisons of putative regulatory regions between species (Berman et al. 2004; Sinha et al. 2004; Wenick and Hobert 2004; Rebeiz et al. 2005; Pennacchio et al. 2006). Here, we combine the power of evolutionary comparisons of the currently available Drosophila genomes with fly transgenesis to identify regulatory sequences and motifs required for gene expression within the CNS.

To study CNS gene regulation, we focus on midline cells that play a central role in the formation of the CNS in both vertebrate and invertebrate species. In Drosophila, these cells provide signaling information to axons during their growth and develop into both neurons and glia themselves (Thomas et al. 1988; Nambu et al. 1990; Nambu et al. 1991; Bossing and Technau 1994; Jacobs 2000; Dickson 2002; Garbe and Bashaw 2004). Over 300 genes have been identified that are expressed in the various midline cell types sometime during fly embryogenesis, making the midline a useful model for understanding transcriptional control of gene regulation within a CNS cell type (Nambu et al. 1991; Jacobs 2000; Kearny et al. 2004; Wheeler et al. 2006).

Genetic experiments indicate that activation of the master control gene, single-minded (sim), leads to CNS midline cell development. Such experiments show that mutations in sim eliminate midline cells (Thomas et al. 1988; Nambu et al. 1990), and ectopic activation of sim in cells of the neuroectoderm can transform cells destined
to other cell fates into midline cells (Nambu et al. 1991). To regulate transcription, the basic helix-loop-helix (bHLH)-PAS transcription factor, Sim, must first form heterodimers with its partner, Tango (Tgo), another bHLH-PAS protein, before binding DNA sequences, called CNS midline elements (CMEs) (ACGTG) to activate transcription (Ohshiro and Saigo 1997; Sonnenfeld et al. 1997). Tgo is ubiquitously expressed in Drosophila embryos, but only located in the nucleus of cells that also express one of its partners, such as sim in midline cells (Ward et al. 1998).

The bHLH-PAS proteins, Sim and Tgo, are critical for the expression of many midline genes, but the presence of these two proteins alone is insufficient to account for the dynamic expression pattern of most genes in the midline. Moreover, Tgo interacts with another bHLH-PAS protein, Tracheless (Trh), to activate a set of genes within the developing respiratory system of the fly, the trachea (Ohshiro and Saigo 1997; Sonnenfeld et al. 1997). The binding site for Trh/Tgo heterodimers appears to be the same as that of Sim/Tgo heterodimers and many genes expressed in midline cells are also expressed in trachea. Additional evidence for shared regulatory properties of midline and tracheal genes comes from multimerizing the CME and fusing it to a reporter gene. Such a reporter is expressed in both midline and tracheal cells of transgenic flies (Ohshiro and Saigo 1997; Sonnenfeld et al. 1997) indicating the importance of this single binding site. However, gene sets exist that are unique to either the midline or trachea and even within midline cells, different gene sets are activated and inactivated at various stages of development. We would like to determine the molecular basis of the unique and shared regulation of various gene sets in midline cells and different midline lineages.

In addition to sim and tgo, the transcription factors Dichaete (D), a Sox HMG protein, and Dfr, a POU domain protein, regulate genes expressed in midline glia (Ma et al. 2000; Bergmann et al. 2002). The D protein directly interacts with the PAS domain of Sim and the POU domain of Dfr and all three genes activate expression of slit in midline glia (Ma et al. 2000).

Most genes expressed in the midline, including sim, are expressed in additional tissues within the developing fly embryo. In contrast, wrapper is restricted largely to midline glia, with a relatively low level of expression in some chordotonal cells of the embryo (Noordermeer et al. 1998). To understand how wrapper is restricted to midline glia during Drosophila embryogenesis, we are studying its regulation, both the transcription factors that activate it and the regulatory sequences controlling its expression. Because wrapper is largely restricted to midline glia, the sequences controlling its expression are predicted to contain motifs for genes that regulate midline gene expression, unencumbered with motifs for factors expressed in other tissues. Moreover, the regions most likely to contain regulatory control elements (motifs) are tractable; the size of the genomic regions flanking the wrapper transcription unit, and the first intron, are relatively small.

The availability of sequenced genomes for 12 Drosophila species provides a unique opportunity for fly geneticists to study the evolution of genes (FlyBase Blast at http://flybase.net/blast/; Stark et al. 2007). While coding regions of genes are conserved and can be compared between quite divergent species, regulatory regions of genes tend to change more rapidly. Genomic sequence comparisons between these 12 Drosophila species should greatly facilitate the identification of particularly important conserved regulatory motifs. Once identified, these regions can be compared in detail between various species to determine if a gene is regulated differently in different species.

Using this approach, we identify conserved sequences upstream of wrapper sufficient to provide midline specific expression of reporter genes in embryos of Drosophila melanogaster. Such wrapper reporter constructs respond to mutations in known regulators of midline cells. The presence of invariant sequences shared by all 12 Drosophila species examined, suggests that any changes within this conserved region might reduce midline expression. To test this, we mutated select nucleotides and demonstrated the importance of four motifs within the conserved region. In contrast to these changes, four other sets of 2- to 3-nucleotide changes within the highly conserved wrapper regulatory region had no deleterious affect on midline expression. In addition to motifs needed for positive regulation in midline glia, we identified a region required to restrict expression to midline glia and prevent expression in a group of midline neurons, including the progeny of the median neuroblast. Taken together with previous studies on transcriptional regulation within midline glia, these results suggest that at least one Sim/Tgo binding site (CME) appears to be critical for expression in midline glia, and at least four additional sites work together with the CME to both positively and negatively regulate expression in midline cells.

