EMBRYONIC REGULATION AND POST-EMBRYONIC FUNCTION OF THE *SINGLE-MINDED* GENE IN THE *DROSOPHILA* CENTRAL NERVOUS SYSTEM

Daniel Christian Lau

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in the Department of Biology

Chapel Hill
2009

Approved by:

Stephen T. Crews, PhD
Mark A. Peifer, PhD
Robert J. Duronio, PhD
Jay E. Brenman, PhD
Robert P. Goldstein, PhD

ABSTRACT

Daniel Christian Lau: Embryonic Regulation and Post-embryonic Function
of the *single-minded* Gene in the *Drosophila* Central Nervous System
(Under the direction of Stephen T. Crews, PhD)

The *single-minded* (*sim*) gene in *Drosophila melanogaster* has long been known to play important roles in specifying the mesectodermal cell fate in the embryonic central nervous system (CNS). Mesectoderm cells differentiate into CNS midline cells by mid-embryogenesis. CNS midline cells contribute both neurons and glia to the developing nervous system, and are a source of both attractive and repulsive axonal guidance cues that combinatorially pattern the bilateral CNS. Removal of *sim* function leads to the failure of midline cell formation, and concomitantly a lack of instructive signal presentation to pathfinding axons from the lateral CNS. As a result, commissural axon tracts that cross the midline are largely absent and parallel longitudinal axon tracts that flank the midline appear fused as a single connective at the embryonic mid-plane. Due to this patterning defect, *sim* mutants are late embryonic lethal. In addition to the CNS midline, Sim can also be found in the developing foregut, posterior terminal structures, and a subset of myoblasts, although its roles in these compartments are less well understood. In collaboration with others, we demonstrated functions for *sim* in developing posterior terminal structures and gonads, patterning the larval cuticle, organizing the adult brain, and in adult behavior and locomotion. Using the MARCM strategy for positively marking *sim* mutant cell clones, we demonstrated that in contrast to its role in neurogenesis in the CNS mesectoderm, *sim* functions to pattern axon fascicles in the larval central brain, a region known

to be important in the interhemispheric communication and the coordination of leg movement. Using RT-PCR, we showed that the *sim* locus yields a third, previously unidentified transcript that is the primary isoform used post-embryonically. Genetic dissection of inter- and intragenic regions from the *sim* locus revealed locations of enhancers that drive expression in the CNS midline, myoblasts, and foregut. Taken together, these results have broadened our understanding of *sim*, an important regulator of development with complex regulation.

To my parents,

without whose love and support this would not be possible

ACKNOWLEDGEMENTS

In different regards, the persons listed here have opened my eyes to the beauty of life:

I thank Victoria Bautch for believing in me at the commencement of this journey, and Rebecca Rapoport who taught me the most elemental skill for molecular biology – how to use a pipettor – in gross detail.

I owe gratitude to Stephen Crews, my research advisor, who exercised patience and support throughout my progress in graduate school, and to my doctorial committee panel of Mark Peifer, Robert Duronio, Jay Brenman, and Robert Goldstein who directed me well. Together, you compose a rigorous and outstanding convocation of top minds and I enjoyed our many meetings.

Thanks also go to previous and current members of the Crews lab: in particular, Patricia Estes, Brian Mitchell, Jack Mosher, and Lan Jiang, who gave me great advice about how to get started in lab; and Joe Kearney, Scott Wheeler, and Amaris Guardiola who mentored me to completion. May your respective scientific careers be as exemplary as they were during my association with you in the Crews lab. Thank you for your friendship.

Thanks go to our German collaborators in the Klämbt lab whose initial work on characterizing novel *single-minded* mutant phenotypes became my point of entry into the work presented in this dissertation.

Thanks to all my friends for keeping me sane, and for understanding that distance and time do little to dilute our bonds. You know who you are.

Thanks to Copper, who showed me that perpetual happiness is indeed possible. You are dearly missed.

And finally, highest thanks go to my parents, Anthony and Jacqueline, and my siblings, Terence and Shirley, who together make up the best family without knowing it.

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xiii

LIST OF ABBREVIATIONS xv

Chapter

I. INTRODUCTION 1

References 6

II. NOVEL BEHAVIORAL AND DEVELOPMENTAL DEFECTS ASSOCIATED WITH *DROSOPHILA SINGLE-MINDED* 8

Preface 8

Abstract 9

Introduction 9

Materials and Methods 12

Genetics 12

Behavioral analyses 13

Molecular analyses 14

Sequence analysis of sim^{J147} mutant *sim* gene 15

Immunohistochemistry 15

Histology 16

Results 16

Identification of a temperature-sensitive *sim* allele 16

The *sim*^{1147} mutant has a mutation in the Sim dimerization domain 17

Requirement of *single-minded* during the development of the larval cuticle 18

sim is required for the development of the genital discs 19

Adult phenotypes associated with *sim* 20

sim affects adult behavior 21

Histological analyses of *sim*^{ts} animals 23

Postembryonic expression of *sim* in the CNS 24

Discussion 26

Embryonic function of *sim* 27

sim function during adult stages 28

Acknowledgements 31

Figures 32

References 53

III. THE ROLE OF *DROSOPHILA SINGLE-MINDED* IN CONTROLLING BRAIN INTERHEMISPHERIC CONNECTIVITY 59

Preface 59

Abstract 59

Introduction 60

Materials and Methods 63

Drosophila stocks 63

Immunostaining 64

MARCM 64

Sequencing of *sim* mutant DNA 65

Results 65

Interhemispheric crossing and fasciculation of Sim$^+$ central brain cell axons....65

Developmental progression of Sim$^+$ medial brain expression....67

sim is expressed in brain neurons but not neuroblasts....68

sim mutations affect axon morphology but not neurogenesis....68

Discussion....71

Acknowledgements....76

Figures....77

Tables....87

References....89

IV. ANALYSIS OF TRANSCRIPTIONAL REGULATION AT THE *SINGLE-MINDED* LOCUS OF *DROSOPHILA MELANOGASTER*....93

Preface....93

Abstract....93

Introduction....94

Materials and Methods....98

Fly stocks....98

Molecular cloning....98

Immunohistochemistry....99

RT-PCR....99

Sequence analysis....100

Results....100

Sequence analysis reveals the absence of recognizable TATA and DPE motifs....100

Detection of a third *sim* isoform that is expressed post-embryonically....101

Discrete enhancers lie within the intergenic and intronic regions of the *sim* locus 103

Discussion 105

Acknowledgements 109

Figures 110

Tables 120

References 122

V. DISCUSSION 125

Summary of results 125

Weaknesses and their solutions 128

What have we learned? 130

Future directions 130

LIST OF TABLES

CHAPTER 3

1. Molecular alterations of *sim* alleles 87

CHAPTER 4

1. Summary of driver (GAL4) lines created and their expression domains 120

LIST OF FIGURES

CHAPTER 2

1. Identification of a temperature-sensitive *sim* allele....32
2. Alignment of Sim proteins reveals that *sim*$^{J1\text{-}47}$ has a mutation in a conserved residue in helix 2....34
3. Larval phenotypes associated with *sim*....36
4. Formation of the anal pad....38
5. Formation of the genital disc....40
6. Adult phenotypes associated with *sim*....42
7. Walking and orientation behavior phenotype....44
8. Structural defects in the adult *sim* brain....46
9. Postembryonic expression of *sim*....48
10. Sim protein expression in the larval CNS....50

CHAPTER 3

1. Interhemispheric crossing and fasciculation of Sim^{+} secondary axon tracts....77
2. Developmental expression of *sim* in the central brain....79
3. *sim* is not required for the differentiation of post-mitotic central brain neurons....81
4. Structure of Sim wild type and mutants proteins....83
5. *sim* mutant clones show axon fasciculation defects....85

CHAPTER 4

1. RT-PCR results show a third *sim* isoform expressed during embryogenesis....110
2. The 1.0 kb fragment drives expression in most Sim^{+} cells of the embryonic midline....112

3. The 2.8 kb fragment drives expression
in a subset of midline glia and in muscle precursors (myoblasts).....114

4. The 3.1 kb fragment drives expression
in the developing foregut and in cells caudally.....116

5. The 5.5 kb fragment drives expression
in a subset of pleural muscles and CNS midline glia.....118

LIST OF ABBREVIATIONS

A	adults
aa	amino acid
AEL	after egg laying
aPV	anterior periventricular nucleus
β-gal	*beta-galactosidase* (gene)
bHLH	basic-helix-loop-helix protein domain
bp	base pair(s)
C	chicken
cd m^{-2}	candela(s) per square meter
cDNA	complementary DNA
CNS	central nervous system
CX	central complex
D	*dichaete* (gene)
D	dorsal cluster
DAPI	4',6-diamidino-2-phenylindole
DDB1	*damaged DNA binding protein 1* (gene)
dfr	*drifter* (gene)
dNTP	deoxynucleotriphosphate
DPE	downstream promoter element
Dv	*Drosophila virilis*
E	embryos
E	esophagus

ed	ejaculatory duct
ELAV	*embryonic lethal, abnormal vision* (gene)
EMS	ethyl methane sulfonate
esg	*escargot* (gene)
eve	*even-skipped* (gene)
fb	fan-shaped body
GFP	*green fluorescent protein* (gene)
GMC	ganglion mother cell
h	hour(s)
H	human
hg	hindgut
HRP	horseradish peroxidase
IPC	inner proliferative center
L	third instar larvae
la	lamina
LB	third instar larval brains
lo	lobula
LPC	laminar proliferative center
M	medial cluster
M	murine
MARCM	mosaic analysis with a repressible cell marker
me	medulla
min	minute(s)
mm	millimeter(s)

MTEG	mammillotegmental tract
MTT	mammillothalamic tract
n	number of data points in a sample
NB	neuroblast
nls	nuclear localization signal
od	oviduct
OPC	outer proliferative center
P	pupae
PAS	period-Arnt-sim protein domain
pb	protocerebral bridge
PBT	phosphate buffered saline solution plus Tween-20 detergent
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
P_E	early promoter
pg	paragonium
phe	phenylalanine (amino acid)
P_L	late promoter
PMT	principal mammillary tract
PVN	paraventricular nucleus
r	rectum
repo	*reverse polarity* (gene)
robo	*roundabout* (gene)
rs	seminal receptacle
RT-PCR	reverse transcription polymerase chain reaction

s	second(s)
SAT	secondary axon tract
SC	supraesophageal commissure
SD	standard deviation
ser	serine (amino acid)
sim	*single-minded* (gene)
SON	supraoptic nucleus
sp	sperm pump
st	spermatheca
T_a	annealing temperature
tdPCR	touchdown PCR
tgo	*tango* (gene)
ts	temperature sensitive
tub	*tubulin* (gene)
u	uterus
UAS	upstream activation sequence
UNC-CH	The University of North Carolina at Chapel Hill
UTR	untranslated region
V	ventral cluster
vd	vas deferens
VNC	ventral nerve cord
X	*Xenopus*
Z	zebrafish

"All the efforts of the human mind cannot exhaust the essence of a single fly."

St. Thomas Aquinas

CHAPTER 1:

INTRODUCTION

One hundred years ago, in a culture bottle on a shelf in the "fly room" of Schermerhorn Hall at Columbia University, Thomas Hunt Morgan noticed a single male fly with white eyes rather than the customary red. When mated to a wild-type sister, all the progeny had red eyes. When Morgan performed a second round of brother-sister matings, he observed that some males of this second generation had white eyes. His persistence in studying what he termed "sex-limited" characters (today called sex-linked) led to the widespread adoption of the term "gene", and the finding that genes occupied specific, precise, and linear positions on chromosomes which could be mapped. Thus grew the importance of using *Drosophila melanogaster* as a model for understanding the principles of genetics as we know them today.

Step back to the mid-1880s, and on a different continent. Spain, France, and Italy were battling a cholera epidemic. A medical officer in the Spanish army was given a gift of a Zeiss microscope by the Zaragoza provincial government in return for his volunteer efforts in combating the outbreak. Using this, and a method of staining brain tissue with silver chromate solution developed by Camillo Golgi, Santiago Ramón y Cajal was able to resolve the fine structure of neurons and conclude that the nervous system was composed of autonomous cells as opposed to a continuous web. Ramón y Cajal formulated the neuron doctrine, which states that neurons are metabolically distinct cells which make up the basic structural and functional

units of the nervous system. The neuron doctrine has become a central tenet of modern neuroscience.

In 1980, while trying to ascertain whether *rosy* was the only gene in the chromosomal micro-region of the right arm of the *Drosophila* third chromosome that encoded xanthine dehydrogenase activity, Hilliker et al. identified the l(3)S8 complementation group (Hilliker et al., 1980). Thomas et al. later discovered that the l(3)S8 deletion harbored a gene that, when mutated, led to fusion of longitudinal axon tracts that normally flank the midline of the embryonic central nervous system (CNS) (Thomas et al., 1988). Following the convention of naming *Drosophila* genes based on their mutant phenotypes, *single-minded (sim)* was adopted as an appropriate descriptor. Further analysis revealed that mutant embryos were missing the cells that occupy positions along the midline of the CNS, and that the *sim* gene was expressed in a number of non-neural tissues. Antibody studies localized Sim protein to the nucleus of cells in which the gene is expressed (Crews et al., 1988); however, the sequence of *sim* shared no homology with any known transcription factors at the time. Not long thereafter, *sim* was found to regulate the expression of genes important for the normal development of CNS midline cells (Nambu et al., 1990). This, coupled with the identification of a basic-helix-loop-helix (bHLH) domain at the amino terminus of the Sim protein (Nambu et al., 1991), cemented its role as a transcription factor that associates in a sequence specific manner with DNA through its basic region and interacts with co-factors through the newly identified and evolutionarily conserved PAS domain. The PAS domain is so-called because its existence was deduced from comparison of the *Drosophila* Sim and Period proteins, and vertebrate Arnt (Aryl hydrocarbon receptor nuclear translocator) (Nambu et al., 1991).

The Sim DNA recognition site was identified (Wharton et al., 1994). Subsequently, the *Drosophila* homolog of *arnt*, *tango*, was identified (Sonnenfeld et al., 1997) and shown to regulate Sim's entry into the nucleus (Ward et al., 1998).

From dorsal to ventral, the monolayer of the embryonic cellular blastoderm is fated to become extraembryonic amnioserosa, dorsal ectoderm, neuroectoderm, mesectoderm, and mesoderm (endoderm gastrulates separately as invaginations near both termini of the embryo). The *sim* gene is initially expressed in the 8 mesectodermal cells per segment beginning shortly before gastrulation. This highly refined expression pattern is the result of the combinatorial action of genetic activation and repression. Activation is performed by the transcription factors Dorsal and Twist along with the Daughterless:Scute heterodimer (Crews, 1998; Jacobs, 2000). The Snail zinc finger protein, localized in the mesoderm, represses *sim* in that tissue, thereby establishing the ventral boundary of *sim* expression (Crews, 1998; Kasai et al., 1992). The Suppressor of Hairless transcription factor represses *sim* in the neuroectoderm (Morel and Schweisguth, 2000), forming the dorsal boundary of *sim* expression.

Drosophila Sim protein contains several major regions that confer specific properties, presented here from amino-terminus to carboxy-terminus: the bHLH DNA binding domain, 2 PAS domains that enable co-factor binding which thereby confers transcriptional specificity, 10 alanine-alanine-glutamine repeats of unelucidated function (Estes et al., 2001), and several transcriptional activation regions (Franks and Crews, 1994).

In addition to the well-characterized role of *sim* as a CNS midline selector gene, there are other non-midline roles. A subset of somatic muscle precursor cells transiently express *sim* prior to their migration dorsally away from their site of birth close to the ventral CNS (Lewis and Crews, 1994). However, genetic abrogation of *sim* expression specifically in these cells revealed no gross abnormalities in the morphology or final position of these muscle cells. Therefore, *sim*

does not seem to play a major role in the development of somatic muscles, in contrast to its keystone role in the CNS midline.

The *sim* gene is expressed in foregut cells adjacent to the brain, which is undergoing morphogenetic and proliferative processes during mid-embryonic development (Page, 2003). Removal of *sim* function in these endodermal cells results in a lack of Egfr signaling from the foregut cells to the midbrain neuroblasts, which results in retarded cell proliferation rates and, concomitantly, reduced size of the brain lateral to the foregut. Thus, *sim* plays a role in patterning the *Drosophila* midbrain.

Further, in this study I present collaborative research that demonstrates a role for *sim* in patterning the anus, patterning male and female genital structures, and a possible role in sterility (see Chapter 2).

All studies to date have focused on determining embryonic roles for *sim*. Clearly, *sim* plays disparate roles in the different tissues that express the gene. How, then, can we broaden our understanding of this developmentally important selector gene? In this dissertation, I demonstrate that *sim* is expressed in the larval brain and nerve cord, suggesting possible roles for *sim* post-embryonically. In collaboration with the Klämbt group at the University of Münster, Germany, *sim* mutant larvae and adults were analyzed in an effort to understand these newly discovered functions. Mutant embryos displayed abnormal genital disc development; mutant larvae displayed abnormal cuticle development; mutant adult flies exhibited morphological and behavioral defects. These issues are presented and discussed in Chapter 2.

I sought to better understand the role of *sim* specifically in the larval brain. The formal possibility exists that the adult behavioral (locomotor) defects described in Chapter 2 are the result of removing *sim* function in the larval brain, since the neurons that express *sim* are located in a brain compartment thought to be important in the coordination of interhemispheric

connectivity. In Chapter 3, I present data that support an axonogenesis function for *sim*. These data suggest a causal relationship between the proper development of Sim-positive central brain cells in the larva and the proper coordination of locomotion in the adult.

Finally, in an effort to better understand genetic regulation at the *sim* locus, I show that there exists a previously unknown third mRNA isoform, and that this isoform is the major species that is expressed by the locus throughout post-embryonic life. Because the expression of *sim* is dependent on enhancers that turn the gene on in discrete tissues and at discrete times during development, I identify the rough locations of several of these enhancers by dissecting the regulatory regions of the locus. These data are presented in Chapter 4.

REFERENCES

Crews, S. T. (1998). Control of Cell Lineage-Specific Development and Transcription by bHLH-PAS Proteins. *Genes Dev.* **12,** 607-620.

Crews, S. T., Thomas, J. B. and Goodman, C. S. (1988). The *Drosophila Single-Minded* Gene Encodes a Nuclear Protein with Sequence Similarity to the *Per* Gene Product. *Cell* **52,** 143-151.

Estes, P., Mosher, J. and Crews, S. T. (2001). *Drosophila Single-Minded* Represses Gene Transcription by Activating the Expression of Repressive Factors. *Dev. Biol.* **232,** 157-175.

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.

Hilliker, A. J., Clark, S. H., Chovnick, A. and Gelbart, W. M. (1980). Cytogenetic Analysis of the Chromosomal Region Immediately Adjacent to the Rosy Locus in *Drosophila Melanogaster*. *Genetics* **95,** 95-110.

Jacobs, J. R. (2000). The Midline Glia of *Drosophila*: A Molecular Genetic Model for the Developmental Functions of Glia. *Prog. Neurobiol.* **62,** 475-508.

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-Minded* Gene. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 3414-3418.

Lewis, J. O. and Crews, S. T. (1994). Genetic Analysis of the *Drosophila Single-Minded* Gene Reveals a Central Nervous System Influence on Muscle Development. *Mech. Dev.* **48,** 81-91.

Morel, V. and Schweisguth, F. (2000). Repression by *Suppressor of Hairless* and Activation by *Notch* are Required to Define a Single Row of *Single-Minded* Expressing Cells in the *Drosophila* embryo. *Genes Dev.* **14,** 377-88.

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, K. A.,Jr 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.

Page, D. T. (2003). A Function for Egf Receptor Signaling in Expanding the Developing Brain in *Drosophila*. *Curr. Biol.* **13,** 474-482.

Sonnenfeld, M., Ward, M., Nystrom, G., Mosher, J., Stahl, S. and Crews, S. (1997). The *Drosophila Tango* Gene Encodes a bHLH-PAS Protein that is Orthologous to Mammalian *Arnt* and Controls CNS Midline and Tracheal Development. *Development* **124,** 4583-4594.

Thomas, J. B., Crews, S. T. and Goodman, C. S. (1988). Molecular Genetics of the *Single-Minded* Locus: A Gene Involved in the Development of the *Drosophila* nervous System. *Cell* **52,** 133-141.

Ward, M. P., Mosher, J. T. and Crews, S. T. (1998). Regulation of *Drosophila* bHLH-PAS Protein Cellular Localization during Embryogenesis. *Development* **125,** 1599-1608.

Wharton, K. A.,Jr, Franks, R. G., Kasai, Y. and Crews, S. T. (1994). Control of CNS Midline Transcription by Asymmetric E-Box-Like Elements: Similarity to Xenobiotic Responsive Regulation. *Development* **120,** 3563-3569.

CHAPTER 2:

NOVEL BEHAVIORAL AND DEVELOPMENTAL DEFECTS ASSOCIATED WITH *DROSOPHILA SINGLE-MINDED*

PREFACE

The contents of this chapter were published on pages 283-299 in volume 249, issue 2 of Developmental Biology on September 15, 2002. Reproduction permission was granted via a limited license to Daniel C. Lau by Elsevier, Ltd. on September 18, 2008.

The original authors of this chapter are Jan Pielage, Georg Steffes, Daniel C. Lau, Beth A. Parente, Stephen T. Crews, Roland Strauss, and Christian Klämbt. The Klämbt group identified the $sim^{J1\text{-}47}$ allele and performed histological analysis of the embryo, larva, and adult. Roland Strauss analyzed and quantified the adult behavioral phenotypes. Beth Parente, a Research Technician in the Crews group, sequenced the portions of the $sim^{J1\text{-}47}$ locus encoding the bHLH domain. The author of this dissertation performed RT-PCR analysis of *sim* expression and immunohistochemical analysis of the larval brain.

ABSTRACT

In *Drosophila*, the development of the midline cells of the embryonic ventral nerve cord depends on the function of the bHLH-PAS transcription factor Single-minded (Sim). The expression domain of *sim*, however, is also found anterior and posterior to the developing ventral cord throughout the germ band. Indeed, mutations in *sim* were identified based on their characteristic cuticle phenotype. Eight abdominal segments (A1-A8) can be easily seen in the larval cuticle, while three more can be identified during embryogenesis. Cells located in A8-A10 give rise to the formation of the genital imaginal discs, and a highly modified A11 segment gives rise to the anal pads that flank the anus. *sim* is expressed in all these segments and is required for the formation of both the anal pads and the genital imaginal discs. A new temperature-sensitive *sim* allele allowed an assessment of possible postembryonic function(s) of *sim*. Reduction of *sim* function below a 50% threshold leads to sterile flies with marked behavioral deficits. Most mutant *sim* flies were only able to walk in circles. Further analyses indicated that this phenotype is likely due to defects in the brain central complex. This brain region, which has previously been implicated in the control of walking behavior, expresses high levels of nuclear Sim protein in three clusters of neurons in each central brain hemisphere. Additional Sim localization in the medullary and laminar neurons of the optic lobes may correlate with the presence of ectopic axon bundles observed in the optic lobes of *sim* mutant flies.

INTRODUCTION

First signs of *Drosophila* nervous system development are evident only a few hours after fertilization during the cellular blastoderm stage. At this time point, maternal gene functions

have divided the embryo into the three germ layers (Campos-Ortega and Hartenstein, 1997). The mesodermal anlage is separated from the lateral neurogenic region by a very special set of cells, which have been recognized as an important part of the ventral nerve cord some 100 years ago (Escherich, 1902). Based on morphological criteria, these cells were later called mesectodermal cells (Poulson, 1950). These are the first neural cells to be specified (Crews et al., 1988) and are initially arranged in a single cell-wide row with about four cells per hemisegment.

Following gastrulation, when the mesoderm has invaginated into the interior of the embryo, the mesectodermal cells intermingle at the ventral midline and move into the interior of the embryo. Here, they form a mitotic domain and generate a small number of neuronal and glial cells located at the midline of the developing ventral cord (Bossing and Technau, 1994; Foe, 1989; Klämbt et al., 1991; Thomas et al., 1988). A number of studies have shown that the midline cells are distinct from the remaining neural cells in a number of ways (Crews, 1998; Jacobs, 2000).

The CNS midline cells exert many prominent functions during CNS development. The loss of all CNS midline cells, which are the source of attractive (Netrins) and repulsive (Slit) axonal guidance cues, leads to a dramatic axonal patterning phenotype (Brose and Tessier-Lavigne, 2000; Kidd et al., 1999; Mitchell et al., 1996; Rothberg et al., 1988; Rothberg et al., 1990; Thomas et al., 1988). In addition, the CNS midline cells regulate directed cell migration toward and away from the midline (Bashaw and Goodman, 1999; Kidd et al., 1999; Kinrade et al., 2001; Kramer et al., 2001). Besides directing the migration of growth cones and cells, inductive signals emanating from the CNS midline regulate the development of cortical neurons and certain mesodermal cells (Chang et al., 2000; Luer et al., 1997; Menne et al., 1997; Zhou et al., 1997). These findings underpin the role of the CNS midline as an important organizing center during normal embryonic development.

The special appearance and the strategic position of CNS midline cells are reflected by the fact that their cell fate is determined very early by the action of neurogenic genes (Menne and Klämbt, 1994; Morel and Schweisguth, 2000). The activation of *Notch* results in the expression of the gene *sim*, which subsequently serves as a master regulatory gene of CNS midline development (Crews et al., 1988; Muralidhar et al., 1993; Nambu et al., 1990; Nambu et al., 1991; Thomas et al., 1988). Loss of *sim* function results in a loss of all CNS midline cells, whereas ectopic expression of *sim* within the nervous system is able to induce the midline differentiation program (Nambu et al., 1991; Sonnenfeld and Jacobs, 1994). Depending on the segmental position, *sim* is able to specify glial as well as neuronal midline cell types (Menne et al., 1997).

Two promoters direct the expression of two alternative *sim* transcripts. *sim* expression starts at the onset of gastrulation in the mesectodermal cells flanking the presumptive mesoderm. Initially, *sim* is expressed by all midline cells; however, during midembryogenesis, expression becomes restricted to the midline glia. In addition, a complex pattern of *sim* expression has been described in the embryonic brain (Therianos et al., 1995). Outside the nervous system, *sim* expression has been reported in a subset of ventral muscle precursor cells (Lewis and Crews, 1994). *sim* expression extends beyond the developing ventral nerve cord to the abdominal-most segmental units. The function of these *sim*-expressing cells is unknown.

sim encodes a basic–helix-loop-helix–PAS (bHLH-PAS) protein that, when binding to an appropriate interaction partner, directly activates transcription. Sim can also repress gene expression in the midline cells by activating the transcription of repressive factors (Estes et al., 2001). To date, two direct interaction partners have been described: Dichaete (Fish-hook), which associates with the PAS domain of Sim, and the bHLH-PAS protein encoded by *tango* (*tgo*) (Ma et al., 2000; Ohshiro and Saigo, 1997; Sanchez-Soriano and Russell, 1998; Sonnenfeld et al.,

1997). Mutant *tgo* embryos display only relatively mild defects during embryonic CNS development (Ohshiro and Saigo, 1997; Sonnenfeld et al., 1997). Unlike Sim, however, Tgo is deposited maternally in the egg, and this strong maternal contribution is likely to compensate for the early requirement of *tgo*. No germline clones have yet been described. Clonal analyses demonstrate that Tgo is required for adult antennal and tarsal development. Here, it does not interact with Sim but with the bHLH-PAS protein Spineless (Emmons et al., 1999). To date, no function of *sim* has been described during postembryonic development.

Here, we report the identification of a temperature sensitive *sim* mutation. Our data show that *sim* expression within the developing brain is important to correctly specify the formation of the central complex, a part of the brain required to control the walking behavior of the fly. In addition, we show that *sim* is required outside the CNS to correctly pattern the genital discs as well as the anal pad anlage.

MATERIALS AND METHODS

Genetics

Among the mutations identified in a large-scale EMS mutagenesis (Hummel et al., 1999), we identified the sim^{ts} mutation. The mutation was subsequently separated from other lethal hits found on this particular chromosome by recombination against *rucuca* chromosomes. The amorphic sim^{H9} allele and the enhancer trap line P[*lacW*]$esc^{B7\text{-}2\text{-}22}$ were obtained from the Bloomington Stock Center.

Behavioral analyses

Walking was quantified in Buridan's paradigm (Gotz, 1980). Wings were clipped under cold anesthesia (4°C) in order to ensure that flies will only walk. Flies were then given at least 4 h to recover in single-fly containers with access to water before they were placed on an elevated circular disc (diameter 85 mm) between two opposing and inaccessible landmarks in an otherwise uniform bright white surrounding (3000 cd m^{-2}). The dark vertical landmarks were 100 mm away from the center of the disc and appeared under viewing angles of 11° horizontally and 58° vertically. The disc was surrounded by a water-filled moat. The walking-trace of each fly was recorded for 15 min by using a video-based computerized tracking system (time resolution 5 frames s^{-1}). The traces were evaluated off-line with regard to walking speed, activity, walked distance, and orientation toward the landmarks as described (Strauss and Pichler, 1998). *Walking speed* is calculated for every transition of the fly from one landmark to the other. Start and end of a transition are defined by the crossing of parallel lines which are perpendicular to the connecting axis of the two landmarks and which intersect with it at +33 mm and -33 mm as seen from the center of the disk. The mean speed of all transitions within 15 min is called the walking speed of the individual fly. *Walking activity* is defined as the fraction of time spent walking instead of resting or grooming. *Walked distance* is the total length of the piecewise linear interpolation of the fly's track given by successive positions sampled every 0.2 s. *Orientation*: For each path increment, also the angular deviation from the direct path toward each of the two landmarks is calculated. The smaller of the two angular values is always integrated into a frequency histogram. At the given sampling rate, each fly contributes 4500 orientation values in a 15-min measurement, that were integrated in a frequency histogram of 5° bin width. For a direct

comparison, the same four evaluations were also applied to random search measurements, where the flies saw no landmarks.

Leg coordination of *sim* flies was inspected on a walking analyzer as described (Strauss, 1998). Briefly, the fly walks on a glass plate which is overlaid with a layer of red laser light invisible to the fly. The light carpet is so thin that only legs are illuminated which are either in contact to the ground or near touch-down. The points of ground contact are registered by cameras underneath the glass plate and the temporal and spatial aspects of stepping are analyzed off-line on a PC.

Molecular analyses

Total RNA was extracted from different stages of *y* w^{67} by using QIAshredder and RNeasy kits (Qiagen). RNA was treated with RNase-free DNase I (RQ1; Promega) to remove contaminating genomic DNA. Stages and dissected tissues were: (1) 0-18 h (AEL) embryos, (2) wandering third instar larvae, (3) wandering third instar larval brains, (4) pupae, and (5) adults (male and female combined). Synthesis of cDNA was performed by using SuperScript reverse transcriptase and the primer, 5'-CTGGTTGATGTGCGGATG-3', which corresponds to the 3' end of *sim* exon 4. PCR was carried out by using the primer pair, 5'-GCCTGGGGCTCATCGCCT-3' (5' end of exon 3) and 5'-CAGCGACAAAATGGCATTC-3' (region of exon 4 just 5' to the primer used for cDNA synthesis). The primer pair was derived from two exons, so that amplification of contaminating genomic DNA could be distinguished from the amplified products derived from RT-PCR of RNA. Two controls were included. Genomic DNA was PCR amplified by using the primers described above to yield a 558-bp fragment. Presence of this band in RT-PCR-amplified

Drosophila RNA samples would indicate the presence of contaminating genomic DNA. Positive control sample involved RT-PCR amplification of RNA synthesized from a full-length *sim* cDNA clone transcribed *in vitro* with SP6 RNA Polymerase. This yielded a DNA fragment of 237 bp and corresponds to the RT-PCR-amplified products derived from *Drosophila* RNA. PCR products were subjected to agarose gel electrophoresis, stained with ethidium bromide, and visualized for fluorescence.

