Development 124, 4571-4582 (1997)
Printed in Great Britain © The Company of Biologists Limited 1997
DEV5152

The *Drosophila tango* gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development

Margaret Sonnenfeld, Mary Ward, Gerald Nystrom, Jack Mosher, Stephanie Stahl and Stephen Crews

Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA
*Author for correspondence (e-mail: steve_crews@unc.edu)

SUMMARY

The *Drosophila single-minded* and *trachealess* bHLH-PAS genes control transcription and development of the CNS midline cell lineage and tracheal tubules, respectively. We show that Single-minded and Trachealess activate transcription by forming dimers with the *Drosophila* Tango protein that is an orthologue of the mammalian Arnt protein. Both cell culture and in vivo studies show that a DNA enhancer element acts as a binding site for both Single-minded::Tango and Trachealess::Tango heterodimers and functions in controlling CNS midline and tracheal transcription. Isolation and analysis of *tango* mutants reveal CNS midline and tracheal defects, and gene dosage studies demonstrate in vivo interactions between *single-minded::tango* and *trachealess::tango*. These experiments support the existence of an evolutionarily conserved, functionally diverse bHLH-PAS protein regulatory system.

Key words: *Arnt*, bHLH, development, PAS, *single-minded*, *tango*, *trachealess*, transcription factor, *Drosophila*

INTRODUCTION

The PAS domain is a large, multifunctional interaction domain found within a family of proteins that control a wide variety of biological processes in vertebrates, invertebrates, plants, and prokaryotes. These functions include neurogenesis, tubule formation, carcinogen metabolism, response to hypoxia, and biological rhythms. Most of the PAS proteins studied in mammals and *Drosophila* also contain a basic-helix-loop-helix (bHLH) domain that implicates them as DNA-binding transcription factors. One important aspect of PAS protein function is that some members can act as receptors and mediate cell signaling pathways. The mammalian aromatic hydrocarbon receptor complex (AHRC; also dioxin receptor) is one of the paradigmatic members of the bHLH-PAS family (Hankinson, 1995). The aromatic hydrocarbon receptor (Ahr) bHLH-PAS protein is localized in the cytoplasm in the ligand-free state. When ligand (e.g. dioxin) enters the cell, it binds Ahr, which then enter the nucleus. The Ahr-ligand complex binds another bHLH-PAS protein, Aromatic hydrocarbon receptor nuclear translocator (Arnt). The Ahr::Arnt heterodimer binds DNA sequence elements with a core GCGTG sequence, referred to as the xenobiotic response element (XRE), that reside within target genes such as that encoding Cytochrome P_1450, and activates their transcription. One of the key questions regarding PAS protein function is the generality of this paradigm. In particular, is the function of Arnt evolutionarily conserved, and does it function as a dimerization partner for multiple bHLH-PAS proteins?

The *Drosophila single-minded* (*sim*) bHLH-PAS gene is a master regulatory gene that controls the development of the neurons and glia that lie along the midline of the central nervous system (CNS). Genetic and ectopic expression studies reveal that *sim* function is required in the midline cells for activation of CNS midline gene expression (Nambu et al., 1990, 1991) and repression of lateral neuroectodermal expression (Chang et al., 1993; Mellerick and Nirenberg, 1995; Xiao et al., 1996). Thus, it acts as a genetic switch. When turned on in the neuroectoderm as cell lineages are being specified, *sim* causes the cells to enter the CNS midline cell lineage. Recently, two mammalian orthologues of *sim* (*Sim1* and *Sim2*) have been discovered (Michaud and Fan, 1997). Both are expressed in the developing CNS. *Sim2* is a candidate gene for Down Syndrome, based on its expression pattern and localization to chromosome 21. Transgenic and mutational analysis of target genes of *Drosophila sim* have identified a *cis*-regulatory CNS midline element (CME) that controls transcription of CNS midline cell transcription by *sim* (Wharton and Crews, 1993; Wharton et al., 1994). The CME has a core sequence of ACGTG. When multimerized, it is sufficient for CNS midline transcription in vivo. It was hypothesized, based on analogy to the AHRC, that Sim forms a dimer with a *Drosophila* Arnt-related protein to bind the CME and activate CNS midline transcription (Wharton et al., 1994). The work described in this paper describes a *Drosophila* Arnt-like protein and demonstrates that it functions as predicted.

More recently, several novel *Drosophila* and mammalian bHLH-PAS proteins have been identified. These include *Drosophila trachealess* (*trh*) (Isaac and Andrew, 1996; Wilk et al., 1996), the mammalian hypoxia inducible factor (HIF) (Wang et al., 1995), *Drosophila similar* (*sima*) (Nambu et al., 1996), and murine *Clock* (King et al., 1997). The *trh* gene plays a prominent role in tracheal development, by controlling the

expression of genes involved in tracheal tubule formation. It also is involved in formation of the posterior spiracle and salivary gland duct. *trh* is specifically expressed in tracheal cells and the salivary gland. HIF controls the cellular response to hypoxia by activating the transcription of genes that protect the organism against oxygen deprivation. HIF consists of a heterodimer in which one subunit (HIF-1α) is related to Sim and the other (HIF-1β) is Arnt. *Drosophila sima* is most closely related to HIF-1α and is broadly expressed in the embryo. This suggests that Sima may play a role in *Drosophila* hypoxia induction, but *sima* mutations have not been isolated and its biological function remains unknown. The *Clock* gene controls the periodicity of biological rhythms in mice, and is postulated to activate circadian rhythm gene transcription. Although the Clock, Sim, Sima, Trh proteins are closely related, their interaction partners and mechanisms of gene control are not well understood.

In this paper, we report the cloning of the *Drosophila tango* (*tgo*) gene. Sequence, biochemical, and expression data indicate that *tgo* is highly related to mammalian *Arnt*. Tgo can dimerize strongly with both Sim, Sima and Trh and is able to activate gene transcription via the CME. Analysis of transgenic fly strains indicate that the CME acts in vivo as both a CNS midline and tracheal enhancer element, consistent with it being a regulatory element for Sim::Tgo and Trh::Tgo heterodimer binding. Mutations in the *tgo* gene reveal both CNS midline and tracheal defects, and genetic interaction studies suggest that *sim* and *trh* interact with *tgo* in vivo. These results provide in vivo evidence that Tgo/Arnt functions as a broadly expressed dimerization partner for a number of specifically expressed bHLH-PAS proteins that control a variety of developmental and physiological processes.

MATERIALS AND METHODS

Drosophila strains

The following mutant loci were used: sim^{H9} (null), trh^{8} (severe), neu^{j6B12} (P[lacW] lethal insertion), and *Df(3R)p-XT9* (84F14; 85C-D). Enhancer trap and P[lacZ] transgenes used were: AA142 and X55 enhancer trap lines (Klämbt et al., 1991), P[w^{+}; 4xCME-lacZ] transgenic line that contains a 4-fold multimerized *Toll* site 4 CME cloned into the C4PLZ enhancer tester vector (Wharton et al., 1994). The third chromosome sim^{H9}, tgo^{1}, and trh^{8} mutations were recombined individually onto a strain containing a P[4xCME-lacZ] that genetically maps to the distal tip of 3R.

Isolation of *Drosophila tgo* genomic and cDNA clones

The probe used to isolate the *Drosophila tgo* gene was derived from the bHLH region of the human *ArntM1* cDNA clone (Hoffman et al., 1991). PCR was used to generate a fragment of 206 bp (nt 208-415) that included all of the *Arnt* bHLH region. Genomic clones were isolated from an EMBL3 bacteriophage library of isogenic *dp cl cn bw* DNA (R. Blackman, University of Illinois) using moderately stringent hybridization conditions consisting of incubation in 6× SSPE/50% formamide/10% dextran sulfate/0.1% SDS/0.1% NaPPi/5× Denhardt's at 42°C for 18 hours with a 10^{7} disints per minute/ml probe followed by 3 washes at 25°C in 2× SSC/0.1% SDS for 15 minutes each and 2 washes at 55°C in 1× SSC/0.1% SDS for 30 minutes/each. Twelve clones (λTG-1-12) were analyzed by restriction enzyme cleavage and shown to correspond to a single gene. The human *ArntM1* bHLH probe was also used to screen a cDNA library derived from 4- to 8-hour embryonic RNA (N. Brown). Twelve cDNA clones (pCT-1-12) corresponding to *tgo* were isolated from 500,000 cDNA clones screened.

Sequence analysis of *tgo* wild-type and mutant cDNA and genomic clones

DNA was sequenced at the UNC-CH Automated DNA Sequencing Facility on a Model 373A DNA Sequencer (Applied Biosystems) using the Taq DyeDeoxy™ Terminator Cycle Sequencing Kit (Applied Biosystems). Four *tgo* cDNA clones, pCT-5, 7, 8, and 12, were completely sequenced. The 1.8 kb *Hin*dIII-*Eco*RI and 2.6 kb B*gl*II-*Hin*dIII restriction fragments that contain the entire *tgo* gene were also sequenced (data not shown). DNA was sequenced from two EMS-generated *tgo* mutants (tgo^{1} and tgo^{2}) and the parental strain they were generated from. Each *tgo* mutant was balanced with TM3 P[*Ubx*-lacZ]. Embryos were collected and stained for β-galactosidase expression using the X-gal reaction without fixation. Embryos that failed to turn blue were considered homozygous mutant for *tgo*, DNA was extracted, the entire *tgo* gene amplified by PCR, and sequenced. Three independent PCR reactions were subcloned and sequenced for each mutant strain to correct for PCR-generated artifacts.

Production of Tgo monoclonal antibodies

To generate antibodies against Tgo, the bHLH PAS domain was cloned by PCR from the pCT-8 cDNA clone and inserted in-frame into pGEX-2T′6, a bacterial, glutathione S-transferase (GST) fusion protein expression plasmid. The Tgo fragment encodes amino acids 1-393 that includes the complete bHLH and PAS domains. Mice were immunized by subcutaneous injection of 50 μg of fusion protein, diluted in Ribi Adjuvant, every 21 days. Monoclonal antibodies were generated and screened by ELISA and embryo staining. Three different anti-Tgo monoclonal antibodies (mAbs) were isolated and designated mAb-Tgo1-3.