**MATERIALS AND METHODS**

**Fly lines:** The line yw® was used for wild-type embryos. To facilitate identification of the homozygous mutant embryos, the D (fsh)® null allele (Nambu and Nambu 1996) and the dfr® allele (Salzberg et al. 1994) were placed over the TM3, Ubx, trithorax-lacZ third chromosome balancer and the spi® allele (Tearle and Nusslein-Volhard 1987) was balanced over the CyO, wingsless-lacZ second chromosome.

Rat α-sim antibody staining (Ward et al. 1998) was used to identify homozygous sim® mutant embryos. The following fly lines were used for ectopic expression studies: UAS-sim-GFP (Estes et al. 2001), UAS-sph/sh (Schweitzer et al. 1995), and da-GAL4 (Gierel et al. 1997).

**Germline transformation:** P element-mediated germ-line transformation was carried out as previously described (Rubin and Spradling 1982).
<table>
<thead>
<tr>
<th>Primers used to amplify <strong>wrapper</strong> upstream sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA</td>
</tr>
<tr>
<td>WB</td>
</tr>
<tr>
<td>K1</td>
</tr>
<tr>
<td>K2</td>
</tr>
<tr>
<td>K3</td>
</tr>
<tr>
<td>K4</td>
</tr>
<tr>
<td>K5</td>
</tr>
<tr>
<td>K7</td>
</tr>
<tr>
<td>X1</td>
</tr>
<tr>
<td>X2</td>
</tr>
<tr>
<td>X3</td>
</tr>
<tr>
<td>X4</td>
</tr>
<tr>
<td>X5</td>
</tr>
<tr>
<td>X6</td>
</tr>
</tbody>
</table>

Engineered KpnI and XhoI sites used to insert the fragments into **pHstinger** are indicated in italics.

**Immunohistochemistry:** Antibody staining of embryos was carried out essentially as described (Patel 1994). The primary antibodies used in this study were mouse monoclonal **a**-**wrapper** (1:5), a-**engrailed** (undiluted) obtained from the Developmental Studies Hybridoma Bank in Iowa, rat **a**-**sim** antibody (Ward et al. 1998), rabbit **a**-**β-galactosidase** (1:1000 or 1:3000; Cappel), and rabbit **a**-**GFP** (1:50; Invitrogen, Carlsbad, CA). All secondary antibodies (**a**-mouse-488, **a**-rabbit-Texas red, **a**-rabbit-488, and **a**-mouse 568) were used at 1:200. Confocal images were obtained on a Zeiss 410 microscope at the University of North Carolina in Chapel Hill and a Zeiss Pascal microscope at North Carolina State University.

**Generation and injection of **wrapper** reporter constructs:** Various fragments from the region 5′ of the **wrapper** transcription unit (Figure 2B) were amplified using the polymerase chain reaction with the primers listed below and genomic DNA from ywp2 flies. After amplification, the fragments were first inserted into the **pSTBlue** vector (Novagen, San Diego) and subsequently cassetted into **pHstinger** (Barolo et al. 2000) using KpnI/XhoI digestion. To generate the **wrapper:**W construct, a 884-bp BglII–KpnI fragment was inserted into BglII–KpnI digested **pHstinger**. The reporter constructs were then injected into ywp2 embryos using P element-mediated transformation. For each construct, at least three fly lines were examined.

**Primers used to amplify **wrapper** upstream sequences:** The following **wrapper** genomic fragments (Figure 2B) were amplified using the primers indicated in parentheses and listed in Table 1: W (5′WA 3′WB), A (5′K1 3′X5), C (5′K4 3′X2), D (5′K5 3′X1), E (5′K1 3′X3), F (5′K4 3′X1), G (5′K3 3′X4), K (5′K3 3′X6), and L (5′K7 3′X4).

**Site-directed mutagenesis:** To test the function of sequence motifs within the **wrapper** regulatory sequences, two or three nucleotides were changed within the **wrapper** G fragment in the **pSTBlue** vector, using the QuickChange Site-directed Mutagenesis Kit (Stratagene, La Jolla, CA) and the primers listed in Table 2. Selection of mutagenized sequences was facilitated by the generation of the following restriction enzyme recognition sites after mutagenesis: G1, NalI; G5, XhoI; G6, PstI; G2, EcoRI; G3, AclI; G7, EcoRI; G8, Smal; and G4, Nhel. Each mutant was subsequently sequenced and then cassetted into the **pHstinger** vector using KpnI/XhoI digestion for injection into flies.