Sequence analysis of sim^{J1-47} *mutant* sim *gene*

The complete coding sequence of the *sim* gene from *sim*$^{J1-47}$ homozygous mutant adult flies was determined by using PCR amplification of isolated genomic DNA sequence followed by direct sequencing of the PCR products. Multiple primer pairs (details provided upon request) and *Taq* polymerase were used to PCR amplify DNA containing exons 2-8, which contain all of the *sim* coding sequence. Fragments were gel purified and sequenced by using an ABI automated sequencer. Each fragment was independently amplified multiple times, and both strands were sequenced.

Immunohistochemistry

Embryos were collected and stained as described previously (Hummel et al., 1997). Wandering third instar larvae were dissected in PBT and fixed in 4% formaldehyde on ice for 1 h. Antibody staining was performed as described (Patel et al., 1987). Antibody dilutions were used as follows: rat anti-Sim, 1:100; mAb anti-Tgo, 1:1; mAb anti-Eve, 1:5; mAb BP102, 1:50; mAb anti-ELAV (9F8A9; from Developmental Studies Hybridoma Bank), 1:100; and rabbit

anti-β-galactosidase (Cappel), 1:2000. Larval brains were dissected in PBS after antibody staining, mounted in Aqua-Poly/Mount (Polysciences, Inc.), and visualized on a Zeiss LSM510 confocal microscope. Images were processed by using the Zeiss LSM Browser and Adobe Photoshop.

Histology

Serial 7-μm-thick paraffin sections of adult heads were prepared in frontal orientation by using the collar method (Ashburner, 1989; Heisenberg and Bohl, 1979). The brains were inspected under a fluorescence microscope.

RESULTS

Identification of a temperature-sensitive sim *allele*

Mutations in the gene *sim* result in the loss of all CNS midline cells. Subsequently, all CNS axon tracts collapse at the CNS midline (Crews et al., 1988; Klämbt et al., 1991; Sonnenfeld and Jacobs, 1994; Thomas et al., 1988) (Fig. 1A and 1B). All *sim* mutations isolated to date lead to this typical CNS phenotype (for review, see Jacobs, 2000). We fortuitously identified a weak *sim* allele in the mutant collection established recently in the lab (Hummel et al., 1999). The $sim^{J1\text{-}47}$ mutation was subsequently isolated by standard recombination techniques. In order to avoid background effects, we generally analyzed *ru h th st cu* $sim^{J1\text{-}47}$ *e/st* $sim^{J1\text{-}47}$ *e ca* transheterozygous embryos (hereafter referred to as homozygous $sim^{J1\text{-}47}$ embryos). To test whether the weak CNS phenotype associated with $sim^{J1\text{-}47}$ results from a temperature-sensitive

mutation, we analyzed the embryonic CNS phenotypes of homozygous *sim*$^{J1\text{-}47}$ and *sim*$^{J1\text{-}47}$/*sim*^{H19} embryos at 17 and 29°C (Fig. 1).

At the permissive temperature (17°C), the CNS of homozygous *sim*$^{J1\text{-}47}$ embryos appeared indistinguishable from wild-type embryos (compare Fig. 1A and 1C). Furthermore, homozygous *sim*$^{J1\text{-}47}$ flies eclosed (see below). However, *sim* function was not completely restored at 17°C since, in trans to the null allele *sim*^{H19}, a moderate CNS phenotype was detected (Fig. 1D) that is never seen in *sim*^{H19}/+ embryos. At the restrictive temperature (29°C), homozygous *sim*$^{J1\text{-}47}$ embryos developed a strong midline defect. Commissures appeared fused and the connectives were found closer to the midline (Fig. 1E). The embryonic axon pattern phenotype became more severe in *sim*$^{J1\text{-}47}$/*sim*^{H19} embryos raised at 29°C and resembled the amorphic *sim* phenotype (compare Fig. 1B and 1F). Thus, we conclude that the allele *J1-47* represents a temperature-sensitive *sim* mutation.

The sim$^{J1\text{-}47}$ *mutant has a mutation in the Sim dimerization domain*

The *sim* gene from *sim*$^{J1\text{-}47}$ homozygous mutant flies was sequenced and compared with the wild-type *sim* gene (Fig. 2). There was a single amino acid change that occurred at residue 41, resulting in a Ser > Phe substitution. The Sim protein contains a bHLH domain, in which the basic region is required for DNA binding and the HLH domain is required for dimerization to the Tgo bHLH-PAS protein. Ser41 lies within helix 2 of the HLH domain. This residue is conserved among all Sim proteins, including two *Drosophila* species and a variety of vertebrate species (Fig. 2). It is also commonly conserved between related invertebrate and vertebrate bHLH-PAS proteins, including Trachealess, Hypoxia-inducible factors, and the Aryl hydrocarbon receptor. The high degree of conservation in a known functional domain strongly

suggests that the Ser > Phe substitution at residue 41 is responsible for the mutant phenotype. The position of Ser41 in the protein suggests a role in influencing protein dimerization, DNA binding, or both (Ferre-D'Amare et al., 1993).

Requirement of single-minded *during the development of the larval cuticle*

Mutations in the *sim* gene were also identified based on their larval cuticle phenotype, which is characterized by defective formation of the ventral-most denticles and abnormal anal pad formation (Mayer and Nusslein-Volhard, 1988).

In the larval cuticle, eight abdominal segments can be easily seen. In the phylotypic stage, the primordia of the abdominal segments A9 and A10 can be identified. Cells located in A8-A10 contribute to the development of the genital imaginal discs. A highly modified A11 segment gives rise to the anal pads that flank the anus (Jurgens and Hartenstein, 1993). In *sim* null mutants, the anal pads formed; however, their size appeared reduced compared with wild-type, and the anal slit is not developed (Fig. 3B). In homozygous $sim^{J1\text{-}47}$ embryos grown at the permissive temperature, the anus always formed normally (Fig. 3C). In trans to the amorphic mutation sim^{H9}, the anal slit occupied only half of the anal pad (Fig. 3D, arrow). When raised at the restrictive temperature, homozygous $sim^{J1\text{-}47}$ larvae as well as $sim^{J1\text{-}47}/sim^{H9}$ larvae completely lacked the anal slit (Fig. 3E and 3F, arrow).

How is the defect in anal pad development mediated? During development, *sim* expression in the midline extends to the posterior end of the germ band, where it demarcates the anterior boundary of the proctodeum (Fig. 4A-4F). As the proctodeum lies within the anal pad, we analyzed possible co-localization with the anal pad marker Even skipped (Eve) (Gorfinkiel et al., 1999). In stage 11 embryos, a crescent of Eve localization was seen abutting the domain of

sim expression at the midline (Fig. 4D-4F). Confocal analyses demonstrated that Sim and Eve are never co-localized. In *sim* null mutant embryos, the onset of Eve localization in the anal pad anlage was normal; however, Eve localization extends across the ventral midline (Fig. 4G). By the end of embryogenesis, the anal pad is reduced in size in *sim* mutants compared with wild-type (Fig. 4H and 4I). Thus, both the anal pad phenotype of *sim* mutant larvae and the expression pattern described above suggest that *sim*-expressing cells contribute to the formation of the anal pads.

sim *is required for the development of the genital discs*

In the abdominal segments A9 and A10, just anterior to the forming anal pads and thus within the expression domain of *sim*, lies the unpaired genital disc primordium (Hartenstein and Jan, 1992; reviewed in Jurgens and Hartenstein, 1993). To investigate whether *sim* also affects the formation of these ectodermal derivatives, we used an enhancer trap insertion in the *escargot* (*esg*) gene, which labels all imaginal disc anlagen by the end of embryogenesis (Hayashi et al., 1993). In wild-type embryos, a field of *esg*-expressing cells was seen posterior to the forming nervous system (not shown). Following germ band retraction, these cells invaginate into the interior of the embryo. Epidermal cells close the gap left by the delaminated disc progenitor cells, and in stage 16 embryos, a narrow strip of *esg*-positive cells was seen (Fig. 5A, 5C, and 5E, arrows). In *sim* mutant embryos, specification of the genital disc anlage is normal as judged by the onset of *esg* expression. The invagination of the anlage starts normally as well but cannot proceed to its final state. Instead, the presumptive disc cells were found in a large ectodermal fold at the posterior end of the embryo (Fig. 5B, 5D, and 5F, arrows). Interestingly, this phenotype correlates with a defect observed in the condensation of the ventral nerve cord. *sim* expression

overlaps with *esg* expression, which initially is expressed in the CNS midline as well (data not shown). At present, we cannot tell whether the genital disc phenotype is due to disruption of nerve cord condensation or whether it is due to a loss of midline cells in the disc anlage.

Adult phenotypes associated with sim

At the permissive temperature, homozygous $sim^{J1\text{-}47}$ flies survived to adulthood; however, both females and males were sterile. A similar phenotype was observed in $Df(3R)ry^{614}/sim^{J19}$ or $Df(3R)ry^{75}/sim^{J19}$ flies (see Fig. 7). In the majority of the flies (>90%), no abnormal external phenotypes were detected. The remaining flies showed dramatic external phenotypes lacking genitalia as well as the anus (Fig. 6B and 6D). Instead of forming the anal plate and the external genital organs (vulva, clasper, and penis apparatus), the tergit and sternit derived from the A8 segment close the flies posteriorly (Fig. 6B' and 6D'). We also analyzed the internal morphology of the female and male gonads of these flies. In both sexes, only rudiments were found. In $sim^{J1\text{-}47}$ females, tiny gonads developed that were generally not connected to the vulva. Despite the abnormal morphology, oogenesis apparently started normally, but arrested at an early stage (Fig. 6A" and 6B"). This indicates that $sim^{J1\text{-}47}$ does not affect the migration of the primordial germ cells into the developing gonads during embryogenesis. In males, only testis rudiments were found (Fig. 6C" and 6D").

Thus, the main defect in $sim^{J1\text{-}47}$ flies raised at 17°C appeared to be due to defective genital disc development. The missing anus leads to a blind ending hindgut, whereas gut development itself was not severely affected and malpighian tubules form normally (Fig. 6D"). In addition, the gonads were not connected to the external excretory organs. Due to the lack of any opening, the gut swelled dramatically, which after a few days, led to the death of the fly.

sim *affects adult behavior*

To even further reduce the function of *sim* during development, we analyzed $sim^{J1\text{-}47}/sim^{H9}$ animals. At 29°C, such flies never appeared, which was expected given the severe embryonic CNS phenotype (Fig. 1F). At 17°C, however, very rare escapers appeared (0.1% of the expected flies; the frequency depends on the exact culture conditions). These flies were sterile and frequently lacked external genitalia and the anal plate, similar to the phenotypes shown by homozygous mutant $sim^{J1\text{-}47}$ flies. The sterility phenotype may be explained not only by defects in external and internal morphology, but also by misbehavior, as mutant males failed to perform the normal courtship behavior and mutant females ignored wild-type males (only flies without external defects were examined).

In addition to abnormal courtship behavior, surviving $sim^{J1\text{-}47}/sim^{H9}$ flies showed an extreme walking phenotype. The majority of the flies analyzed only walked in circles (see Fig. 7A-7D for walking traces). The phenotype was variable, possibly due to the changing levels of residual *sim* gene activity due to slight differences in the culture conditions. We further analyzed the behavioral deficits and studied 18 of these flies in Buridan's paradigm, which allows evaluation of many parameters of the flies' walking ability (Gotz, 1980; Strauss and Heisenberg, 1993; Strauss and Pichler, 1998) (Fig. 7, and see Materials and Methods).

In Buridan's paradigm, wild-type flies ran in relatively straight lines between the two stripes and showed no preferred sense of rotation when they turned in front of the landmarks (Fig. 7E). Eleven $sim^{J1\text{-}47}/sim^{H9}$ flies were tested in the first experiment; they generally turned either right or left (see Fig. 7A-7C for representative tracings). Interestingly, for a given fly, the turning direction usually did not change. Compared with wild-type, all mutant flies showed a

markedly reduced walking speed, and in most cases, the activity period was shorter (Fig. 7H and 7I). Only one of the *sim* flies showed some visible interest in the landmarks (see Fig. 7D).

In order to quantify a possibly existing residual orientation toward the landmarks, which might have been obscured by the circling behavior, the same individuals were monitored for an additional 15 min in the same arena but without landmarks. Normal flies then showed area-covering random search behavior with an increasing mean free path (Schuster and Gotz, 1994). Again, $sim^{J1\text{-}47}/sim^{H19}$ flies individually preferred either a clockwise or counterclockwise sense and, compared with wild-type, showed reduced walking activity and speed (Fig. 7I and 7J). In Fig. 7F, the mean difference in orientation behavior between the Buridan behavior and random search behavior is shown for *sim* and wild-type flies. Whereas wild-type flies showed the expected strong preference for the landmarks (i.e., small error angles are found more frequently than large error angles; see Materials and Methods), there was no detectable difference between Buridan and random search data in $sim^{J1\text{-}47}/sim^{H19}$ flies. In summary, we find no measurable influence from visual landmarks on the orientation behavior of *sim* mutant flies.

We next addressed the important question whether $sim^{J1\text{-}47}/sim^{H19}$ flies are blind by exposing them to optomotor stimuli. A striped drum consisting of six dark and six interspaced bright stripes rotated around the walking platform. All stripes were equally broad and equally spaced (i.e., pattern wavelength 60°; contrast 0.94; 30 full rotations were shown in 150 s). Wild-type flies turned in the sense of the pattern rotation in an attempt to compensate for the seen rotation (which is >90% under these conditions) (Strauss et al., 1997). In most *sim* flies, the spontaneous circling behavior was weakly modulated by the optomotor stimulus. Pattern rotation in their preferred direction enhanced their turning tendency, and stimulation against their preferred direction either decreased their turning tendency or even reversed it into the stimulus direction. For example, two spontaneously counterclockwise circling individuals

showed 14 (16) counterclockwise rotations for 30 counterclockwise rotations of the pattern, whereas they performed 8 counterclockwise (2 clockwise) rotations for 30 rotations in the nonpreferred clockwise sense. Two of 11 inspected $sim^{J1\text{-}47}/sim^{H19}$ flies did not react to optomotor stimuli at all. We thus conclude that most of the *sim* flies were not entirely blind.

The mean track length of $sim^{J1\text{-}47}/sim^{H19}$ flies decayed significantly with every 3-min bin (Fig. 7G). Regardless of any landmarks, *sim* flies started out with about half the wild-type track length and then walked progressively less. Wild-type flies, in contrast, tended to produce a more constant track length, often over many hours, after their initial arousal from handling has decayed. Interestingly, females appeared to be affected more severely by the *sim* mutation. Detailed analysis of the walking traces using a laser carpet (Strauss, 1998) suggested that leg movement is coordinated normally; however, step length differed between the right and left body side (data not shown).

Histological analyses of sim^{ts} *animals*

Consistent with the behavioral defects, we noted alterations in the normal brain structure. Serial frontal sections were examined for all of the $sim^{J1\text{-}47}/sim^{H19}$ flies in which behavior was assessed. The series were inspected by using autofluorescence conditions under which all neuropil have a blue-green and the pericarya a bright green to yellow appearance (Fig. 8). In particular, we noted defects associated with the inner chiasm of the optic lobes and in the central complex (CX), a structure that spans the protocerebral hemispheres and develops from the larval interhemispheric commissure during larval and pupal stages (Hanesch et al., 1989). The CX is composed of the four neuropilar regions called the protocerebral bridge (pb), fanshaped body (fb), ellipsoid body, and paired noduli. In 70% of the 19 inspected *sim* brains, the pb as the

posterior-most CX neuropil was markedly thinner at the sagittal midplane and generally less compact than the intact wild-type pb (Fig. 8A and 8B). The pb consists of a linear array of 8 bilaterally paired glomeruli. Fibers might be missing in *sim* flies, which normally run along the pb and connect the paired glomeruli in a highly ordered fashion (Hanesch et al., 1989). The fb, the largest of the CX neuropils, was divided sagittally in its posterior shell in 15% of the *sim* brains (Fig. 8C and 8D). The ellipsoid-body and noduli were not affected by the *sim* mutation.

The inner chiasm of the optic lobes connects the medulla with the lobula and the lobula plate. In 40% of the inspected *sim* flies, ectopic fiber bundles were found to cut into the lobula either unilaterally or bilaterally (Fig. 8E and 8F). Such ectopic bundles are never observed in wild-type optic lobes (Brunner et al., 1992). We were unable to directly correlate the gross anatomical salience of the CX or optic lobe defects with the severity of the walking problems or the degree of circling. Flies with a unilateral inner chiasm defect circled as consistently as flies with no visible defect. As an exception, one of the putatively blind flies mentioned above had the most severe optic lobe defects on both sides.

Postembryonic expression of Sim *in the CNS*

The above described phenotypes suggest that *sim* is expressed postembryonically. This was initially examined by RT-PCR experiments (see Materials and Methods), which demonstrated that *sim* transcripts were present throughout fly development (Fig. 9). Transcripts were barely detected in whole third instar larvae, but were abundant in dissected larval brains (Fig. 9).

Spatial expression of *sim* was further examined by immunostaining larvae and adult flies with Sim antibodies, followed by examination with confocal microscopy. In the larval ventral

nerve cord, *sim* is expressed in the midline glial cells (Fig. 10A). These cells, which can easily be identified by their characteristic morphology (Awad and Truman, 1997; Stollewerk et al., 1996), express *sim* during embryonic CNS development (Crews et al., 1988; Thomas et al., 1988).

Similar to its distribution in embryonic cells, Sim protein was confined to cell nuclei (Ward et al., 1998). During embryonic development, nuclear localization of the Sim protein requires heterodimerization with the Tgo bHLH-PAS protein. Tgo is found in all cells and is localized to the cytoplasm in the absence of a partner bHLH-PAS protein, such as Sim (Ward et al., 1998). In the presence of Sim, the Sim:Tgo complex translocates to the nucleus, and immunostaining with Anti-Tgo shows prominent nuclear staining in those cells. To analyze whether Sim may interact with Tgo in the larval midline cells and to obtain a validation control for the Sim immunostaining, we double-stained larval nerve cords with anti-Sim and anti-Tgo antibodies. The results revealed a complete colocalization of nuclear Sim and Tgo expression in the CNS midline (Fig. 10C).

Examination of larval brain staining revealed additional sites of Sim localization. Sim was prominently expressed in the lamina and the medulla, which are synaptic targets of photoreceptor axons (Fig. 10D). Again, a colocalization of Tgo was found (Fig. 10E and 10F). Examination of the central brain, which has been implicated in the control of the walking behavior, identified two clusters of Sim-positive cells, one on each side of the midline (Fig. 10I and 10J). Each cluster has three groups of cells that are not contiguous. The two anterior-most clusters showed higher levels of Sim protein than the posterior-most cluster (Fig. 10J). These cells may be important for the locomotor defects of mutant $sim^{J1\text{-}47}/sim^{H19}$ flies. Despite the fact that $sim^{J1\text{-}47}/sim^{H19}$ flies are probably able to see (see above), *sim* expression in the optic lobe might nevertheless contribute to the behavioral deficits. There also exist smaller, less distinct, groups of *sim*-expressing cells in the anterior of the brain (not shown).

Thus, the temporal and spatial expression patterns of *sim* support a postembryonic function of the gene. The availability of the temperature-sensitive *sim* allele could in principle be used to address this question. However, since only very few $sim^{J1\text{-}47}/sim^{H19}$ flies eclosed even at the permissive temperature, we could only analyze $sim^{J1\text{-}47}/sim^{J1\text{-}47}$ flies. Flies were crossed at the permissive temperature of 17°C and were transferred to fresh food vials every 24 h. The animals were subjected to the restrictive temperature (29°C) in 24-h intervals. As long as the shift to the restrictive temperature occurred after embryogenesis, normal numbers of flies eclosed, indicating that postembryonic functions of *sim* do not affect viability.

DISCUSSION

The gene *sim* acts as a master regulatory gene controlling midline development in the ventral nerve cord of *Drosophila*. In fact, it was one of the first genes associated with such an important function during cell-type specification. The function of *sim* outside the nerve cord has been known for a long time but was never investigated in much detail. Here, we report the identification of a temperature-sensitive *sim* mutation that demonstrates a requirement of *sim* during genital imaginal disc and anal pad development. In addition, behavioral deficits are associated with *sim* function. Most importantly, we noted a walking defect that is likely due to a disruption of the central brain complex development in conjunction with a developmental defect found in the inner chiasm of the optic lobes.

Embryonic function of sim

sim expression is first evident during the cellular blastoderm stage in a strip of cells flanking the mesodermal anlage. The majority of these cells will later divide to generate the neurons and glial cells found at the midline of the ventral nerve cord. It is important to note, however, that *sim* expression exceeds the neurogenic region from which the nerve cord will form. At the posterior, *sim* expression extends into abdominal segment 10, where it can be detected until end of stage 11. The fate of these cells is presently unknown. Possibly, the ectodermal midline cells provide inductive signals influencing the developing neighboring tissues, which appears to be a more general feature of the midline cells. Within the CNS, the midline cells act as an organizing center controlling the patterning of axons by providing attractive and repulsive cues. Furthermore, the midline cells regulate the number and differentiation of cortical neurons and mesodermal cells (Chang et al., 2000; Luer et al., 1997; Menne et al., 1997; Zhou et al., 1997).

The anal pad anlage can be labeled by Eve localization (Gorfinkiel et al., 1999) and forms immediately posterior to *sim*-expressing cells. In *sim* mutants, the *eve* expression domain shifts toward the midline and meets at the midline. Possibly, the gap in *eve* expression in wild-type embryos allows the formation of the anal slit. In *sim* mutants, the posterior midgut invaginates normally and the proctodeum initially forms. However, during later stages, the cells that will give rise to the anus will die and thus prevent the external opening of the hindgut.

What is the function of the ventral midline during genital disc development? In both female and male flies, the sexually dimorphic terminalia are formed by a common genital disc comprising three primordia (Sanchez and Guerrero, 2001). The female genital primordium is derived from the 8th abdominal segment, the male genital primordium from the 9th abdominal

segment, and the anal primordium from the 10th and 11th abdominal segments. In both sexes, the anal primordium will develop, whereas depending on the sex of the animal, either the female or the male primordium will develop (Sanchez and Guerrero, 2001). The definition of the genital disc anlage does not appear to be affected in *sim* mutants, but the subsequent delamination from the ectoderm is abnormal. In wild-type embryos, the genital disc anlage forms just posterior to the developing ventral nerve cord. Following germ band retraction, the ventral nerve cord retracts, and concomitantly, the genital disc anlage delaminates from the ectoderm. At present, our analyses do not allow us to discriminate whether the genital disc phenotype found in mutant *sim* embryos is an indirect consequence of the nerve cord condensation defect or whether it is due to the loss of *sim* expression in the genital disc primordium.

sim *function during adult stages*

The temperature-sensitive mutation $sim^{J1\text{-}47}$ allowed us to address the question whether *sim* is required in larval or adult stages. Following the reduction of *sim* function, a number of interesting phenotypes emerged. The sterility phenotype displayed by the hypomorphic *sim* allele *J1-47* as well as the amorphic mutation $sim^{1\text{-}19}$ in trans to deficiencies affecting only one of the two promoters (Fig. 8) demonstrated that these phenotypic traits are indeed due to a reduction in the level of *sim* function. The sterility phenotype is likely to be a direct consequence of abnormal genital disc development during embryonic stages. Interestingly, these flies also showed abnormal courtship behavior, suggesting a requirement of *sim* in larval/adult neurogenesis. A similar conclusion has to be drawn by the walking defects of mutant *sim* flies. Mutant flies were only able to walk in circles. This phenotype could be due to a loss of motoneurons in the ventral nerve cord or it could be due to disruption of higher centers that coordinate walking.

An extended analysis of the walking behavior in different structural brain mutants showed that the central complex (CX) between the protocerebral brain hemispheres serves as such a higher center (Heisenberg and Bohl, 1979; Strauss and Heisenberg, 1993). A hallmark of mutants affecting the CX was a slower mean and maximum walking speed and a decaying locomotor activity in Buridan's paradigm, which is also shown by $sim^{J1\text{-}47}/sim^{H19}$ flies. In the majority of these mutants, the protocerebral bridges and their fan-shaped bodies were affected. The remaining flies with no gross morphological CX defects may well have defects that may be detectable only at the single cell level.

Why do $sim^{J1\text{-}47}/sim^{H19}$ flies circle? First of all, circling in only one direction is not the normal behavior of blind flies in the arena situation. They would show random-search behavior with a balanced frequency of left and right turns (e.g., *no-receptor-potential AP24*; data not shown). Secondly, we assume that most *sim* flies are not entirely blind since most of them reacted, at least weakly, to optomotor stimuli. *sim* flies are not the first example of circling flies. In the screen for locomotor mutants, the CX-defective mutant *C31* was isolated that frequently walked in wavy lines (Strauss and Trinath, 1996). *C31* function was subsequently studied in mosaics that were generated by using the gynandromorph technique. When one half of the body, including the head, was mutant, the flies were unable to walk straight and persistently turned toward the defective body side. However, unilateral mutant flies with an intact brain could walk straight pointing toward the role of the CX in balancing left-right motocontrol (Strauss and Trinath, 1996; R.S., unpublished observation). In support of the notion that the CX controls locomotive behavior is the finding that *Pax-6/eyeless* mutants cause gross morphological CX defects and, concomitantly, severe locomotor deficits (Callaerts et al., 2001).

A mutation in the gene *pirouette,* which was identified in a screen for genes affecting auditory behavior, shows a similar walking phenotype as described here for the hypomorphic *sim*

mutation (Eberl et al., 1997). Within the CNS, the optic lobes degenerate but no information about the development of the CX is available. We have not observed any genetic interaction between the two loci (data not shown). To date, only few other mutations have been described that specifically affect the development of the CX. The transcription factor AP2 is not required during embryonic development; however, adult flies display severe disruptions in the CX. It is unknown whether AP2 mutations affect behavior similar to *sim* (Monge et al., 2001). Other mutants affecting the formation and connectivity of the CX have been described, but no information is available on walking abilities of the different mutant flies (Boquet et al., 2000; Hitier et al., 2001; Simon et al., 1998).

Beside expression within the central complex, we noted high levels of Sim localization in the optic lobes, the lamina, and the medulla, which is in agreement with the mutant phenotype. During larval development, the optic lobes undergo extensive rounds of cell proliferation to give rise to the mature neurons and glia. DNA replication and cell division occur at several discrete sites: the inner proliferative center (IPC), the outer proliferative center (OPC), and the laminar precursor center (LPC) (Hofbauer and Campos-Ortega, 1990; Selleck and Steller, 1991). Since *sim* expression was observed during the proliferative phase of optic lobe development, we addressed whether *sim* was expressed in proliferating cells or the postmitotic cells. The proliferative zones were visualized by expression of *GFP* from a *PCNA-GFP* transgenic strain (R. Duronio, personal communication). The *Drosophila* Proliferating Cell Nuclear Antigen (PCNA) protein is encoded by the *mus209* locus and is localized in replicating cells. Since PCNA-GFP and Sim do not co-localize (Fig. 10G), it appears that *sim* is only expressed in postmitotic cells in the optic lobes. Double labeling experiments with glial and neuronal antigens indicate that, within the brain, *sim* is expressed only in neuronal cells (Fig. 10J-10L). The optic lobes of *sim* mutant flies show aberrant axonal projects, but the medullary and laminar neurons

are present. This suggests that the role of *sim* in optic lobe development may be different from its role in controlling formation of the CNS midline cells in embryonic development.

ACKNOWLEDGMENTS

We thank R. Nothiger for advice and help, John Nambu who carried-out early experiments on *sim* sterile phenotypes, Mary Ward who performed initial Sim larval brain staining, D. Eberl for mutant *pirouette* flies, C. S. Goodman for antibodies, Bob Duronio, Craig Montell, and Andrew Tomlinson for advice and reagents, and members of the Klämbt lab for help and discussions. This work was supported through a predoctoral fellowship of the Boehringer Ingelheim Fonds (to J.P.), a BMBF-BioFuture grant (to R.S.), a grant from NICHD (to S.T.C.), and a grant of the DFG (to C.K.).

FIGURES

Figure 1. Identification of a temperature-sensitive *sim* allele. Frontal views of dissected stage 16 ventral nerve cords of the indicated genotype stained for the overall axon pattern using the mAb BP102. The breeding temperature is indicated. Anterior is to the left. (A) In wild-type embryos, a ladder-like axon pattern can be recognized. Axons running in the longitudinal connectives connect the different neuromeres along the A/P axis. In every neuromere, two commissures are found: the anterior commissure and the posterior commissure. (B) In sim^{H19} null mutants, all CNS axons collapse at the midline. (C) The ventral nerve cord of homozygous $sim^{J1\text{-}47}$ embryos appears wild-type and no abnormalities can be detected. (D) In $sim^{J1\text{-}47}/sim^{H19}$ embryos grown at low temperature, a mild fused commissure phenotype develops, which is indicative of midline glial cell defects. (E) When homozygous $sim^{J1\text{-}47}$ embryos are raised at the restrictive temperature, a strong fused commissure phenotype develops. Note that the connectives are also affected. (F) Mutant $sim^{J1\text{-}47}/sim^{H19}$ embryos grown at the restrictive temperature show a complete collapse of the nervous system similar to the phenotype caused by the complete loss of *sim* function (B).

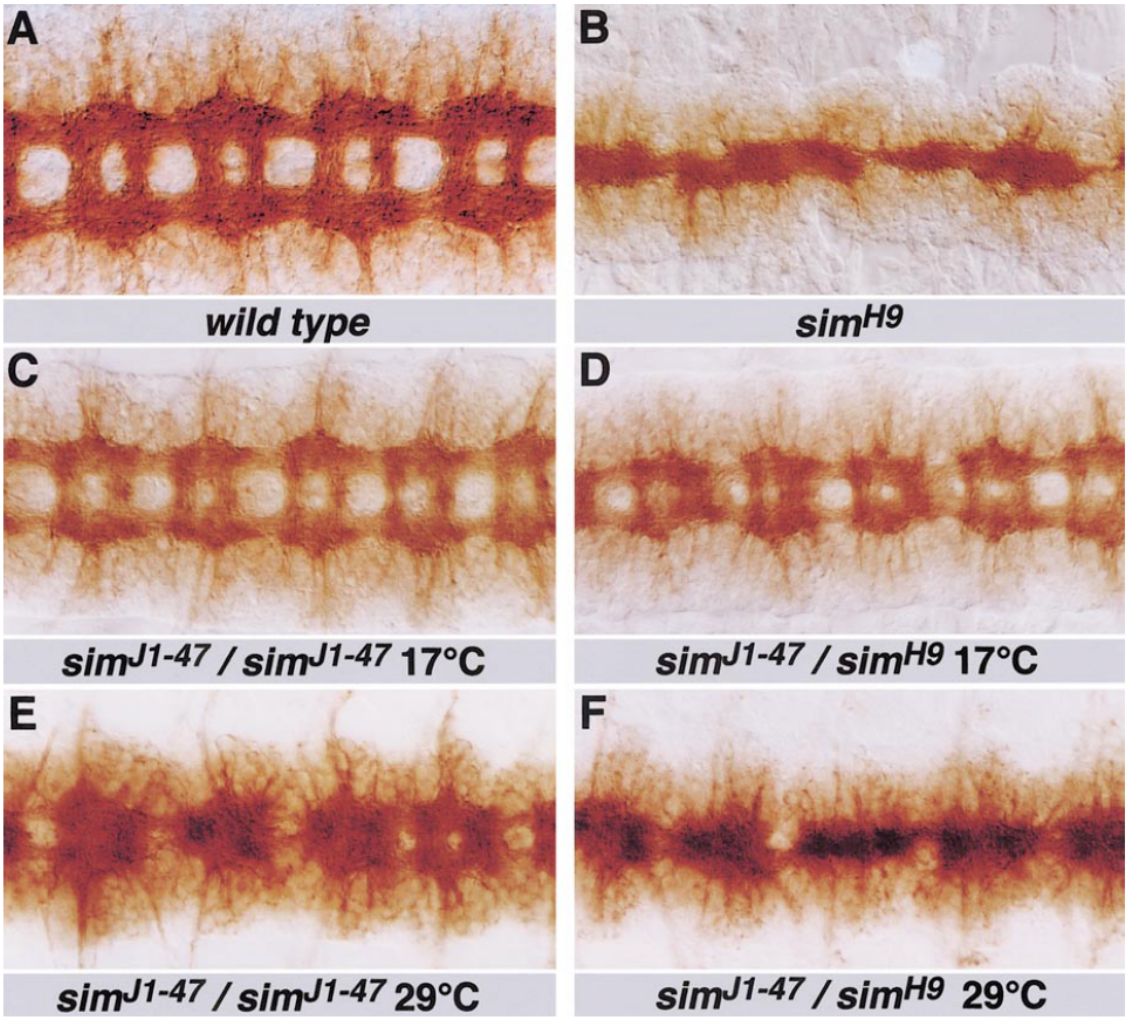
A
wild type
B
simH9
C
simJ1-47 / simJ1-47 17°C
D
simJ1-47 / simH9 17°C
E
simJ1-47 / simJ1-47 29°C
F
simJ1-47 / simH9 29°C

Figure 1

Figure 2. Alignment of Sim proteins reveals that *sim*$^{J1\text{-}47}$ has a mutation in a conserved residue in helix 2. Sequence of the entire *sim*$^{J1\text{-}47}$ *sim* gene revealed only a single amino acid replacement compared with the wild-type *sim* sequence: Ser41 was changed to a Phe in *sim*$^{J1\text{-}47}$. Alignment of all of the known Sim proteins, both insect and vertebrate, indicate that all contain a Ser at residue 41. Humans, mice, and other vertebrates contain at least two *Sim* genes: *Sim1* and *Sim2*. D, *Drosophila melanogaster*; Dv, *Drosophila virilis*; C, chicken; H, humans; M, murine; X, *Xenopus*; Z, zebrafish.