Immunostaining

Antibody staining of embryos using HRP/DAB histochemistry was carried out according to standard protocols (Patel et al., 1987). Antibodies used were: mAb anti-β-galactosidase (Promega), mAb BP102, mAb 2A12, and mAb-Tgo1.

Yeast two-hybrid analysis of bHLH-PAS protein interactions

The yeast interaction system used in this study was the LexA system of Brent and colleagues (Brent, 1994). All bait constructs were cloned into the pEG202 vector, which fuses the sequence to be tested onto a LexA DNA binding domain. The prey plasmids were cloned into pJG4-5, which fuses the protein to be tested onto an acid blob transcriptional activation domain. Both bait and prey plasmids were transformed into the yeast strain EGY48, which contains the pSH18-34 *lacZ* reporter gene preceded by eight LexA binding sites. At least three independent colonies were picked for each bait-prey pair and assayed in the presence or absence of galactose for β-galactosidase expression using either solution or colony enzyme assays. β-galactosidase enzyme activity of transformants was quantitated using the chromogenic substrate o-nitrophenyl-β-D-galactoside (ONPG) in a standard enzyme assay, and the colony assay was a modification of the standard freeze-thaw assay.

Baculoviral expression of bHLH-PAS proteins and immunoprecipitation assays

Sf9 cells were infected individually or in combination with recombinant viruses containing full-length cDNA sequences of *sim*, *tgo* and *trh* using the Bac-to-Bac system (Gibco-BRL). The coinfection experiments were carried out at a m.o.i. of 2 for cells co-infected with Bac-sim and Bac-tgo. Cells co-infected with Bac-trh and Bac-tgo, used a m.o.i. of 0.2 for Bac-trh and 20 for Bac-tgo. Individually infected cells utilized the same m.o.i. as coinfected cells. Cytoplasmic extracts were prepared and subjected to immunoprecipitation assays (Gradin et al.,

1996) using mAb-Tgo1, rabbit α-Sim, rat α-Trh and Protein-A Sepharose (Pharmacia) followed by western blot analysis using the same antibodies except α-Sim, which was a rat antibody.

SL2 cell transient expression assays

cDNA clone fragments containing the complete coding sequences of *sim*, *tgo* and *trh* were PCR-cloned into the vector pAct5C, which contains the *Drosophila actin 5C* promoter and poly(A) site (Han et al., 1989). The reporter plasmid contained 6 *Toll* site 4 CMEs fused to *lacZ* in C4PLZ. Standard transfections of *Drosophila* SL2 cells contained 10 mg copia-LTR-CAT as a transfection control and 10 mg of each transfected DNA. Each transfection was done in triplicate. 48 hours after transfection, cells were lysed and b-galactosidase and CAT activity assayed. CAT activity was quantitated using FAST-CAT (Molecular Probes) and b-galactosidase activity was quantitated using a LacZ/Galactosidase Quantitation Kit (Molecular Probes). Fluorescence units derived from the b-galactosidase assay were standardized using percentage conversion values of CAT assays as a control for transfection efficiency.

Transposase-mediated local P-element hopping

The starter strain for the local P-element transposition was neu^{j6B12} which contained a P[w^+; lacW] P-element in the 5′ region of the *neu* gene (Berkeley *Drosophila* Genome Project). The P[w^+; lacW] element was mobilized in the male germline by mating *w*; neu^{j6B12}/TM3 females to males containing the *w*; Δ2-3 *Sb*/TM6 transposase source. Mosaic male progeny (*w*; neu^{j6B12}/Δ2-3 *Sb*) were mated 2 per vial with 3 *w*; TM3/TM6B *Tb Hu* females to stabilize the mobilization events. Male or female progeny were scored for eye color changes from the original orange-eyed starter strain. The eye color changes can signify alterations in the copy number and/or positions of the P[w^+; lacW] element with respect to its original site within the genome. An inverse PCR screen was performed to identify P-element insertions into the *tgo* gene according to the method of Dalby et al. (1995) with modifications by Bob Nelson (personal communication). Identification of the genomic DNA sites flanking the novel P-element insertions was achieved by plasmid rescue and subsequent hybridization and DNA sequence analyses.

EMS mutagenesis

EMS mutations in *tgo* were identified using an F_2 lethal screen in which EMS mutations were screened over the R65 P-element insertion line for lethality. Male w^{1118} flies were fed 25 mM EMS in a 1% sucrose solution (Lewis and Bacher, 1968), and mated en masse to *w*; TM3 *Sb e*/TM6 *Tb e* female flies. Male progeny (2092) of the genotype *w*; */TM6 *Tb e* were mated individually with *w*; R65/TM6 *Tb e* females. Complementation to R65 was assayed by scoring the vials for the presence of Tb^+ pupae. Vials containing Tb^+ pupae were discarded and stocks established from w^- flies in the vials with only *Tb* flies. Flies from the seven stocks selected were tested for complementation inter se and with *neu* mutations. Four of the mutations failed to complement *neu* and the other 3 made up a novel complementation group shown to be *tgo*.

RESULTS

Isolation of a *Drosophila* orthologue of the mammalian *Arnt* gene

Sequence comparison of bHLH proteins between divergent species indicates that the bHLH region is often highly conserved with respect to the rest of the protein. Cloning of a *Drosophila Arnt*-related gene was carried out by screening a *Drosophila* genomic library with a human *Arnt* bHLH probe (Hoffman et al., 1991). Positive clones were identified, restriction mapped, and shown to correspond to a single gene, subsequently named '*tango*'. Sequence analyses of the complete gene and corresponding embryonic cDNA clones indicate that *tgo* is highly related to mammalian *Arnt* (Fig. 1), both in sequence and predicted structure. The bHLH region is near the N terminus, a feature common to all bHLH-PAS proteins, followed closely by the PAS domain, and then glutamine-rich C-terminal domains. The bHLH regions are 92% identical and PAS domains 53% (Fig. 1B). The C-terminal regions are generally unrelated in primary sequence, but share the occurrence of glutamine-rich sequences (18% in Tgo), which are activation domains in mammalian Arnt (Jain et al., 1994; Li et al., 1994), *Drosophila* Sim (Franks and Crews, 1994), and many other transcription factors (Mitchell and Tjian, 1989), and the entire region is proline-rich (15% in Tgo). In particular, the poly[glutamine] regions are closely flanked by proline residues. One interesting feature of the *Drosophila* Tgo C terminus not found in other Arnt proteins is the presence of a histidine-proline-rich region of unknown function found in a small number of other *Drosophila* transcription factors including the Paired segment polarity and Bicoid homeobox proteins. This region is called the Paired (PRD)-repeat (Frigerio et al., 1986), and is distinct from the 'Paired domain'.

The *tgo* gene maps to 85C adjacent to the *neuralized* gene

Sequence analysis of *tgo* genomic and cDNA clones revealed a relatively small gene of 2.9 kb with a single intron of 142 bp within the 5′-untranslated region (Fig. 2; nucleotide sequence data not shown). The simplicity of the exon-intron structure of the *tgo* gene is in contrast to the mammalian *Arnt* gene that is much larger with a more complex exon-intron structure (Maltepe et al., 1997). The sequence upstream of *tgo* revealed that the 5′-end of the longest *tgo* cDNA clone is 328 bp 3′-to the *neuralized* (*neu*) gene in the same orientation (Fig. 2). Since *neu* has been mapped by polytene chromosome in situ hybridization to 85C (Boulianne et al., 1991), *tgo* resides in this cytological vicinity. Furthermore, the PRD repeat sequence of *tgo* is identical to the previously reported *prd7* gene sequence homology that was mapped cytologically to 85C (Frigerio et al., 1986; Price et al., 1993).

tgo is broadly expressed in the embryo

Embryonic expression of the *tgo* gene was examined by staining embryos with a monoclonal antibody generated against a bacterial Tgo fusion protein. Specificity of the reagent was indicated by the similarity between in situ hybridization (data not shown) and antibody staining with monoclonal and polyclonal antibodies, and reduction of expression in embryos homozygous for a deficiency of the *tgo* gene (Fig. 3F). Tgo protein is found in cell nuclei or cytoplasm depending on cell type and time of development (detailed description of subcellular Tgo localization will be published elsewhere). In the precellular blastoderm, Tgo protein is uniformly distributed (Fig. 3A), presumably due to maternal contribution. Staining is more intense in the cytoplasm than the nucleus. During the extended germband stage, Tgo protein is detected in all three germ layers (Fig. 3B). As tracheal pits form, the cells surrounding the pits show enhanced levels of Tgo protein (Fig. 3C) and RNA (data not shown) compared to surrounding cells. Amounts of Tgo protein above the ubiquitous levels continue to be observed in the tracheal cells including the posterior spiracles from stage

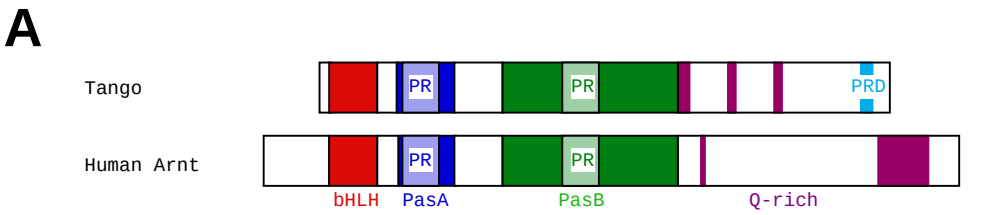
11 to the end of embryogenesis at stage 17 (Fig. 3D,E). As the CNS forms, uniformly high levels of Tgo protein are found in the brain (supraesophageal ganglia) and ventral nerve cord (Fig. 3E). In summary, Tgo protein is found broadly distributed throughout embryogenesis, although certain cell types including trachea and CNS have enhanced levels.