**RESULTS**

To facilitate the identification of sequences responsible for **wrapper** expression in the midline glia of Drosophila, we first examined the genomic region flanking the **wrapper** transcription unit using Fly BLAST to determine the degree of conservation between the 12 available Drosophila species. The regions most likely to contain regulatory control elements (motifs) of **wrapper** are tractable; the genomic regions flanking the transcription unit and the first intron are relatively small. The results of this analysis highlighted a region between −492 and −326 upstream of the transcription start site of **wrapper** that is highly conserved in all Drosophila species examined (**melanogaster**, **simulans**, **sechelia**, **yakuba**, **erecta**, **anannassae**, **pseudobscura**, **persimilis**, **willistoni**, **yo67**).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>GTAATGAGTTTATCTGCAATCCAGATATTTGCTTGTCCG</td>
</tr>
<tr>
<td>G1R</td>
<td>CCGAGATTAATGATGATGACATACCACATGAGCTGGATAC</td>
</tr>
<tr>
<td>G2</td>
<td>CTGCGTATGCTGCTGACAAGATTCGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G2R</td>
<td>CACTCCCTCGTGGGATGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G3</td>
<td>CGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G3R</td>
<td>CGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G4</td>
<td>CGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G4R</td>
<td>CCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G5</td>
<td>CATCAGTGGGGGAAAGAGCTGCTAGTGGGGGAAAGAGCT</td>
</tr>
<tr>
<td>G5R</td>
<td>CAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G6</td>
<td>CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G6R</td>
<td>CACTCCCTCGTGGGATGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G7</td>
<td>GACTTTGATGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G7R</td>
<td>CACTCCCTCGTGGGATGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G8</td>
<td>CGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
<tr>
<td>G8R</td>
<td>CGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG</td>
</tr>
</tbody>
</table>

The specific nucleotides changed are underlined. Both the forward and the reverse (R) primers are shown.
A Midline glia activation

pseudoobscura  -TGTGCCACAATGCACTTTAATCTCAAAAATAAGCAGCAGCCACAGCTTCCTGAAGCGA
willistoni   -TTGCACACATGCACTTTAATCTCAAAAATAAGCAGCAGCCACAGCTTCCTGAAGCGA
ananassae   ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
yakuba       ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
erecta       ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
melanogaster ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
sechellia    ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
simulans     ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
mojavensis  ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT
virilis      ----------AATAACCTGATCTGCGGCCCTAATGAGATCTGGTTCT

* ************************************************** * ************************************************** *

pseudoobscura   GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
persimilis   GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
willistoni   GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
ananassae   GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
yakuba       GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
erecta       GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
melanogaster GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
sechellia    GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
simulans     GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
mojavensis  GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT
virilis      GAGGTGTAATAA-GATCAGC-**********-TACGGG-AATCTCGATTCT

B Midline neuron silencer

pseudoobscura   ---------AGAATTCGAA-AAAAACTACCGATCTGGCGCCTAATGAGATCTGGTTCT
persimilis   ---------AGAATTCGAA-AAAAACTACCGATCTGGCGCCTAATGAGATCTGGTTCT
sechellia    ---------AGAATTCGAA-AAAAACTACCGATCTGGCGCCTAATGAGATCTGGTTCT
simulans     ---------AGAATTCGAA-AAAAACTACCGATCTGGCGCCTAATGAGATCTGGTTCT
mojavensis  ---------AGAATTCGAA-AAAAACTACCGATCTGGCGCCTAATGAGATCTGGTTCT
virilis      ---------AGAATTCGAA-AAAAACTACCGATCTGGCGCCTAATGAGATCTGGTTCT

* ************************************************** * ************************************************** *

mojavensis, virilis, and grimshawi: http://flybase.net/blast/, particularly a 70-bp region (Figures 1 and 2). To test if these sequences are responsible for the \textit{w}rapper expression pattern in embryos, we first amplified this genomic region within a 884-bp fragment (W; Figure 2B), and then fused it to the green fluorescent protein (GFP) reporter gene within the \textit{pHstinger} vector, which contains a minimal Hsp70 promoter (BAROLO et al. 2000). This DNA construct (\textit{w}rapper \textit{W}:GFP) was injected into \textit{D. melanogaster} embryos using \textit{P} element-mediated transformation (RUBIN and SPRADLING 1982) to generate stable fly lines. Embryos containing this construct express GFP in midline glia (Figure 3, A–I) beginning at stage 12 of embryogenesis and throughout larval stages (not shown). We confirmed that GFP was expressed in midline glia by staining embryos simultaneously with either (1) \textit{w}rapper and GFP (Figure 3, A–C) or (2) \textit{s}im and GFP (Figure 3, D–I). Because \textit{w}rapper protein is found at the surface of midline glial cells, but the GFP produced by \textit{pHstinger} localizes to the nucleus, \textit{w}rapper protein encircles the GFP in these cells (Figure 3A). The \textit{w}rapper \textit{W}:GFP reporter construct also drives expression in a few additional cells within the lateral CNS (Figure 3E) and muscles (data not shown), a pattern that differs from the endogenous \textit{w}rapper expression pattern. This suggests that the \textit{W} fragment, although sufficient to drive high levels of expression in midline glia, lacks certain sequences that exclude expression in lateral CNS cells. To confirm the midline expression pattern generated by the reporters, all subsequent experiments were performed by staining embryos with both \textit{s}im and GFP at stage 16 of embryogenesis. These experiments revealed that GFP generated by the \textit{w}rapper \textit{W}:GFP reporter gene was indeed expressed in the midline glia, but not in the cells that develop into midline neurons (Figure 3, G–I). Next, to determine the minimal sequences required to provide expression in midline glia, we divided this 884-bp region into several subregions, fused them to GFP under the \textit{pHstinger} vector and tested their ability to drive midline expression in transgenic embryos. Region \textit{E} (Figure 3, J–L), extending from sequences –756 to –286, is sufficient to drive high levels of GFP expression in midline glia (Figure 3K). Moreover, a smaller 166-bp
To deter-
tentions in midline transcriptional activators:

To determine if previously identified midline transcriptional activators:

D. melanogaster E

The midline expression pattern

D. virilis

D. virilis

D. melanogaster

D. melanogaster (and function of the regulatory sequences of (Figure 3, J–L). These results suggest that the location

The fragments tested for the ability

to drive expression in midline glia are shown. Fragments W, E, G, and K drove midline expression, while fragments A, C, D, F, and L did not. The positions of the highly conserved sequences shown in Figure 1 are indicated with the gray line upstream of wrapper. The scale is at the bottom and represents base pairs. Whether or not each construct was expressed in midline glia and midline neurons is indicated.

viralis sequences can drive midline expression of a GFP reporter gene in melanogaster: Next, to determine

if the observed conservation at the sequence level

between Drosophila species reflects conservation in function, we tested if the corresponding E region from D. virilis could drive GFP reporter expression in the midline glia of D. melanogaster. The E region is also located upstream of wrapper in D. virilis and is 476 bp in length, while it is 462 bp in melanogaster. The entire E region is 58.4% identical in the two species, and the 70-bp highly conserved section differs by only six nucleotides (Figure 1). The midline expression pattern provided by the D. virilis wrapper E:GFP construct (Figure 3, M–O) in D. melanogaster flies is indistinguishable from that of the corresponding D. melanogaster E region (Figure 3, J–L). These results suggest that the location and function of the regulatory sequences of wrapper have been conserved between D. melanogaster and D. virilis.

The wrapper reporter genes are sensitive to reductions in midline transcriptional activators: To determine if previously identified midline transcription factors affect wrapper through these regulatory sequences, we tested the wrapper W:GFP reporter gene in a number of mutant backgrounds. First, we tested the effect of sim mutations on the reporter gene by placing the 884-bp wrapper W:GFP transgene into a sim mutant background (Figure 4, D–F), a mutation that eliminates Sim protein expression (Nambu et al. 1990). In this background, GFP expression was abolished in most cells, suggesting that sim expression is required for wrapper transcriptional activation in the midline. A few remaining cells did express GFP and these are likely lateral CNS cells also observed in wild-type embryos containing the wrapper W:GFP reporter (see Figure 3E).

Next, we tested the reporter gene in a spitz (spi) mutant background (Figure 4, G–I). Spi is a signaling molecule that plays multiple roles during Drosophila development (for a review, see Shilo 2005). Wrapper protein is normally found on the surface of midline glia where it mediates direct contact with the lateral CNS axons that cross the midline and promotes survival of midline glia (Noordermeer et al. 1998). In wrapper mutant embryos, this intimate interaction cannot occur and additional midline glia die. The amount of spi signaling provided by lateral CNS axons determines how many midline glia survive in each segment (Bergmann et al. 2002). The spi mutation severely disrupted CNS development so that the sim positive cells remained on the ventral surface of the embryo (Figure 4, G–I). Only a few of the sim positive cells also express GFP driven by wrapper regulatory sequences, suggesting these are the remaining midline glia. The cells expressing sim, but not GFP, are likely midline neurons (see Figure 4A), while cells expressing GFP and not sim are lateral glia, because they also express reversed polarity (repo; data not shown), a marker of lateral CNS glia (Campbell et al. 1994; Xiong

Figure 2.—Wrapper genomic region and fragments that provide midline glial expression. (A) The genomic region 2R:18,267,017–18,270,700 of D. melanogaster is shown schematically and wrapper exons are indicated with boxes and introns with lines. The region indicated with the dotted lines is expanded and shown in B. (B) The fragments tested for the ability to drive expression in midline glia are shown. Fragments W, E, G, and K drove midline expression, while fragments A, C, D, F, and L did not. The positions of the highly conserved sequences shown in Figure 1 are indicated with the gray line upstream of wrapper. The scale is at the bottom and represents base pairs. Whether or not each construct was expressed in midline glia and midline neurons is indicated.

