Remembrance of things PAS: regulation of development by bHLH–PAS proteins

Stephen T Crews* and Chen-Ming Fan†

Strange fits of passion I have known...
(W Wordsworth, 'Strange fits of passion'.)

bHLH–PAS proteins are regulators of developmental and physiological events that are well conserved between vertebrates and invertebrates. Recent studies using mouse knockouts of bHLH–PAS genes have provided novel insight into the roles of hypoxia inducible factors in controlling oxygen-regulated development and homeostasis, and the role of *Single-minded-1* in regulating development and transcription in the hypothalamus. The *Drosophila spineless* and vertebrate *Aryl hydrocarbon receptor* bHLH–PAS orthologs both function in chemosensory processes, but in fundamentally different ways. *Spineless* controls antennal, limb, and sensory cell development, whereas the Aryl hydrocarbon receptor regulates the response to toxin metabolism. Structural analyses of the PAS domain provide insight into how this interaction domain can act as ligand-binding environmental sensor and signal transducer.

Addresses
*Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7260, USA; e-mail: steve_crews@unc.edu
†Carnegie Institute of Washington, Department of Embryology, Baltimore, Maryland 21210, USA; e-mail: fan@mail1.ciwemb.edu

Current Opinion in Genetics & Development 1999, **9**:580–587

0959-437X/99/$ – see front matter

Abbreviations

Ahr	Aryl hydrocarbon receptor
AIP	Ahr-interacting protein
aPV	anterior periventricular nucleus
Arnt	Aryl hydrocarbon receptor nuclear translocator
bHLH	basic helix-loop-helix
Bmal1	brain and muscle Arnt-like protein 1
CRH	corticotropin-releasing hormone
EPAS1	endothelial PAS domain protein 1
HERG	human Ether-a-Go-Go
HIF	hypoxia inducible factor
HRE	hypoxia-response element
Hsp90	heat-shock protein 90
NPAS	neuronal PAS domain protein
OT	oxytocin
OZ	Organ of Zuckerkandl
PAS	Per, Arnt, Sim
Per	Period
PVN	paraventricular nucleus
PYP	photoactive yellow protein
Sim	single-minded
SON	supraoptic nucleus
Ss	Spineless
SS	somatostatin
TRH	thyrotropin-releasing hormone
VEGF	vascular endothelial growth factor
Vnd	ventral nervous system defective
VP	vasopressin

Introduction

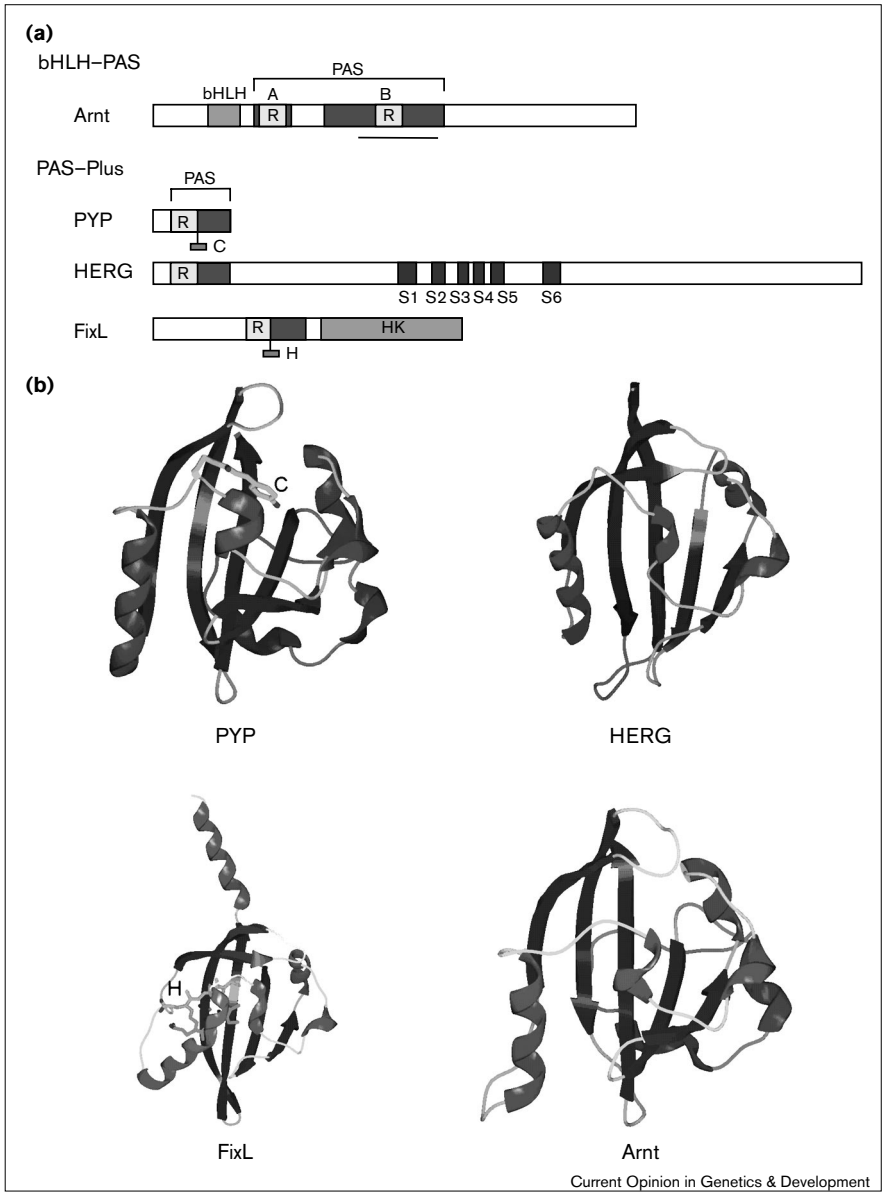
The PAS domain is a multi-functional interaction domain found on proteins from bacteria to humans. There are two distinct groups of PAS proteins (Figure 1a). One group has basic helix-loop-helix (bHLH) motifs in addition to the PAS domains. These bHLH–PAS proteins generally form heterodimeric transcription factors (the Period proteins, which only have PAS domains, are the exception within this group) (reviewed in [1•]). The second group (referred to here as the PAS–Plus group) has a PAS domain associated with a variety of functional domains, including kinase, zinc finger, transmembrane, and chromophore-binding domains [2–4]. The PAS–Plus group of proteins are often direct sensors of environmental and physiological signals.