Interaction assays show that Tgo dimerizes with Sim, Trh and Sima

Direct evidence that Tgo can dimerize with *Drosophila* bHLH-PAS proteins was obtained using the yeast two-hybrid system. The full-length Tgo bHLH-PAS protein was used as a prey with bHLH-PAS protein baits of *Drosophila* Sim, Sima, and Trh. The results, shown in Fig. 4A, indicate that Tgo interacts strongly with all three bHLH-PAS proteins, and the levels of activation are comparable. Additional evidence for biochemical associations between Sim and Tgo and Trh and Tgo were revealed in co-immunoprecipitation experiments employing Sf9 cells infected with recombinant baculoviruses that express Sim, Tgo and Trh. These experiments showed that Sim and Tgo and Trh and Tgo formed stable complexes when co-infected (Fig. 4B). Since the cytoplasmic fraction, which does not include DNA, was analyzed, the associations are due to direct protein-protein interactions and not to independent binding of each protein to the same DNA molecules.

Evolutionary conservation of Tgo/Arnt function

Additional two-hybrid interaction assays were carried out to determine the evolutionary conservation of Tgo/Arnt in its ability to interact with bHLH-PAS proteins, and to examine the



B

```
                                        N-terminus
Tango                                                                    MDEANIQD     8
H Arnt   MAATTANPEMTSDVPSLGPAIASGNSGPGIQGGGAIVQRAIKRRPGLDFDDDGEGNSKFLRCD•DQMSN•    70

                    Basic          Helix 1        Loop         Helix 2
Tango    KERFA SRENHCEIERRRR NKMTAYITELSDMVP TCSALARKPD KLTILRMAVAHMKAL RGTGNTS    73
H Arnt   ••••- A••••S••••••• •••••••••••••• •••••••••• •••••••••S•••S• •••••••   134

             Spacer 1                                PasA
Tango    SDGTYKPSFLTDQEL KHLILEAADGFLFVVSCDSGRVIYVSDSVTPVLNYTQSDWYGTSLYEHIHPDDR   142
H Arnt   T••S•••••••••• •••••••••••••I•••ET•••V••••••••••QP••E•F•ST••DQV••••V   203
                                  F-------------------PR-----------------H--D

                                        Spacer 2
Tango    EKIREQLS TQESQNAGRILDLKSGTVKKEGHQSSMRLSMGARRGFICRMRVGNVN--PESMV SGHLNR   208
H Arnt   D•L••••• •S•NALT•••••••T•••••••Q•••••MC••S••S••••••C•SSSVD•E•VN RLSFV•   271

                                     PasB
Tango    LKQRNSLGPSRDG-TNYAVVHCTGYIKNWPPTDMFPNMHMERDVDDMSSHCCLVAIGRLQVTSTAA-NDM   276
H Arnt   NRC••G••SVKE•EPHFV•••••••••A•••AGV--SLPDDDPEAGQG•KF••••••••••••SPNCT••   339

Tango    SGSNNQSEFITRHAMDGKFTFVDQRVLNILGYTPTELLGKICYDFFHPEDQSHMKESFDQVLKQKGQMFS   346
H Arnt   •NVCQPT•••S••NIEGI•••••H•CVATV••Q•Q•••••NIVE•C•••••QLLRD••Q••V•L•••VL•   409
                  F---------------- PR -----------------H--D

Tango    LLYRARAKNSEYVWLRTQAYAFLNPYTDEVEYIVCTNSSGKTMHGAPLDAAAA HTPEQVQQQQQQQQEQ   415
H Arnt   VMF•F•S•NQ•WL•M••SSFT•Q•••S••I•••I•••TNV•NSSQE•RPTLSN TIQRPQLGPTANLPLE   478

Tango    HVYVQAAPGVDYARRELTPVGSATNDGMYQTHMLAMQAPTPQQQQQQQQRPGSAQTTPVGYTYDTTHSPY   485
H Arnt   MGSGQLAPRQQQQQTELDMVPGRDGLASYNHSQVVQPVTTTGPEHSKPLEKSDGLFAQDRDPRFSEIYHN   548

                                  tgo2  X       tgo1  X
Tango    SAGGPSPLAKIPKSGTSPTPVAPNSWAALRPQQQQQQQQPVTEGYQYQQTSPARSPSGPTYTQLSAGNGN   555
H Arnt   INADQSKGISSSTVPATQQLFSQGNTFPPTPRPAENFRNSGLAPPVTIVQPSASAGQMLAQISRHSNPTQ   618

                                           PRD repeat
Tango    RQQAQPGAYQAGPPPPPNAPGMWDWQQAGGHPHPPHPTAHPHHPHAHPGGPAGAGQPQGQEFSDMLQMLD   625
H Arnt   GATPTWTPTTRSGFSAQQVATQATAKTRTSQFGVGSFQTPSSFSSMSLPGAPTASPGAAAYPSLTNRGSN   688

Tango    HTPTTFEDLNINMFSTPFE                                                     644
H Arnt   FAPETGQTAGQFQTRTAEGVGVWPQWQGQQPHHRSSSSEQHVQQPPAQQPGQPEVFQEMLSMLGDQSNSY   758


H Arnt   NNEEFPDLTMFPPFSE                                                        774
```

Fig. 1. *Drosophila tgo* is highly related to mammalian *Arnt*. (A) The sequence organization of the *Drosophila* Tango and Human Arnt proteins derived from cDNA clone sequences are shown along with important sequence motifs. The N terminus is to the left. The bHLH domains are shown in red, and the PAS domains have been subdivided into PasA (blue) and PasB (green) subdomains. Within each Pas subdomain is a Pas repeat (PR) illustrated in lighter colors. In the C-terminal regions are glutamine (Q)-rich aa stretches shown in purple. The Tgo protein has a histidine-proline-rich PRD-repeat near its C terminus (cyan). (B) The protein sequences of *Drosophila* Tango and human Arnt (H Arnt) are compared. Amino acid numbering is shown on the right for both proteins. Sequence identities are indicated by dots and deletions by dashes. Different sequence regions are divided by spaces. The two PRs, shown below the sequences, are 44 aa long and bounded by invariant phenylalanine (F) and histidine (H) X X aspartic acid (D) residues. Three poly[glutamine] clusters in the Tgo C-terminal region that are likely transcriptional activation domains are shown in purple, as are two glutamine-rich clusters in human Arnt. Shown in cyan is the *prd7* PRD repeat. The position of the premature translational stops (X) corresponding to the tgo^1 and tgo^2 EMS mutations are shown above the Tgo protein sequence. This sequence has been deposited with GenBank under accession number AF020426.

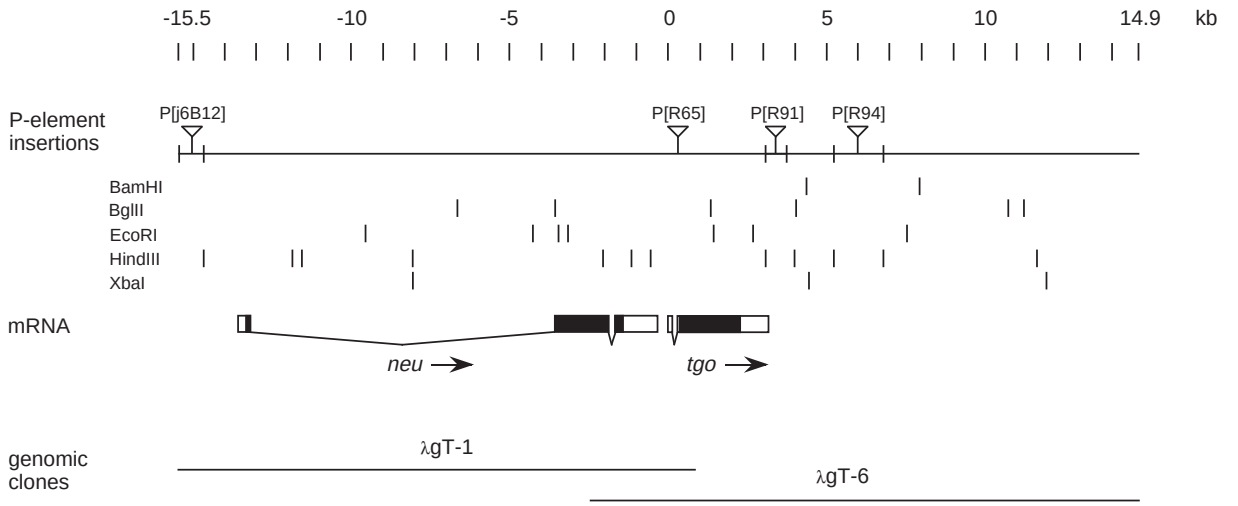


Fig. 2. Structure of the *tgo* genomic region. Shown is the genomic interval that contains the *tgo* and *neu* genes. The scale in kb is shown at the top with the number '0' denoting the 5′-end of the longest *tgo* cDNA clone. The locations of the four P-element insertions, described in the text, that reside in this interval are indicated. The location of P[R65] was determined by sequence analysis. The other three insertions, P[j6B12], P[R91] and P[R94] were mapped by Southern hybridization of genomic DNA and the limits of their positions are indicated between the flanking vertical lines. The restriction enzyme cleavage sites for *Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, and *Xba*I are indicated. The structure of *neu* and *tgo* mRNAs as determined by analysis of cDNA clones are shown. Transcriptional orientations (5′ to 3′) are indicated by arrows. Unfilled blocks indicate untranslated regions, filled boxes are coding sequences, and adjoining lines indicate introns. Two genomic clones, λgT-1 and λgT-6, that cover the 30.4 kb interval are shown, although a total of 12 genomic clones in this region were mapped.