(−492 to −327) G fragment (Figure 3, P–R), and an even smaller 119-bp (−492 to −374) internal K fragment (Figure 3, S–X), that both include the highly conserved region, are also sufficient to drive GFP expression in midline glia, but the level of expression is reduced compared to that of the E fragment and the intact 884-bp W fragment. None of the other reporter constructs drove GFP expression in the midline (wrapper A, C, D, F, or L; Figure 2B). The K fragment is also expressed in a subset of midline neurons (Figure 3, V–X), including progeny of the median neuroblast (Wheeler et al. 2006), suggesting that the larger W, E, and G fragments contain a silencer, which is absent from the K fragment and normally represses expression in these midline neurons.
These results indicate $spi$ mutations reduce the number of midline glia in the embryo and also reduce expression of the $wrapper$ $W:GFP$ reporter gene. In addition to $sim$ and $tgo$, the transcription factors $Dichaete$ ($D$), a Sox HMG protein (Ma et al. 2000), and $Dfr$, a POU domain protein, regulate genes expressed in midline glia (Bergmann et al. 2002). The $D$ protein directly interacts with the PAS domain of $Sim$ and the POU domain of $Dfr$ and all three genes activate expression of $slit$ in midline glial (Ma et al. 2000). We tested the $wrapper$ $W:GFP$ construct in both a $D$ (Figure 4, J–L) and $dfr$ (Figure 4, M–O) mutant background. In both cases, the number and behavior of midline cells was altered and they did not migrate to the dorsal region of the ventral nerve cord, as they normally do. While development of midline cells was disrupted in these mutant backgrounds as has been previously reported (Cerlet et al. 1996; Nambu and Nambu 1996; Ma et al. 2000) and fewer midline glia were present, robust $GFP$ expression was still observed from the reporter construct in the midline cells that remained, suggesting that (1) $D$ and $Dfr$ do not directly activate $wrapper$ via these regulatory sequences, (2) additional, redundant factors exist that can substitute for them, or (3) they can substitute for one another, as suggested by previous studies (Ma et al. 2000).

In summary, midline cell development was disrupted in $sim$, $spi$, $D$, and $dfr$ mutant backgrounds. The $sim^{trp}$ mutation eliminated midline glia and neurons, while a mutation in $spi$ eliminated most midline glia. As predicted, both $sim$ and $spi$ mutations severely reduced the number of cells expressing $GFP$ driven by the $wrapper$ $W:GFP$ reporter gene. In the $D$ and $dfr$ mutants, the number of midline glia was reduced and the remaining midline glia expressed high levels of $GFP$. 

Figure 3.—Sequences upstream of $wrapper$ from both $D. melanogaster$ and $D. virilis$ drive midline glial expression of a reporter gene in $D. melanogaster$. The $wrapper$ $W:GFP$ fusion construct, containing a 884-bp genomic fragment consisting of the sequences upstream of $wrapper$ (735 bp) and within the first exon (149 bp), drives expression of $GFP$ in midline glia. (A–F) Transgenic embryos containing the $wrapper$ $W:GFP$ reporter gene were stained with $\alpha$-$GFP$ (green, B and E) and $\alpha$-$wrapper$ (red, C) or $\alpha$-$sim$ antibody (red, F) during stage 13 of embryogenesis. The overlap in expression between the reporter gene and $wrapper$ (A) or the reporter gene and $sim$ (D) is shown. Embryos containing the $wrapper$ $W:GFP$ construct had $GFP$ expression in midline glia as well as cells of the lateral CNS cells (arrows). (G–X) The $wrapper$ reporter genes drive $GFP$ expression in midline glia during stage 16 of embryogenesis. Transgenic embryos containing the $wrapper$ $W:GFP$ (G–I), $wrapper$ $E:GFP$ (J–L), $wrapper$ $E:GFP$ from $D. virilis$ (M–O), $wrapper$ $G:GFP$ (P–R), $wrapper$ $K:GFP$ (S–X) reporter genes were stained with $\alpha$-$GFP$ (green; H, K, N, Q, T, and W) and $\alpha$-$sim$ antibody (red; I, L, O, R, and U) or $\alpha$-engrailed antibody (red, X) and analyzed by confocal microscopy. The overlap in expression between the reporter gene and $sim$ (G, J, M, P, and S) or between the reporter gene and $engrailed$ (V) is shown. Embryos containing the $wrapper$ $K:GFP$ construct had $GFP$ expression in midline glia as well as midline neurons (arrowheads, S–X). The expression pattern driven by $wrapper$ $K:GFP$ in midline neurons partially overlaps with $engrailed$ expression in midline neurons, suggesting it is in the progeny of the median neuroblast (Wheeler et al. 2006). Ventrolateral views of embryos are shown: anterior is to the left.
Ectopic expression of sim expands the wrapper expression domain: Ectopic sim expression converts neuroectodermal cells into midline cells and activates downstream, midline genes (Nambu et al. 1991; Kearney et al. 2004). To test the effect of ectopic sim on wrapper expression, we overexpressed sim using the UAS/GAL4 system (Brand and Perrimon 1993) and found that wrapper was expressed in neuroectodermal cells outside of the midline (Figure 5D), but not in all cells that overexpress sim (data not shown). In the UAS-sim/da-GAL4 embryos, wrapper is activated in cells that correspond to the lateral edges of the CNS and the cells in the anterior of each segment, with gaps in the expression pattern. Next, we tested if overexpression of the secreted form of spi (Schweitzer et al. 1995) could expand wrapper to cells outside the midline. Ectopic expression of secreted spi with the da-GAL4 driver also expanded wrapper expression (Figure 5G). To determine if it is possible to expand the expression domain of wrapper further, we overexpressed sim together with spi. This caused additional expansion of the wrapper domain into broad stripes within ectodermal cells (Figure 5J). In addition, overexpression of either sim or spi causes severe disruption in embryonic development.