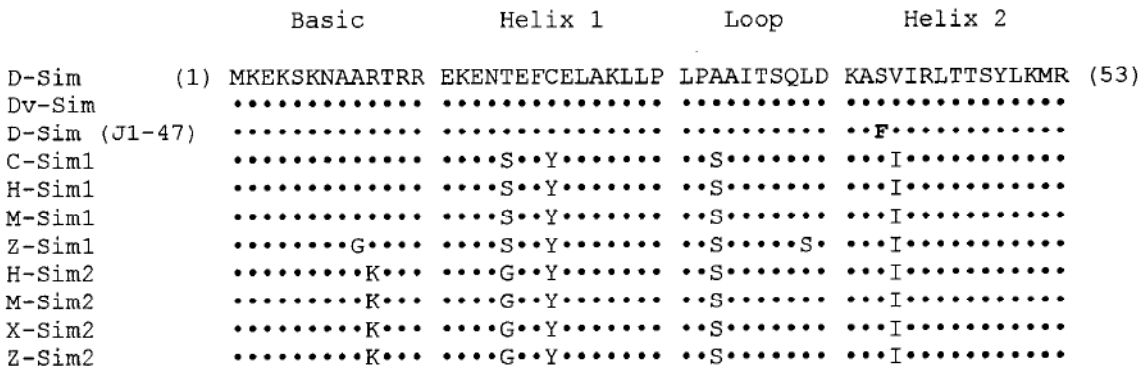
Basic
Helix 1
Loop
Helix 2
D-Sim (1) MKEKSKNAARTRR EKENTEFCELAKLLP LPAAITSQLD KASVIRLTTSYLKMR (53)
Dv-Sim
D-Sim (J1-47)
C-Sim1
H-Sim1
M-Sim1
Z-Sim1
H-Sim2
M-Sim2
X-Sim2
Z-Sim2

Figure 2

Figure 3. Larval phenotypes associated with *sim*. Cuticle preparations of wild-type and mutant *sim* larvae. Only the tail region is shown. Anterior is up. (A) Wild-type, the arrow points to the anterior origin of the anal slit in the anal pad (ap). The A8 segment is indicated. (B) Homozygous mutant sim^{H9} larvae lacking all *sim* function. The anal pad is smaller compared with wild-type and no anal slit can be detected (arrow). (C) Homozygous mutant $sim^{J1\text{-}47}$ larvae grown at the permissive temperature. The sizes of the anal pad and the anal slit are normal. (D) In $sim^{J1\text{-}47}/sim^{H9}$ larvae grown at the permissive temperature, the anal slit is shortened (arrow). (E) Homozygous mutant $sim^{J1\text{-}47}$ larvae grown at the restrictive temperature. The size of the anal slit is greatly reduced. (F) In $sim^{J1\text{-}47}/sim^{H9}$ larvae grown at the restrictive temperature, the anal slit is absent.

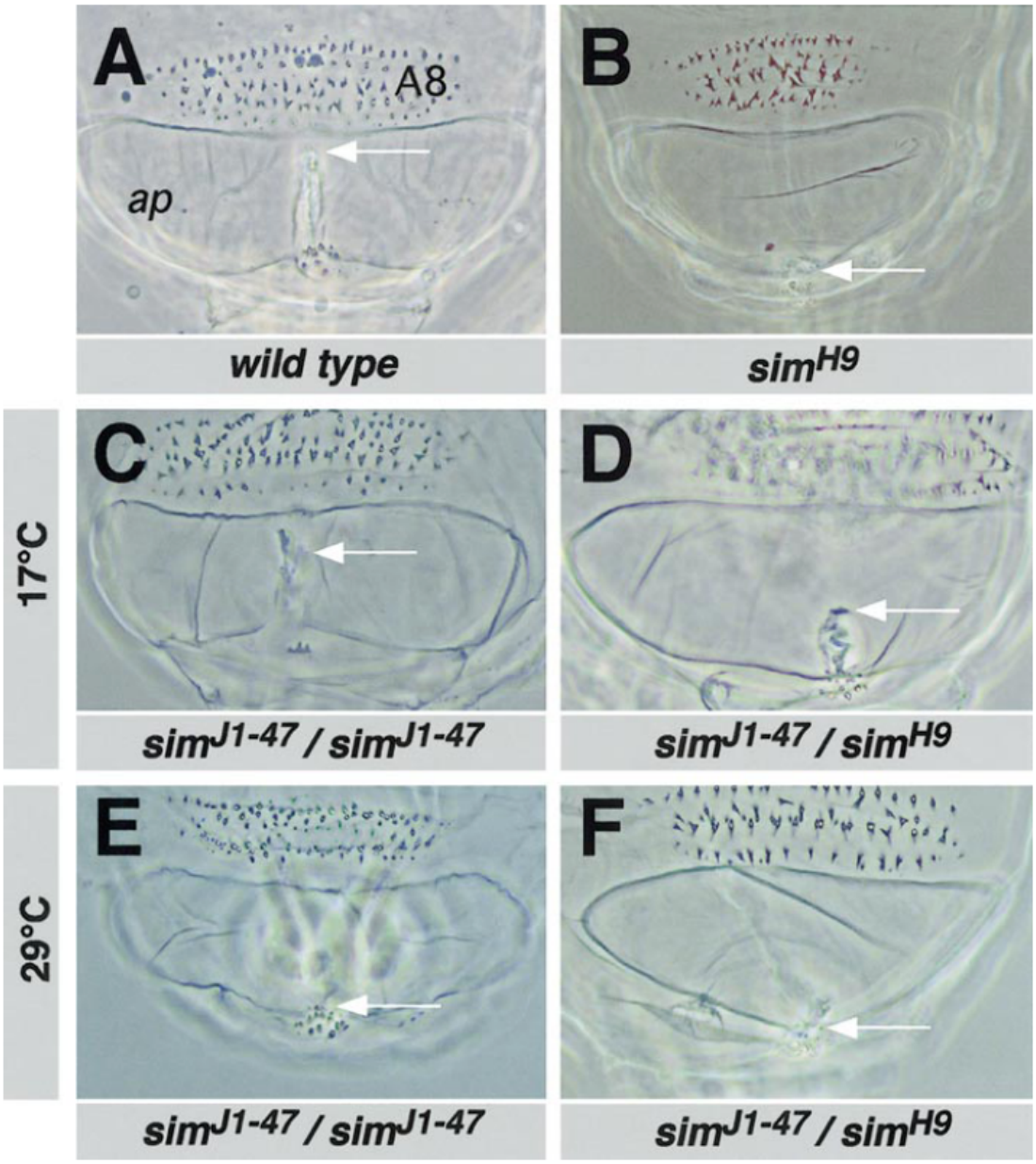
A
A8
ap
wild type
B
simH9
17°C
C
simJ1-47 / simJ1-47
D
simJ1-47 / simH9
29°C
E
simJ1-47 / simJ1-47
F
simJ1-47 / simH9

Figure 3

Figure 4. Formation of the anal pad. The formation of the anal pad was followed by using anti-Eve antibodies and confocal microscopy. Sim localization (green) was detected by using anti-Sim antibodies. Mutant *sim* embryos (G, H) were recognized by the lack of Sim localization and the use of a labeled balancer chromosome. (A-C) Three consecutive focal planes from dorsal (A) to ventral (C) of a stage 8 embryo. Sim localization (green) extends to the posterior end of the germ band abutting the anterior end of the proctodeum. Arrows point to the proctodeum. (D-F) Three consecutive focal planes from dorsal (D) to ventral (F) of a stage 11/12 embryo. A crescent of Eve-positive (red) cells is found at the posterior end of the germ band. At the ventral midline, Eve-positive cells abut Sim-positive cells; no co-localization is found. The Sim-positive cells demarcate the anterior end of the proctodeum (arrow). (G) In *sim* mutant embryos (stage 11/12), Eve localization is found across the ventral midline. Abnormal formation of the proctodeum is already evident. (H) In stage 16 *sim* mutant embryos, the anal pad is reduced in size compared with wild-type embryos (I). Asterisks indicate Eve-positive neurons.

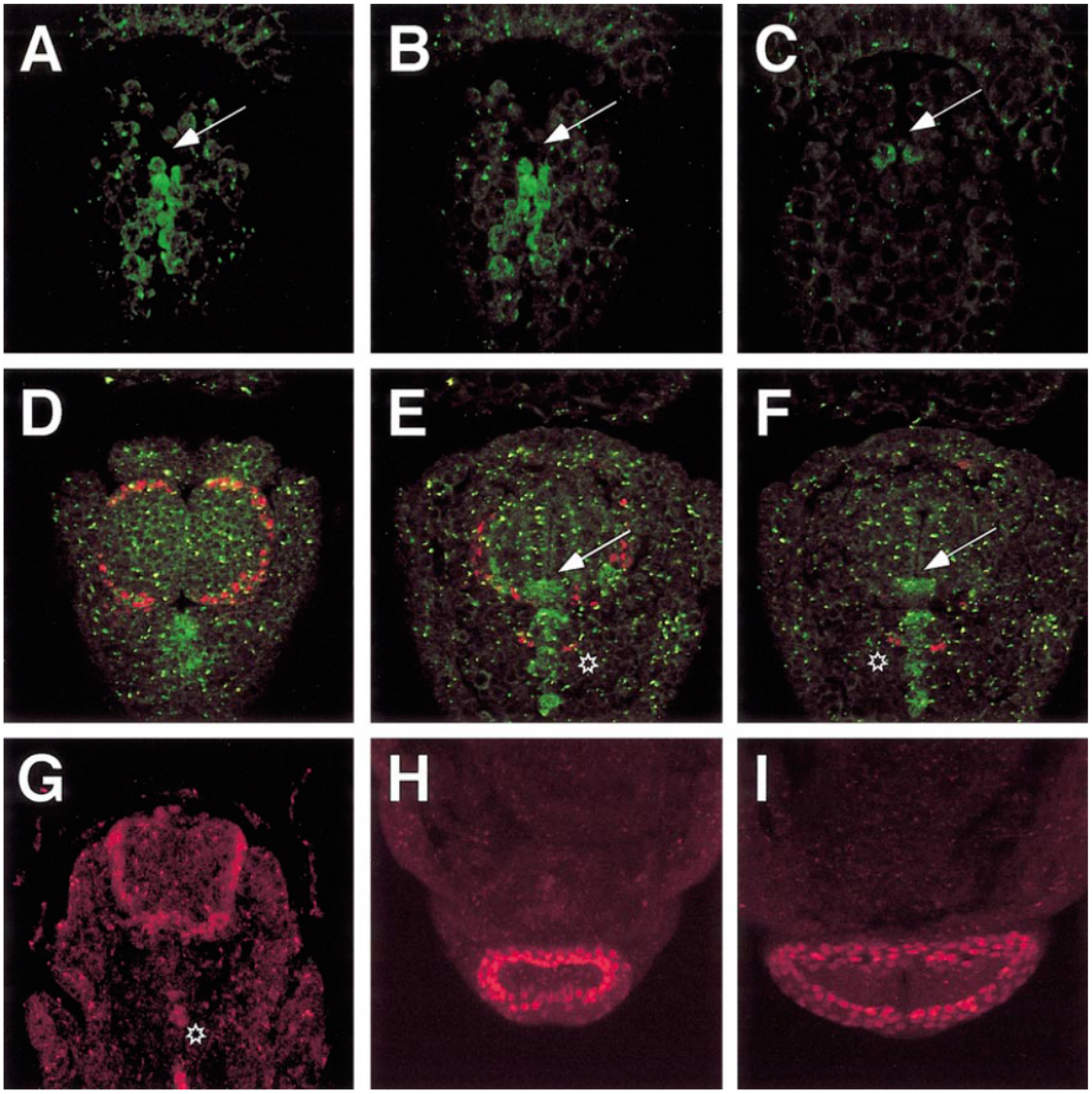

Figure 4

Figure 5. Formation of the genital disc. Preparations of wild-type (A, C) and mutant *sim*^{H9} (B, D) stage 16 embryos. The genital disc anlage has been labeled by Esg localization (arrows). Axons in the ventral nerve cord were visualized by using BP102 and subsequent HRP immunohistochemistry to determine the genotype. (A) In a lateral view, the genital disc anlage can be seen just anterior to the hindgut. (C) In dissected embryos, the disc cells are closely associated with the ectoderm. (E) Schematic drawing: the genital disc anlage is indicated in red. (B, D) In mutant *sim* embryos, Esg localization is initiated normally. However, in stage 16 embryos, a deep indentation (arrow) can be seen instead of an invaginated disc anlage. (F) Schematic drawing.

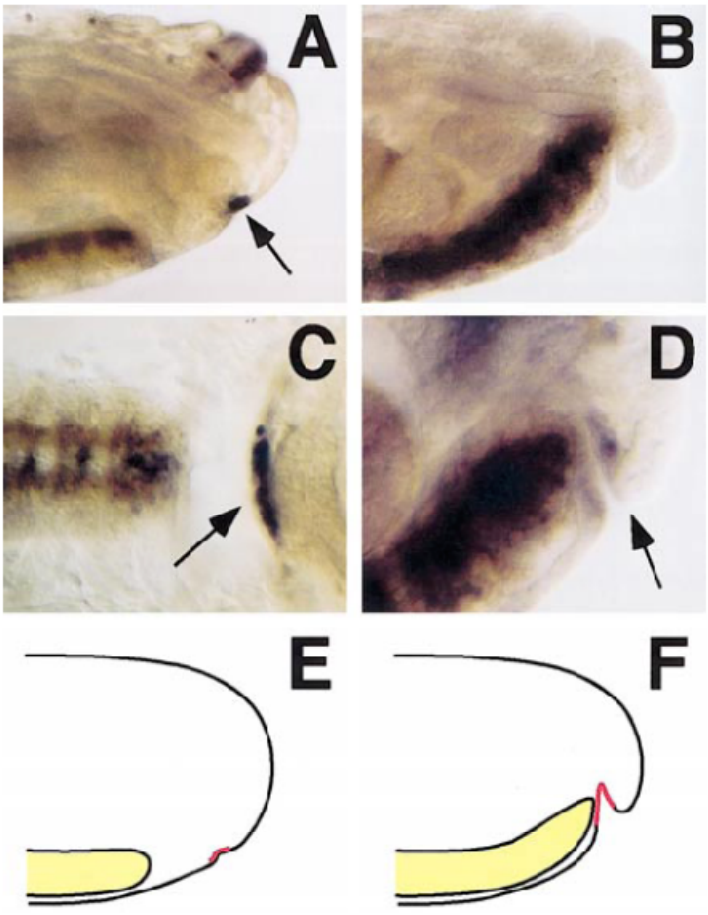

Figure 5

Figure 6. Adult phenotypes associated with *sim*. Preparations of wild-type and mutant *sim* flies. The tail regions and the gonads are shown. (A, A') *Drosophila* wild-type females can be easily recognized by the prominent anal pad (ap) and the thorn bristles (tb) around the vulva (v). (B, B') In homozygous mutant $sim^{J1\text{-}47}$ females, the anal plate cannot be recognized. The thorn bristles and the vulva appear to be missing. (A") Wild-type ovaries. The ovaries (o) are well developed and are connected via the oviduct (od) and the uterus (u) to the vulva [hindgut (hg); rectum (R); spermathecae (st); seminal receptacle (rs)] (B") In mutant $sim^{J1\text{-}47}$ female flies, the ovaries are present although poorly developed. Oogenesis apparently starts normally but stops prematurely. The hindgut is not connected to the rectum (r) and is frequently swollen. (C, C') Wild-type males have a flat anal plate (ap) and a dark pigmented clasper (c) surrounding the penis apparatus (p). (D, D') In homozygous mutant $sim^{J1\text{-}47}$ males, the anal plate, the clasper, and the penis are missing. (C") In wild-type males, the testis is well developed. Paragonium (pg), vas deferens (vd), ejaculatory duct (ed), and sperm pump (sp) can be detected. (D") In homozygous mutant $sim^{J1\text{-}47}$ males, testis development is severely impaired. Only rudiments (asterisk) that are not connected to the outside are found. As in females, the hindgut (hg) is not properly connected to the rectum (r).

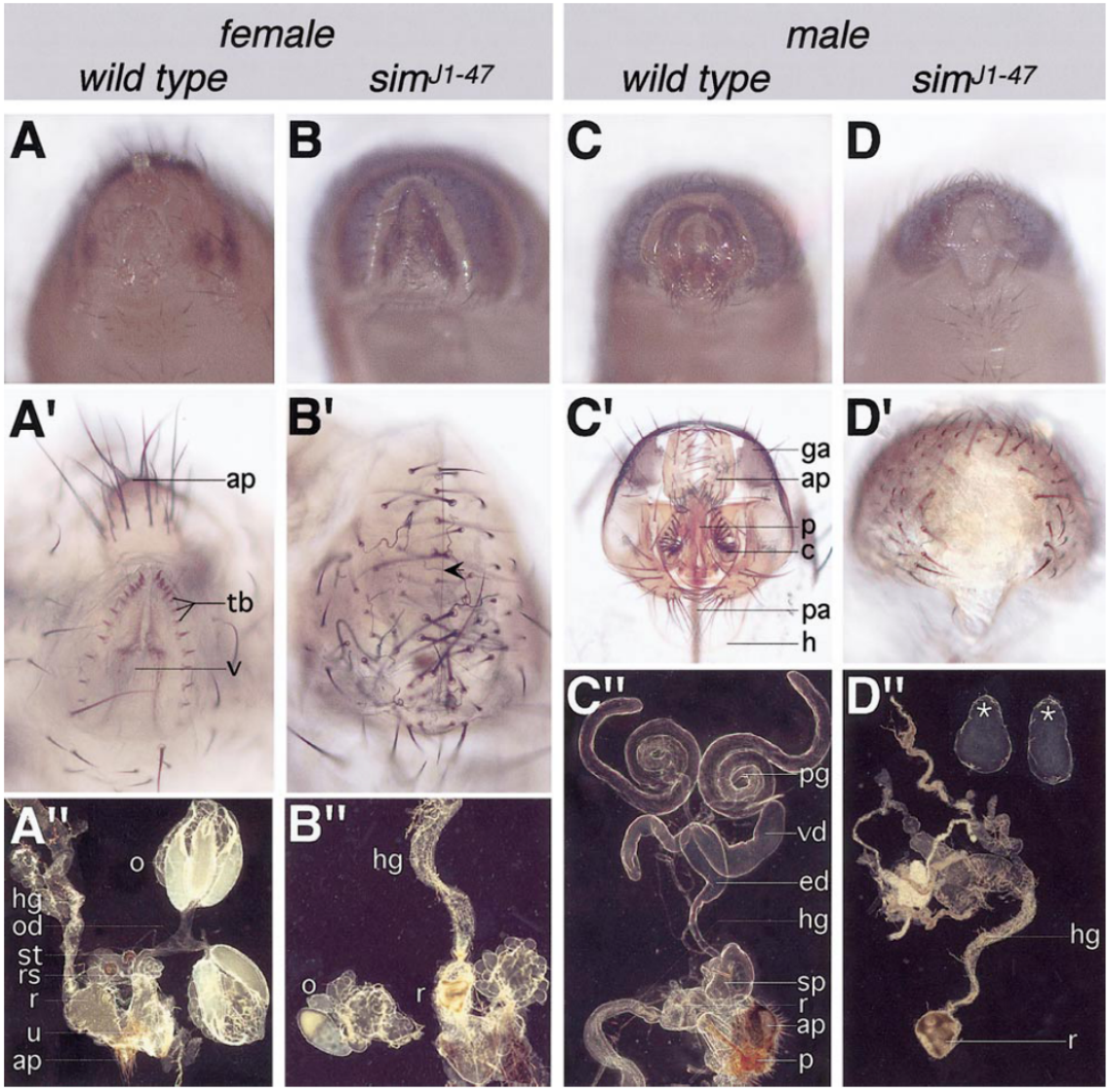
female
wild type
sim J1-47
male
wild type
sim J1-47
A
B
C
D
A'
B'
C'
D'
ap
tb
v
ga
ap
p
c
pa
h
A''
B''
C''
D''
o
hg
od
st
rs
r
u
ap
hg
o
r
pg
vd
ed
hg
sp
r
ap
p
hg
r

Figure 6

Figure 7. Walking and orientation behavior phenotype. (A-E) Walking traces. Single flies walked for 15 min on an elevated water-surrounded platform of 85 mm diameter. $sim^{J1\text{-}47}/sim^{H19}$ flies walk in circles (A-D) despite the two inaccessible landmarks (indicated by the vertical bars) which keep wild-type flies alternating between them for many hours (E, wild-type: Oregon-R). $sim^{J1\text{-}47}/sim^{H19}$ flies have a preferred side to which they almost always turn. (A) Five-minute walking traces from an extremely tight turning *sim* fly (arrowheads point to very high frequency turning points). (B, C) Medium wide and a wide turning mutant *sim* fly. (D) Exceptional example of a $sim^{J1\text{-}47}/sim^{H19}$ fly producing a measurable orientation component toward the landmarks. (F) Mean occurrence frequency of certain error angles between the actual walking direction (obtained every 0.2 s) and the straight direction toward the nearer, in angular terms, of the two landmarks (Buridan's paradigm) minus the respective frequency distribution for random search behavior in the absence of landmarks. $N = 11$ flies per test group ($sim^{J1\text{-}47}/sim^{H19}$, wild-type Oregon-R) were measured in random order each for 15 min in the Buridan situation and for 15 min in the empty arena. On average, there is no measurable influence from the landmarks on the orientation behavior of mutant *sim* flies. (G) Mean track length per 3-min bin. Same experimental groups as in (F). Regardless of the presence or absence of landmarks, *sim* mutant flies start with about half the wild-type track length and decay significantly, whereas wild-type flies tend to produce a more constant track length as soon as their initial arousal from handling has decayed. The overall deficits in the track length of mutant *sim* flies are due to a drastically reduced walking speed (H) and to a bisected activity (I; mean percentage of time spent walking). (J) Mutant *sim* female flies are more strongly affected than mutant *sim* male flies. An insignificantly lower walking speed and lower activity in females combine to a significant deficit in mean walked distance when compared with male *sim* flies (*t*-test, two-tailed: $P < 0.05$). All error bars indicate SEMs.

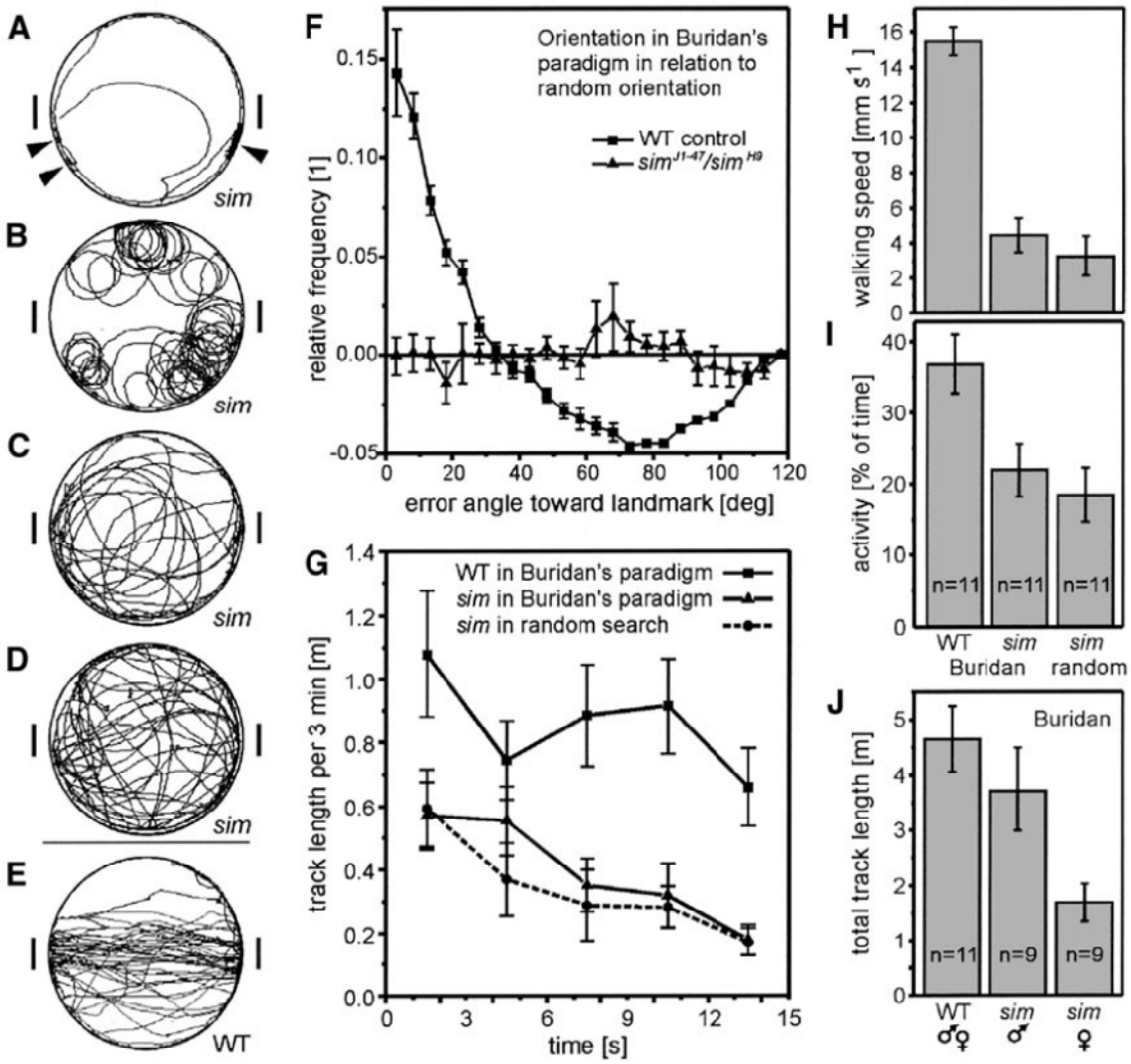
A
sim
B
sim
C
sim
D
sim
E
WT
F
Orientation in Buridan's paradigm in relation to random orientation
WT control
sim J1-47/sim H9
relative frequency [1]
error angle toward landmark [deg]
G
WT in Buridan's paradigm
sim in Buridan's paradigm
sim in random search
track length per 3 min [m]
time [s]
H
walking speed [mm s-1]
I
acivity [% of time]
n=11
n=11
n=11
WT Buridan
sim random
J
Buridan
total track length [m]
n=11
n=9
n=9
WT ♂♀
sim ♂
sim ♀

Figure 7

Figure 8. Structural defects in the adult *sim* brain. Frontal 7-μm-thick sections through wild-type brains (A, C, E) and $sim^{J1\text{-}47}/sim^{I19}$ brains at comparable levels (B, D, F). Dorsal is up. Under autofluorescence conditions, all neuropil appears in darker green and the pericarya in bright green-yellow. (A, B) In 70% of the inspected *sim* brains, the protocerebral bridge (pb) of the central complex was markedly thinner and less compact than a wild-type pb. Arrowheads indicate the thinnest part at the sagittal midplane. (C, D) The fan-shaped body (fb) of the central complex was divided posterior-sagittally in 15% of the *sim* brains. Arrowheads indicate the dorsal-most extents of the fb. (E, F) The inner chiasm of the optic lobes (only right side is shown) is situated between the medulla (me) and the lobula (lo). The ventral and dorsal ends of the chiasm region are indicated by open brackets. The inner chiasm was disordered either unilaterally or bilaterally in 40% of the *sim* brains. Axon bundles take an abnormal path into the distal lobula, cutting into this neuropil to varying depths (arrowheads). Such bundles are never observed in wild-type optic lobes. Scale bars indicate 20 μm; (A-D) are on the same scale as (C); the bar in (E) applies also to (F).