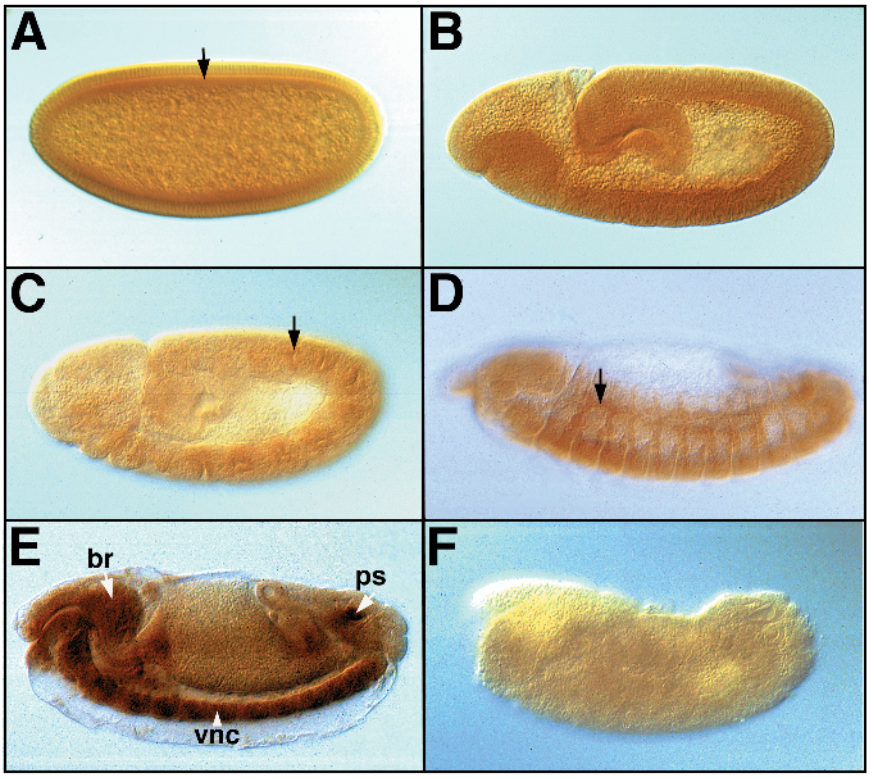


Fig. 3. Antibody staining of wild-type embryos with anti-Tgo monoclonal antibody shows that Tgo protein is broadly distributed throughout development. Embryos are sagittal views oriented with the anterior to the left and ventral at the bottom. (A) Stage 5 precellular blastoderm embryo. Arrow denotes cytoplasm internal to the nuclei. (B) Germband-extended stage 9 embryo. (C) Stage 11 embryo. Tgo is concentrated around the tracheal pits (arrow). (D) Stage 13 embryo. Enhanced *tgo* staining is found in tracheal branches (arrow). (E) Stage 15 embryo with high levels of Tgo in the brain (br), ventral nerve cord (vnc) and posterior spiracles (ps). (F) Specificity of anti-Tgo antibody staining is revealed by a strong reduction of staining in embryos (stage 13 or older) that are homozygous mutant for the *Df(3R)p-XT9* chromosomal deficiency that deletes the *tgo* gene.

rules that govern *Drosophila* bHLH-PAS protein interactions. Trans-species interactions with Tgo can occur since Tgo interacts strongly with murine Ahr (Fig. 4A). Additional experiments were carried out using *Drosophila* bHLH-PAS proteins and human Arnt in all pairwise combinations, both as bait and prey. The results (Table 1) indicate that the *Drosophila* Sim, Sima and Trh proteins interact strongly with human Arnt, as did murine Ahr. Homodimerization was observed with human Arnt, a result observed in some published accounts (Sogawa et al., 1995; Swanson et al., 1995), but not others (Reisz-Porszasz et al., 1994). However, no other bHLH-PAS protein examined was able to form homodimers. Heterodimerization was only

Table 1. Yeast two-hybrid analysis indicates that Tgo/Arnt function is evolutionarily conserved and that bHLH-PAS protein heterodimerization with Tgo/Arnt is the predominant mode of interaction

	Sim	Sima	Trh	Ahr	Arnt	Tgo	Prey
Sim	–	–	–	–	+	+	
Sima	–	–	–	–	+	+	
Trh	–	–	–	–	+	+	
Ahr	–	–	–	–	+	+	
Arnt	+	+	+	+	+	nd	
Bait							

Each bHLH-PAS protein (*Drosophila* proteins: Sim, Sima, Tgo, and Trh; mammalian proteins: murine Ahr and human Arnt), was tested pairwise as both bait and prey, except Tgo, which was used only as prey, in the two-hybrid interaction assay using a *lacZ* reporter. All proteins contained intact bHLH and PAS domains, but had the C-terminal activation domain removed, except the Tgo prey, which is a full length protein. Proteins are the same as described in Fig. 4 with the addition of the human Arnt bHLH-PAS bait (67-458) and prey (67-589) proteins. β-galactosidase activity was assessed using either liquid assays or Bluo-Gal colorometric assay on yeast colonies. Results were comparable between the two assays and only strong interactions were considered positive (+). Absence of consistent β-galactosidase expression was considered negative (–). (nd) indicates that assay was not done.

observed between bHLH-PAS proteins and Tgo/Arnt, no heterodimers were formed between pairwise combinations of Sim, Sima, Trh and Ahr. These results show that both *Drosophila* and mammalian Arnt are not only highly related in sequence and expression, but can form heterodimers with other bHLH-PAS proteins, regardless of species.

Transient expression studies in *Drosophila* tissue culture indicate that Tgo interacts with Sim and Trh to bind DNA and activate transcription

Direct evidence that Tgo and other bHLH-PAS proteins can dimerize, bind DNA, and activate transcription through the CME was obtained using transient transfection assays with *Drosophila* SL2 cell culture. Western blot analysis indicated that Tgo protein is present in SL2 cells, but Sim and Trh are absent (data not shown). Full length cDNAs encoding each of these proteins were cloned into vectors that express them constitutively under the control of the *actin5C* promoter (Han et al., 1989), yielding the plasmids pAct-sim, pAct-tgo, and pAct-trh. The reporter gene (P[6xCME-lacZ]) contained six copies of the CME (the putative Sim::Tgo and Trh::Tgo binding sites) fused to a P-element promoter driving *lacZ*. This is the same DNA construct that functions as a midline (Wharton et al., 1994) and tracheal (see below) enhancer in vivo, except that the in vivo studies were carried out with a reporter containing only 4 copies of the CME.

When pAct-sim and pAct-tgo were cotransfected with the P[6xCME-lacZ] reporter, *lacZ* transcription was induced to high levels (Fig. 5). Control experiments indicated that high levels of *lacZ* activation required the presence of the CME and overexpression of both Sim and Tgo. The magnitude of reporter gene activation in cultures transfected with pAct-sim and pAct-tgo compared to those transfected with only pAct-sim correlates with the magnitude of increased Tgo protein detected by western analysis (data not shown). The enhanced levels of *lacZ* reporter expression that were obtained when both Sim and Tgo were co-expressed at high levels suggests that the two proteins interact to activate CME-regulated transcription.

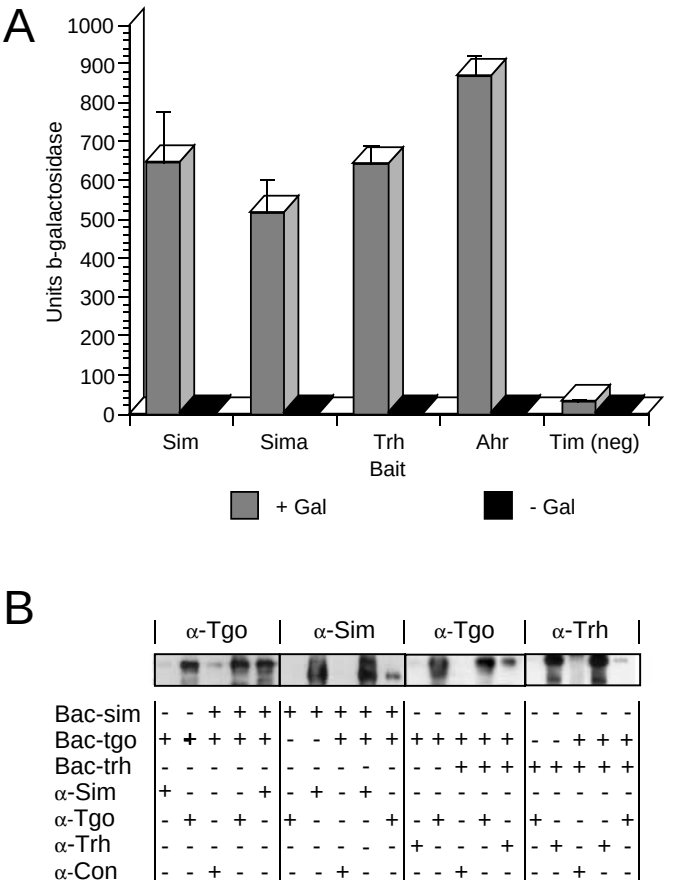


Fig. 4. Yeast two-hybrid and Sf9 cell co-precipitation assays show that Tgo dimerizes with Sim, Trh, and other bHLH-PAS proteins.
(A) Liquid cultures of Yeast containing bHLH-PAS protein baits (*Drosophila* Sim, Sima, Trh, and murine Ahr) and Tgo full-length bHLH-PAS prey were assayed in the presence or absence of galactose for β-galactosidase activity expressed in units activity ± s.e.m. The pSH18-34 lacZ gene was used as reporter. Galactose induces the expression of the Tgo prey construct. All bait constructs contained intact bHLH and PAS domains, but had C-terminal regions removed, which contain transcriptional activation domains. The amino acid sequences of each protein are the following: Sim (1-461); Sima (32-488); Trh (1-698); Ahr (6-420). The Tgo prey was full length (644 aa). The reciprocal experiment using Tgo bHLH-PAS baits could not be performed since these proteins self-activated in the absence of prey. The *Drosophila* Timeless (Tim) protein was used as a negative control; results showing low β-galactosidase activity were similar to those using no bait or other control baits. The bHLH-PAS protein baits also showed low β-galactosidase activity when tested in the presence of galactose with either no prey or unrelated preys (also see Table 1).
(B) Sf9 cells were singly infected or coinfected with baculovirus Bac-sim, Bac-tgo, and Bac-trh expression plasmids, cytoplasmic extracts prepared, and proteins immunoprecipitated with either control mouse or rat pre-immune antibody (α-Con), polyclonal rabbit anti-Sim (α-Sim), polyclonal rat anti-Trh (α-Trh), or monoclonal anti-Tgo (α-Tgo). Immunoprecipitates were electrophoresed on SDS/PAGE and immunoblotted with mAb α-Tgo, rat α-Sim, or rat α-Trh. Shown above the autoradiogram are the antibodies used for immunoblotting; shown below are baculoviral expression plasmids used for infection and antibodies used for immunoprecipitation. The α-Sim and α-Trh antibodies were more efficient in immunoprecipitating heterodimeric complexes than α-Tgo. Bac-sim infected cells routinely had two Sim protein bands of 84 and 92×10^3 M_r. The baculoviral Tgo protein was 75×10^3 and Trh was 120×10^3 M_r.