Next, we tested the ability of sim and spi, either alone or together, to expand expression of the wrapper reporter genes. Expression from both the full-length reporter construct, wrapper W:GFP (Figure 5E), and the smaller wrapper G:GFP construct (Figure 5F) expanded in the UAS-sim/da-GAL4 embryos to a greater
extent than the endogenous wrapper gene. The expression pattern provided by the reporter constructs differs from the endogenous wrapper expression pattern, suggesting that either (1) some of the sequences that normally repress wrapper in tissues outside the midline glia may be missing in these wrapper W and G constructs, or (2) ectodermal cells overexpressing sim may undergo cell death and the GFP marker may be more stable in these dying cells compared to wrapper. Overexpression of spi alone also expanded reporter gene expression driven by both the wrapper W:GFP and wrapper G:GFP constructs (Figure 5, H and I). The GFP expression domain was expanded to a greater extent in embryos overexpressing sim together with spi (Figure 5, K and L) compared to those overexpressing either gene alone (Figure 5, E, F, H, and I). Taken together, the results indicate that (1) limiting the wrapper regulatory sequences and (2) increasing the cells that express sim and spi converts the highly specific expression pattern of wrapper from a single strip of CNS cells to a more general pattern throughout the ectoderm of the embryo. In addition, these results suggest that both the sim transcription factor and spi signaling molecule can activate transcription through these sequences derived from the regulatory region of wrapper.

Identification of sequence motifs required for wrapper expression: To both (1) identify functionally important motifs needed for wrapper expression and (2) determine if all the invariant nucleotides within the conserved 70-bp region of wrapper are essential for the observed midline glial expression pattern, we tested effects of select mutations within the wrapper G region. Previous studies have demonstrated the importance of sim/tgo, D, dfr, and spi for the expression of midline glial genes and, therefore, we first searched for possible binding sites for these factors. To examine both predicted binding sites, as well as other conserved sequences that may contain binding sites for novel factors, we divided the region into eight motifs that were tested for their effect on midline glia expression (Table 3 and Figure 6).

Each of these conserved motifs was tested by changing 2–3 nucleotides in the context of the D. melanogaster G fragment (Figure 6). The altered G fragments were then inserted independently into the pHstinger vector and injected into fly embryos to test their ability to drive midline expression.

Despite the high degree of conservation within this region, only four of the eight mutations that we tested (G1, G2, G5, and G7) caused a noticeable reduction in
reporter expression. Two of the mutation sets destroyed midline expression of the G reporter construct. The putative Sim/Tgo binding site (G2, CACGT) was needed for midline expression, because changing this sequence to GAAGT eliminated midline glial expression (Figure 7, D–F). In addition, another sequence, ATTTTATC (G5), located upstream of the G2, was required for expression of the reporter gene in wild-type embryos and changing this sequence to ATTGGATC eliminated midline glial expression (Figure 7, M–O). Two additional sites within the G fragment of wrapper are needed for midline expression: CGGAGAG (G7; Figure 7, S–U) and CACAAT (G1; Figure 7, A–C). If either of these motifs is altered, midline glial expression is greatly reduced, but not completely eliminated.

In contrast, the other four sets of mutations had no detectable negative effect on midline glial expression of the reporter gene, even though these sequences are conserved in all 12 Drosophila species. Mutation sets G4 (Figure 7, J–L), G6 (Figure 7, P–R), and G8 (Figure 7, V–X) did cause a low level of reporter gene activation in some midline neurons, suggesting that repressor proteins present in midline neurons may interact with these regions of the wrapper regulatory region. Finally, mutation G3 had no detectable positive or negative effect on expression of the reporter gene (Figure 7, G–I), despite being conserved in all 12 Drosophila species. In summary, the various mutations had three different effects on expression driven by the wrapper regulatory sequences: (1) some reduced midline glial expression, (2) some caused the inappropriate activation of the wrapper reporter in midline neurons, and (3) one was conserved, but apparently had no effect on wrapper regulation, in the context of the experiments presented here.

**DISCUSSION**

Through the comparison of genomic sequences flanking the wrapper gene in 12 Drosophila species, we identified a 119-nucleotide sequence that can drive high levels of transcription in midline glia. This region contains a sector in which 61/70 (87%) nucleotides are conserved within all Drosophila species examined, relatively high for noncoding genomic sequences. Prior genomic comparisons between two species, D. melanogaster and pseudoobscura, demonstrated the presence of short, yet highly conserved regulatory regions of genes, including giant, forkhead, m7, snail, even-skipped, and sloppy-paired (Papatsenko et al. 2006). Because consensus binding sites of most transcription factors generally can vary at several positions, it is surprising to see such a high level of conservation in regulatory sequences among all 12 fly species studied here. This may be a consequence of interactions between various factors that bind and regulate transcription in these regions that dictate a particular spatial requirement and order (Zinzen et al. 2006). Previous studies have identified a regulatory network consisting of sim and tgo,

### TABLE 3

<table>
<thead>
<tr>
<th>Motif</th>
<th>Sequence</th>
<th>Possible transcription factor</th>
<th>Midline glia expression</th>
<th>Midline neuron expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>CACAAT</td>
<td>Sox</td>
<td>+/-</td>
<td>—</td>
</tr>
<tr>
<td>G2</td>
<td>CACGT</td>
<td>Sim/Tgo</td>
<td>—</td>
<td>+</td>
</tr>
<tr>
<td>G3</td>
<td>TGTAAT</td>
<td>Sox</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>G4</td>
<td>AGGGATAT</td>
<td>POU</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>G5</td>
<td>ATTTTATC</td>
<td>homedomain</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G6</td>
<td>AGGGATAC</td>
<td>POU</td>
<td>—</td>
<td>+</td>
</tr>
<tr>
<td>G7</td>
<td>CGGAGAG</td>
<td>pointed</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G8</td>
<td>ATGGGTGG</td>
<td>POU</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