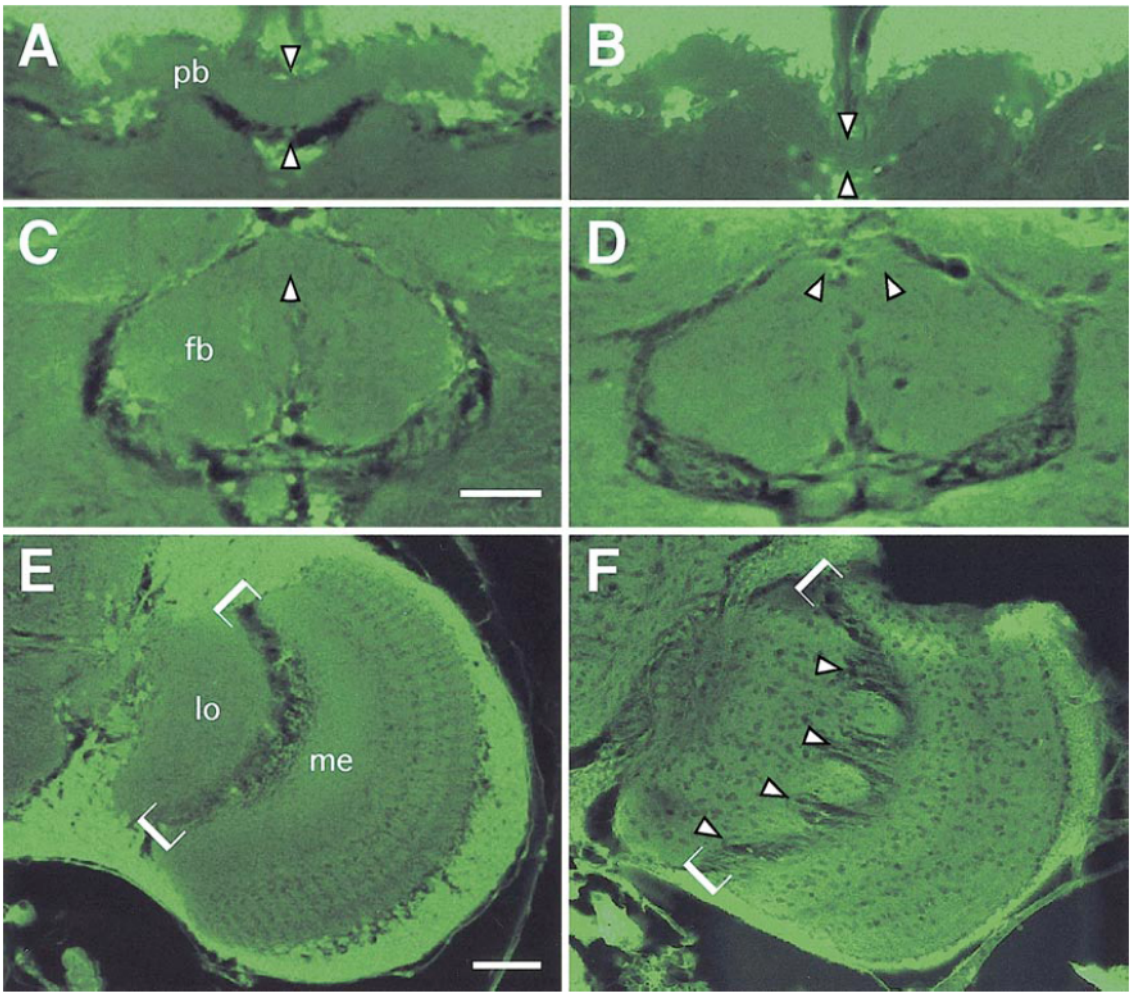

Figure 8

Figure 9. Postembryonic expression of *sim*. (A) The schematic depicts the *sim* gene structure. Exons are numbered 1-8 with untranslated regions shown as unfilled boxes and the coding sequence depicted as filled boxes. The early promoter (P_E) drives expression in the CNS midline precursor cells, whereas the late promoter (P_L) drives expression in embryonic midline glia. Two deletions, $Df(3R)ry^{75}$ and $Df(3R)ry^{614}$, remove P_L, but leave early expression intact. The breakpoints of the deletions are within the boxed regions and DNA to the left of the box is deleted. Scale in kb is shown at the bottom. (B) Flies that are either $Df(3R)ry^{75}/sim^{H19}$ or $Df(3R)ry^{614}/sim^{H19}$ presumably have sufficient embryonic *sim* expression for viability. However, the resulting male and female adults are sterile. (C) RT-PCR was carried out on total RNA from different *Drosophila* stages, including 0- to 18-h embryos (E), third instar larvae (L), dissected third instar larval brains (LB), pupae (P), and adults (A). Total RNA was converted to cDNA by using a gene-specific primer and reverse transcriptase. PCR was performed by using a primer pair corresponding to the *sim* coding sequence. After PCR, the DNA products were electrophoresed on an agarose gel and visualized by ethidium bromide staining. The primer pair spanned an intron, so that the predicted size of the mRNA amplification product is 237 bp and the predicted size of the amplification product from genomic DNA is 558 bp. This is confirmed by *in vitro* transcribing a *sim* cDNA clone, followed by RT-PCR and electrophoresis. The 237-bp product is shown (R). Amplification of *Drosophila* genomic DNA shows the 558-bp predicted band (G). The presence of a 237-bp product in the developmentally staged RNA lanes indicates that the amplification product is derived from RNA and not from contaminating genomic DNA.

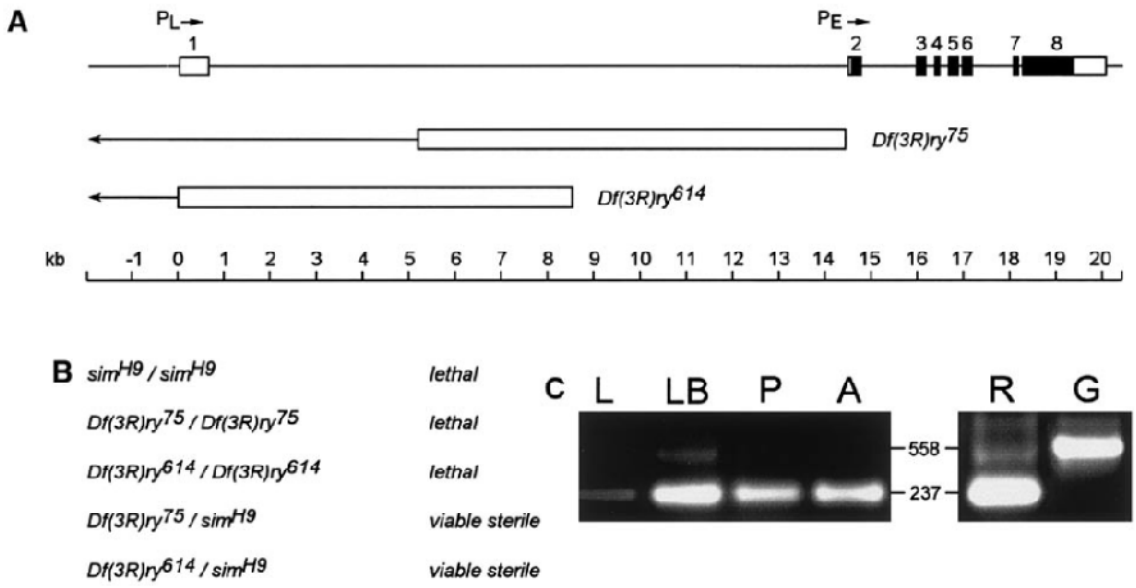
A
PL
1
PE
2
3 4 5 6
7
8
Df(3R)ry75
Df(3R)ry614
kb
-1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
B
simH9 / simH9
lethal
Df(3R)ry75 / Df(3R)ry75
lethal
Df(3R)ry614 / Df(3R)ry614
lethal
Df(3R)ry75 / simH9
viable sterile
Df(3R)ry614 / simH9
viable sterile
C
L
LB
P
A
R
G
558
237

Figure 9

Figure 10. Sim protein expression in the larval CNS. Third instar larval brains and ventral nerve cord (vnc) were dissected from third instar larvae, stained with antibodies, and examined by confocal microscopy. (A-C) Larval CNS was double-stained with anti-Sim (A; red) and anti-Tgo (B; green). Merge image is shown in (C). Horizontal view of the dorsal surface of the vnc is shown. (A) Sim is found in nuclei of cells that lie along the midline of the vnc (arrow). (B) Tgo is present in the cytoplasm of all vnc cells, but accumulates in nuclei of only the midline cells (arrow). (C) Merge image shows overlap of Sim and Tgo nuclear staining in the midline cells (arrow). (D-F) Dissected optic lobe of larval brain double-stained with anti-Sim (D) and anti-Tgo (E). Merge image is shown in (F). (D) Nuclear Sim staining is observed in cells of the lamina (arrows) and medulla (arrowheads). (E) Tgo is present in the nuclei of the lamina (arrows) and medulla (arrowheads) and present in the cytoplasm elsewhere. (F) The merge image shows that nuclear Sim and Tgo completely overlap. (G) Optic lobe from a PCNA-GFP larva stained with anti-Sim (red) and visualized for GFP (green). The GFP is expressed in the proliferation centers of the optic lobe, including outer proliferation center (O), lamina precursor center (L), and inner proliferation center (I). Sim protein is shown in the lamina (arrow) and medulla (arrowhead). There is no overlap between *GFP* and *sim* expression, indicating that *sim* is not expressed in proliferating cells of the optic lobe. (H) Brain from a *repo-lacZ* third instar larva double-stained with anti-Sim (green) and anti-β-gal (red). The *repo-lacZ* enhancer trap line expresses *lacZ* in glia. Sim is observed in the lamina (unfilled arrows) and three paired regions in the central brain complex (arrowheads). The two most anterior clusters (filled and unfilled arrowheads) stain more intensely than the most posterior clusters (filled arrow). The Sim-positive cells in the central complex are nonglial, since they do not overlap with *repo-lacZ*-expressing glial cells. (I) Higher magnification of the Sim-positive central brain clusters showing the three paired clusters. (J–L) Central brain Sim-positive cells are neuronal. Larval brain was

double-stained with anti-Sim (J; green) and anti-ELAV (K; red). Merge is shown in (L). ELAV is specifically expressed in postmitotic neurons. The Sim-positive cells are also ELAV-positive.

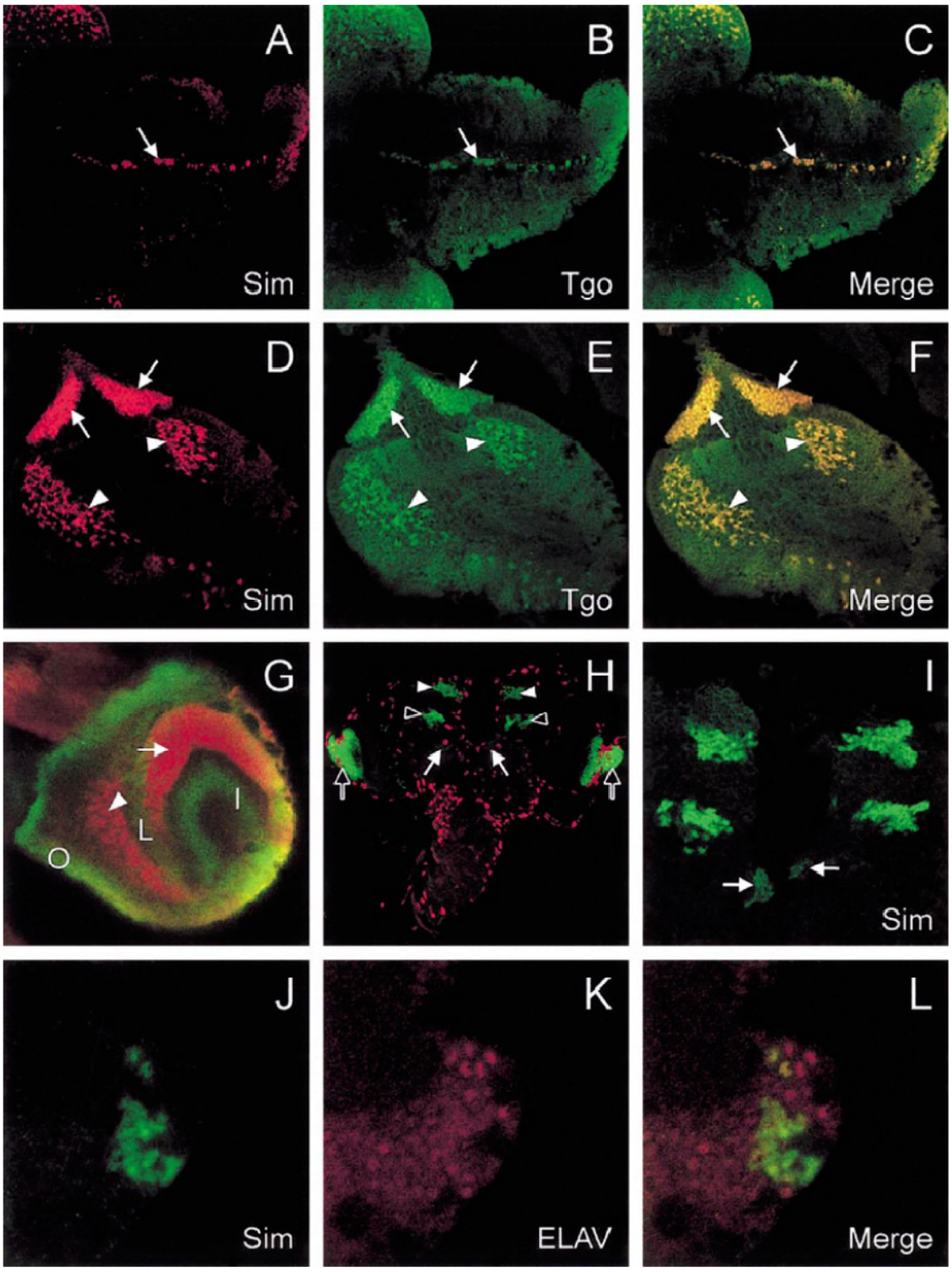
A
Sim
B
Tgo
C
Merge
D
Sim
E
Tgo
F
Merge
G
I
L
O
H
I
Sim
J
Sim
K
ELAV
L
Merge

Figure 10

REFERENCES

Ashburner, M. (1989). *Drosophila: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Awad, T. A. and Truman, J. W. (1997). Postembryonic Development of the Midline Glia in the CNS of *Drosophila*: Proliferation, Programmed Cell Death, and Endocrine Regulation. *Dev Biol* **187,** 283-97.

Bashaw, G. J. and Goodman, C. S. (1999). Chimeric Axon Guidance Receptors: The Cytoplasmic Domains of Slit and Netrin Receptors Specify Attraction Versus Repulsion. *Cell* **97,** 917-26.

Boquet, I., Hitier, R., Dumas, M., Chaminade, M. and Preat, T. (2000). Central Brain Postembryonic Development in *Drosophila*: Implication of Genes Expressed at the Interhemispheric Junction. *J Neurobiol* **42,** 33-48.

Bossing, T. and Technau, G. M. (1994). The Fate of the CNS Midline Progenitors in *Drosophila* as Revealed by a New Method for Single Cell Labelling. *Development* **120,** 1895-1906.

Brose, K. and Tessier-Lavigne, M. (2000). Slit Proteins: Key Regulators of Axon Guidance, Axonal Branching, and Cell Migration. *Curr Opin Neurobiol* **10,** 95-102.

Brunner, A., Wolf, R., Pflugfelder, G. O., Poeck, B. and Heisenberg, M. (1992). Mutations in the Proximal Region of the Optomotor-Blind Locus of *Drosophila Melanogaster* reveal a Gradient of Neuroanatomical and Behavioral Phenotypes. *J Neurogenet* **8,** 43-55.

Callaerts, P., Leng, S., Clements, J., Benassayag, C., Cribbs, D., Kang, Y. Y., Walldorf, U., Fischbach, K. F. and Strauss, R. (2001). *Drosophila Pax-6/eyeless* is Essential for Normal Adult Brain Structure and Function. *J Neurobiol* **46,** 73-88.

Campos-Ortega, J. A. and Hartenstein, V. (1997). *The Embryonic Development of Drosophila Melanogaster*. Berlin, Heidelberg: Springer Verlag.

Chang, J., Kim, I. O., Ahn, J. S., Kwon, J. S., Jeon, S. H. and Kim, S. H. (2000). The CNS Midline Cells Coordinate Proper Cell Cycle Progression and Identity Determination of the *Drosophila* ventral Neuroectoderm. *Dev Biol* **227,** 307-23.

Crews, S. T. (1998). Control of Cell Lineage-Specific Development and Transcription by bHLH-PAS Proteins. *Genes Dev.* **12,** 607-620.

Crews, S. T., Thomas, J. B. and Goodman, C. S. (1988). The *Drosophila Single-Minded* Gene Encodes a Nuclear Protein with Sequence Similarity to the *Per* Gene Product. *Cell* **52,** 143-151.

Eberl, D. F., Duyk, G. M. and Perrimon, N. (1997). A Genetic Screen for Mutations that Disrupt an Auditory Response in *Drosophila Melanogaster*. *Proc Natl Acad Sci U S A* **94,** 14837-42.

Emmons, R. B., Duncan, D., Estes, P. A., Kiefel, P., Mosher, J. T., Sonnenfeld, M., Ward, M. P., Duncan, I. and Crews, S. T. (1999). The Spineless-Aristapedia and Tango bHLH-PAS Proteins Interact to Control Antennal and Tarsal Development in *Drosophila*. *Development* **126,** 3937-3945.

Escherich, K. (1902). Zur Entwicklung Des Nervensystems Der Musciden, Mit Besonderer Berucksicktigung Des Sog Mittelstranges. *Zeitschrift fur Wissensch. Zoologie* **LXXI,** 525-549.

Estes, P., Mosher, J. and Crews, S. T. (2001). *Drosophila Single-Minded* Represses Gene Transcription by Activating the Expression of Repressive Factors. *Dev. Biol.* **232,** 157-175.

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.

Foe, V. E. (1989). Mitotic Domains Reveal Early Commitment of Cells in *Drosophila* Embryos. *Development* **107,** 1-22.

Gorfinkiel, N., Sanchez, L. and Guerrero, I. (1999). *Drosophila* terminalia as an Appendage-Like Structure. *Mech Dev* **86,** 113-23.

Gotz, K. G. (1980). Visual Guidance in *Drosophila*. *Basic Life Sci* **16,** 391-407.

Hanesch, U., Fischbach, K. -. and Heisenberg, M. (1989). Neuronal Architecture of the Central Complex in *Drosophila Melanogaster*. *Cell and Tissue Research* **257,** 343-366.

Hartenstein, V. and Jan, Y. N. (1992). Studying *Drosophila* embryogenesis with P-lacZ Enhancer Trap Lines. *Development Genes and Evolution* **201,** 194-220.

Hayashi, S., Hirose, S., Metcalfe, T. and Shirras, A. D. (1993). Control of Imaginal Cell Development by the *Escargot* gene of *Drosophila*. *Development* **118,** 105-15.

Heisenberg, M. and Bohl, K. (1979). Isolation of Anatomical Brain Mutants of *Drosophila* by Histological Means. *Z Naturforsch* **34,** 143-147.

Hitier, R., Chaminade, M. and Preat, T. (2001). The *Drosophila Castor* gene is Involved in Postembryonic Brain Development. *Mech Dev* **103,** 3-11.

Hofbauer, A. and Campos-Ortega, J. A. (1990). Proliferation Pattern and Early Differentiation of the Optic Lobes in *Drosophila Melanogaster*. *Roux's Arch. Dev. Biol.* **198,** 264-274.

Hummel, T., Schimmelpfeng, K. and Klambt, C. (1997). Fast and Efficient Egg Collection and Antibody Staining from Large Numbers of *Drosophila* strains. *Development Genes and Evolution* **207,** 131-135.

Hummel, T., Schimmelpfeng, K. and Klambt, C. (1999). Commissure Formation in the Embryonic CNS of *Drosophila* I. Identification of the Required Gene Functions. *Dev. Biol.* **209,** 381-398.

Jacobs, J. R. (2000). The Midline Glia of *Drosophila*: A Molecular Genetic Model for the Developmental Functions of Glia. *Prog. Neurobiol.* **62,** 475-508.

Jurgens, G. and Hartenstein, V. (1993). The Terminal Regions of the Body Pattern. In *The Development of Drosophila Melanogaster* (ed. M. Bate and A. Martinez-Arias), pp. 687-746. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Kidd, T., Bland, K. S. and Goodman, C. S. (1999). Slit is the Midline Repellent for the Robo Receptor in *Drosophila*. *Cell* **96,** 785-94.

Kinrade, E. F., Brates, T., Tear, G. and Hidalgo, A. (2001). Roundabout Signalling, Cell Contact and Trophic Support Confine Longitudinal Glia and Axons in the *Drosophila* CNS. *Development* **128,** 207-16.

Klämbt, C., Jacobs, J. R. and Goodman, C. S. (1991). The Midline of the *Drosophila* central Nervous System: A Model for the Genetic Analysis of Cell Fate, Cell Migration, and Growth Cone Guidance. *Cell* **64,** 801-815.

Kramer, S. G., Kidd, T., Simpson, J. H. and Goodman, C. S. (2001). Switching Repulsion to Attraction: Changing Responses to Slit during Transition in Mesoderm Migration. *Science* **292,** 737-40.

Lewis, J. O. and Crews, S. T. (1994). Genetic Analysis of the *Drosophila Single-Minded* Gene Reveals a Central Nervous System Influence on Muscle Development. *Mech. Dev.* **48,** 81-91.

Luer, K., Urban, J., Klambt, C. and Technau, G. M. (1997). Induction of Identified Mesodermal Cells by CNS Midline Progenitors in *Drosophila*. *Development* **124,** 2681-2690.

Ma, Y., Certel, K., Gao, Y., Niemitz, E., Mosher, J., Mukherjee, A., Mutsuddi, M., Huseinovic, N., Crews, S. T., Johnson, W. A. et al. (2000). Functional Interactions between *Drosophila* bHLH/PAS, Sox, and POU Transcription Factors Regulate CNS Midline Expression of the *Slit* gene. *J. Neurosci.* **20,** 4596-605.

Mayer, U. and Nusslein-Volhard, C. (1988). A Group of Genes Required for Pattern Formation in the Ventral Ectoderm of the *Drosophila* embryo. *Genes Dev.* **2,** 1496-1511.

Menne, T. V. and Klämbt, C. (1994). The Formation of Commissures in the *Drosophila* CNS Depends on the Midline Cells and on the *Notch* gene. *Development* **120,** 123-133.

Menne, T. V., Luer, K., Technau, G. M. and Klambt, C. (1997). CNS Midline Cells in *Drosophila* Induce the Differentiation of Lateral Neural Cells. *Development* **124,** 4949-4958.

Mitchell, K. J., Doyle, J. L., Serafini, T., Kennedy, T., Tessier-lavigne, M., Goodman, C. S. and Dickson, B. J. (1996). Genetic Analysis of *Netrin* Genes in *Drosophila*: Netrins Control Guidance of Commissural Axons and Peripheral Motor Axons. *Neuron* **17,** 203-215.

Monge, I., Krishnamurthy, R., Sims, D., Hirth, F., Spengler, M., Kammermeier, L., Reichert, H. and Mitchell, P. J. (2001). *Drosophila* transcription Factor AP-2 in Proboscis, Leg and Brain Central Complex Development. *Development* **128,** 1239-52.

Morel, V. and Schweisguth, F. (2000). Repression by *Suppressor of Hairless* and Activation by *Notch* are Required to Define a Single Row of *Single-Minded* Expressing Cells in the *Drosophila* embryo. *Genes Dev.* **14,** 377-88.

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.

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, K. A.,Jr 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.

Ohshiro, T. and Saigo, K. (1997). Transcriptional Regulation of *Breathless* FGF Receptor Gene by Binding of TRACHEALESS/dARNT Heterodimers to Three Central Midline Elements in *Drosophila* Developing Trachea. *Development* **124,** 3975-3986.

Patel, N. H., Snow, P. M. and Goodman, C. S. (1987). Characterization and Cloning of *Fasciclin III*: A Glycoprotein Expressed on a Subset of Neurons and Axon Pathways in *Drosophila*. *Cell* **48,** 975-988.

Poulson, D. F. (1950). Histogenesis, Organogenesis, Differentiation in the Embryo of *Drosophila Melanogaster*. In *Biology of Drosophila* (ed. M. Demerec), pp. 168-274. New York: Wiley.

Rothberg, J. M., Hartley, D. A., Walther, Z. and Artavanis-Tsakonas, S. (1988). *Slit*: An EGF-Homologous Locus of *D. Melanogaster* Involved in the Development of the Embryonic Central Nervous System. *Cell* **55,** 1047-1059.

Rothberg, J. M., Jacobs, J. R., Goodman, C. S. and Artavanis-Tsakonas, S. (1990). Slit: An Extracellular Protein Necessary for Development of Midline Glia and Commissural Axon Pathways Contains both EGF and LRR Domains. *Genes & Dev.* **4,** 2169-2187.

Sanchez, L. and Guerrero, I. (2001). The Development of the *Drosophila* genital Disc. *Bioessays* **23,** 698-707.

Sanchez-Soriano, N. and Russell, S. (1998). The *Drosophila* SOX-Domain Protein Dichaete is Required for the Development of the Central Nervous System Midline. *Development* **125,** 3989-3996.

Schuster, S. and Gotz, K. G. (1994). Adaptation of Area Covering Random Walk in *Drosophila*. In *In Sensory Transduction* (ed. N. Elsner and H. Breer), pp. 304. Gottingen: Thieme.

Selleck, S. B. and Steller, H. (1991). The Influence of Retinal Innervation on Neurogenesis in the First Optic Ganglion of *Drosophila*. *Neuron* **6,** 83-99.

Simon, A. F., Boquet, I., Synguelakis, M. and Preat, T. (1998). The *Drosophila* putative Kinase Linotte (Derailed) Prevents Central Brain Axons from Converging on a Newly Described Interhemispheric Ring. *Mech Dev* **76,** 45-55.

Sonnenfeld, M., Ward, M., Nystrom, G., Mosher, J., Stahl, S. and Crews, S. (1997). The *Drosophila Tango* Gene Encodes a bHLH-PAS Protein that is Orthologous to Mammalian *Arnt* and Controls CNS Midline and Tracheal Development. *Development* **124,** 4583-4594.

Sonnenfeld, M. J. and Jacobs, J. R. (1994). Mesectodermal Cell Fate Analysis in *Drosophila* Midline Mutants. *Mech. Dev.* **46,** 3-13.

Stollewerk, A., Klambt, C. and Cantera, R. (1996). Electron Microscopic Analysis of *Drosophila* midline Glia during Embryogenesis and Larval Development using Beta-Galactosidase Expression as Endogenous Cell Marker. *Microsc Res Tech* **35,** 294-306.

Strauss, R. (1998). Automatische Diagnose Genetisch Bedingter Laufanomalien Der Fliege *Drosophila* bei Freier Bewegung in Realer Un Virtueller Umgebung. In *In Forschung Und Wissenschaftliches Rechnen. GWDG-Bericht Nr. 51* (ed. Plesser, T., and Wittenburg,P.), pp. 53-78. Gottingen: Gesellschaft fur wissenschaftliche Datenverarbveitung mbH Gottingen.

Strauss, R. and Heisenberg, M. (1993). A Higher Control Center of Locomotor Behavior in the *Drosophila* brain. *J Neurosci* **13,** 1852-61.

Strauss, R. and Pichler, J. (1998). Persistence of Orientation Toward a Temporarily Invisible Landmark in *Drosophila Melanogaster*. *J Comp Physiol [A]* **182,** 411-23.

Strauss, R., Schuster, S. and Gotz, K. G. (1997). Processing of Artificial Visual Feedback in the Walking Fruit Fly *Drosophila Melanogaster*. *J Exp Biol* **200,** 1281-96.

Strauss, R. and Trinath, T. (1996). Is Walking in a Straight Line Controlled by the Central Complex? Evidence from a New *Drosophila* mutant. In *Gottingen Neurobiology Report* (ed. N. Elsner and H. -. Schnitzler). Gottingen: Thieme.

Therianos, S., Leuzinger, S., Hirth, F., Goodman, C. S. and Reichert, H. (1995). Embryonic Development of the *Drosophila* brain: Formation of Commissural and Descending Pathways. *Development* **121,** 3849-60.

Thomas, J. B., Crews, S. T. and Goodman, C. S. (1988). Molecular Genetics of the *Single-Minded* Locus: A Gene Involved in the Development of the *Drosophila* nervous System. *Cell* **52,** 133-141.

Ward, M. P., Mosher, J. T. and Crews, S. T. (1998). Regulation of *Drosophila* bHLH-PAS Protein Cellular Localization during Embryogenesis. *Development* **125,** 1599-1608.

Zhou, L., Xiao, H. and Nambu, J. R. (1997). CNS Midline to Mesoderm Signaling in *Drosophila*. *Mech. Dev.* **67,** 59-68.

CHAPTER 3:

THE ROLE OF *DROSOPHILA SINGLE-MINDED* IN CONTROLLING BRAIN INTERHEMISPHERIC CONNECTIVITY

PREFACE

Kristin Benjamin, an undergraduate student in the Crews group, sequenced *sim* alleles. The author of this dissertation performed all other work presented in this chapter.

ABSTRACT

The function and evolution of complex behaviors requires the construction of sophisticated neural circuits. Neurons of the *Drosophila* larval central brain express the *single-minded* gene. Previous analysis of a temperature-sensitive *single-minded* allele revealed defects in controlling locomotory behavior, potentially explained by defects in brain interhemispheric communication. Here I analyzed *sim* expression and function in the larval brain. Both immunostaining experiments and visualization of GFP^+ clones indicate that Single-minded$^+$ neurons of the central brain extend axons across the midline and fasciculate. Null mutants of *single-minded* were identified, and the MARCM technique was used to analyze brain defects. The results indicated that *single-minded* mutants do not exhibit aberrant neurogenesis, which is its

major role during embryonic CNS development. Instead, *single-minded* mutants display fasciculation defects, in which the axon bundles split into two tracts as they cross the midline. This likely results in incorrect axon connectivity between the two brain hemispheres, and provides a possible explanation for the *single-minded* behavioral defects. Relevant to the evolution of the CNS, both *Drosophila* and mammalian *single-minded* play roles in both neurogenesis and axonogenesis.

INTRODUCTION

The formation of functional central nervous system (CNS) neural circuits consists of a series of events beginning with neurogenesis, followed by axonogenesis, synaptic connectivity, and remodeling of the juvenile brain into its adult form. These circuits underlie the complex behaviors found throughout the animal kingdom. Key to CNS development is the action of transcriptional regulatory proteins. Only about 700 of the close to 14,000 genes in the *Drosophila* genome encode DNA sequence-specific transcription regulatory proteins (Adams et al., 2000). However, the roles of these proteins are wide-ranging; they are often used multiple times during development to regulate different sets of genes and developmental processes. Thus, understanding how transcriptional regulation controls neurodevelopment will ultimately provide insight into the evolutionary basis for species differences in neural circuitry and behavior.

The *Drosophila single-minded* (*sim*) and mammalian *Single-minded 1* and *2* genes provide good examples of how transcription factors perform different functions throughout development. *Drosophila sim* encodes a basic-helix-loop-helix-PAS domain (bHLH-PAS) protein that forms a DNA binding heterodimer with the Tango (Tgo) bHLH-PAS protein (reviewed in Crews, 2003). During embryogenesis, cells that lie along the midline of the *Drosophila* CNS prominently express

sim. These cells include ~22 mature midline neurons and glia (Wheeler et al., 2006). The *sim* gene acts as a master regulator of CNS midline cell development. In *sim* mutants, the midline cells fail to form and all midline-specific transcription is absent (Nambu et al., 1990). Inducing ectopic expression of *sim* throughout the neuroectoderm prior to embryonic stage 10 results in the transformation of lateral CNS cells into only midline cells (Nambu et al., 1991). These lateral CNS cells abnormally express *rhomboid*, *slit*, and *center divider*, genes that are prominently expressed in midline cells during early neurogenesis. These cells adopt morphology more characteristic of midline cells, with extended nuclei and cytoplasmic projections, in lieu of the usual appearance of lateral neuroblasts and neurons.

In the larval brain, cells in the lamina and medulla of the optic lobes express *sim*, and the gene plays a role in the differentiation of laminar precursor cells into mature neurons (Umetsu et al., 2006). Mammals have two *sim* genes, *Sim1* and *Sim2* (Chen et al., 1995; Chrast et al., 1997; Dahmane et al., 1995; Ema et al., 1996a; Ema et al., 1996b). Cells in the developing hypothalamus, including the paraventricular nucleus (PVN), anterior periventricular nucleus (aPV), and supraoptic nucleus (SON), express *Sim1* (reviewed in Fan, 2003; Fan et al., 1996). Cells of the aPV and a subset of PVN cells also express *Sim2*. Genetic analysis of *Sim1* homozygous mutant mice revealed an absence of the PVN and SON, and implicated *Sim1* in controlling the terminal differentiation or migration of PVN and SON anterior neuroendocrine hypothalamus (Michaud et al., 1998). Further investigation showed *Sim1* mutant cells are born normally in the PVN/SON progenitor region and appear to maintain their PVN/SON progenitor fate, but fail to migrate to their normal positions and are likely arrested at a stage before hormone gene expression (Xu and Fan, 2007). Thus, both *Drosophila* and mammalian *sim* genes play important roles in generating functional neurons.