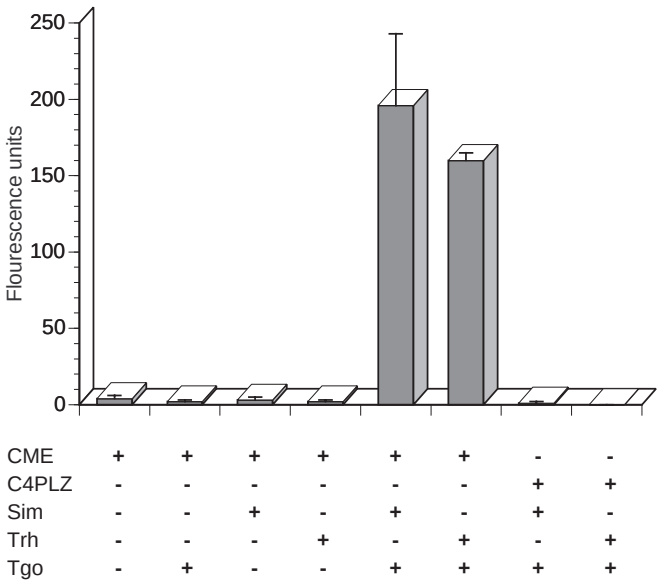


Fig. 5. Sim and Trh interact with Tgo in SL2 cells to activate transcription from CME-bearing genes. *Drosophila* SL2 cells were transfected with expression plasmids that act as a source of Sim, Tgo, and Trh proteins, reporter plasmids, and a transfection normalization plasmid. The expression plasmids are pAct-sim, pAct-tgo, and pAct-trh. The reporter plasmids are either: (1) P[6xCME-lacZ] that contains six copies of the CME fused to C4PLZ, which contains a P-element promoter fused to the *lacZ* gene (CME), or (2) C4PLZ alone. The copia-LTR-CAT plasmid was used to normalize transformation efficiency between samples. Normalized β-galactosidase is expressed in arbitrary fluorescence units as the mean ± s.e.m. of multiple assays.

Similar experiments were performed with pAct-trh and pAct-tgo; Trh and Tgo also interact to activate CME transcription.

Creation of *Tgo* mutations by P-element and EMS mutagenesis

tgo maps to 85C on the third chromosome. Mutations in *tgo* were generated using a two-step strategy. The first step involved hopping P-elements from the adjacent *neu* gene into the *tgo* gene (Dalby et al., 1995; Tower et al., 1992; see Materials and Methods). The P-element mutant *tgo* strain was then used to screen for EMS-induced *tgo* mutations. The local hop involved crossing P-element transposase into neu^{j6B12} and identifying progeny flies with an eye color darker or lighter than the orange color of the parental strain. It is expected that some of these strains will have novel P-element insertions. Flies from 120 distinct lines that had altered eye color were screened using an inverse PCR/Southern blot procedure (Dalby et al., 1995). Three lines (R65, R91, and R94), all with red eyes, were found to have novel P-element insertions in or near the *tgo* locus (Fig. 2).

Sequence analysis indicated that R65 was located in exon 2 in the 5′-untranslated region 43 bp 5′ to the proposed initiator methionine (Fig. 2). This insertion is likely to disrupt transcription of the *tgo* gene, resulting in a severe mutation. The R65 strain retained the original P-element near the *neu* gene as well as the newly generated P-element insertion into *tgo*. Cellular analysis of the *neu tgo* double insertion strain consistently showed more severe embryonic defects than neu^{j6B12} in the structure of the trachea (data not shown). The *neu tgo* double insertion line was used to screen 2092 EMS-mutagenized third chromosomes. Seven EMS mutations were obtained that were lethal over both the *neu tgo* double insertion strain and *Df(3R)p-XT9*, a chromosomal deletion of 85C. Complementation experiments revealed that 4 were new mutations of *neu* and the other 3 corresponded to a novel complementation group shown to be *tgo*.

Confirmation that the new lethal complementation group corresponded to the *tgo* gene was obtained by DNA sequence analysis of two of the *tgo* mutants (Fig. 1). The tgo^1 mutant has a termination codon at aa 532 that deletes a proline-rich region, the PRD repeat, and a glutamine-rich region from the protein. The tgo^2 mutant has a termination codon at aa 518 that deletes another poly[glutamine] stretch in addition to the region deleted in tgo^1. These experiments confirm that the *tgo* mutations correspond to the *Arnt*-related transcription unit and that *tgo* is a vital locus.

Mutations in Tgo result in CNS midline and tracheal defects

Mutant *tgo* embryos were examined for defects in CNS midline and tracheal development, which would be expected if Tgo is a dimerization partner for Sim and Trh. The tgo^1 mutant was used in all of the experiments described below. The tgo^2 and tgo^3 mutants qualitatively show similar phenotypes to tgo^1, but quantitatively are more severe. tgo^1 and tgo^2 are likely to be hypomorphic, since they cause quantitatively more severe phenotypes when analyzed in *trans* to the *Df(3R)p-XT9* chromosome that is deficient for *tgo*. Defects in CNS midline neurons and glia were examined using the AA142 and X55 enhancer trap reporters, respectively (Klämbt et al., 1991); expression of *lacZ* in both of the genes is absent in *sim* mutant embryos (Sonnenfeld and Jacobs, 1994). In wild-type embryos, the AA142 enhancer trap gene is expressed in an average of 3.5 midline glia/segment by stage 14 of embryogenesis (Fig. 6A). In tgo^1 mutant embryos, there was a reduction in the number of stained midline glia to approximately 1 cell per/segment (Fig. 6B). The X55 enhancer trap gene stains the ventral unpaired median neurons (VUMs) and the median neuroblast (MNB) and its progeny in the ventral region of the CNS (Fig. 6C). In tgo^1 mutant embryos, the number of VUM neurons and MNB progeny were reduced in number (60% of wild-type) and did not migrate into the ventral regions of the VNC (Fig. 6D). Mutations in *trh* result in severe defects in tracheal tubule formation (Younossi-Hartenstein and Hartenstein, 1993; Isaac and Andrew, 1996; Wilk et al., 1996). The role of *tgo* in tracheal development was examined by staining tgo^1 mutant embryos with monoclonal antibody 2A12 that stains the lumen of the tracheal tubes (Fig. 6E). The results revealed that tgo^1 mutant embryos have a variety of tracheal defects, some weak (Fig. 6F) and others more severe (Table 2). The relatively weak phenotypes of tgo^1 mutations compared to *sim* and *trh* mutations are likely due to phenotypic rescue by maternal *tgo* (Fig. 3A) and the hypomorphic nature of tgo^1.

Sim::Tgo, and Trh::Tgo control transcription in vivo through a common enhancer element

Previous work on putative target genes that are expressed in the CNS midline cells revealed the existence of the CME, an enhancer element both required and sufficient for CNS midline

transcription (Wharton et al., 1994). Multimerization (4×) of the CME drives CNS midline cell expression from a heterologous promoter (Fig. 7A). Further characterization of the P[4xCME-lacZ] transgene shows that it is also expressed in the developing and mature trachea, posterior spiracles, and salivary ducts (Fig. 7A,B). This expression pattern resembles the combined *sim* and *trh* expression patterns, and suggests that the CME is an in vivo target element of Sim, Trh and Tgo.

Tests of the genetic control of CME expression by *trh*, *sim*, and *tgo* were performed by crossing P[4xCME-lacZ] into the three mutant backgrounds. All CNS midline *lacZ* expression from P[4xCME-lacZ] was abolished in *sim* mutant embryos, although tracheal expression was normal (Fig. 7C). Similarly, all tracheal *lacZ* expression from P[4xCME-lacZ] was abolished in *trh* mutant backgrounds (Fig. 7D), while CNS midline expression was unaffected. In tgo^1 mutant embryos, both CNS midline and tracheal expression were reduced (Fig. 7E,F), although not as severely as observed in *sim* and *trh* mutant embryos, perhaps due to residual maternal *tgo* and the hypomorphic nature of tgo^1.

Gene dosage interactions suggest that Sim and Trh interact with Tgo in vivo

The CNS midline and tracheal defects observed in *tgo* mutant embryos are consistent with Tgo being a dimerization partner for Sim and Trh and binding to the CME in vivo. The less severe phenotypes observed in tgo^1 mutants compared to *sim* and *trh* are interpreted as being due to the occurrence of maternal *tgo* and the hypomorphic nature of tgo^1. We have used the relatively weak phenotypes observed with the tgo^1 mutant strain to genetically test for in vivo interactions between *sim-tgo* and *trh-tgo*. Both *sim* and *trh* were individually recombined onto tgo^1, and crosses done to generate embryos with variable copies of mutant and wild genes. Embryos were then assayed for CNS and tracheal defects.

mAb BP102 was used to examine the CNS for indication of midline defects (Thomas et al., 1988). 100% of the embryos that are homozygous mutant for *sim* show a severe