See Figure 6 for mutations created to test the importance of each site.
*Certain positions within these three motifs vary in certain Drosophila species (see Figure 1), although the sequences tested are conserved in all species. The results obtained from mutagenesis of each motif in the wrapper G fragment are indicated.*

---

**Figure 6.**—Site-directed mutagenesis of eight motifs within the wrapper genomic sequences conserved in 12 Drosophila species. The 70-bp conserved region is shown with the various mutations (G1–8) tested indicated at the top of the sequences. Sequences invariant within all 12 Drosophila species are shaded and the CME is underlined.
D, dfr, and spi signaling that impacts midline glial development and gene expression (Certel et al. 1996; Nambu and Nambu 1996; Ohshiro and Saigo 1997; Sonnenfeld et al. 1997; Ma et al. 2000; Bergmann et al. 2002) and the wrapper conserved element includes potential binding sites for these transcriptional regulators (Table 3).

Regulation by sim: Several experiments described here suggest that Sim/Tgo heterodimers may directly regulate wrapper gene expression. First, activity of the wrapper W:GFP reporter gene is severely reduced in a sim mutant background, suggesting sim is necessary for expression of this transgene and that sim regulates wrapper by activating transcription through these sequences. Second, midline activity of the wrapper reporter gene is abolished by eliminating the single CME (CACGT) present within this region. Third, wrapper reporter gene expression is expanded in sim overexpression embryos. Future biochemical studies will determine if Sim/Tgo heterodimers directly interact with the wrapper regulatory motif identified here.

Spi signaling in midline glia: The studies described here demonstrate that the wrapper reporter genes are sensitive to levels of spi signaling. Mutations in spi reduce wrapper reporter gene expression and overexpression of the secreted form of spi, together with sim expands, not only the expression domain of the endogenous wrapper gene, but the wrapper reporter genes as well. Spi binds the Epidermal Growth Factor Receptor in midline glia, leading to MAPK activation (Gabay et al. 1997a,b) and subsequent activation of the ETS transcription factor, pnt (Klämbt, 1993). Therefore, it may be Pnt that directly activates wrapper transcription through the regulatory sequences studied here. One of the identified motifs needed for transcriptional activity of wrapper is: CGGAGAG, which loosely conforms to the consensus binding site for ETS transcription factors (C/A)GGA(A/T) (A/G)(C/T) (Sharrock et al. 1997). However, further experiments are needed to determine if Pnt directly interacts with these regulatory sequences, as well as the precise mechanism whereby spi signaling regulates wrapper. Taken together with previous studies, these results suggest that the spi signaling pathway may play at least two roles in promoting survival of midline glia: (1) activating wrapper, needed for neuron–glial interactions and (2) phosphorylating, thereby inactivating head involution defective (Bergmann et al. 2002), which would otherwise cause programmed cell death in midline glia.

Figure 7.—Identification of four motifs required for midline glial expression within the wrapper genomic sequences conserved in 12 Drosophila species. (A–X) Eight different sets of mutations were made within the context of the 166-bp G fragment located upstream of the D. melanogaster wrapper gene that contains a 70-bp region highly conserved with 11 other Drosophila species. Transgenic embryos containing the wrapper G1:GFP (A–C), wrapper G2:GFP (D–F), wrapper G3:GFP (G–I), wrapper G4:GFP (J–L), wrapper G5:GFP (M–O), wrapper G6:GFP (P–R), wrapper G7:GFP (S–U), and wrapper G8:GFP (V–X) reporter genes were stained with α-GFP (green; B, E, H, K, N, T, and W) and α-sim antibody (red; C, F, I, L, O, R, U, and X) during stage 16 of embryogenesis and analyzed by confocal microscopy. The overlap in expression between the reporter gene and sim (A, D, G, J, M, P, S, and V) is shown. Ventrolateral views of embryos are shown; anterior is to the left. Mutation wrapper G7:GFP causes expression of GFP in the gut (T; arrow) and mutations wrapper G4:GFP, wrapper G6:GFP, and wrapper G8:GFP cause a low level of expression in some midline neurons (K, Q, and W; arrowheads).
Sox, POU, and homeodomain proteins in CNS transcriptional regulation: Many genes expressed in the CNS of metazoan organisms are regulated through synergistic interactions between Sox HMG-containing proteins and POU domain proteins (Ambrosutta et al. 2000; Ma et al. 2000; Tanaka et al. 2004; Bailey et al. 2006). Recently, many vertebrate genes expressed in the developing CNS have been shown to contain highly conserved noncoding DNA regions enriched for binding sites for three classes of transcription factors: Sox, POU, and homeodomain proteins (Bailey et al. 2006). Experiments indicated that Sox and POU proteins work together to activate, while homeodomain proteins repress and limit expression of CNS genes. Interestingly, several motifs identified here as important for regulation in midline glia of Drosophila resemble binding sites for Sox (G1: CACAAT; Wengner 1999), POU (G4: ATGCAAAT, G6: ATGCAACA, and G8: ATGCGTGG; Pankratova and Polanovsky 1998), and homeodomain proteins (G5: ATTTTATC; Kalionis and O’Farrell 1993).