Additional roles of both *Drosophila* and mammalian *sim* have emerged. *Drosophila sim* is expressed in the larval brain in 3 paired clusters of cells in the central brain, a region implicated in coordinating movement in the adult (Pielage et al., 2002). Behavioral analysis of individuals harboring a temperature-sensitive *sim* allele revealed that when shifted to the non-permissive temperature after embryonic neurogenesis was complete, adult flies showed locomotory defects (Pielage et al., 2002). Thus, when tested in a paradigm designed to measure coordination, mutant flies walked in circles rather than the straight lines characteristic of wild type flies. Morphological analysis of the adult brain indicated a disorganization of the central complex (CX) neuropil. These results suggested a defect in interhemispheric communication, and a subsequent inability to coordinate movement. Recent work on murine *Sim1* and *Sim2* revealed that they play a role in controlling axonogenesis of mammillary body axons (Marion et al., 2005). The mammillary body is long known to be important in the processing of recognition memory (reviewed in Vann and Aggleton, 2004). Thus, results from both mammals and *Drosophila* indicate that *sim* functions at multiple times during CNS development and participates in processes including cell fate specification, terminal differentiation, and axonogenesis.

The disorganized neuropil in the adult brain of *sim* mutants suggested that developmental defects might underlie the *sim* walking defect. However, a mechanistic understanding of how loss of *sim* leads to the altered behavior was unclear. Does *sim* control neurogenesis of the central brain, similar to its role in embryonic development, or does it regulate another process, such as axonogenesis? Determining the Sim^+ central brain cell axon trajectories and using clonal analysis of multiple *sim* null mutant strains to assay neurogenesis and axonogenesis directly addressed this issue. We found that some Sim^+ cells normally extend their axons across the supraesophageal commissure connecting the right and left brain hemispheres, and they fasciculate. Mutant analysis indicated that *sim* does not play a role in central brain

neurogenesis, but is required for correct axonogenesis. In *sim* mutant clones, axons defasciculated and formed multiple tracts. This likely leads to defects in interhemispheric communication and locomotory coordination. These results also revealed further similarities in the function of *Drosophila* and mammalian *sim* genes, since both contribute to axonogenesis as well as neurogenesis.

MATERIALS AND METHODS

Drosophila *stocks*

The wild type strain was w^{1118}. Twelve *sim* alleles were analyzed from 5 primary sources: sim^{1}, sim^{2}, sim^{5}, sim^{6}, sim^{7} (Hilliker et al., 1980); sim^{8} (Mayer and Nusslein-Volhard, 1988), sim^{BB68}, sim^{JJ22}, sim^{M55}, sim^{TT63}, sim^{W3} (J. Skeath and C. Doe; unpublished), sim^{J1-47} (Pielage et al., 2002) and sim^{K} (C. Klämbt, unpublished). All strains were tested for noncomplementation with the sim^{2} null mutant at 25°C. Mutant stocks were balanced over *TM3 Sb P[w+; Krüppel-Gal4] P[w+; UAS-GFP]* (Casso et al., 2000), or *TM3 Sb P[ry+; ftz-lacZ]*. The marked balancer chromosomes allowed identification of homozygous mutant embryos. The putative amorphic alleles sim^{2}, sim^{8}, and sim^{BB68} were recombined onto a chromosome bearing FRT^{82B} for use in MARCM, in combination with *w+; tub-Gal4* FRT^{82B} *tub-Gal80/TM3* (J. Treisman) and *w+, hsFlp, elav-Gal4, UAS-mCD8::GFP* (Bloomington *Drosophila* Stock Center).

Immunostaining

Embryos were collected, fixed, and stained using standard procedures (Patel et al., 1987). Larval brains were dissected in 1X PT buffer (1X phosphate buffered saline solution containing 0.1% Triton X-100), fixed in 4% methanol-free formaldehyde (Ted Pella) in 1X PEM buffer (0.1 M PIPES, 2 mM EGTA, 1 mM $MgSO_4$), and stained using the following antibodies and reagents: rat anti-Sim, murine mAb anti-Tgo, rabbit anti-β-galactosidase (Cappel), murine mAb 9F8A9 anti-Elav (Developmental Studies Hybridoma Bank), rabbit anti-GFP (Abcam), Cy3-, Cy5-, Alexa350-, Alexa568-conjugated secondary antibodies (Invitrogen), HRP-conjugated secondary antibodies (Jackson ImmunoResearch), and 1.0 mg/ml DAPI (Sigma) used at a 1:1000 dilution. Chromogenically stained specimens were mounted in 70% glycerol and imaged on a Zeiss Axiophot microscope, and fluorescently stained specimens were mounted in Aqua Poly/Mount (Polysciences) and imaged on Zeiss LSM510 confocal microscope.

MARCM

w, hsFlp, elav-Gal4, UAS-mCD8::GFP; tub-Gal4, FRT^{82B}*, tub-Gal80/+* males were crossed to virgin females bearing either FRT^{82B} sim^{2}; FRT^{82B} sim^{8}; FRT^{82B} sim^{BB68}; or FRT^{82B} $P[w^{+}]$ chromosomes (Lee and Luo, 1999). Embryos were collected for 3 h and aged for 24 h at 25°C, then heat shocked for 1.5 h at 37°C, followed by aging at 25°C until they became wandering third instar larvae. Brains were isolated, fixed, stained, and analyzed as described above.

Sequencing of sim *mutant DNA*

Genomic DNA was isolated from stage 14-15 homozygous *sim* mutant embryos that were identified based on the absence of balancer chromosome *GFP* expression. Sequencing of the *sim* gene was performed using DNA fragments isolated by touchdown PCR (tdPCR) (Don et al., 1991; Hecker and Roux, 1996). Seven sets of primer pairs were used to amplify exons 2-8, which comprise the complete coding sequence and corresponding splice sites (Table 1). The 50 µl reaction mix consisted of 2 µl Taq DNA polymerase (Invitrogen), 1X PCR buffer (Invitrogen) with 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 µM primer, and 5 µl template DNA. The tdPCR used an annealing temperature (T_a) 5°C more stringent than the normal T_a (~5°C below the melting temperature). For ten cycles, the T_a was decreased by 0.5°C/cycle. The reaction then proceeded at the normal T_a for 25 additional cycles. The denaturation step was at 94°C for 45 sec (3 minutes at the beginning of the reaction), annealing for 45 sec at a specific temperature for each primer pair, and elongation for 1.5 min at 72°C (7 min at the end of the 35 cycle reaction). PCR products were purified and sequenced at the UNC-CH Genome Analysis Facility.

RESULTS

Interhemispheric crossing and fasciculation of Sim$^+$ central brain cell axons

The Sim protein is localized to nuclei in 3 paired clusters (dorsal, medial, and ventral) of central brain cells in the third instar larval CNS (Fig. 1A and 1B). This suggests that each Sim^+ cell cluster is composed of neurons either born from a single neuroblast or from multiple adjacent neuroblasts, and whose cells share a common lineage (Dumstrei et al., 2003; Pereanu

and Hartenstein, 2006; Truman et al., 2004; Younossi-Hartenstein et al., 2006). These clusters putatively correspond with previously identified *Drosophila* neural lineages as follows: DAMv1, DAMv2 (dorsal clusters); BAmas1, BAmas2 (medial clusters); and TRdm (ventral clusters) (Pereanu and Hartenstein, 2006). Double staining larval brains with anti-Sim and monoclonal antibody (mAb) BP106, which reacts with Neurotactin and stains neurons and axons, allows visualization of the axons emanating from each Sim^+ cluster. The dorsal cluster consists of 2 adjacent subgroups of cells, indicating these neurons are born from 2 neighboring neuroblast lineages. The cells comprising each subgroup extend their axons in a discrete bundle (Fig. 1C). Both fascicles then merge to comprise a single tract that extends anteriorly (with respect to the coordinates of the neuraxis) to the supraesophageal commissure, crosses the midplane, and fasciculates with the Sim^+ axons from the contralateral side (Fig. 1E and 1F). Using MARCM (Lee and Luo, 1999) to label clones of Sim^+ brain cells and visualize their axon trajectories by green fluorescent protein (GFP) fluorescence, confirmed these results. In one example (Fig. 1G and 1H), simultaneous GFP^+ clones were induced in each dorsal Sim^+ cluster from the 2 brain hemispheres. The axons from the 2 clusters followed similar trajectories and fasciculated. Additionally, GFP^+ MARCM clones were observed to encompass approximately half the Sim^+ cells within the dorsal cluster in each instance. Taken together, these data indicate that each dorsal Sim^+ cluster is composed of two neighboring neuroblast lineages, and each cluster exhibits bilateral symmetry.

The medial cluster also consists of 2 subgroups of neurons (Fig. 1D). Each subgroup, as visualized by BP106 staining, produces an axon bundle; the 2 bundles join and project centro-dorsally towards the supraesophageal commissure (data not shown). However, it is difficult to ascertain from the BP106 staining whether their axons cross the midline. Analysis of wild type MARCM clones in third instar larval brains showed that the axons remained on the ipsilateral

side at the same time that the dorsal cluster axons had crossed the midline (Fig. 1I and 1J). Thus, in the brains examined, axons from the medial Sim^+ clusters did not cross the midline or fasciculate as a conjugate pair, although it is possible that this occurs later in development.

The ventral cluster of Sim^+ cells consists of a single group of cells and extends a singular axon tract which projects ipsilaterally near the ventral side of the esophageal opening (Fig. 1K and 1L).

Developmental progression of Sim^+ medial brain expression

Brain expression of *sim* can be seen in late stage 17 embryos in 6-14 cells that are adjacent to the esophagus ($\bar{x}$=11, n=4, SD = 3.4) (Fig. 2A). The cells can be observed in dissected brains of first and second instar larvae as discrete cell clusters, but the increase in cell numbers is low during these stages (Fig. 2B and 2C). The dorsal cluster pair contains about 8 cells/cluster (n=10, SD=1.4), the medial pair 7 (n=10, SD=1.3), and the ventral pair 3 (n=7, SD=0.5) in the first instar larval CNS. The second instar larval CNS contains 12 cells/cluster in the dorsal pair (n=10, SD=2.1), 8 in the medial pair (n=10, SD=2.4), and 4 in the ventral pair (n=9, SD=0.7). However, there is considerable expansion of Sim^+ cells during third instar larval development, such that by the end of larval development, the clusters have grown in size and consist of an average of 85 cells in each dorsal cluster (n=10, SD=12) and 62 cells in each medial cluster (n=10, SD=13) (Fig. 2D). The ventral clusters each contain 38 cells, on average (n=8, SD=14). There are additional Sim^+ cells in the third instar larval brain but these cells are located distal from the central brain region, site of the presumptive CX in adults.

sim *is expressed in brain neurons but not neuroblasts*

Both CNS midline precursor cells, including the median neuroblast, and their neuronal and glial progeny, express *sim* in the embryo (Crews et al., 1988; Thomas et al., 1988). Previously, we showed that most, if not all, Sim^+ cells were also Embryonic lethal, abnormal vision$^+$ ($Elav^+$). Elav immunostaining marks post-mitotic neurons in larvae (Robinow and White, 1991), although recent evidence has shown transient localization of Elav protein in early born glial cells of the embryo and detection of *elav* transcripts in mitotically active embryonic neuroblasts (Berger et al., 2007). Unresolved was whether neuroblasts and ganglion mother cells also expressed *sim*. This was addressed by using MARCM coupled with Sim and Elav immunostaining to visualize neuroblasts associated with Sim^+ brain cells. In Fig. 3, a GFP^+ MARCM clone that overlaps with Sim^+ cells shows strong nuclear $Elav^+$ localization in neurons. However, the large GFP^+ neuroblast and 2-3 adjacent cells, likely to be recently born, weakly $Elav^+$ ganglion mother cells (GMCs), do not express *sim*. Thus, in the third instar larval brain, *sim* expression is associated with post-mitotic neurons and not neural precursor cells, unlike its expression profile in both precursor and progeny cells in the embryonic CNS midline cells. This argues against a role for *sim* in brain neurogenesis, and is more consistent with a role in axonogenesis and differentiated neuronal properties.

sim *mutations affect axon morphology but not neurogenesis*

The role of *sim* in larval central brain development was assayed by examining *sim* mutant MARCM clones that overlap with the Sim^+ cell clusters. To validate any phenotypes observed, MARCM was conducted using multiple *sim* null mutant alleles. Since *sim* mutant strains (with the

exception of *sim*$^{J1\text{-}47}$) had not previously been sequenced and their corresponding molecular defects identified, we sequenced all 7 coding sequence exons of 12 additional *sim* alleles (Table 1). In total, 11 were EMS-induced, 1 was X-ray induced, and 1 allele was derived using an unknown mutagen. All 13 mutants showed a sequence alteration in either the coding sequence or splice site that could lead to loss of *sim* function (Fig. 4A, Table 1). Ten mutants have premature stop codons and 3 have missense mutations in critical protein domains. The *sim*^{7} missense mutation changes a conserved amino acid (aa) in the basic region, and may affect the ability of Sim to bind DNA. The *sim*$^{J1\text{-}47}$ strain, as noted previously in Pielage et al., 2002, has a mutation in helix 2 and likely affects the ability of Sim to dimerize with Tgo, which is required for DNA binding. The *sim*^{K} strain has a mutation in PAS-2 that alters a residue conserved in most Sim proteins. Three mutants predicted to produce truncated proteins were selected for MARCM, since they are likely to be amorphic. *sim*^{8} is predicted to produce a protein only 12 aa long, and *sim*^{BB68} and *sim*^{2} are both predicted to generate proteins of 290 and 291 aa, respectively, less than half the size of full-length Sim protein (673 aa) (Nambu et al., 1991). Previous work has shown that a truncated Sim protein less than 462 aa, and lacking known activation domains (Franks and Crews, 1994), was unable to activate midline cell transcription in vivo (Estes et al., 2001). Mutant *sim* embryos from the 3 selected strains were stained with anti-Sim raised against a bacterially synthesized protein fragment that is predicted to be lacking in all 3 mutant proteins (Fig. 4A). No Sim immunoreactivity was observed in homozygous mutant embryos (Fig. 4B), as predicted. After each strain was recombined with *FRT*^{82B}, the *sim* gene was resequenced to confirm that the appropriate mutation was present.

We directed our focus of *sim* genetic analysis on the development of the dorsal cluster of Sim^{+} cells, since their axon crossing and fasciculation is most consistent with a defect in interhemispheric coordination and with the adult behavioral and morphological phenotypes. As

a control we generated FRT^{82B} lines that did not harbor a mutant *sim* allele. All wild type clones contained only Sim^+ neurons with the exception of the neuroblast and ~3 nascent cells next to the neuroblast. This indicated that all cells have the potential to be phenotypically mutant (Fig. 3E, 5A, 5C). All wild type and *sim* mutant clones were comprised of ~50% of the cells that normally constitute a dorsal Sim^+ cluster (Fig. 5). Clones that were mutant for *sim* showed wild type numbers of neurons ($\bar{x}$=36, n=8, SD=4.0 in wild type; $\bar{x}$=41.4, n=10, SD=5.2 in sim^2, sim^8, and sim^{BB68} mutants) and contained a neuroblast and associated GMCs (Fig. 3F-3J). The neurons were $Elav^+$ (Fig. 3G and 3J), and were able to extend axons (Fig. 5E-5P). Similar results were observed for *sim* mutant medial cluster cells (data not shown). These results indicate that neurogenesis occurs normally in *sim* mutant central brain cells.

All wild type MARCM clones within the dorsal Sim^+ cell cluster (n=8) extended axons with a similar morphology (Fig. 1G, 1H, 5A-5D). The GFP^+ axons leave the soma dorsally as a common fascicle, elaborate filopodia ipsilaterally (Pereanu and Hartenstein, 2006), and make a sharp turn towards the contralateral side via the supraesophageal commissure. In the supraesophageal commissure, contralateral Sim^+ axons form a common tract. MARCM *sim* mutant clones often showed axon defects (Fig. 5E-5N). Of 12 MARCM clones in the Sim^+ dorsal cluster, 7 showed a clear mutant phenotype, and 5 appeared wild type. All 3 alleles had at least one clone with a mutant phenotype and all alleles showed a similar defect. In general, the axons of *sim* mutant clones extended a tract centro-dorsally, formed an ipsilateral filopodial protuberance, then turned towards the supraesophageal commissure, resembling wild type axons. However, a subset of mutant axons aberrantly defasciculated from the main bundle and continued across the midline. While the phenotypes differed in individual clones, they consistently revealed fasciculation defects. Two of 3 sim^{BB68} clones examined showed fasciculation defects and showed a complete lack of Sim immunoreactivity (Fig. 5E-5H). Four of

5 *sim*^{2} clones had fasciculation defects, and Sim immunoreactivity was greatly reduced but not completely absent (Fig. 5I-5L). One of 4 *sim*^{8} clones showed a fasciculation defect, and there was some Sim immunoreactivity in mutant clones, although levels were strongly reduced (Fig. 5M and 5N). In summary, mutant clones of all 3 *sim* alleles showed similar axon fasciculation defects, while neurogenesis was unaffected.

DISCUSSION

The insect CX has been proposed to be important in fine-tuning behavioral outputs, particularly in locomotion, although it may play additional roles, including learning and memory (reviewed in Davis, 1996). Previous analysis of the *Drosophila sim* gene provided novel insight into CX development and function (Pielage et al., 2002). The *sim* gene is expressed in 3 paired clusters of cells in the larval central brain. Genetic analysis of *sim* indicated disorganization in the adult central complex neuropil. Finally, *sim* mutant adult flies were unable to coordinate their walking behavior. These results suggested that *sim* plays a role in specification or function of axons that connect the two hemispheres, thus controlling a key aspect of interhemispheric communication required for proper motion control.

Analysis of the *sim*^{1147} temperature-sensitive allele at the non-permissive temperature previously revealed the *sim* behavioral and adult brain defects. However, it is unknown to what extent this mutation removes *sim* function. In this paper, we sequenced 12 alleles of *sim* to identify severe mutants to be used for MARCM. The 3 mutants selected, *sim*^{2}, *sim*^{8}, and *sim*^{BB68}, are likely to be severe, probably null mutants. All 3 mutants possess in-frame stop codons that should produce truncated Sim proteins. *sim*^{8} is predicted to produce a protein only 12 aa long, whereas *sim*^{2} and *sim*^{BB68} should produce proteins that terminate in the PAS-2 domain and lack

Sim activation domains. Previous work has shown that the absence of the Sim activation domains results in a protein that cannot activate transcription in vivo (Franks and Crews, 1994). Staining embryos with anti-Sim revealed that full-length Sim protein was absent (Fig. 4). Consistent with these molecular defects, all 3 mutants show similar axon defects.

In this chapter, we correlate the identities of the three paired Sim^+ clusters in the larval central brain with neuronal lineages previously identified and catalogued (Pereanu and Hartenstein, 2006). Since limited but informative sets of markers for these lineages have been identified (Sprecher et al., 2007), immunochemical assays can be conducted to determine these correlations with certainty. For example, the DAMv1 and DAMv2 (dorsal) clusters express *fork head* and *twin of eyeless*, while DAMv1 expresses *empty spiracles* and DAMv2 expresses *eyeless*; Sim should co-localize with these markers in these clusters. Similarly, the BAmas1 and BAmas2 (medial) clusters express *hedgehog*, *eyeless*, *dachshund*, and *extradenticle*, with BAmas2 additionally expressing *engrailed* and *even skipped*.

Here, we extend our analysis for the role of *sim* in interhemispheric communication. We show that the Sim^+ cells from conjugate dorsal central brain clusters extend axons across the midline via the supraesophageal commissure and fasciculate. Thus, the Sim^+ neurons are part of a direct circuit between both brain hemispheres. Genetic analysis of *sim* null mutants further demonstrates that *sim* function is required for proper axon outgrowth. The axons in over half the mutant clones showed a fasciculation defect, which likely leads to incorrect interhemispheric axon connectivity. This result is consistent with the previous observation regarding the adult brain CX axon disorganization and behavioral defects. However, there are a number of relevant points to consider. The first is that while axon guidance defects are observed for the dorsal cluster, it is unknown whether the behavioral defects are the result of this defect, since the $sim^{J1\text{-}47}$ mutation may also affect *sim* function in the medial and ventral central brain clusters, the optic

lobes, and the midline cells of the ventral nerve cord. This issue can be addressed in future experiments by targeting disruptions of *sim* function specifically to central brain cells. Another point is that since the clonal analysis of *sim* mutants targeted only one of the 2 dorsal clusters, the extent of how well the *sim* mutant fibers connected with their contralateral partners is unknown. Finally, since only about half the neurons in each cluster were mutant, the presence of genetically wild type axons mixed with *sim* mutant axons could reduce the severity of the phenotype. Nevertheless, what is clear is that *sim* does not control the neurogenesis of the Sim^+ dorsal central brain neurons; rather, it controls proper axon outgrowth.

There are multiple interpretations of the sim^{1147} behavioral phenotype. In *sim* mutants, axons from the dorsal pair of Sim^+ central brain clusters may not fasciculate properly, resulting in a lack of proper interhemispheric communication. This interpretation implies that the Sim^+ cells play a role in controlling locomotory coordination. It is also is possible that reductions in *sim* could control additional aspects of terminal differentiation and neurotransmission in addition to the axon guidance defects. This could also contribute to the behavioral phenotype. Another possible developmental role is that the Sim^+ cells themselves do not physiologically contribute to locomotory coordination, but their axons may act to pioneer the axons of other neurons. These other neurons may control movement, whereas the Sim^+ cells only contribute the developmental pioneering role. These issues can be resolved using targeted expression of various transgenes affecting *sim* function and neurotransmission.

One of the earliest morphological events in CNS circuit formation occurs when pioneer neurons lay a path with their axons, followed by the creation of primary axon tracts by the end of embryogenesis (reviewed in Hartenstein et al., 2008). Neurons born of the same lineage contribute to primary axon tracts. After a period of quiescence which starts prior to larval eclosion and lasts for the first 8 hours of the first instar larval stage (Ito and Hotta, 1992), central

brain neuroblasts give rise to GMCs and neurons whose axons project centripetally from the soma-filled cortex of the brain toward the neuropil to form the secondary axon tracts (SATs). At the glial boundary surrounding the neuropil, SATs travel a short distance along the interface then enter the neuropil, or immediately do so to form secondary axon systems with SATs from other lineages. The pattern of SATs is highly stereotypic. In the mosaic analysis performed here, mutant Sim^+ neurons with reduced or eliminated Sim function are projecting aberrant diffuse SATs which sometimes bifurcate to form ectopic or supernumerary secondary tract systems with SATs from incorrect lineages. This would indicate that at least one function of Sim in this context is to provide identity cues to the cells. Whether the association with incorrect SATs is directed (i.e. an alternate identity is imparted in mutant Sim^+ cells) or promiscuous (i.e. no specific identity is imparted to mutant Sim^+ cells) remains unelucidated.

Both mammalian and *Drosophila sim* play multiple roles in development in the CNS and in other cell types (Crews, 2003; Fan, 2003). Within the CNS, each plays a role in neurogenesis and later in axonogenesis (Marion et al., 2005). *Drosophila sim* controls neurogenesis of embryonic CNS midline cells, and differentiation of the optic lobe laminar neurons (Umetsu et al., 2006). In mammals, *Sim1* plays a prominent role in neurogenesis of the hippocampus (Michaud et al., 1998). Additionally, the murine *Sim1* and *Sim2* genes are expressed in the mammillary body (Ema et al., 1996a; Fan et al., 1996), and control axonogenesis (Marion et al., 2005). In wild type mice, the $Sim1^+$ $Sim2^+$ mammillary body cells extend axons along the principal mammillary tract (PMT) that project to the thalamus and tegmentum via the mammillotegmental (MTEG) and mammillothalamic (MTT) tracts. Genetic experiments indicated that the MTEG and MTT are reduced in *Sim1 Sim2* double mutant embryos and, to a lesser degree, in *Sim1* single mutant embryos. Normally the PMT extends along the ipsilateral side of the developing brain, but in *Sim1 Sim2* mutant embryos, the axons are abnormally targeted across the midline. This suggests

that the mammillary body axons no longer respond to a midline-directed repellent in *Sim* mutant embryos. Consistent with this interpretation, *Sim1* and *Sim2* were shown to normally repress expression of *Rig-1/Robo3*, a gene that antagonizes *Slit*-mediated repulsion (Marillat et al., 2004; Sabatier et al., 2004). Consequently, upregulation of *Rig-1/Robo3* in *Sim* mutant embryos results in loss of PMT repulsion by the midline.

The *Drosophila sim* central brain axon defect differs from the mammalian *Sim1 Sim2* mutant defect in that the central brain axons show fasciculation defects. Significantly, targeting appears unaffected since mutant axons branch and are attracted to the midline. Presumably, *sim* regulates expression of one or more genes involved in controlling axon fasciculation, although the identities of those genes are unknown. There are a number of *Drosophila* cell adhesion proteins that have been implicated in axon fasciculation (Van Vactor, 1998), including Fasciclin II, Roughest, and Cadherin-N. One possible explanation for the *sim* phenotype is that *sim* positively regulates levels of cell adhesion or fasciculation proteins, and when their levels drop below a threshold level, defasciculation can occur. Conversely, there exists a class of genes that are anti-adhesive, such as *beaten path* (Fambrough and Goodman, 1996) and protein tyrosine phosphatases (Desai et al., 1996), which could be normally repressed or silenced by *sim*, and become active in *sim* mutants and promote defasciculation. However, it is currently not known whether any of these genes are direct or indirect targets of *sim* regulation. It will also be interesting to see whether the fasciculation of the Sim^+ axons at the midline is also affected in *sim* mutants. Further insight into the mechanisms that govern axon guidance of Sim^+ cells will require identifying the relevant transcriptional targets of Sim. Additional insight into understanding the genetic basis of locomotory coordination will require identifying the neural inputs and outputs in the brain CX that drive these behaviors.

ACKNOWLEDGEMENTS

The authors would like to thank Jim Skeath, Chris Doe, Christian Klämbt, Jessica Treisman, FlyBase, the Developmental Studies Hybridoma Bank, and the Bloomington *Drosophila* Stock Center for fly stocks and advice. This work was supported by NIH grant 5R37HD025251 to STC.

FIGURES

Figure 1. Interhemispheric crossing and fasciculation of Sim^+ secondary axon tracts. Frontal sections of dissected, stained brains of wild type third instar larvae are shown except where indicated. (A) Brain stained with anti-Sim (magenta) showing the 3 paired clusters of Sim^+ central brain cells. They are the dorsal (D), medial (M), and ventral (V) clusters. The location of the supraesophageal commissure (SC), which consists of crossing, interhemispheric axons, is indicated but beyond the imaged volume. The central dark space represents the esophagus (E) that runs through the central brain. (B) Schematic of a third instar larval brain shown in sagittal and frontal views indicating the locations of the Sim^+ central brain clusters. (C, D) High magnification views of a dorsal cluster (C) and medial cluster (D) stained with anti-Sim (magenta) and mAb BP106 (green) showing that each cluster consists of 2 populations of cells whose axons (arrowheads) fasciculate. (E, F) Brain stained with anti-Sim (magenta) and BP106 (green) showing that the dorsal clusters (arrowheads) extend axons across the supraesophageal commissure and fasciculate (arrow). (F) Merge image. (G, H) Brain visualized for Sim immunoreactivity (magenta) and GFP (green) showing 2 GFP^+ MARCM clones, one in each dorsal Sim^+ cluster (arrowheads), that cross and fasciculate in the supraesophageal commissure (arrow). (H) Merge image. (I, J) Brain visualized for Sim immunoreactivity (magenta) and GFP (green) showing a GFP^+ MARCM clone in the Sim^+ medial cluster (arrowhead) that extends a secondary axon tract dorsally (arrows), but does not cross the midline. (J) Merge image. (K, L) Brain stained with anti-Sim (magenta) and BP106 (green) showing that the ventral cluster (arrowhead) extends a short, ipsilateral axon tract (arrow) toward the neuropil near the ventral esophagus (E).

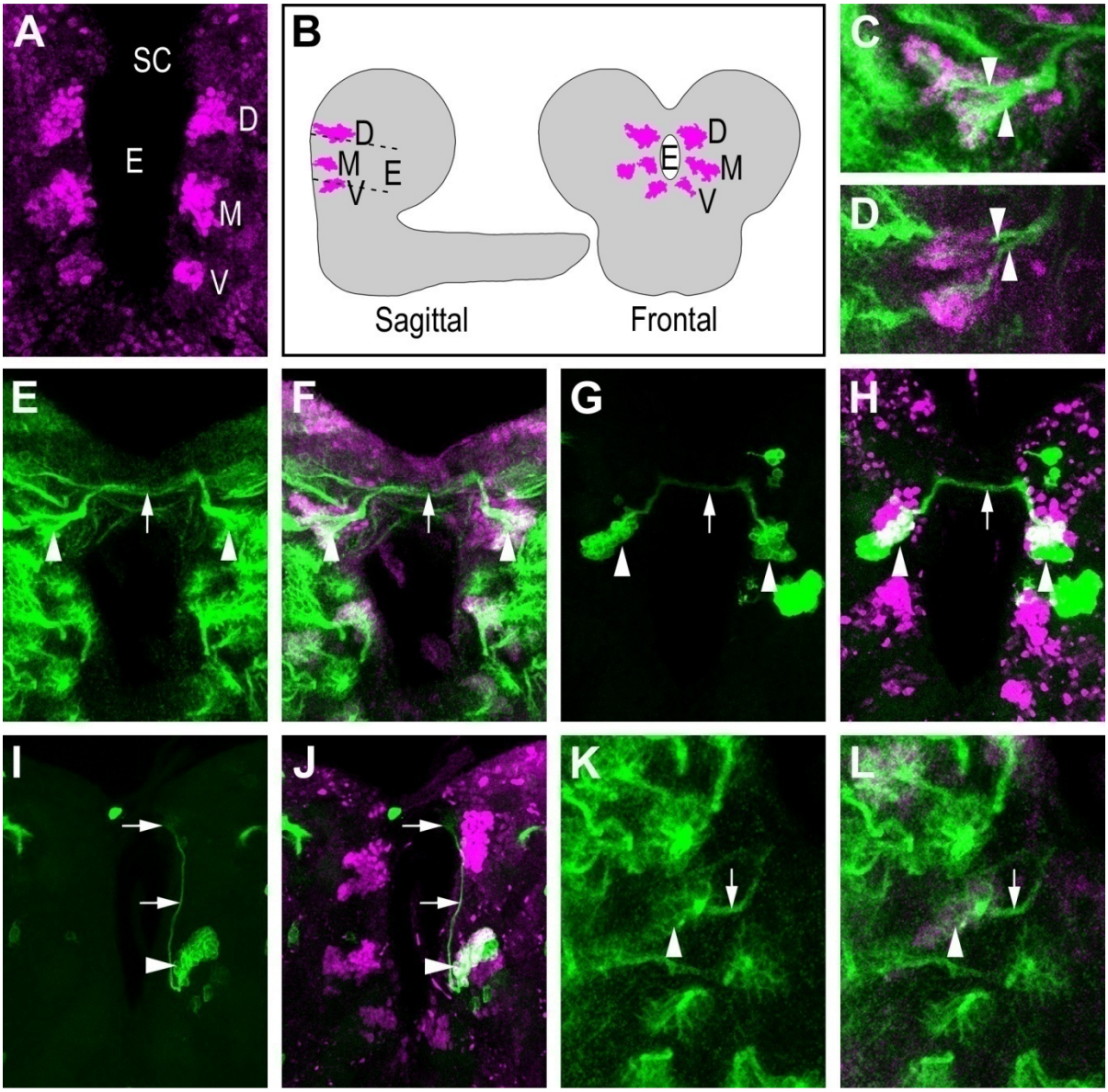

Figure 1

Figure 2. Developmental expression of *sim* in the central brain. Embryos and larvae were dissected and immunostained with anti-Sim (red). Dorsal is up in B, C, and D; anterior is up in A. (A) Embryonic stage 17 brain showing Sim localization in one domain located posteriorly and medially within the left and right hemispheres. Because of the bent neuraxis at the level of the supraesophageal ganglion, the brain is shown in transverse section while the rest of the larva is shown in dorsal view. Dark circles immediately laterad of Sim^+ cells in the central brain are the paired cervical connective (Nassif et al., 1998; Nassif et al., 2003). (B-D) Dissected brains from first instar (B), second instar (C), and third instar (E) larvae show Sim localization in paired clusters of cells in the central brain throughout larval development: dorsal cluster (D), medial cluster (M), and ventral cluster (V). The number of Sim^+ central brain cells is relatively similar in first and second instar larvae, but increases significantly between the second instar and late third instar larval stages. Due to a specimen positioning artifact, the first instar larval brain shows Sim^+ cells located medially within the left brain hemisphere. However, these cells occupy a similar anatomical position to their counterparts from the right hemisphere; these cells directly abut the esophagus as they do in later larval stages. Note only the left ventral cluster of Sim^+ cells is visible in (B); the contralateral cluster and both clusters in (C) lie outside the imaged volume. Also note strong Sim localization in the nuclei of cells of the developing optic lobes in the third instar larval brain as these compartments undergo differentiation from the placode stage, present during first and second instar larval stages, to become the lamina (La) and medulla (Me) of the adult visual system.