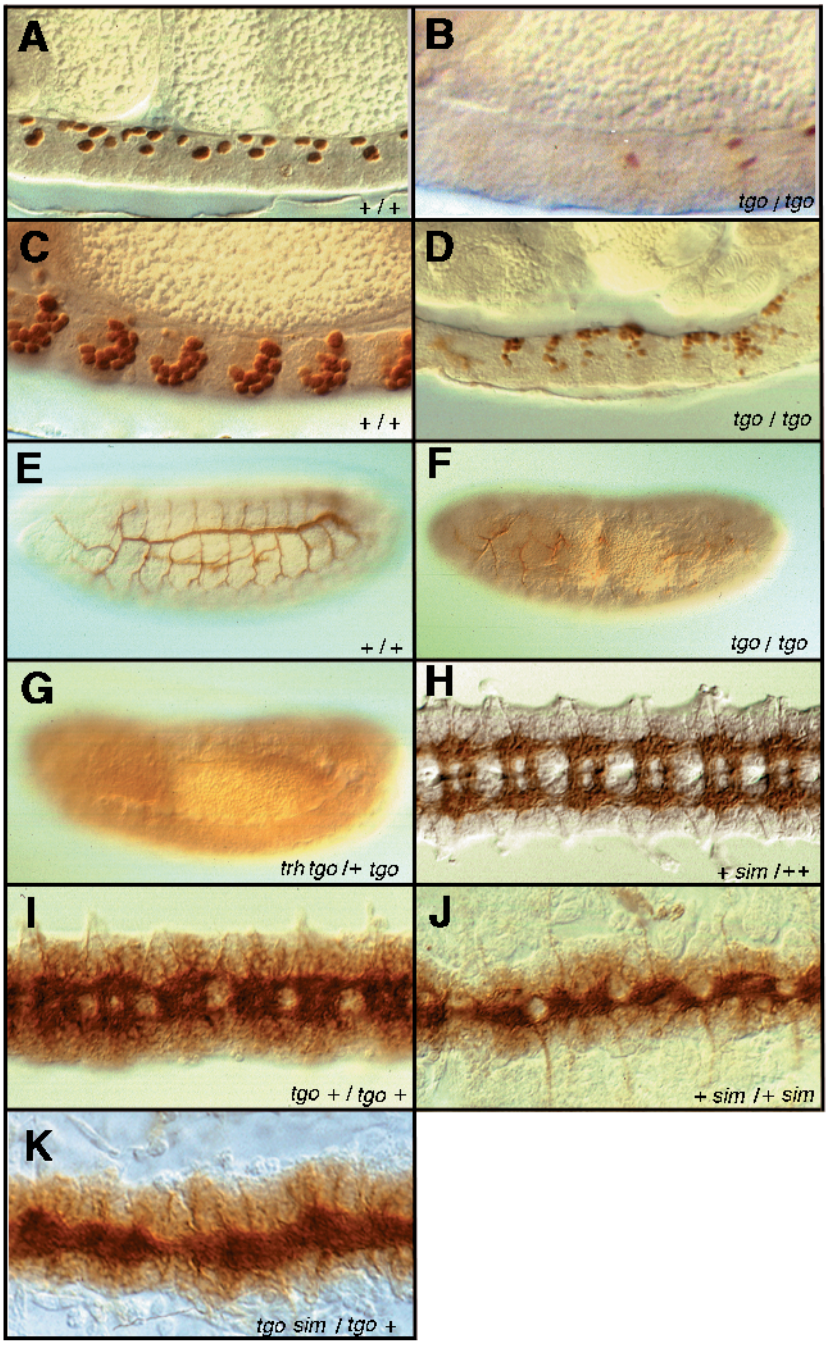


Fig. 6. *tgo* mutant embryos show defects in CNS midline cell and tracheal development and double mutants suggest in vivo interactions between *tgo* and *sim* and *tgo* and *trh*. Anterior is to the left in all panels. (A-G) Sagittal views. (A) Stage 15 wild-type embryo containing the AA142 enhancer trap gene that stains midline glia present in the dorsal aspect of the CNS. (B) There is a strong reduction in the number of AA142-positive midline glia in tgo^1 mutant embryos. (C) Stage 15 wild-type embryo containing the X55 enhancer trap gene that stains midline neurons present in the median and ventral regions of the CNS. (D) In many tgo^1 mutant ganglia, there is a reduction in the number of X55-expressing neurons and they are misplaced to the dorsal side of the CNS instead of the median and ventral regions. (E) Stage 15 wild-type embryo stained with mAb 2A12 that visualizes the tracheal lumen. (F) Stage 15 tgo^1 mutant embryo stained with mAb 2A12 showing a severe phenotype in which only a few tracheal branches are present. (G) Stage 15 *trh* tgo^1/+ tgo^1 embryo stained with mAb 2A12 showing a complete absence of staining. (H-K) Dissected stage 15 embryos are stained with mAb BP102 to reveal the CNS axon scaffold. (H) Embryo heterozygous for *sim* showing a wild-type CNS with orthogonal axon scaffold consisting of anterior and posterior commissures and longitudinal connectives. (I) The CNS axonal phenotype in tgo^1 mutant embryos shows disorganized commissures and thin connectives, but is not severely collapsed. (J) The axon scaffold in sim^{H9} mutant embryos shows a severe collapsed CNS phenotype. (K) The axon scaffold phenotype of embryos that are reduced 50% in *sim* dosage in a homozygous mutant tgo^1 background (tgo^1 *sim*/tgo^1 +) show a severe collapsed CNS phenotype identical to homozygous *sim* mutant embryos.

Table 2. Phenotypic analysis of mutant combinations show genetic interactions between *sim* and *tgo* and *trh* and *tgo*

	% Axon phenotypes of mutant embryo population			
Genotype	Wildtype	Severe collapsed	Mild collapsed	*n*
+ *sim* / + +	100	0	0	96
+ *sim* /+ *sim*	0	100	0	29
tgo^1 + / tgo^1 +	58	0	42	65
tgo^1 *sim* / tgo^1 +	43	57	0	46

	% Tracheal phenotypes of mutant embryo population					
		Severe		Mild		
Genotype	Wildtype	No trachea	Small patches	No dlt branches present	Branches present but broken	*n*
+ *trh* / + +	100	0	0	0	0	65
+ *trh* / + *trh*	4	96	0	0	0	71
tgo^1 + / tgo^1 +	14	3	11	11	61	84
tgo^1 *trh* / tgo^1 +	18	42	19	9	12	115

Mutant combinations of *sim*, *tgo*, and *trh* were analyzed with mAb BP102 for axonal defects and mAb 2A12 for tracheal defects. Only embryos of the indicated mutant genotype were considered; they were identified using a *lacZ*-marked balancer chromosome. For *sim-tgo* combinations, the severe collapsed phenotype corresponds to a '*sim*' phenotype and the mild collapsed phenotype has a narrow axon scaffold, thin longitudinal connectives, and incompletely separated commissures. The tracheal phenotypes include severe phenotypes in which there are no detectable trachea or only small patches. Milder phenotypes include: (1) the presence of the dorsal longitudinal trunk (dlt) but no other branches present, and (2) the presence of tracheal branches that are incompletely joined (broken). *n*=number of mutant embryos examined of the indicated genotype.

collapsed phenotype (Fig. 6J; Table 2; Thomas et al., 1988). Embryos that are *sim*/+ showed 100% wild-type CNS (Fig. 6H; Table 2). Mutant embryos that are homozygous for tgo^1 are either wild-type (58%) or have a mild defect (42%) in axon morphology (Fig. 6I). None of the tgo^1 mutant embryos show a severe collapsed '*sim*' phenotype (Table 2). In contrast, 57% of the embryos that are tgo^1 *sim*/tgo^1 + have a severe collapsed CNS (Fig. 6K; Table 2). Corresponding results were obtained using P[4xCME-lacZ] as a CNS midline marker: most tgo^1 *sim*/tgo^1 + embryos had a complete absence of CNS midline β-gal staining (Fig. 7G), whereas tgo^1 +/tgo^1 + embryos generally retained staining although it was reduced from wild-type (Fig. 7E).

Similar results were obtained for mutant embryos that were homozygous for tgo^1 and heterozygous for *trh*. When stained

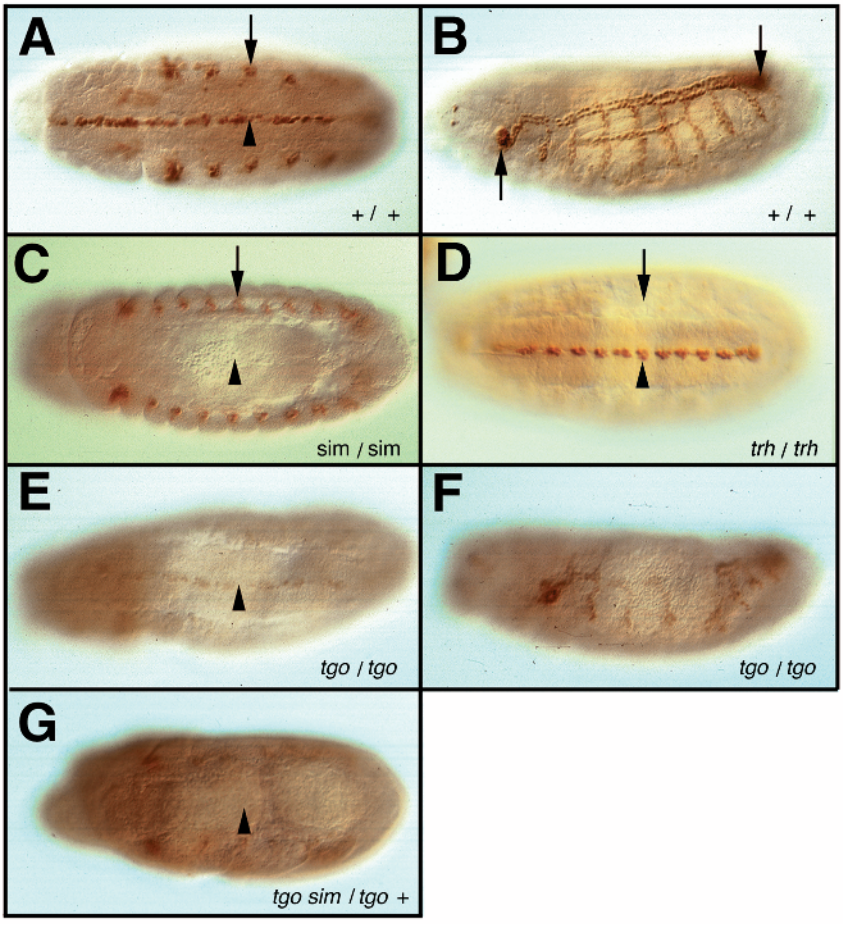


Fig. 7. The CME acts as a CNS midline and tracheal enhancer and its function is controlled by *sim*, *tgo* and *trh*. Ventral views (A,C,D,E,G). Sagittal views (B,F). All embryos contain a P[4xCME-lacZ] transgene and are stained with anti-β-galactosidase. (A) Wild-type stage 11 embryo shows that P[4xCME-lacZ] is expressed in both the CNS midline cells (arrowhead) and cells surrounding the tracheal pits (arrow). (B) P[4xCME-lacZ] expression in wild-type stage 15 embryo can be observed in the tracheal cells including the anterior and posterior spiracles (arrows). (C) Stage 13 embryo mutant for sim^{H9} shows an absence of P[4xCME-lacZ] expression in the CNS midline cells (arrowhead), but shows normal expression in the trachea (arrow). (D) Stage 15 embryo mutant for trh^8 shows an absence of P[4xCME-lacZ] expression in the trachea (arrow), but expression is present in the CNS midline cells (arrowhead). (E) Stage 11 embryo mutant for tgo^1 reveals that P[4xCME-lacZ] expression in the midline cells is often present but the number of stained cells and staining intensity is reduced (arrowhead). (F) Stage 15 embryo mutant for tgo^1 shows reduced expression in trachea. (G) CNS midline expression (arrowhead) is absent in P[4xCME-lacZ] embryos that are tgo^1 *sim*/tgo^1 +.