Reporter gene expression in midline neurons: That the wrapper K:GFP gene, but not the wrapper G:GFP construct is expressed in certain midline neurons, identifies a midline neural silencer in the 43-bp region present in the G fragment, but absent in the K fragment. Within this region, 27 bp are highly conserved in all 12 Drosophila species (Figure 1B) and two of the three mutations in the G fragment that cause slight activation of reporter gene expression in midline neurons (wrapper G4:GFP and wrapper G8:GFP; Figure 7) are found within the 43-bp region. All three sites that lead to activation in midline neurons, G4, G6, and G8, conform to a POU domain binding site (Pankratova and Polanovsky 1998), suggesting a POU domain protein expressed in midline neurons may bind to one or more of these sites to keep the wrapper gene silent.

One POU domain protein, Dfr, binds to the sequence ATGCAAAT in other gene regulatory regions to activate transcription, including those of two genes expressed in midline glia: dfr itself and slit (Ceret et al. 1996; Ma et al. 2000). This sequence is found at site G8 in the wrapper regulatory region, but when changed to ATGC TAGC, caused a low level of activation in midline neurons (Figure 7X), rather than reducing expression in midline glia. Although the number of midline glia is reduced in a dfr mutant background, those that remain express high levels of reporter gene expression driven by wrapper sequences and the results suggest dfr is not absolutely required for wrapper reporter gene expression in midline glia.

Mutations in the POU domain motifs within the wrapper regulatory sequences suggest a notable difference between the CNS genes studied previously in vertebrates (Bailey et al. 2006) and the midline glial gene studied here. The POU domain binding sites appear to limit expression in midline neurons (rather than activate expression as in vertebrate CNS genes), and it is the Sox and homeodomain binding sites that are needed for activation. This may reflect a key difference in regulatory control of glial vs. neural genes and it is plausible that other midline glial genes excluded from midline neurons will contain silencer elements similar to the one identified here, but further experiments are needed to confirm this.

Reporter gene expression in other tissues: Some of the wrapper reporter gene constructs are expressed in other tissues during embryogenesis. In addition to midline glia, the wrapper W:GFP construct is expressed in a few cells within the lateral CNS (Figure 3, A–I) and a subset of somatic muscle cells (not shown). The wrapper G:GFP construct is also expressed in some somatic muscles and in the salivary glands (not shown), the wrapper G7:GFP construct is expressed in cells of the gut (Figure 7T), and the smaller wrapper K:GFP construct is expressed in certain CNS midline neurons (including progeny of the median neuroblast; Figure 3, V–X), unlike the endogenous wrapper gene.

These results indicate genes expressed within midline glia must share motifs closely related to those found within genes expressed in tissues such as (1) the lateral CNS glia, (2) midline neurons, (3) trachea, (4) muscles, (5) salivary glands, and (6) gut, and slight changes in these sequences can switch expression from midline glia to one or more of these tissues.

Summary: These results demonstrate that certain Drosophila CNS genes contain short, ~30–80-bp highly conserved genomic signatures indicative of regulatory function. Within the conserved regulatory region of wrapper, a combination of a minimum of four sites (CAGCG, GCGGAGA, CACAA, and the T-rich motif) is required for transcriptional activation in midline glia. In addition, a neuron silencer is required to repress expression of a midline glial gene in midline neurons. Finally, these experiments also highlight dfr/tgo and the spi signaling pathway as key components in the regulation of wrapper.

Future experiments are needed to determine (1) if Sim and/or Pnt directly bind the sequences identified here, (2) which proteins expressed in midline neurons repress wrapper through the midline neuron silencer identified here, and (3) if other genes expressed in midline glia also contain the conserved motifs identified here. These motifs and other conserved sequence signatures should be valuable for studying both (1) conservation in regulatory regions to identify transcription factor binding sites and/or possible structural components of regulatory DNA and (2) any variations within these otherwise conserved blocks in divergent species. Relatively small changes within these regions could lead to a change in the spatial or temporal expression pattern of a gene that may ultimately lead to novel functions within various fly species.
We thank John Nambu for the Dac3 mutant flies and Ben-Zion Shilo for the UAS-specta flies. We are grateful to Michael Purugganan, Trudy Mackay, Greg Gibson, and other members of the North Carolina State University Genetics Department for help with sequencing and to Trudy Mackay and Faye Lawrence for fly food. We thank Tony Perdue for help with confocal microscopy, Laura Mathies for the use of her microscopes, Amanda Welty, Anna Edens, and Kelly Daigle for help with the injections of fly embryos, and Trudy Mackay and Laura Mathies for comments on the manuscript. The dfl and spi mutant lines were obtained from the Bloomington Stock Center and the wrapper antibody was obtained from the Iowa Hybridoma Center. This work was supported by National Science Foundation grant no. 0416812 to P.E.

**LITERATURE CITED**


Communicating editor: K. Golic