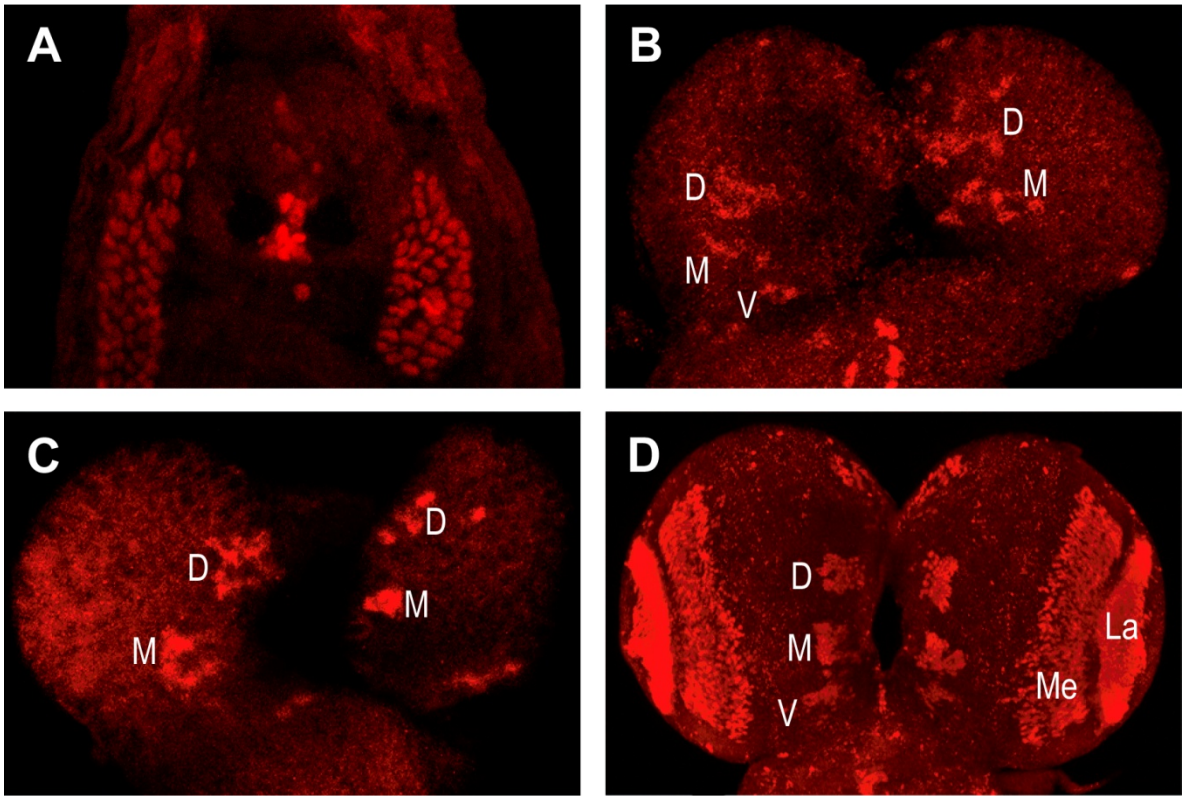

Figure 2

Figure 3. *sim* is not required for the differentiation of post-mitotic central brain neurons. (A-E) Wild type third instar larval brain showing that Sim protein is not present in neural precursor cells, but only in post-mitotic neurons. Shown is a GFP^+ MARCM clone (D; green) stained with anti-Sim (A; red), Elav (B; magenta), and DAPI (C; blue); merge is shown in E. The GFP^+ cells include a neuroblast (NB; arrowhead), 2-3 putative GMCs (white arrowheads), and ~10 neurons (yellow arrows). Elav immunoreactivity is absent from the NB, GMCs have a low level of Elav, and neurons have high Elav levels. Sim protein is only detectable in neurons and is absent from the NB and GMCs. (F-J) sim^8 MARCM clone showing that NB (arrowhead), associated GMCs, and neurons (yellow arrow) are present. The neurons (F, J) show relatively weak Sim staining (red) and are $Elav^+$. While the Elav staining in sim^8 (G) appears considerably less intense than in wild type (B), multiple experiments show variability in Elav staining and the sim^8 Elav protein levels are comparable to that seen in most wild type brains.

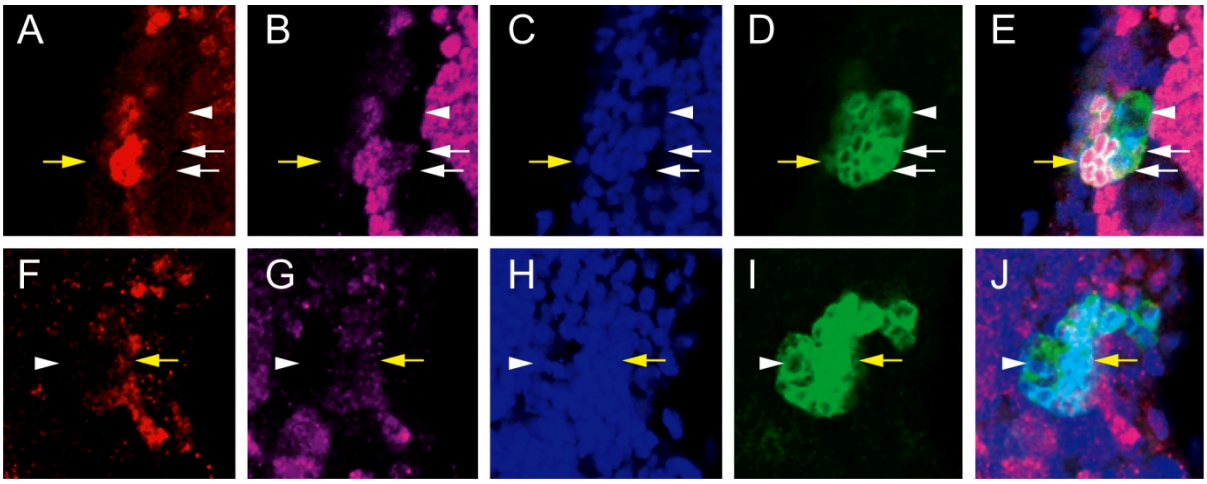

Figure 3

Figure 4. Structure of Sim wild type and mutants proteins. (A) The sequence structure of the Sim protein is shown. The length of wild type Sim is 673 aa (Nambu et al., 1991); the Sim-PB isoform is shown. At the top are the boundaries between defined regions: basic region (b), HLH domain, PAS-1, PAS-2, AQ-rich region, and 3 transcriptional activation domains (A1-3). The sites of exon junctions are shown below the structure, with the exons (E) separated by a vertical line and the residue where the splice occurs below the line. The antigen used to generate the Sim antibody used in this work spans aa residues 413 to 650 and is indicated by a brown box. Predicted proteins of each mutant strain are shown to the right of each allele name. The green bar indicates the size of the protein, if translated, and the length of the predicted protein in amino acids is indicated to the right. The orange vertical lines for sim^{7}, $sim^{J1\text{-}47}$ and sim^{K} indicate the sites of their missense mutations. The orange region for sim^{W3} indicates distance from the splice site mutation to the first in-frame termination codon. The orange region for sim^{1} indicates the distance from the site of the frameshift mutation to the first in-frame termination codon. (B) Stage 10 embryos from wild type, sim^{2}, sim^{8}, and sim^{BB68} were stained with anti-Sim to assay expression. None of the mutants had embryonic Sim protein.

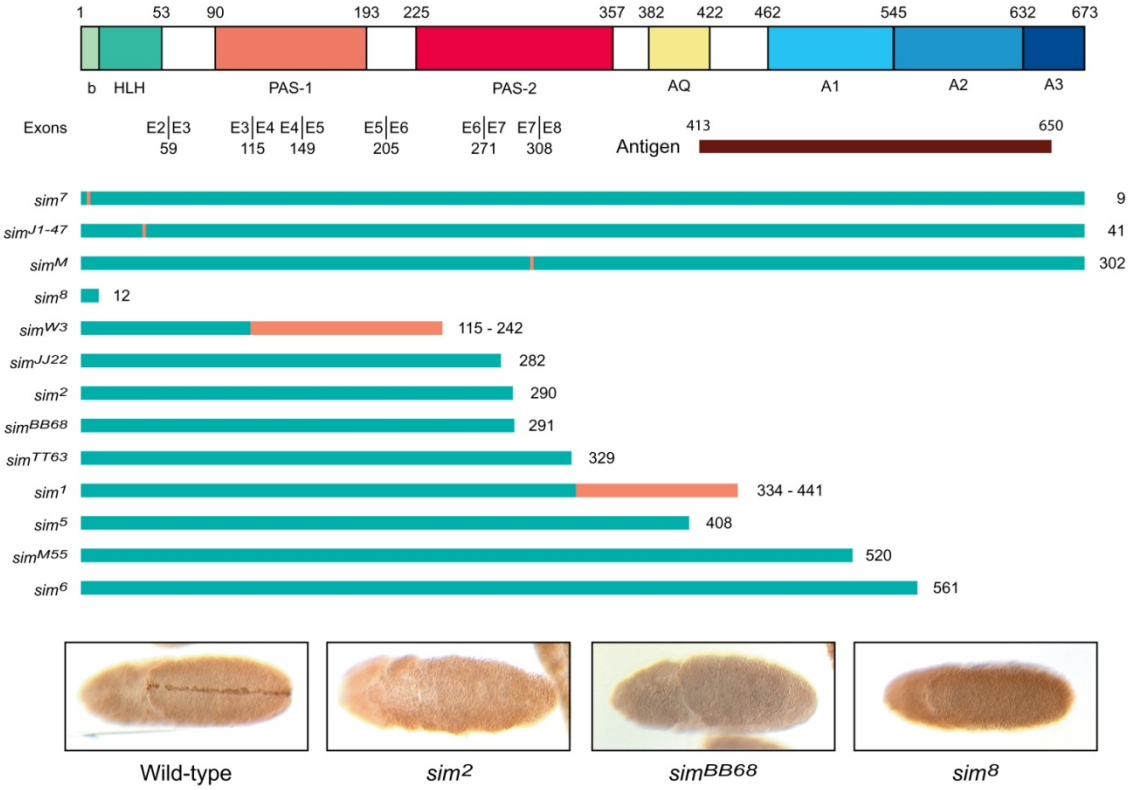
1
53
90
193
225
357
382
422
462
545
632
673
b
HLH
PAS-1
PAS-2
AQ
A1
A2
A3
Exons
E2|E3
59
E3|E4
115
E4|E5
149
E5|E6
205
E6|E7
271
E7|E8
308
413
650
Antigen
sim7
9
simJ1-47
41
simM
302
sim8
12
simW3
115 - 242
simJJ22
282
sim2
290
simBB68
291
simTT63
329
sim1
334 - 441
sim5
408
simM55
520
sim6
561
Wild-type
sim2
simBB68
sim8

Figure 4

Figure 5. *sim* mutant clones show axon fasciculation defects. Wild type and mutant MARCM GFP^+ (green) dorsal cluster Sim^+ clones were stained with anti-Sim (magenta). For each pair of images, the left panel shows a single optical slice showing the GFP^+ cell bodies, and the right panel is a projection. (A-D) Two wild type GFP^+ clones showing the characteristic axon tract that extends anterior-dorsally toward the neuropil, then elaborates filopodia (arrow) before projecting contralaterally across the supraesophageal commissure on a horizontal trajectory (arrowhead). (E-H) Two sim^{BB68} mutant clones, in which the axons split into multiple fascicles, rather than traverse the supraesophageal commissure as a singular, tight fascicle. (I-L) Two sim^{2} clones that also show multiple branches. (M-N) The sim^{8} clone also shows multiple branches.

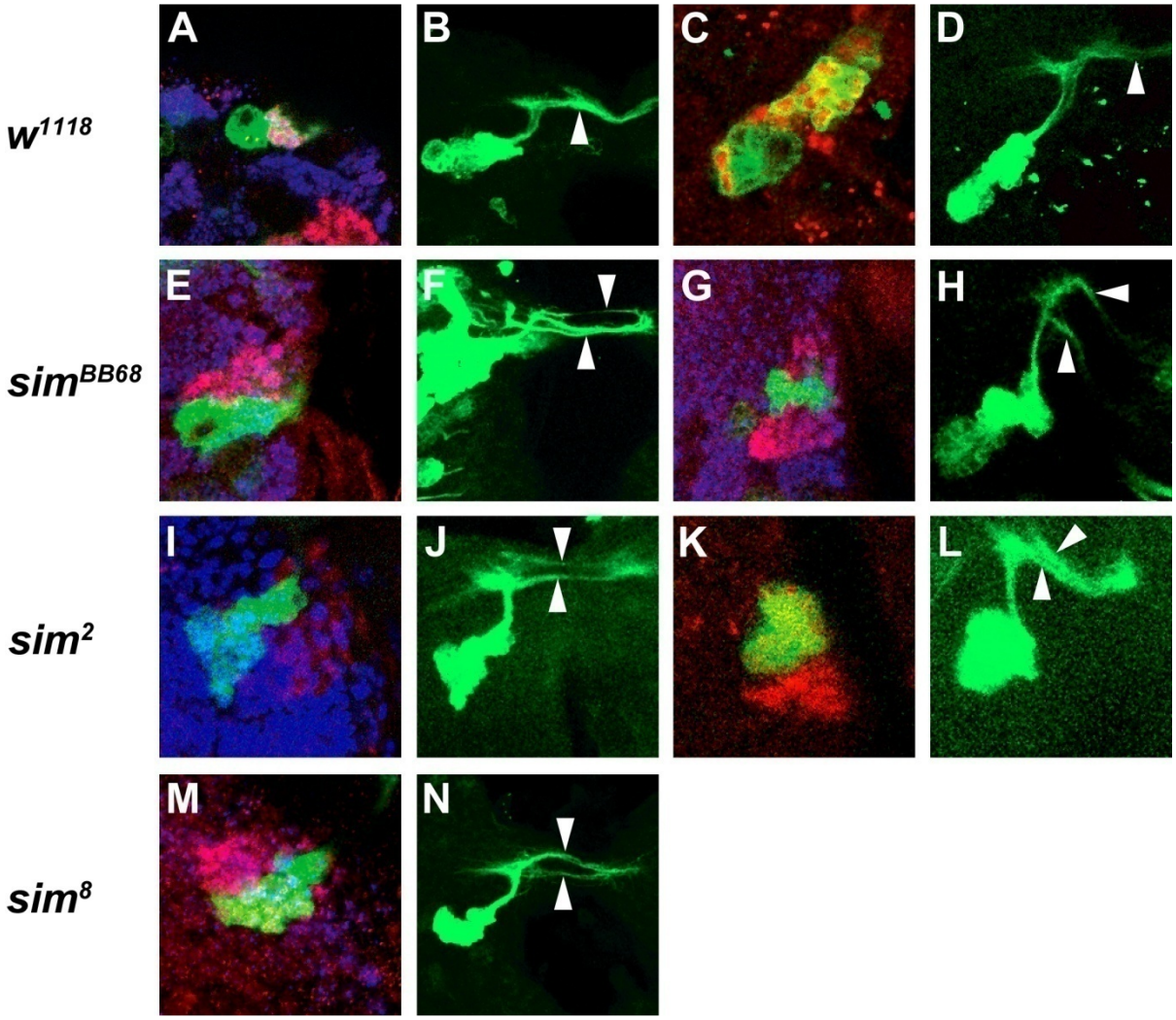

Figure 5

TABLES

Table 1. Molecular alterations of *sim* alleles.

Allele	Synonym	Mutagen	DNA Mutation (WT/Mut)	Location	Protein Change
sim^{1}	*sim*^{S8}	X-ray	CGCGACACTGG CGC- - - - -TGG	exon 8	Frameshift at D334, STOP at 441
sim^{2}	*sim*^{H9}	EMS	TAT → TAG	exon 7	Y290 → STOP
sim^{5}	*sim*$^{B13-4}$	EMS	CAG → TAG	exon 8	Q408 → STOP
sim^{6}	*sim*$^{B21-2}$	EMS	CAG → TAG	exon 8	Q561 → STOP
sim^{7}	*sim*$^{B30-1}$	EMS	GCA → ACA	exon 2	A9 → T
sim^{8}	*sim*^{E320}	EMS	CGA → TGA	exon 2	R12 → STOP
sim^{BB68}		EMS	CAG → TAG	exon 7	Q291 → STOP
sim^{J1-47}		EMS	TCC → TTC	exon 2	S41 → F
sim^{K}		Unknown	CGC → CCC	exon 7	R302 → P
sim^{JJ22}		EMS	CAG → TAG	exon 7	Q282 → STOP
sim^{M55}		EMS	TAT → TAA	exon 8	Y520 → STOP
sim^{TT63}		EMS	TGG → TGA	exon 8	W329 → STOP
sim^{W3}		EMS	CCA|GTGA CCA|ATGA	exon 4 | intron 4	Splice mutant after exon 4, STOP at 242

Table 1

REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The Genome Sequence of *Drosophila Melanogaster* . *Science* **287,** 2185-2195.

Berger, C., Renner, S., Luer, K. and Technau, G. M. (2007). The Commonly used Marker ELAV is Transiently Expressed in Neuroblasts and Glial Cells in the *Drosophila* embryonic CNS. *Dev. Dyn.* **236,** 3562-3568.

Casso, D., Ramirez-Weber, F. and Kornberg, T. B. (2000). GFP-Tagged Balancer Chromosomes for *Drosophila Melanogaster* . *Mech. Dev.* **91,** 451-454.

Chen, H., Chrast, R., Rossier, C., Gos, A., Antonarakis, S. E., Kudoh, J., Yamaki, A., Shindoh, N., Maeda, H. and Minoshima, S. (1995). *Single-Minded* and Down Syndrome? *Nat. Genet.* **10,** 9-10.

Chrast, R., Scott, H. S., Chen, H., Kudoh, J., Rossier, C., Minoshima, S., Wang, Y., Shimizu, N. and Antonarakis, S. E. (1997). Cloning of Two Human Homologs of the *Drosophila Single-Minded* Gene SIM1 on Chromosome 6q and SIM2 on 21q within the Down Syndrome Chromosomal Region. *Genome Res.* **7,** 615-624.

Crews, S. T. (2003). *Drosophila* bHLH-PAS Developmental Regulatory Proteins. In *PAS PROTEINS: Regulators and Sensors of Development and Physiology* (ed. S. T. Crews), pp. 51-68. Norwell, MA: Kluwer Academic Publishers.

Crews, S. T., Thomas, J. B. and Goodman, C. S. (1988). The *Drosophila Single-Minded* Gene Encodes a Nuclear Protein with Sequence Similarity to the *Per* Gene Product. *Cell* **52,** 143-151.

Dahmane, N., Charron, G., Lopes, C., Yaspo, M. L., Maunoury, C., Decorte, L., Sinet, P. M., Bloch, B. and Delabar, J. M. (1995). Down Syndrome-Critical Region Contains a Gene Homologous to *Drosophila Sim* Expressed during Rat and Human Central Nervous System Development. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 9191-9195.

Davis, R. L. (1996). Physiology and Biochemistry of *Drosophila* learning Mutants. *Physiol. Rev.* **76,** 299-317.

Desai, C. J., Gindhart, J. G.,Jr, Goldstein, L. S. and Zinn, K. (1996). Receptor Tyrosine Phosphatases are Required for Motor Axon Guidance in the *Drosophila* embryo. *Cell* **84,** 599-609.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S. (1991). 'Touchdown' PCR to Circumvent Spurious Priming during Gene Amplification. *Nucleic Acids Res.* **19,** 4008.

Dumstrei, K., Wang, F., Nassif, C. and Hartenstein, V. (2003). Early Development of the *Drosophila* brain: V. Pattern of Postembryonic Neuronal Lineages Expressing DE-Cadherin. *J. Comp. Neurol.* **455,** 451-462.

Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., Kikuchi, Y. et al. (1996a). Two New Members of the Murine *Sim* gene Family are Transcriptional Repressors and show Different Expression Patterns during Mouse Embryogenesis. *Mol. Cell. Biol.* **16,** 5865-5875.

Ema, M., Suzuki, M., Morita, M., Hirose, K., Sogawa, K., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H. et al. (1996b). CDNA Cloning of a Murine Homologue of *Drosophila Single-Minded*, its mRNA Expression in Mouse Development, and Chromosome Localization. *Biochem. Biophys. Res. Commun.* **218,** 588-594.

Estes, P., Mosher, J. and Crews, S. T. (2001). *Drosophila Single-Minded* Represses Gene Transcription by Activating the Expression of Repressive Factors. *Dev. Biol.* **232,** 157-175.

Fambrough, D. and Goodman, C. S. (1996). The *Drosophila Beaten Path* Gene Encodes a Novel Secreted Protein that Regulates Defasciculation at Motor Axon Choice Points. *Cell* **87,** 1049-1058.

Fan, C. (2003). Hormones, Obesity, Learning, and Breathing - the Many Functions of Mammalian *Single-Minded* Genes. In *PAS PROTEINS: Regulators and Sensors of Development and Physiology* (ed. S. T. Crews), pp. 205-230. Norwell, MA: Kluwer Academic Publishers.

Fan, C. M., Kuwana, E., Bulfone, A., Fletcher, C. F., Copeland, N. G., Jenkins, N. A., Crews, S., Martinez, S., Puelles, L., Rubenstein, J. L. et al. (1996). Expression Patterns of Two Murine Homologs of *Drosophila Single-Minded* Suggest Possible Roles in Embryonic Patterning and in the Pathogenesis of Down Syndrome. *Mol. Cell. Neurosci.* **7,** 1-16.

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.

Hartenstein, V., Spindler, S., Pereanu, W. and Fung, S. (2008). The Development of the *Drosophila* Larval Brain. *Advances in Experimental Medicine and Biology* **628,** 1-31.

Hecker, K. H. and Roux, K. H. (1996). High and Low Annealing Temperatures Increase both Specificity and Yield in Touchdown and Stepdown PCR. *BioTechniques* **20,** 478-485.

Hilliker, A. J., Clark, S. H., Chovnick, A. and Gelbart, W. M. (1980). Cytogenetic Analysis of the Chromosomal Region Immediately Adjacent to the Rosy Locus in *Drosophila Melanogaster* . *Genetics* **95,** 95-110.

Ito, K. and Hotta, Y. (1992). Proliferation Pattern of Postembryonic Neuroblasts in the Brain of *Drosophila Melanogaster* . *Dev. Biol.* **149,** 134-148.

Lee, T. and Luo, L. (1999). Mosaic Analysis with a Repressible Cell Marker for Studies of Gene Function in Neuronal Morphogenesis. *Neuron* **22,** 451-461.

Marillat, V., Sabatier, C., Failli, V., Matsunaga, E., Sotelo, C., Tessier-Lavigne, M. and Chedotal, A. (2004). The Slit Receptor Rig-1/Robo3 Controls Midline Crossing by Hindbrain Precerebellar Neurons and Axons. *Neuron* **43,** 69-79.

Marion, J. F., Yang, C., Caqueret, A., Boucher, F. and Michaud, J. L. (2005). *Sim1* and *Sim2* are Required for the Correct Targeting of Mammillary Body Axons. *Development* **132,** 5527-5537.

Mayer, U. and Nusslein-Volhard, C. (1988). A Group of Genes Required for Pattern Formation in the Ventral Ectoderm of the *Drosophila* embryo. *Genes Dev.* **2,** 1496-1511.

Michaud, J. L., Rosenquist, T., May, N. R. and Fan, C. M. (1998). Development of Neuroendocrine Lineages Requires the bHLH-PAS Transcription Factor SIM1. *Genes Dev.* **12,** 3264-3275.

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, K. A.,Jr 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.

Nassif, C., Noveen, A. and Hartenstein, V. (1998). Embryonic Development of the *Drosophila* Brain. I. Pattern of Pioneer Tracts. *J. Comp. Neurol.* **402,** 10-31.

Nassif, C., Noveen, A. and Hartenstein, V. (2003). Early Development of the *Drosophila* Brain: III. the Pattern of Neuropile Founder Tracts during the Larval Period. *J. Comp. Neurol.* **455,** 417-434.

Patel, N. H., Snow, P. M. and Goodman, C. S. (1987). Characterization and Cloning of *Fasciclin III*: A Glycoprotein Expressed on a Subset of Neurons and Axon Pathways in *Drosophila*. *Cell* **48,** 975-988.

Pereanu, W. and Hartenstein, V. (2006). Neural Lineages of the *Drosophila* brain: A Three-Dimensional Digital Atlas of the Pattern of Lineage Location and Projection at the Late Larval Stage. *J. Neurosci.* **26,** 5534-5553.

Pielage, J., Steffes, G., Lau, D. C., Parente, B. A., Crews, S. T., Strauss, R. and Klambt, C. (2002). Novel Behavioral and Developmental Defects Associated with *Drosophila Single-Minded* . *Dev. Biol.* **249,** 283-299.

Robinow, S. and White, K. (1991). Characterization and Spatial Distribution of the ELAV Protein during *Drosophila Melanogaster* Development. *J. Neurobiol.* **22,** 443-461.

Sabatier, C., Plump, A. S., Le, M., Brose, K., Tamada, A., Murakami, F., Lee, E. Y. and Tessier-Lavigne, M. (2004). The Divergent Robo Family Protein Rig-1/Robo3 is a Negative Regulator of Slit Responsiveness Required for Midline Crossing by Commissural Axons. *Cell* **117,** 157-169.

Sprecher, S. G., Reichert, H. and Hartenstein, V. (2007). Gene Expression Patterns in Primary Neuronal Clusters of the Drosophila Embryonic Brain. *Gene Expr. Patterns* **7,** 584-595.

Thomas, J. B., Crews, S. T. and Goodman, C. S. (1988). Molecular Genetics of the *Single-Minded* Locus: A Gene Involved in the Development of the *Drosophila* nervous System. *Cell* **52,** 133-141.

Truman, J. W., Schuppe, H., Shepherd, D. and Williams, D. W. (2004). Developmental Architecture of Adult-Specific Lineages in the Ventral CNS of *Drosophila*. *Development* **131,** 5167-5184.

Umetsu, D., Murakami, S., Sato, M. and Tabata, T. (2006). The Highly Ordered Assembly of Retinal Axons and their Synaptic Partners is Regulated by Hedgehog/Single-Minded in the *Drosophila* visual System. *Development* **133,** 791-800.

Van Vactor, D. (1998). Adhesion and Signaling in Axonal Fasciculation. *Curr. Opin. Neurobiol.* **8,** 80-86.

Vann, S. D. and Aggleton, J. P. (2004). The Mammillary Bodies: Two Memory Systems in One? *Nat. Rev. Neurosci.* **5,** 35-44.

Wheeler, S. R., Kearney, J. B., Guardiola, A. R. and Crews, S. T. (2006). Single-Cell Mapping of Neural and Glial Gene Expression in the Developing *Drosophila* CNS Midline Cells. *Dev. Biol.* **294,** 509-524.

Xu, C. and Fan, C. M. (2007). Allocation of Paraventricular and Supraoptic Neurons Requires Sim1 Function: A Role for a Sim1 Downstream Gene PlexinC1. *Mol. Endocrinol.* **21,** 1234-1245.

Younossi-Hartenstein, A., Nguyen, B., Shy, D. and Hartenstein, V. (2006). Embryonic Origin of the *Drosophila* brain Neuropile. *J. Comp. Neurol.* **497,** 981-998.

CHAPTER 4:

ANALYSIS OF TRANSCRIPTIONAL REGULATION AT THE *SINGLE-MINDED* LOCUS OF *DROSOPHILA MELANOGASTER*

PREFACE

All work presented in this chapter was performed by the author of this dissertation.

ABSTRACT

The *sim* gene in *Drosophila* acts as the master regulator of the midline cell fate in the developing CNS. However, its expression in the embryonic midline is complex, and it is also expressed in non-neural tissues such as the foregut, hindgut, myoblasts, as well as post-embryonically in the larval central brain. It is therefore important to understand the transcriptional regulatory elements that drive *sim* expression in these discrete tissue compartments. Bioinformatic analysis revealed the absence of canonical TATA sequences, DPE motifs, and *Drosophila*-specific Inr sequences in the expected regions upstream of the transcriptional start sites. This suggested that the *sim* locus utilizes cryptic core promoters. Using RT-PCR, we demonstrated the existence of a third *sim* isoform (*sim-RC*) that is present after the onset of midline neurogenesis and persists throughout post-embryonic life. Sequence analysis revealed that this isoform shares an alternatively spliced first exon with *sim-RB*, and that these

isoforms together utilize a first exon different from the first exon contained within *sim-RA*, the early *sim* transcript. This indicated the presence of 2 promoters at the *sim* locus. We employed molecular genetics to dissect the upstream intergenic sequence as well as the intronic sequence from the *sim* locus in an effort to uncover the locations of enhancers which drive tissue-specific expression and found the enhancer for midline, foregut, and myoblast expression. Taken together, these results have increased our knowledge of the expectedly complex transcriptional regulation at the *sim* locus, given its complex spatial and temporal expression.

INTRODUCTION

A common biological theme in the development of metazoans is parsimony of gene usage. One locus can be co-opted several times during development and its product used in different contexts to perform different roles in different tissues. One example of this parsimony can be found in the *single-minded* (*sim*) locus. The Sim protein is a basic helix-loop-helix (bHLH) transcription factor which acts as an activator in *Drosophila melanogaster* (Estes et al., 2001), although its orthologs in mammals may act as both activators and repressors (Ema et al., 1996; Moffett and Pelletier, 2000; Probst et al., 1997). One important role for this gene during embryonic development of the fruit fly is in the midline cells of the central nervous system (CNS). During early neurogenesis, cells that express the *sim* gene will take on a mesectodermal cell fate. This is accomplished by the Sim protein, in concert with its ubiquitously localized dimerization partner Tango (Tgo) (Sonnenfeld et al., 1997; Ward et al., 1998), regulating genes such as *slit*, *Toll*, *center divider*, and *rhomboid* (Nambu et al., 1990) which promote the midline cell fate while concomitantly activating repressors of genes which promote lateral CNS cell fate (Nambu et al., 1990; Nambu et al., 1991; Xiao et al., 1996). Further, Sim:Tgo interacts with

cofactors to adopt greater transcriptional specificity. For example, biochemical experiments have shown that the Sim:Tgo heterodimer cooperates with Dichaete through the Sim PAS domain and Drifter through its POU domain (Ma et al., 2000) to effect midline glia-specific gene transcription. Mutations in either *drifter (dfr)* or *Dichaete (D)* result in relatively normal expression of midline glial markers and, while midline glia are present, they fail to migrate properly. Analysis of *D dfr* double mutants, however, showed that midline glia were largely absent and concomitantly, midline glia-specific gene expression was greatly reduced (Ma et al., 2000; Sanchez-Soriano and Russell, 1998). Thus, the action of Sim as a transcriptional regulator is dependent on its dimerization with Tgo, and its tissue-type specificity is conferred by selective interactions with co-regulatory proteins.