with mAb 2A12, which visualizes the tracheal lumen, 96% of the embryos homozygous mutant for *trh* show a complete absence of trachea, and 100% of *trh* heterozygotes have wild-type trachea (Table 2). Most tgo^1 homozygous mutant embryos (86%) have mutant trachea, but the phenotypes are generally weak (Fig. 6F); 97% have some mAb 2A12-staining material, and only 14% have severe defects. In contrast, 42% of embryos that are *trh* $tgo^1/+$ tgo^1 completely lack tracheal staining (Fig. 6G) and another 19% had a severe phenotype. Thus, the loss of a single copy of *trh* in the tgo^1 homozygous mutant background resulted in greater than a 4-fold increase in severe tracheal defects. In summary, the results of the genetic experiments provide strong evidence that *sim* and *trh* interact with *tgo* in vivo to control CNS midline and tracheal transcription and development.

DISCUSSION

Tgo/Arnt Acts as an evolutionarily conserved, broadly expressed bHLH-PAS protein dimerization partner

These results show that Arnt is a highly conserved protein that functions as a dimerization partner for a number of bHLH-PAS proteins. The *Drosophila tgo* gene is orthologous to mammalian *Arnt*, as indicated by the following criteria. (1) The primary sequence and protein domain organization of the two proteins are strongly conserved. (2) The expression patterns of *tgo* and mammalian *Arnt* (Abbott et al., 1995; Abbott and Probst, 1995) are also similar. They are both transcribed in most if not all cells of the embryo. Another mammalian *Arnt* gene, *Arnt2*, is expressed in a subset of cells in the brain (Hirose et al., 1996), and, for this reason, *tgo* resembles *Arnt* more than *Arnt2*. (3) Interaction studies indicate that Tgo and mammalian Arnt act interchangeably as dimerization partners for other phylogenetically diverse bHLH-PAS proteins. These observations demonstrate that Tgo/Arnt acts as a broadly expressed dimerization platform for bHLH-PAS proteins in both mammals and *Drosophila*. Recently, *Arnt* has been identified in fish (Pollenz et al., 1996) and *C. elegans* (Genome Project). Thus, *Arnt* constitutes an evolutionarily conserved transcriptional regulator found in most, if not all, multicellular animals.

Together, Arnt/bHLH-PAS protein complexes control a wide variety of developmental and physiological processes. Given the strong evolutionary conservation between *Arnt*, *sim*, their genomic target DNA sequence elements, and their accessory proteins including Hsp90, it is apparent that *Arnt* represents a central element in an evolutionarily conserved gene regulatory pathway. In some cases, these transcriptional regulatory pathways may be responsive to ligand::bHLH-PAS protein interactions (e.g. AHRC). Identification of *tgo* in *Drosophila* opens up the possibility of sophisticated genetic analysis of these regulatory and signaling pathways. These possibilities are realized in this work in which genetic interactions between *tgo* and other bHLH-PAS genes are clearly revealed.

Biochemical studies in mammals have implicated Arnt as a partner for Ahr and HIF-1α. We have provided evidence that Tgo associates in vivo with Sim and Trh using several criteria. (1) Tgo protein is expressed in all cells in the embryo and thus overlaps in expression with Sim in the CNS midline cells and Trh in the tracheal cells. (2) Studies using the yeast two-hybrid assay indicate that Tgo can form dimers with Sim, Trh, and Sima. (3) Co-immunoprecipitation experiments with baculoviral-expressed proteins indicates that Tgo forms dimers with Sim and Trh. (4) SL2 transient expression assays suggest that Sim and Tgo and Trh and Tgo interact to bind DNA and activate transcription. (5) Mutations in *tgo* result in CNS midline and tracheal defects. (6) Gene dosage experiments further suggest that *sim* and *tgo* and *trh* and *tgo* interact to control CNS midline and tracheal cell transcription and development, respectively. This work firmly establishes that *tgo* functions in CNS cell fate specification and formation of a functional respiratory system.

The interaction experiments described in this paper and others indicate that Tgo/Arnt is a common dimerization partner for bHLH-PAS proteins. These bHLH-PAS proteins form heterodimers exclusively with Tgo/Arnt and not with each other. In addition, they do not form homodimers while Arnt does. Even though Arnt homodimers can be detected, no in vivo function has been reported. As more PAS proteins are discovered, exceptions to these rules may emerge (e.g. Hogenesch et al., 1997; Tian et al., 1997). Never-the-less, it is convenient to think of Arnt as a common PAS dimerization partner playing a role similar to that observed for E2A and Max for other classes of bHLH proteins.

The CME controls CNS midline and tracheal transcription and is a target of Sim::Tgo and Trh::Tgo

The CME acts as an enhancer element for both CNS midline and tracheal transcription. CME transcription within the CNS midline is dependent on *sim* function, CME transcription in the trachea is dependent on *trh* function, and *tgo* function is required for CME transcription in both cell types. The CME is sufficient for transcription in these cell types since multimerized *Toll* site 4 is sufficient to drive transcription specifically within the CNS midline and tracheal cells. These results suggest that Sim::Tgo heterodimers control CNS midline transcription through the CME, and that Trh::Tgo heterodimers control tracheal transcription through the CME. Consistent with this view are the transient transfection results indicating that co-introduction of either Sim and Tgo or Trh and Tgo induce high levels of transcription from a multimerized CME reporter gene. Nevertheless, complete molecular dissection and identification of all possible protein::DNA interactions within the 20 bp *Toll* site 4 have not been attempted, and it is possible that regulatory proteins in addition to Sim::Tgo and Trh::Tgo may bind to this DNA and help control its CNS midline and tracheal enhancer function.

Our results suggest that for a gene to be expressed in both CNS midline precursor and tracheal cells, the only requirements are a promoter and multiple copies of the CME. Although synthetic genes of this type are expressed in both cell types, expression of authentic target genes of *sim*, *trh*, and *tgo* are more complex. Minimally, there are six classes of putative target genes that may utilize a CME. Their expression in vivo (including an example) are: (1) CNS midline and trachea (MT) – *breathless* (Klämbt et al., 1992), (2) CNS midline only (M) – *sim* autoregulation (Nambu et al., 1990), (3) trachea only (T) – *trh* autoregulation (Wilk et al., 1996), (4) CNS midline glia (MG) – *slit* (Wharton et al., 1994), (5) *sim* midline repression

(MR) – *ventral nervous system defective* (Mellerick and Nirenberg, 1995; Xiao et al., 1996), and (6) hypoxia (H) – HIF-1α (Wang et al., 1995; Nagao et al., 1996). Since the CME by itself drives expression in both the CNS midline and trachea, CME-controlled expression would acquire specificity via the addition of repressor elements. We would predict that the MT class of genes would require only multiple copies of the CME to be expressed in both cell types. In contrast, the M and T classes may require the presence of cell-specific repressors to restrict the expression of CME-driven processes to only a single cell type. For example, the regulatory regions of genes such as *sim* and *Toll*, in which the CME has been directly implicated in controlling CNS midline, but not tracheal, transcription (Wharton et al., 1994) may possess binding sites for tracheal repressors. The presence of midline precursor and midline neuron repressors may also restrict expression of the *slit* MG enhancer to midline glia (Wharton and Crews, 1993).

Developmental and physiological roles of Tgo

The studies described in this paper establish that Tgo plays a role in CNS midline cell development and tracheal tubule formation. In particular, Tgo interacts with Sim just after gastrulation to activate midline precursor gene transcription. All genes expressed in the CNS midline cells that have been tested require *sim* (and presumably *tgo*) function. In this sense, *tgo* and *sim* control specification of the CNS midline cell lineage. *trh* and *tgo*, on the other hand, are not required for transcription of all tracheal genes, but only a subset that are hypothesized to control tubule formation (Isaac and Andrew, 1996).

There are likely to be additional roles for *tgo*. It is shown here that it interacts biochemically with the *Drosophila* Sima protein. The function of Sima is currently unknown, but its ubiquitous expression pattern and primary sequence suggest it may be related functionally to mammalian HIF-1α and control the *Drosophila* response to hypoxia. *Drosophila* cell culture experiments have revealed the existence of a CME-binding factor that is induced under hypoxic conditions (Nagao et al., 1996). Genetic analysis of *tgo* mutants should confirm whether *tgo* is implicated in controlling the hypoxia response, regardless of the role of *sima*.

Consistent with the possibility that Tgo has additional partners, *tgo* mutants are abnormal in their general embryonic morphology distinct from *sim* and *trh* defects; we are currently analyzing the defects in more detail. Models have also been proposed that suggest that bHLH-PAS proteins, including Clock and Arnt, may interact with the Per PAS protein to control the periodicity of biological rhythms (Huang et al., 1993; Lindebro et al., 1995; King et al., 1997). Thus, it seems likely that additional roles for *tgo* will be uncovered upon further analysis and with the generation of germline clones and postembryonic mosaics. In summary, it is proposed that *Arnt*/*tgo* is the centerpiece of an ancient metazoan regulatory/signaling pathway that has diverged to carry out numerous biological roles.