The *sim* gene is strongly expressed in the mesectoderm immediately prior to gastrulation of the developing embryo, and this expression persists in all midline precursor cells until after embryonic stage 11 (Ashburner, 1989; Wheeler et al., 2006). By stage 17, the end of embryogenesis, the midline precursors have developed into their respective terminal cell fates: 3 midline glia, 2 MP1 neurons, 2 MP3 interneurons (the H-cell and H-cell sib), 3 iVUM interneurons, 3 mVUM interneurons, 1 median neuroblast, and 7-8 progeny of the median neuroblast (Bossing and Technau, 1994; Kearney et al., 2004). The midline glia, H-cell sib, iVUMs, median neuroblast, and its progeny all show continued strong expression from the *sim* locus while expression in the MP1 neurons, H-cell, and mVUMs is reduced to detectable but considerably lower levels by the end of embryogenesis (Kearney et al., 2004). This raises the possibility that the gene product may be performing different functions in these cell types at these developmental stages.

There are also non-midline functions of *sim*. The *sim* gene is expressed in a subset of foregut cells which abut the developing CNS at the level where the ventral nerve cord joins the

developing brain during embryonic stages 10-15 (Crews, unpublished). In *sim* zygotic mutants, neuroblasts proximal to the neighboring foregut do not proliferate to the extent seen in wild type individuals and thus the CNS at this level appears less developed (Page, 2003). This effect is presumably mediated by the absence of Egf receptor signaling. Thus, one function of *sim* expression in the foregut appears to be the expansion of the midbrain lateral to this structure.

In addition to the CNS, Sim protein can be found transiently in a subset of myoblasts during embryonic stages 11-13 (Lewis and Crews, 1994). It is thought that the protein is required to mediate repulsion of the developing body wall muscles away from the CNS (and anatomical) midline via the *slit* signaling pathway.

Sim can also be found in the hindgut, and one function of this tissue-specific localization is to create the anal slit, later to become the adult anus. Embryos homozygous mutant for the amorphic sim^{2} allele have reduced anal pads and undeveloped anal slits (Pielage et al., 2002). Transheterozygous embryos bearing one copy of the sim^{2} allele and one copy of a temperature sensitive *sim* allele ($sim^{J1\text{-}47}$) were observed to have anal slits which occupied only half of the anal pad when reared at the permissive temperature of 17°C and lacked the anal slit entirely when reared at the restrictive temperature of 29°C, similar to homozygous $sim^{J1\text{-}47}$ animals. Homozygous $sim^{J1\text{-}47}$ animals survive to adulthood when reared at the permissive temperature but were sterile. Although most individuals displayed no abnormal external phenotypes, less than 10% of these adults lack visible anal openings and genital structures. These hindgut-defective individuals displayed normorexia and normal feeding behavior but died after a few days. One possible explanation for this adult lethality is the affected individual's inability to eliminate metabolic waste, although this has not been formally tested.

Currently there are two known splice variants that are produced by the *sim* locus. These isoforms differ in the 5' untranslated region (UTR), encoded by alternate first exons. Protein

primary sequence is the same between both variants (Crews, unpublished). The early promoter, immediately upstream of the early first exon, drives *sim* expression in CNS midline cells during the primordium and mesectodermal stages, prior to the onset of midline cell specialization. The late promoter is putatively responsible for driving expression in all non-CNS tissues as well as later CNS midline expression in terminally differentiated midline glia and neurons.

Given the large variety of tissues in which the *sim* gene is expressed, it is predicted that the locus contains large regulatory regions which are required for temporal and tissue-specific patterns of expression since discrete enhancer elements are required for separate components of its expression pattern. It is of interest to understand the disposition of regulatory elements which govern the expression of this developmentally important gene. In this paper we present an analysis of the two known promoters within the locus and show they are not of either the canonical TATA box or downstream promoter element (DPE) types. We identify a third splice variant for *sim* via reverse transcription-PCR and show that this is the major *sim* isoform expressed throughout the animal's life, and that this isoform is transcribed from the late promoter. Finally, we dissect the large first intron of the *sim* locus using the GAL4/UAS system (Brand and Perrimon, 1993) and show that this first intron, as well as the intergenic region between the *sim* locus and the gene immediately upstream, harbor enhancers that drive gene expression in different tissues. Greater understanding of the *sim* locus will facilitate the construction of mutant loci, which can be used to further understand the disparate functions of this developmentally important gene.

MATERIALS AND METHODS

Fly stocks

The w^{1118} strain served as wild type and was used to create transformant lines. w; P{w^+; UAS-GFP.nls} and w; P{w^+; UAS-τ-GFP} stocks were obtained from the Bloomington *Drosophila* Stock Center.

Molecular cloning

1.0, 2.8, 3.1, and 5.5 kb genomic fragments were amplified from wild type genomic DNA template using standard PCR and cloned into either pCR II-TOPO (1.0, 2.8; Invitrogen) or pGEM-T Easy (3.1, 5.5; Promega) shuttle vectors. The subcloned 1.0 kb fragment was excised from pCR II-TOPO using EcoRI and cloned into a unique EcoRI site in the polylinker region of pPTGAL. Both forward and reverse oriented insertions were constructed in this manner. The 2.8 kb fragment was excised from pCR II-TOPO using EcoRI and cloned into the EcoRI site in pPTGAL to create the forward oriented construct. To create the reverse construct, the 2.8 kb fragment was excised from pPTGAL using NotI and KpnI and cloned into pPTGAL previously digested with NotI and KpnI. The 3.1 and 5.5 kb fragments were excised from pGEM-T Easy by NotI digestion and cloned into pPTGAL to create forward and reverse oriented constructs.

Midiprepped constructs were co-injected with pTURBO (encoding Δ2-3 transposase) into w^{1118} precellularized embryos and transformant individuals were identified and isolated as described (Ashburner, 1989).

Immunohistochemistry

Embryos were collected, fixed, and stained as described (Pielage et al., 2002). Larval brains were dissected in 1x PT and fixed in 4% methanol-free formaldehyde (Ted Pella) in 1x PEM. Antibody staining was performed as described (Patel et al., 1987). Antibodies and dilutions were as follows:

rat-anti-Sim 1:200

rabbit-anti-GFP 1:1000 (Abcam)

HRP-conjugated anti-rabbit IgG (Jackson ImmunoResearch) 1:1000

All fluorophore-conjugated secondary antibodies (Molecular Probes, now Invitrogen) were used at 1:200. Fluorescently labeled specimens were mounted in Aqua Poly/Mount (Polysciences) and imaged on a Zeiss LSM510 confocal microscope. Chromogenically labeled specimens were mounted in glycerol and imaged on a Zeiss Axiophot brightfield microscope. Photomicrographs were processed using Photoshop CS and CS4 (Adobe) or LSM Image Browser (Zeiss).

RT-PCR

Total RNA was extracted from 0-3, 3-6, 6-9, 9-12, and 12-15 h (after egg laying, AEL) embryos as well as from first, second, and third instar larvae, light (1-2 day) and dark (3-4 day) pupae, and 2-day-old adults using QIAshredder and RNeasy kits (Qiagen). Samples were treated with RQ1 RNase-free DNase (Promega) and converted to single stranded cDNA using SuperScript II First-Strand Synthesis System for RT-PCR (Invitrogen) and oligo-d(T). Transcripts were then detected by PCR in the linear range using the following primers:

Sim A 5'-TGG ATG CTG GTT GAT GTG CGG -3'

Sim B 5'-CAG GGA TAT GAG CAA GTG CTG AGA A -3'

Sim C 5'-GCC CAA GTG CCA TAA ACG CAA T -3'

RP49F 5'-ATC CGC CCA GCA TAC AGG -3'

RP49R 5'-CTC GTT CTC TTG AGA ACG CAG -3'

Sequence analysis

RT-PCR products were sequenced at the UNC-CH Genome Analysis Facility using automated dye terminator technology, then compared with DNA sequences obtained from FlyBase and analyzed using VectorNTI Advance v10.3.1 (Invitrogen).

RESULTS

Sequence analysis reveals the absence of recognizable TATA and DPE motifs

Basal promoters are often defined as the approximately 100 bp of sequence surrounding the transcription initiation site, which are engaged by the general transcriptional apparatus. TATA boxes, Initiator sequences, and downstream promoter elements (DPE) are often found in this region; however, few basal promoters contain all three elements. Kutach and Kadonaga (Kutach and Kadonaga, 2000) reported that out of 205 arthropod promoters studied, approximately half had a TATA sequence, half had a DPE, and a third had neither. Cherbas and Cherbas (Cherbas and Cherbas, 1993) reported that approximately one quarter of arthropod RNA polymerase II-transcribed promoters contained Initiator or Initiator-like sequences. The

remaining fraction of promoters utilize so-called "cryptic" sequences upon which to assemble the basal transcriptional apparatus.

Analysis of the 200 bp region surrounding the presumptive transcription initiation sites (100 bp upstream and 100 bp downstream) of the early and both late *sim* transcripts, indicate that neither basal promoters contain recognizable TATA sequences using the TATAAA consensus. Further, neither regions contain sequences resembling any permutation of the DPE consensus A/G/T-C/G-A/T-C/T-A/C/G-C/T (Kutach and Kadonaga, 2000), or the *Drosophila*-specific Inr consensus T-C-A-G/T-T-T/C (Arkhipova, 1995; Lagrange et al., 1998). It is therefore likely that the *sim* locus utilizes cryptic core promoters as nucleating sequences upon which the basal transcriptional machinery is recruited and assembled.

Detection of a third sim *isoform that is expressed post-embryonically*

RT-PCR analysis was performed to determine which of the two known promoters within the *sim* locus was transcriptionally active during different stages of the fly life cycle. Total RNA was extracted from embryos in 3 h bins, as well as from the three instar larval stages, light pupae representing the first two days of pupation, dark pupae from the third and fourth days of pupation when wings and eyes are forming, and two day old adults. After DNase treatment, the total RNA samples were reverse transcribed and the products used as a complex template for PCR. Amplification was performed using a common reverse primer, "Sim C", corresponding to the minus strand within the fourth exon, and two forward primers, "Sim A" and "Sim B", corresponding to the two known early and late alternatively spliced first exons.

The early transcript, represented in FlyBase as *sim-RA*, is expressed strongly from 0-6 h post-fertilization, corresponding to embryonic stages 0-mid 11 (Fig. 1A). During these stages the

pronuclei fuse and undergo mitotic cycles to create nuclear content for the cellularizing blastoderm (0-3 h). Gastrulation then ensues in concert with germ band elongation, characterized by the lengthening of the ventral epidermis. Cephalic features such as the gnathal and clypeolabral lobes and stomodeum, form at stage 10 (from 2 h 40 min - 5 h 20 min).With respect to the developing CNS, neuroblasts are segmented and born from stages 9-11 while the presumptive midline cells are transitioning from mesectoderm anlage in statu nascendi (stages 5-6) to mesectoderm anlage (stages 7-8) to midline primordium (stages 9-12) and later fated to become the mature midline cells (stages 13-17) (Kearney et al., 2004). The earliest detection of Sim protein occurs at late stage 5 (approximately 2.5 h) in the nuclei of cells comprising the mesectoderm anlage, arranged in two single cell-wide stripes on either side of the nascent gastrulation furrow. At this stage Sim protein is strongly localized in the cells at the rostral-most and caudal-most poles of the embryo, and together these sites of expression account for the robust RT-PCR product seen in the 0-3 h lane. The 3-6 h lane contains two equally abundant splice variants, possibly due to the onset of *sim* expression in other non-CNS tissues such as the foregut and muscle precursor cells. Alternately, the presumptive midline cells may be switching transcript production from the early promoter to the later promoter during this 3 h bin. The early transcript displays potent attenuation in abundance after 6 h, while a novel short form of the late transcript, hereafter referred to as *sim-RC*, comprises the major species seen through the end of embryogenesis and throughout the rest of the fly life cycle. The long form of the late transcript, *sim-RB*, is almost entirely absent from the 0-3 h lane and is only residually present from 3-15 h. It is undetectable post-embryonically.

Sequencing of the RT-PCR products shows that all three detected isoforms share exons 2-4 in common (and presumably the downstream exons 5-8 as well); however, *sim-RA* contains coding region from the early first exon while *sim-RB* and *sim-RC* contain sequences

corresponding to the late first exon (Fig. 1B). Their sequences extend at least beyond the 5' extent of the Sim C primer but *sim-RC* spliced early within the late first exon (after 485 bp) to exon 2 while *sim-RB* contains the entire 619 bp of the late first exon. Together, these results indicate that transcription from the *sim* locus is driven from at least two discrete promoters, yielding at least three molecularly distinct transcripts. All utilize the same translational start site.

Discrete enhancers lie within the intergenic and intronic regions of the sim *locus*

Located approximately 1 kb upstream of the *sim* locus is the transcriptional unit encoding Damaged DNA Binding Protein 1 (DDB1), transcribed from the plus strand. To better understand the molecular genetics of the *sim* locus outside of the 7.8 kb region previously characterized and shown to harbor regulatory elements responsible for driving expression in the mesectoderm anlage in statu nascendi, mesectoderm anlage, and midline primordium stages, the approximately 1.0 kb region between *DDB1* and *sim* was cloned into pPTGAL, a vector carrying a basal promoter and the coding region for *GALA* flanked by P repeats for ease of integration into the insect genome (Sharma et al., 2002). To cover the large first intron, 2.8, 3.1, and 5.5 kb fragments were also cloned behind the *GALA* basal promoter of pPTGAL. Expression of a reporter from a responding (UAS) strain should reveal whether any enhancers exist within these fragments. The remaining 3' portion of the first intron is the region covered by the previously characterized 7.8 kb (7.8sim) construct.

Embryos transformed with the 1.0 kb fragment (represented by the 1.0FC2 line) showed enhancer-driven *GFP* expression in most (>80%) sim^+ cells of the CNS midline starting at embryonic stage 11 (Fig. 2A and 2B). By stage 12/3 (Fig. 2C and 2D), all sim^+ cells have begun

to express *GFP*, and this expression persists until the end of embryogenesis (Fig. 2E-2I). No GFP^+ cells were detected in the third instar larval brain (Figure 2J, Table 1).

The 2.8 kb fragment (represented by the 2.8FB1 line) yielded expression in myoblast precursor cells immediately laterad to the ventral nerve cord, as well as in midline glia (Crews, personal communication). Onset of expression is at stage 12 (Fig. 3B and 3C). Midline glial expression remained strong until stage 15, when expression started to diminish (Fig. 3F). By stage 16, all CNS expression was absent (Fig. 3G). Somatic mesoderm expression, however, persisted until the end of embryogenesis. At stage 12, myoblasts directly abut the developing ventral nerve cord. By stage 13, these cells have started to migrate away from the ventral nerve cord toward their eventual positions in the lateral body wall, and have undergone characteristic cell shape changes and cell fusion events as they differentiate (Fig. 3D). By stage 15, these syncytia occupied their positions in the larva and have terminally differentiated (Beckett and Baylies, 2006). No GFP^+ cells were detected in the third instar larval brain (Fig. 3H, Table 1).

Embryos transformed with the 3.1 kb fragment (represented by the 3.1FB3 transformant line) showed no expression until stage 12/3 (Fig. 4A and 4B). During this stage of embryonic development, *GFP* expression was observed in ~20 cells at the level of the mandibular and maxillary buds, site of the developing foregut and brain, respectively (Hartenstein, 1993). Comparison of lateral views of *GFP* expression at this stage with previously analyzed Sim localization in the embryonic foregut showed that these GFP^+ cells are the endodermal Sim^+ cells responsible for patterning the CNS at the level of the foregut (Page, 2003). Also at this stage, GFP^+ cells were observed in the caudal region. These cells were possibly components of the trachea in segment A8 or the posterior spiracles (at this stage, the posterior-most segment of the tracheal system has yet to fuse with its neighboring segment but is continuous with the posterior spiracle), although this remains untested. Expression in the foregut and caudal region

persist until the end of embryogenesis (Fig. 4C-4H). No GFP^+ cells were detected in the third instar larval brain (Figure 4I, Table 1).

The 5.5 kb fragment (represented by the 5.5FG1 transformant line) yielded no embryonic expression until stage 15. GFP localization was then observed in the pleural external transverse muscles and a subset of pleural external oblique muscles, as well as 1-2 midline glia/segment (Fig. 5A and 5B). Expression persisted until the end of embryogenesis largely unchanged. Less than 20 GFP^+ cells were detected in the third instar larval brain, none of which corresponded to the location of the Sim^+ cell clusters present in the central brain region (Fig. 5E, Table 1).

These results described the enhancer-driven *GFP* expression of embryos containing forward-oriented constructs; that is, the intergenic or intronic sequence was cloned 5' to 3' as it would occur in the *sim* locus; results from reverse-oriented (antiparallel) constructs were similar (data not shown). Many transgenic lines showed expression in tissues in which *sim* is not known to be expressed; we ascribe these extra sites of expression to position effects. Further, some of the lines tested displayed salivary gland expression, which arises from the presence of a salivary gland enhancer present in the *hsp70* sequences upstream of the GAL4 coding region (Duffy, 2002; Gerlitz et al., 2002; Sharma et al., 2002).

DISCUSSION

Here I report the identification of a third, previously unknown transcript product of the *sim* locus. I show that this third splice variant shares a common, putatively untranslated first exon with *sim-RB*, a previously characterized isoform, and that the third isoform differs from *sim-RB* in the 3' extent of the first exon. I demonstrate through RT-PCR that all three transcripts

are expressed during embryogenesis but the third transcript, *sim-RC*, is the only species transcribed post-embryonically. Sequence analysis of the regions flanking the 2 transcription start sites at this locus indicate the absence of recognizable TATA and DPE sequences. Using the bipartite UAS/GAL4 system, I show that discrete enhancers lie within the 5' intergenic region between the upstream gene *DDB1* and the *sim* locus, and within the large first intron that separates the untranslated first exon behind the late promoter for exons 2-8. Most of these enhancers are active embryonically but appear not to be active during larval development.

The formal possibility exists that more than 3 splice variants are transcribed from the *sim* locus which, due to the experimental design of our RT-PCR experiments, would remain undetected. However, Northern blot experiments previously performed (Crews, unpublished) revealed the presence of only 2 species separated by the molecular weight difference accountable by the size difference between the 2 alternatively spliced first exons.

Given the importance of the *sim* gene during embryonic development, it is perhaps not surprising that 3 splice variants can be detected during these stages. All 3 isoforms are thought to encode the same protein, detectable by both monoclonal and polyclonal antibodies raised against a portion of the unconserved C-terminus (the N-terminus contains the basic and helix-loop-helix domains, the latter of which is highly conserved across bHLH proteins present in the fly proteome). The difference between the 3 isoforms lies in the first exon. 5' untranslated regions have been implicated in mRNA stability, localization, and translation efficiency. Stability of mRNA transcripts may be mediated by the 5' (and 3') UTRs, due to varying affinity for certain RNases which can promote or inhibit the relative stability of the transcript. Since a more stable transcript may allow more protein to be transcribed from its template, expressing splice variants in different temporal and spatial manners would confer varying properties to the same protein. For example, changing the stability of *sim* transcripts from one tissue to the next might

alter the available pool of unbound Sim moiety in order to effect accelerated or retarded binding kinetics with its dimerization partner Tango, as needed. Subcellular localization of transcripts is another function conferred, in part, to untranslated regions. At present, it is unknown whether *sim* transcripts are translated in unique compartments within the cytoplasm, for example, in the soma or at the distal tips of growth cones in the neurons that express the transcription factor. Immunohistochemical evidence has repeatedly shown Sim protein to be detectable only in the nucleus, in agreement with its role as a transcription factor. However, this does not preclude the possibility that Sim is translated at lower levels in specific areas within the cytoplasm. This phenomenon and its importance remain unelucidated. Finally, UTRs may contain binding sites for proteins, or sequences which confer specific secondary structure, both of which may affect translational efficiency by sterically hindering the ribosome's access to the mRNA template.

The rationale for creating transgenic *Drosophila* lines carrying putative *sim* regulatory elements was two-fold. First, we wished to characterize the enhancers that drive *sim* expression in the embryonic and larval tissues in which the gene is expressed. Results presented here demonstrated that *sim* expression in the CNS midline is driven not only by enhancers located within the 1.0 kb upstream of the *sim* locus, but also by enhancers harbored within the 14.3 kb first intron. The set of differentiated midline cells that exist at the end of embryogenesis is composed of midline glia, MP1 neurons, the H-cell MP3 interneuron, the H-cell sib MP3 interneuron, iVUM interneurons, mVUM interneurons, the median neuroblast, and its progeny. Only the extant midline glia, the H-cell sib, the iVUM cells, the median neuroblast, and its progeny continue to express *sim* at the end of embryonic neurogenesis. *sim* expression is absent in the remaining midline cell types. Given the heterochronicity of *sim* expression in the midline cells, as well as the diverse neuronal and glial subtypes in which it is expressed, it is a formal possibility that *sim* is performing different functions in each of these midline cell subtypes. Early

studies on *sim* (Nambu et al., 1990; Nambu et al., 1991; Thomas et al., 1988) have cemented its role as the master regulator of the midline cell fate. However, its role in midline cells after they have been specified remains unelucidated.

We also wished to create GAL4 lines that would direct UAS expression in the Sim^+ cells of the larval central brain, in order develop assays to better understand the function of *sim* in those cells clusters. Previous studies have shown that removal of *sim* function in the central brain of the third instar larva leads to perturbations in axon fasciculation (Lau et al., in preparation; see Chapter 3). It will be useful to possess a reagent that specifically targets expression of transgenes encoding cytotoxins, dominant negative receptors, etc. to these cells to better understand the role of *sim* in the post-embryonic brain. However, none of these intergenic and intronic fragments directed GAL4 expression in the Sim^+ cells of the larval central brain. This is not due to a failure of the UAS/GAL4 system, as the use of this bipartite driver-responder system has been shown to be efficacious during all *Drosophila* life stages. The most obvious explanation is that the enhancers which direct post-embryonic expression are not contained in the regions analyzed in this current study. Another possibility is that the design of the constructs containing noncoding regions split the enhancers responsible for post-embryonic expression of the *sim* gene. However, a more parsimonious explanation lies in the possibility that genes with complex regulation, such as *sim*, often contain regulatory elements harbored within large regions, presumably to shield the effect of one element from another. For example, the *bithorax* complex contains three hox genes in a 300 kb region rife with regulatory elements (Lewis, 1998). Given the relatively large number of discrete tissues and chronologies in which the *sim* gene is expressed, it is possible that enhancers that drive post-embryonic expression lie distal to the locus in regions unexplored in this study. These possibilities may also serve to explain why the design of this study failed to reveal the enhancer that drives *sim* expression in the hindgut.

General properties of transcriptional regulation in eukaryotic systems are highly conserved (Lee and Young, 2000), therefore the findings in this study may be applicable to other insects. Alternate forms of the TBP protein have been found in *Drosophila* that possess tissue-specific expression and sequence specificity (Freiman et al., 2001; Verrijzer, 2001); therefore, different basal transcription complexes may assemble on different basal promoters. That the *sim* locus contains neither canonical TATA or DPE sequences represents yet another variable which may contribute to the specialized expression of multiple isoforms in different tissues and at different time points during the development of the animal.

ACKNOWLEDGEMENTS

The authors would like to thank Joe Kearney for critical reading of the manuscript, Daniel Eberl for the pPTGAL vector, and Bloomington *Drosophila* Stock Center for fly stocks. This work was supported by NIH grant 5R37HD025251 to STC.

FIGURES

Figure 1. (A) RT-PCR results show a third *sim* isoform expressed during embryogenesis. Total RNA was extracted and reverse transcribed from the following animals or tissues and used as a complex template for PCR: 0-3 h AEL, 3-6 h AEL, 6-9 h AEL, 9-12 h AEL, 12-15 h AEL embryos; first, second, and third instar larvae, and third instar larval brains; light (days 1-2 of pupation) and dark (days 3-4) pupae; whole 2-day-old adults (equal numbers of males & females), 2-day-old adult heads, and 2-day-old adult thoraxes and abdomens. No PCR products were observed in the negative controls, as follows: PCR reaction without template (-Temp.), and PCR reaction carried out without reverse transcriptase added to the template preparation. RpL32, a structural constituent of the ribosome, was co-amplified in each reaction and serves as a loading control. M=marker. (B) To-scale schematic of the *sim* locus shows the alternate splice identities of *sim-RA*, *sim-RB*, and *sim-RC* transcripts. PCR products from (A) were sequenced and resulting data compared against the known *sim* sequence. Results indicated that all three transcripts differ only in the first exon; *sim-RA* contains the so-called early first exon, while *sim-RB* and *sim-RC* share the larger late first exon. *sim-RB* contains the entire late first exon, while *sim-RC* contains a truncated version. Also shown are the relative locations of inter- and intragenic sequences cloned into pPTGAL for bipartite expression analysis. Not shown is a previously characterized 7.8 kb fragment whose sequence begins at approximately 11.5 kb as denoted on the locus scale, covering the remainder of the large first intron and terminating within the 8th exon. Thus, the combination of the fragments created in this study (1.0, 2.8. 3.1, and 5.5) and the 7.8 kb fragment ensured total coverage of the upstream intergenic region relative to the *sim* locus as well as the first intron. Also not shown is the DDB1 gene which resides approximately 1.0 kb upstream of the *sim-RB*, *sim-RC* transcription start site.

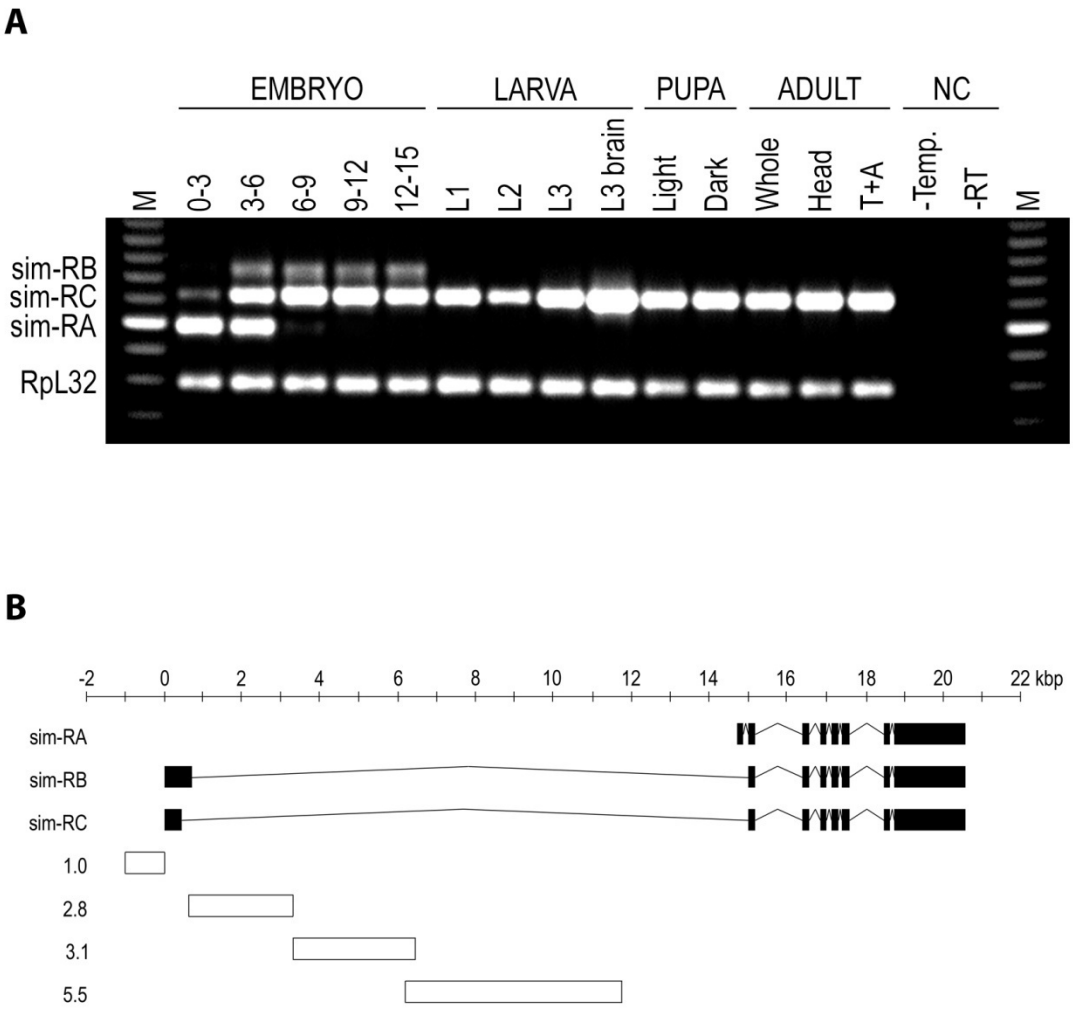
A
EMBRYO
LARVA
PUPA
ADULT
NC
M
0-3
3-6
6-9
9-12
12-15
L1
L2
L3
L3 brain
Light
Dark
Whole
Head
T+A
-Temp.
-RT
M
sim-RB
sim-RC
sim-RA
RpL32
B
-2
0
2
4
6
8
10
12
14
16
18
20
22 kbp
sim-RA
sim-RB
sim-RC
1.0
2.8
3.1
5.5

Figure 1

Figure 2. The 1.0 kb fragment drives expression in most Sim^+ cells of the embryonic midline. (A-I) Transgenic animals were crossed with UAS-τ-GFP responder lines and chromogenically stained for GFP. (A,C,E,G,I) 10x magnification. (B,D,F,H) 40x magnification. Lateral view of stage 11 (A,B), stage 12 (C,D), stage 13 (E,F), stage 14 (G,H), and stage 15 (I) embryos show staining on the cells of the CNS midline. No staining was observed during prior stages (not shown), but transgene expression persists until the end of embryogenesis (not shown). Expression driven by the 1.0 kb fragment phenocopies native *sim* expression in the CNS. Notably absent is non-CNS expression such as that found transiently in the myoblasts, foregut, and hindgut. (J) A compiled confocal stack imaging an unstained third instar larval brain of identical genotype shows no GFP localization in the central brain cells or elsewhere. Rod-like structures are autofluorescent trachea that have ramified the CNS.

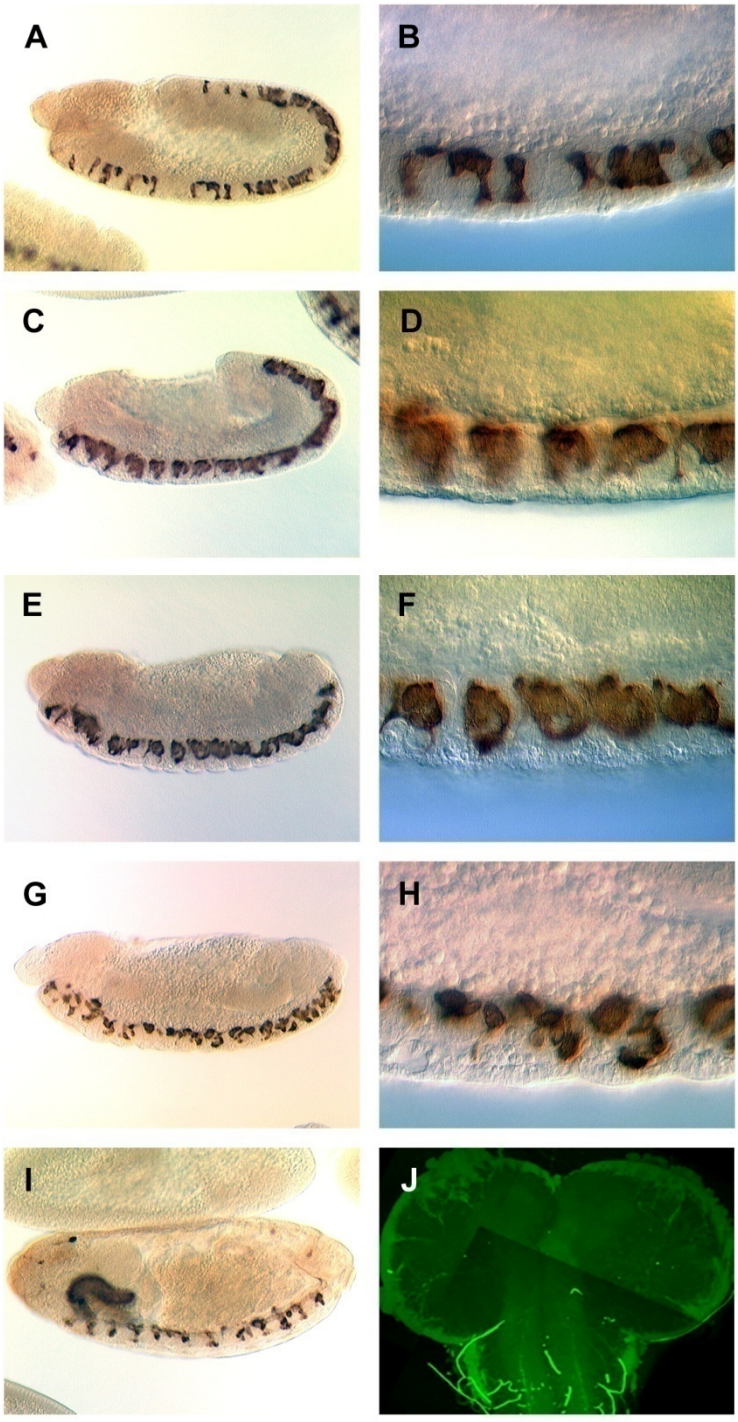

Figure 2

Figure 3. The 2.8 kb fragment drives expression in a subset of midline glia and in muscle precursors (myoblasts). All embryonic images are 10x. (A-G) Transgenic animals were crossed with UAS-τ-GFP responder lines and chromogenically stained for GFP. Ventral view of stage 11 (A), stage 12 (B), stage 13 (D), stage 14 (E), stage 15 (F), and stage 16 (G) embryos show a developmental time-course of transgene expression. Lateral view of stage 12 embryo (C). Prior to stage 12, no expression was observed. Beginning at stage 12, expression was observed bilaterally in differentiating myoblasts as well as in a subset of midline glia. As the myoblasts migrate away from their ventral positions abutting the CNS, they undergo stereotypic cell shape changes until they attain their final positions in the lateral body wall by stage 16. Concomitantly, glial expression begins to diminish starting at stage 15. (H) A compiled confocal stack imaging an unstained third instar larval brain of identical genotype shows no GFP localization in cells of the central brain region or elsewhere. Rod-like structures are autofluorescent trachea. Puncta of autofluorescence observed in the anterior extent (top) are caused by cells of the fat bodies.

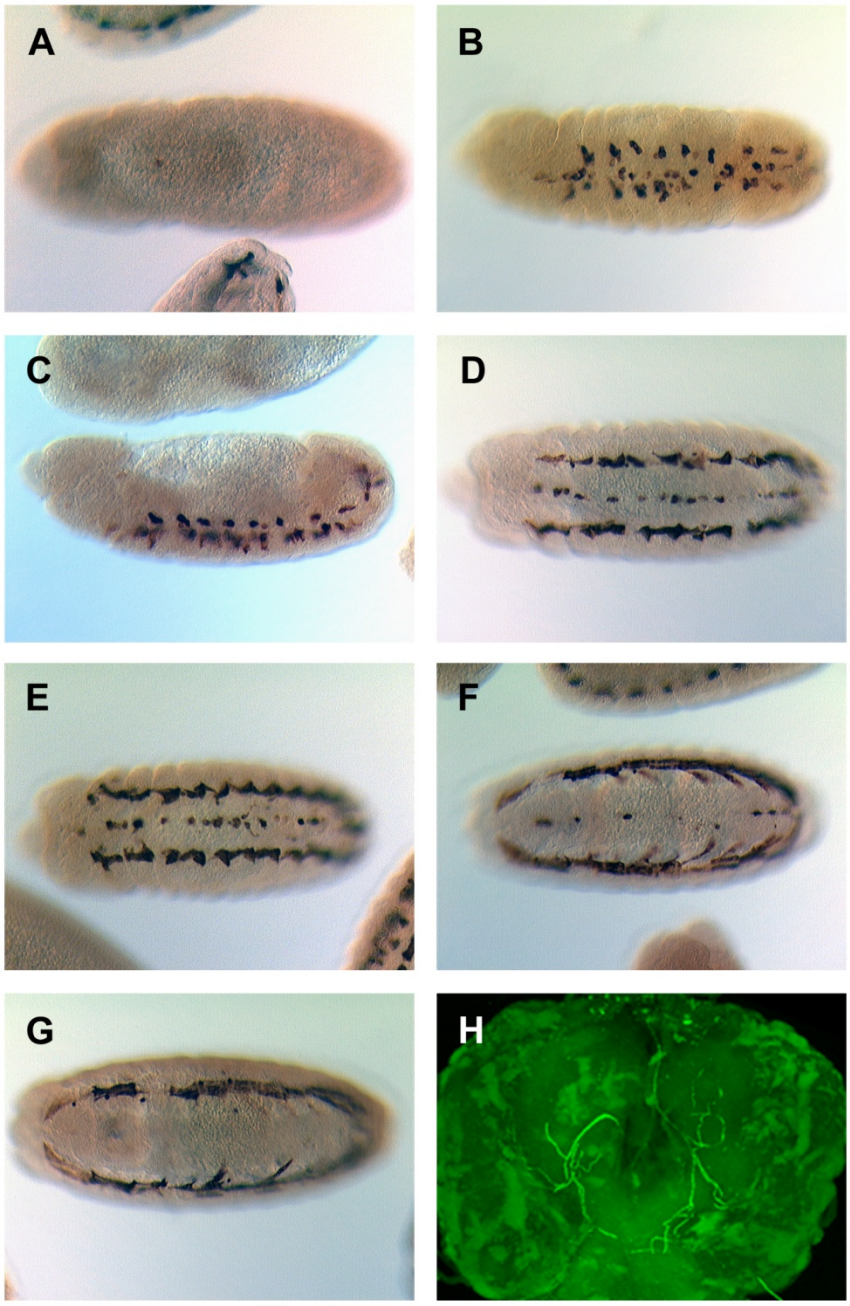

Figure 3

Figure 4. The 3.1 kb fragment drives expression in the developing foregut and in cells caudally. Transgenic animals were crossed with UAS-τ-GFP responder lines and chromogenically stained for GFP. (A,B,C,D,F) 10x magnification. (E,G,H) 40x magnification. (A,B) Dorsal and lateral views, respectively, of stage 12/3 embryos. GFP localization was observed in <20 cells anteriorly and in an unidentified paired structure caudally; the posterior structures are possibly components of the trachea which have not yet fused with their counterparts from the adjacent segment. No expression was observed prior to stage 12/3. (C,D) Ventral and lateral views, respectively, of stage 13 embryos showing continued localization in the mandibular and maxillary buds, and in the paired posterior structures. Transgene expression was observed in discrete cells along the length of the embryo; however, these sites of expression correspond to no known tissues in which *sim* is endogenously expressed and are therefore ascribed to position effects. (E) 40x view of GFP localization in the cells of the embryonic foregut. (F) Ventral view of a stage 17 embryo showing continued localization of GFP in the developing foregut and posterior structures. GFP localization was also observed in cells of the salivary glands, again due to position effects. (G,H) 40x views of lateral foregut and ventral posterior extent at this stage, respectively. (I) A compiled confocal stack imaging an unstained third instar larval brain of identical genotype shows no GFP localization in cells of the central brain region or elsewhere. Rod-like structures are autofluorescent trachea.

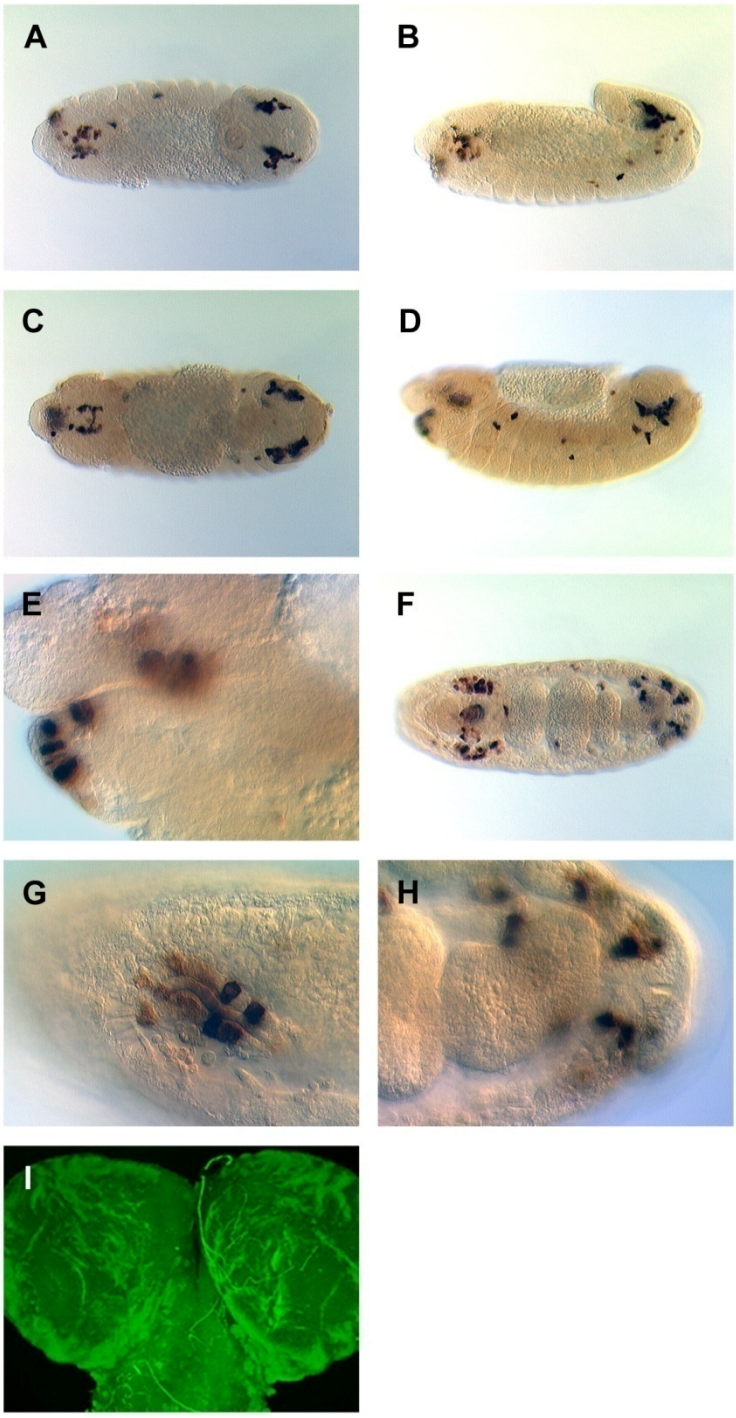

Figure 4

Figure 5. The 5.5 kb fragment drives expression in a subset of pleural muscles and CNS midline glia. Transgenic animals were crossed with UAS-τ-GFP responder lines and chromogenically stained for GFP. All embryos shown in ventral view. (A,C) 10x magnification. (B,D) 40x magnification. (A,B) stage 15 embryos. (C,D) stage 16 embryos. GFP localization was observed in a subset of pleural muscles, as well as 1-2 midline glia/segment. No GFP localization was observed prior to stage 15. (E) A compiled confocal stack imaging an unstained third instar larval brain of identical genotype shows GFP localization in <20 cells of the central brain; however, these cells do not correspond with endogenous sites of *sim* expression in paired clusters during this stage. *GFP* expression was also observed in a subset of longitudinal axons in the ventral nerve cord. Whether these neurites are associated with GFP^+ cells of the brain proper is undetermined.

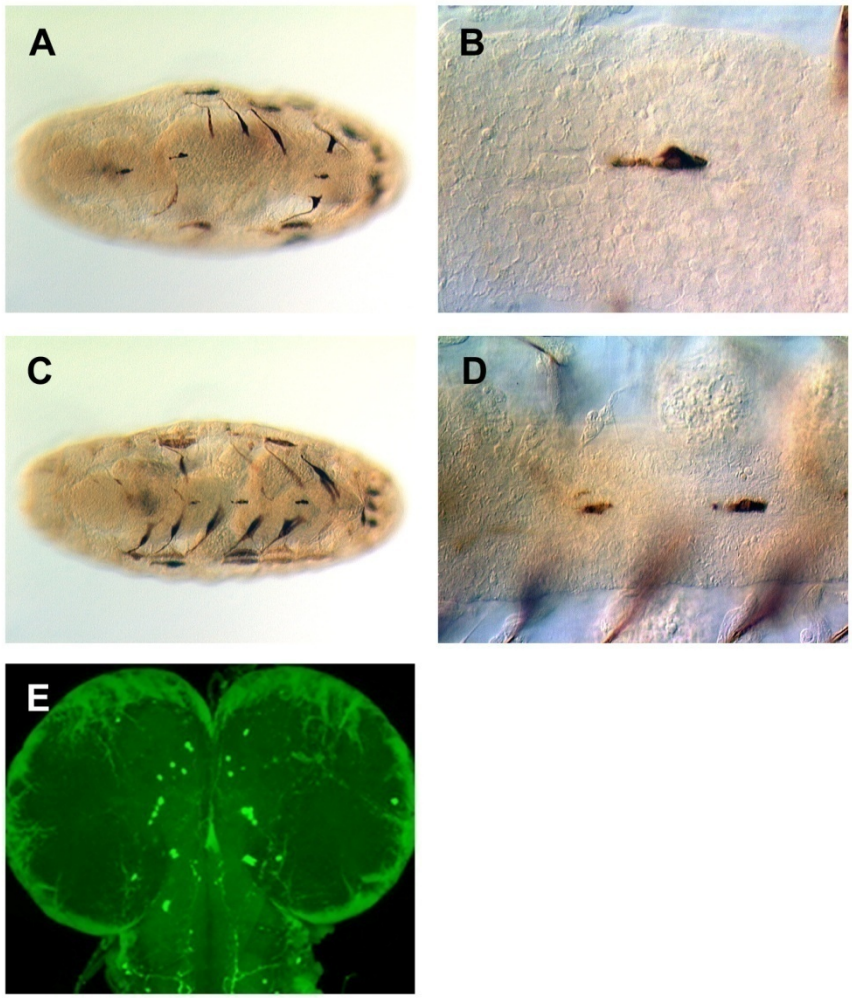

Figure 5

TABLES

Table 1. Summary of driver (GAL4) lines created and their expression domains. Only domains in which *sim* is natively expressed are listed.

	EMBRYO				LARVA		
Fragment	CNS Midline	Foregut	Hindgut	Myoblasts	CNS Midline	Central Brain	Optic Lobes
1.0	X						
2.8	X			X			
3.1		X					
5.5	X			X			

Table 1

REFERENCES

Arkhipova, I. R. (1995). Promoter Elements in *Drosophila Melanogaster* Revealed by Sequence Analysis. *Genetics* **139,** 1359-1369.

Ashburner, M. (1989). *Drosophila: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Beckett, K. and Baylies, M. K. (2006). The Development of the Drosophila Larval Body Wall Muscles. *Int. Rev. Neurobiol.* **75,** 55-70.

Bossing, T. and Technau, G. M. (1994). The Fate of the CNS Midline Progenitors in *Drosophila* as Revealed by a New Method for Single Cell Labelling. *Development* **120,** 1895-1906.

Brand, A. H. and Perrimon, N. (1993). Targeted Gene Expression as a Means of Altering Cell Fates and Generating Dominant Phenotypes. *Development* **118,** 401-415.

Cherbas, L. and Cherbas, P. (1993). The Arthropod Initiator: The Capsite Consensus Plays an Important Role in Transcription. *Insect Biochem. Mol. Biol.* **23,** 81-90.

Duffy, J. B. (2002). GAL4 System in Drosophila: A Fly Geneticist's Swiss Army Knife. *Genesis* **34,** 1-15.

Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., Saijoh, Y., Fujii, H., Hamada, H., Kikuchi, Y. et al. (1996). Two New Members of the Murine *Sim* gene Family are Transcriptional Repressors and show Different Expression Patterns during Mouse Embryogenesis. *Mol. Cell. Biol.* **16,** 5865-5875.

Estes, P., Mosher, J. and Crews, S. T. (2001). *Drosophila Single-Minded* Represses Gene Transcription by Activating the Expression of Repressive Factors. *Dev. Biol.* **232,** 157-175.

Freiman, R. N., Albright, S. R., Zheng, S., Sha, W. C., Hammer, R. E. and Tjian, R. (2001). Requirement of Tissue-Selective TBP-Associated Factor TAFII105 in Ovarian Development. *Science* **293,** 2084-2087.

Gerlitz, O., Nellen, D., Ottiger, M. and Basler, K. (2002). A Screen for Genes Expressed in Drosophila Imaginal Discs. *Int. J. Dev. Biol.* **46,** 173-176.

Hartenstein, V. (1993). *Atlas of Drosophila development*, pp. 57. Plainview, NY: Cold Spring Harbor Laboratory Press.

Kearney, J. B., Wheeler, S. R., Estes, P., Parente, B. and Crews, S. T. (2004). Gene Expression Profiling of the Developing *Drosophila* CNS Midline Cells. *Dev. Biol.* **275,** 473-492.

Kutach, A. K. and Kadonaga, J. T. (2000). The Downstream Promoter Element DPE Appears to be as Widely used as the TATA Box in *Drosophila* core Promoters. *Mol. Cell. Biol.* **20,** 4754-4764.

Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D. and Ebright, R. H. (1998). New Core Promoter Element in RNA Polymerase II-Dependent Transcription: Sequence-Specific DNA Binding by Transcription Factor IIB. *Genes Dev.* **12,** 34-44.

Lee, T. I. and Young, R. A. (2000). Transcription of Eukaryotic Protein-Coding Genes. *Annu. Rev. Genet.* **34,** 77-137.

Lewis, E. B. (1998). The Bithorax Complex: The First Fifty Years. *Int. J. Dev. Biol.* **42,** 403-415.

Lewis, J. O. and Crews, S. T. (1994). Genetic Analysis of the *Drosophila Single-Minded* Gene Reveals a Central Nervous System Influence on Muscle Development. *Mech. Dev.* **48,** 81-91.

Ma, Y., Certel, K., Gao, Y., Niemitz, E., Mosher, J., Mukherjee, A., Mutsuddi, M., Huseinovic, N., Crews, S. T., Johnson, W. A. et al. (2000). Functional Interactions between *Drosophila* bHLH/PAS, Sox, and POU Transcription Factors Regulate CNS Midline Expression of the *Slit* gene. *J. Neurosci.* **20,** 4596-605.

Moffett, P. and Pelletier, J. (2000). Different Transcriptional Properties of mSim-1 and mSim-2. *FEBS Lett.* **466,** 80-86.

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, K. A.,Jr 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.

Page, D. T. (2003). A Function for Egf Receptor Signaling in Expanding the Developing Brain in *Drosophila*. *Curr. Biol.* **13,** 474-482.

Patel, N. H., Snow, P. M. and Goodman, C. S. (1987). Characterization and Cloning of *Fasciclin III*: A Glycoprotein Expressed on a Subset of Neurons and Axon Pathways in *Drosophila*. *Cell* **48,** 975-988.

Pielage, J., Steffes, G., Lau, D. C., Parente, B. A., Crews, S. T., Strauss, R. and Klambt, C. (2002). Novel Behavioral and Developmental Defects Associated with *Drosophila Single-Minded*. *Dev. Biol.* **249,** 283-299.

Probst, M. R., Fan, C. M., Tessier-Lavigne, M. and Hankinson, O. (1997). Two Murine Homologs of the *Drosophila* Single-Minded Protein that Interact with the Mouse Aryl Hydrocarbon Receptor Nuclear Translocator Protein. *J. Biol. Chem.* **272,** 4451-4457.

Sanchez-Soriano, N. and Russell, S. (1998). The *Drosophila* SOX-Domain Protein Dichaete is Required for the Development of the Central Nervous System Midline. *Development* **125,** 3989-3996.

Sharma, Y., Cheung, U., Larsen, E. W. and Eberl, D. F. (2002). PPTGAL, a Convenient Gal4 P-Element Vector for Testing Expression of Enhancer Fragments in *Drosophila*. *Genesis* **34,** 115-118.

Sonnenfeld, M., Ward, M., Nystrom, G., Mosher, J., Stahl, S. and Crews, S. (1997). The *Drosophila Tango* Gene Encodes a bHLH-PAS Protein that is Orthologous to Mammalian *Arnt* and Controls CNS Midline and Tracheal Development. *Development* **124,** 4583-4594.

Thomas, J. B., Crews, S. T. and Goodman, C. S. (1988). Molecular Genetics of the *Single-Minded* Locus: A Gene Involved in the Development of the *Drosophila* nervous System. *Cell* **52,** 133-141.

Verrijzer, C. P. (2001). Transcription Factor IID--Not so Basal After all. *Science* **293,** 2010-2011.

Ward, M. P., Mosher, J. T. and Crews, S. T. (1998). Regulation of *Drosophila* bHLH-PAS Protein Cellular Localization during Embryogenesis. *Development* **125,** 1599-1608.

Wheeler, S. R., Kearney, J. B., Guardiola, A. R. and Crews, S. T. (2006). Single-Cell Mapping of Neural and Glial Gene Expression in the Developing *Drosophila* CNS Midline Cells. *Dev. Biol.* **294,** 509-524.

Xiao, H., Hrdlicka, L. A. and Nambu, J. R. (1996). Alternate Functions of the *Single-Minded* and *Rhomboid* Genes in Development of the *Drosophila* Ventral Neuroectoderm. *Mech. Dev.* **58,** 65-74.

CHAPTER 5:

DISCUSSION

Summary of results

In this dissertation I presented data that showed novel sites of expression and novel functions for *sim*, and have analyzed the *sim* locus to better understand its complex regulation.

My German collaborators created the only known temperature sensitive allele of *sim*, and increased the number of available strains in the allelic series to 13. Sequencing of this novel allele revealed an amino acid residue substitution at a position that is conserved across taxa. The substitution occurs within helix 2 of the bHLH domain, and probably influences protein dimerization, DNA binding, or both. At restrictive temperature, mutant embryos exhibited disruptions in anal pad morphology and in the development of genital discs. Mutant larvae displayed defective cuticle patterning. Mutant adults reared at permissive temperature were sterile due to gonadal aplasia coupled with perturbations in external genitalia formation. In trans to sim^{H9}, a protein null allele, adults failed to display normal courtship behavior. These individuals walked in circles of varying tightness and showed no interest in visual cues; the former is a possible read-out of miscommunication between brain hemispheres, while the latter was not due to blindness. Histological analysis of brains from these transheterozygous mutants showed gross perturbations in the morphology of the optic lobes and in the central complex. The above described phenotypes suggest that *sim* is expressed post-embryonically, and this was

confirmed by immunohistochemical investigation as well as by RT-PCR. In the larval CNS, Sim protein was found in the nuclei of cells in midline glia of the ventral nerve cord, the lamina and medulla of the optic lobes, and in 3 paired clusters of neurons in the central brain.

Taken together, these data indicate a role for *sim* in proper patterning of the larval CNS. I narrowed my focus to the examination of *sim* in the larval central brain.

Further immunohistochemical investigation revealed that the increase in number of Sim-positive cells proceeded saltatorily during development, with gradual expansion during embryonic, first instar, and second instar larval stages, and a significant expansion during the third instar larval stage. This finding was in agreement with previously known brain neuroblast division kinetics. Further, it was found that *sim* expression in the larval brain was confined to post-mitotic neurons and was absent in neuroblasts and ganglion mother cells, in contrast to its expression profile in the embryonic CNS midline where *sim* is expressed in both precursor cells, such as the median neuroblast, as well as in the neuroblast's neuronal and glial progeny. In the third instar larval brain, each Sim-positive central brain cluster of neurons projected axons as a tight fascicle, a hallmark property of cells born from the same neuroblast. The two anterior-most cluster pairs were each found to be composed of contributions from two discrete but abutting sets of Sim-positive cells. The anterior-most cluster's fascicle extended across the supraesophageal commissure to the contralateral neuropil, where it fasciculated with its bilateral partner emanating from the other hemisphere. This finding indicated that the anterior-most Sim-positive neurons in the central brain participate in interhemispheric communication. Using the MARCM strategy for creating marked mutant clones, removal of *sim* function from larval central brain cells was found to cause a marked perturbation in axon morphology. The *sim* phenotype was consistent across 4 different alleles used in the MARCM strategy, indicating a general (as opposed to allele-specific) effect. In contrast to the early function of *sim* in the embryonic CNS

midline, which is to invoke the mesectodermal cell fate by activating mesectoderm-specific genes while concomitantly repressing lateral CNS genes, no differences in cell fate decisions or proliferation kinetics between mutant and wild-type central brain cells were found. Taken together, these data indicate *sim* is performing a role in axonogenesis as opposed to neurogenesis in the larval central brain.

All available *sim* alleles were sequenced and analyzed and were found to contain mutations in either coding region or splice sites that could lead to loss of *sim* function.

To better understand the genetic regulation of *sim*, I performed bioinformatic searches for the TATA motif as well as the downstream promoter element and *Drosophila*-specific Initiator element and found none. I concluded that *sim* utilizes cryptic promoters, previously shown to be utilized in approximately one third of arthropod genes. I performed RT-PCR and discovered a third mRNA species given by the locus, termed *sim-RC*, and found that this species is the major contributor to *sim* expression during late embryogenesis, and is the sole contributor post-embryonically. Sequencing showed that the two mRNA species expressed from the late promoter, *sim-RB* and *sim-RC*, share a common untranslated first exon but differ in the 5' splice site. The *sim-RA* species expressed from the early promoter utilizes a completely different untranslated first exon. Between *sim* and its neighboring upstream gene is approximately 1.0 kb of non-coding sequence; 14.0 kb separates the early and late alternatively spliced first exons. Hypothesizing that these regions harbor regulatory sequences important for *sim* expression, portions of these regions were cloned into a GAL4 driver to assay for the presence of such sequences. The 1.0 kb intergenic fragment was found to contain all the regulatory elements required to drive *sim* expression in the embryonic CNS midline. A 2.8 kb fragment taken from the large first intron drove expression in the developing myoblasts, where *sim* is transiently expressed during early to mid-embryonic development. That construct also drove expression in

midline glia of the CNS. A 3.1 kb fragment, also taken from the large first intron, drove expression in the foregut, where *sim* was previously reported to activate Egfr signaling which in turn induces the midbrain to proliferate. A 5.5 kb fragment, again from the large first intron, drove expression in myoblast precursors and in a small subset of embryonic CNS midline cells. None of these constructs was able to drive expression in Sim-positive cells of the larval brain. It is probable that the post-embryonic enhancers were not captured in the tested fragments; less likely but still a formal possibility is that the fragments split the necessary enhancers.

Weaknesses and their solutions

The sim^{J147} allele was sequenced and found to contain a serine to phenylalanine substitution at residue 41, corresponding to the second helix domain. A residue at this location potentially participates in DNA binding, co-factor binding, or both. Serine is a small, polar (hydrophilic) amino acid while phenylalanine contains a bulky benzyl side chain and is nonpolar (hydrophobic). It is possible that such a substitution creates a kink in the local helical secondary structure. This can be tested by employing site-directed mutagenesis to substitute serine-41 for methionine, a nonpolar amino acid with a bulky side chain, with similar biochemical properties to phenylalanine. If flies harboring this mutation recapitulate the behavioral and morphological deficiencies found in flies transheterozygous for sim^{J147}, the effect can be ascribed to a disruption in protein folding caused by residue 41.

The MARCM strategy for creating marked mutant clones enabled me to circumvent the early requirement for *sim* in embryonic development. Every cell in the animal is heterozygous for *sim* until mitotic recombination is induced via heat shock. However, in practice, targeting the Sim-positive cell clusters in the third instar larval central brain was a low frequency event. A

large quantity of time and reagents was expended in harvesting, preparing, and interrogating brains only to find MARCM clones created in cells other than those under study. Unfortunately, no ready solution is evident. An alternate strategy for targeting Sim-positive cells would require knowledge of enhancers that drive gene expression in specifically those cells and none other, and at the appropriate time in development. No enhancers are known to possess this level of specificity at this time. An effort is underway in the Crews group to find central brain enhancers (performed by Stephanie Freer).

Regarding the hunt for tissue-specific enhancers, the cloning strategy employed was expeditious but came at the cost of resolution. Creating 4 driver lines, as opposed to a greater number, enabled relatively quick progress toward results of "high granularity". For example, the 2.8 kb fragment drove transgene expression in midline glia and myoblasts. Breaking that fragment into smaller sequences may enable the midline glial enhancer to be isolated; however, one benefit to the strategy of using larger sequences is to lower the possibility of splitting enhancer elements. An effort is underway in the Crews group to increase the resolution of the enhancer fragments, again performed by Stephanie Freer. To abrogate position effects, this second effort is utilizing the ΦC31 site-specific transgene integration strategy. Greater effort in identifying the specific cells in which the GAL4 driver is active can yield useful reagents for the Crews group, especially with respect to the embryonic CNS midline. For example, the 1.0 kb, 2.8 kb, and 5.5 kb fragments drove expression in a subset of embryonic midline cells. The exact identity of those midline cells can be elucidated via co-staining with known midline cell-specific markers. The foregut cells in which the 3.1 kb fragment drives expression can be identified unambiguously using mAb BP102 and anti-Sim antibodies in the transgenic background. BP102 will decorate the axon projections in a characteristic ring formation, and will act as a landmark for pharyngeal Sim-positive cells. The expectation is that anti-Sim antibody and the UAS

responder (for example, GFP) co-localize. These driver lines can then be used to drive cytotoxins, dominant negative receptors, etc. to answer questions of cell autonomy and signaling events as they pertain to the CNS midline.

What have we learned?

Parsimony of gene usage is a recurrent principle in metazoan development. Prior to the work presented in this dissertation, most of what was known about *sim* pertained to its role as the master regulator of the embryonic CNS midline cell fate (its role in myoblasts remains uncertain). We can now ascribe post-embryonic functions to what was previously regarded as a developmentally important selector gene. The *sim* gene participates in proper development of embryonic genital disc formation; larval cuticle patterning and axonal pathfinding in the central brain; and adult brain patterning, genital and gonadal development, and behavior. The *sim* gene can properly be considered pleiotropic.

Future directions

The *sim* gene is expressed in the lamina and medulla of the larval optic lobes, as well as in a subset of glia in the adult brain. It would be of interest to elucidate *sim*'s function in those compartments. Some promising results have already been obtained from the Crews group: one of the aforementioned "low granularity" ΦC31 GAL4 strains (from the large first intron) drives expression in the lamina, and another strain (from the third intron) drives expression in a subset of medullary neurons.

It would also be of interest to discover the synaptic targets of the Sim-positive cells in the larval central brain, and the transcriptional targets of *sim* that underlie the axon misguidance and defasciculation phenotype. Clearly, more work remains to be done before we can arrive at a holistic understanding of *sim*; however, the work presented in this dissertation has contributed good progress to that end, and to our understanding of the biology of *Drosophila*.