The assistance of Satish Gopal, Jackie Stilwell, and Brad Tewes is gratefully acknowledged. We would like to thank Chen-Ming Fan, Oliver Hankinson, Bob Nelson, Mark Peifer and Michael Rosbash for stimulating and useful discussions. We would also like to thank Debbie Andrew, Ron Blackman, Gabriele Boulianne, Corey Goodman, Oliver Hankinson, Ruth Lehmann, Jim Manley, Nipam Patel, Michael Rosbash, Susan Shepherd and Benny Shilo for providing essential antibodies, clones and fly stocks, and Russ Finley and Roger Brent for two-hybrid yeast strains and plasmids. We would also like to acknowledge the North Carolina State University monoclonal antibody facility for production of Tango antibodies, the Bloomington *Drosophila* Stock Center for providing *Drosophila* stocks, and the NICHD Developmental Studies Hybridoma Bank for monoclonal antibodies. This work was supported by a Canadian MRC fellowship to M. S. and an American Cancer Society Fellowship to M. W. Research is supported by NIH grant RD25251 to S. T. C.

REFERENCES

Abbott, B. D., Birnbaum, L. S. and Perdew, G. H. (1995). Developmental expression of two members of a new class of transcription factors: I. expression of aryl hydrocarbon receptor in the C57BL/6N mouse embryo. *Dev. Dyn.* **204,** 133-143.

Abbott, B. D. and Probst, M. R. (1995). Developmental expression of two members of a new class of transcription factors: II. Expression of aryl hydrocarbon receptor nuclear translocator in the C57BL/6N mouse embryo. *Dev. Dyn.* **204,** 144-155.

Boulianne, G. L., de la Concha, A., Campos-Ortega, J. A., Jan, L. Y. and Jan, Y. N. (1991). The *Drosophila* neurogenic gene *neuralized* encodes a novel protein and is expressed in precursors of larval and adult neurons. *EMBO J.* **10,** 2975-2983.

Brent, R. (1994). Interaction trap/two-hybrid system to identify interacting proteins. *Curr. Protocols Molec. Biol.* **27,** 1-17.

Chang, Z., Price, B. D., Bockheim, S., Boedigheimer, M. J., Smith, R. and Laughon, A. (1993). Molecular and genetic characterization of the *Drosophila tartan* gene. *Dev. Biol.* **160,** 315-332.

Dalby, B., Pereira, A. J. and Goldstein, L. S. B. (1995). An inverse PCR screen for the detection of P element insertions in cloned genomic intervals in *Drosophila melanogaster*. *Genetics* **139,** 757-766.

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.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. and Noll, M. (1986). Structure of the segmentation gene *paired* and the Drosophila PRD gene set as part of a gene network. *Cell* **47,** 735-746.

Gradin, K., McGuire, J., Wenger, R. H., Kvietikova, I., Whitelaw, M. L., Toftgard, R., Tora, L., Gassmann, M. and Poellinger, L. (1996). Functional interference between hypoxia and Dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor. *Mol. Cell. Biol.* **16,** 5221-5231.

Han, K., Levine, M. S. and Manley, J. L. (1989). Synergistic activation and repression of transcription by *Drosophila* homeobox proteins. *Cell* **56,** 573-583.

Hankinson, O. (1995). The Aryl hydrocarbon receptor complex. *Ann. Rev. Pharmacol. Toxicol.* **35,** 307-40.

Hirose, K., Morita, M., Ema, M., Mimura, J., Hamada, H., Fujii, H., Saijo, Y., Gotoh, O., Sogawa, K. and Fujii-Kuriyama, Y. (1996). cDNA cloning and tissue-specific expression of a novel basic-helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt). *Mol. Cell. Bio.l* **16,** 1706-1713.

Hoffman, E. C., Reyes, H., Chu, F., Sander, F., Conley, L. H., Brooks, B. A. and Hankinson, O. (1991). Cloning of a subunit of the DNA-binding form of the Ah (dioxin) receptor. *Science* **252,** 954-958.

Hogenesch, J. B., Chan, W. K., Kackiw, V. H., Brown, R. C., Gu, Y.-Z., Pray-Grant, M., Perdew, G. H. and Bradfield, C. A. (1997). Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway. *J. Biol. Chem.* **272,** 8581-8593.

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

Isaac, D. D. and Andrew, D. J. (1996). Tubulogenesis in *Drosophila*: a requirement for the trachealess gene product. *Genes Dev.* **10,** 103-117.

Jain, S., Dolwick, K. M., Schmidt, J. V. and Bradfield, C. A. (1994). Potent transactivation domains of the Ah receptor and the Ah receptor nuclear translocator map to their carboxyl termini. *J. Biol. Chem.* **269,** 31518-31524.

King, D. P., Zhao, Y., Sangoram, A. M., Wilsbacher, L. D., Tanaka, M.,

Antoch, M. P., Steeves, T. D. L., Vitaterna, M. H., Kornhauser, J. M., Lowrey, P. L., Turek, F. W. and Takahashi, J. S. (1997). Positional cloning of the mouse circadian Clock gene. *Cell* **89,** 641-653.

Klämbt, C., Glazer, L. and Shilo, B.-Z. (1992). *breathless*, a Drosophila FGF receptor homolog, is essential for migration of tracheal and specific midline glial cells. *Genes Dev.* **6,** 1668-1678.

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.

Lewis, E. B. and Bacher, F. (1968). Methods of feeding ethyl methane sulfonate (EMS) to Drosophila males. *Drosophila Inf. Service* **43,** 193.

Li, H., Dong, L. and Whitlock, J. P. (1994). Transcriptional activation function of the mouse Ah receptor nuclear translocator. *J. Biol. Chem.* **269,** 28098-28105.

Lindebro, M. C., Poellinger, L. and Whitelaw, M. L. (1995). Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. *EMBO J.* 101-112.

Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A. and Simon, M. C. (1997). Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT. *Nature* **386,** 403-407.

Mellerick, D. M. and Nirenberg, M. (1995). Dorsal-ventral patterning genes restrict NK-2 homeobox gene expression to the ventral half of the central nervous system of Drosophila embryos. *Dev. Biol.* **171,** 306-316.

Michaud, J. and Fan, C.-M. (1997). *Single-minded* - two genes, three chromosomes. *Genome Res.* **7,** 569-571.

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

Nagao, M., Ebert, B. L., Ratcliffe, P. J. and Pugh, C. W. (1996). Drosophila melanogaster SL2 cells contain a hypoxically inducible DNA binding complex which recognises mammalian HIF-1 binding sites. *FEBS Lett.* **387,** 161-166.

Nambu, J. R., Chen, W., Hu, S. and Crews, S. T. (1996). The *Drosophila melanogaster similar bHLH-PAS* gene encodes a protein related to human Hypoxia-inducible factor 1α and *Drosophila* Single-minded. *Gene* **172,** 249-254.

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

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.

Pollenz, R. S., Sullivan, H. R., Holmes, J., Necela, B. and Peterson, R. E. (1996). Isolation and expression of cDNAs from Rainbow Trout (Oncorhynchus mykiss) that encode two novel basic helix-loop-helix/PER-ARNT-SIM (bHLH/PAS) proteins with distinct functions in the presence of the aryl hydrocarbon receptor. *J. Biol. Chem.* **271,** 30886-30896.

Price, B. D., Chang, Z., Smith, R., Bockheim, S. and Laughon, A. (1993). The *Drosophila neuralized* gene encodes a C3HC4 zinc finger. *EMBO J.* 2411-2418.

Reisz-Porszasz, S., Probst, M. R., Fukunaga, B. N. and Hankinson, O. (1994). Identification of functional domains of the aryl hydrocarbon receptor nuclear translocator protein (ARNT). *Mol. Cell. Biol.* **14,** 6075-6086.

Sogawa, K. R., Nakano, A., Kobayashi, Y., Kikuchi, N., Ohe, N., Matsushita, N. and Fujii-Kuriyama, Y. (1995). Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation. *Proc. Nat. Acad. Sci USA* **92,** 1936-1940.

Sonnenfeld, M. J. and Jacobs, J. R. (1994). Mesectodermal cell fate analysis in *Drosophila* midline mutants. *Mech. Dev.* **46,** 3-13.

Swanson, H. I., Chan, W. K. and Bradfield, C. A. (1995). DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J. Biol. Chem.* **280,** 26292-26302.

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.

Tian, H., McKnight, S. L. and Russell, D. W. (1997). Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev.* **11,** 72-82.

Tower, J., Karpen, G. H., Craig, N. and Spradling, A. C. (1992). Preferential transposition of Drosophila P elements to nearby chromosomal sites. *Genetics* **133,** 347-359.

Wang, G. L., Jaing, B.-H., Rue, E. and Semenza, G. L. (1995). Hypoxia-inducible factor 1 is a basic helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc. Natl. Acad. Sci. USA* **92,** 5510-5514.

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

Wharton, Jr, K. A., Franks, R. G., Kasai, Y. and Crews, S. T. (1994). Control of CNS midline transcription by asymmetric E-box elements: similarity to xenobiotic responsive regulation. *Development* **120,** 3563-3569.

Wilk, R., Weizman, I., Glazer, L. and Shilo, B.-Z. (1996). *trachealess* encodes a bHLH-PAS protein and is a master regulator gene in the *Drosophila* tracheal system. *Genes Dev.* **10,** 93-102.

Xiao, H., Hrdlicka, L. A. and Nambu, J. R. (1996). Alternate function of the *single-minded* and *rhomboid* genes in development of the *Drosophila* ventral neuroectoderm. *Mech. Dev.* **58,** 65-74.

Younossi-Hartenstein, A. and Hartenstein, V. (1993). The role of the tracheae and musculature during pathfinding of *Drosophila* embryonic sensory axons. *Dev. Biol.* **158,** 430-437.

(Accepted 5 September 1997)