In this review, we focus on the structure of the PAS domain and recent work on the roles of bHLH–PAS proteins during development, with particular emphasis on the mammalian bHLH–PAS proteins and comparisons to their *Drosophila* counterparts. One prominent group of bHLH–PAS proteins controls circadian rhythms. Clock and Bmal1 form a heterodimer that positively regulates circadian rhythm-expressed genes, and are negatively regulated by Period. These proteins have been comprehensively reviewed elsewhere [5•].

The PAS domain: conserved structure of an interaction domain

The PAS domain has a variety of known functions that include ligand and protein–protein interactions. Structural determination of the bacterial photoactive yellow protein (PYP) blue-light photoreceptor [6], the FixL heme-binding oxygen sensor [7•], and the human Ether-a-Go-Go (HERG) voltage-dependent K+ channel [8•], all PAS–Plus proteins have demonstrated that the core region of the PAS domain is structurally conserved and can be superimposed even though the sequence conservation is low (reviewed in [9•]). The structures all consist of a 5- or 6-stranded antiparallel β-sheet flanked by α-helices and loops with a 'vine' running across and connecting the β-sheet (Figure 1b). This structure forms a PAS fold or 'glove' in which ligands such as the 4-hydroxycinnamyl chromophore of PYP [6] and heme of FixL [7•] reside. It is proposed that light stimulation of the chromophore induces a conformational change in PYP that allows interaction of this PAS protein with a downstream signaling component. Similarly, absence of heme-bound O_2 results in a conformational change within FixL triggering a phosphorylation cascade involving the FixL histidine kinase domain. Residues in the HERG PAS domain that are required for its function [8•] are postulated to interact with other protein surfaces. Recent molecular modeling reveals that the PAS-B domain of Aryl hydrocarbon receptor nuclear translocator (Arnt, a

Figure 1



Structures of PAS proteins. **(a)** Schematic of representative bHLH–PAS and PAS–Plus proteins. Sequence structure of human Arnt, a typical bHLH–PAS protein, showing the bHLH region and the PAS domain. The PAS domain consists of two conserved regions, PAS-A and PAS-B (dark shading) separated by a spacer. Within both PAS-A and PAS-B is a repeated segment (R). The region of PAS-B modeled in the three-dimensional structure shown in (b) is underlined. All animal bHLH–PAS proteins have a similar domain organization. Three PAS–Plus proteins are shown: photoactive yellow protein (PYP) from purple bacteria, human Ether-a-go-go (HERG), and rhizobial FixL. The PAS domain is shown as a darkly shaded box and the locations of sequences corresponding to the PAS repeat (R) are indicated. PYP has a 4-hydroxycinnamyl chromophore (C) bound to the protein, HERG has six transmembrane domains (S1–6), and FixL is a heme-binding protein (H) that also contains a histidine kinase domain (HK). **(b)** Shown are protein structures either determined by analysis of crystals (PYP, HERG, and FixL) or modeled based on the PYP structure (Arnt). Note the common structure consisting of a β-sheet flanked by α-helices. Fitting within the PAS folds are the chromophore (C) for PYP, and heme (H) for FixL.

bHLH–PAS protein) can be modeled into a structure resembling PYP [10•], further supporting the general structural similarity of PAS domains. This conserved surface may mediate protein–protein interactions of bHLH–PAS proteins [10•].

Some bHLH–PAS and PAS–Plus proteins have sensing and signaling regions within the PAS domain (e.g. Aryl hydrocarbon receptor [Ahr], PYP, FixL) and others have these functional regions outside the PAS domain (e.g. hypoxia inducible factor-1α [HIF-1α], phytochromes). Nevertheless, the structure of the PAS core provides a general framework for examining how bHLH–PAS and PAS–Plus proteins interact with other signaling components and proteins.

Hypoxia inducible factors mediate O_2 homeostasis during development

bHLH–PAS proteins play an important role in responding to low oxygen levels in vertebrates and probably invertebrates [11•,12]. There are three members of this family (referred to as hypoxia inducible factors [HIFs]): HIF-1α, endothelial PAS domain protein 1 (EPAS1), and HIF-3α [13–15], and possibly additional members such as neuronal PAS domain proteins NPAS1 and NPAS2 [16,17]. These proteins can dimerize with Arnt [13–15], Arnt2, and Bmal1 [18,19]. The potential of HIF proteins to dimerize with both Arnt and Bmal1 [18,19] raises the possibility of cross-regulation of O_2 homeostasis and circadian rhythms, as Bmal1 is a component of the transcription complex that regulates circadian rhythms.

HIF-1α protein is stabilized under hypoxic conditions by a poorly characterized O_2 sensing pathway. It then dimerizes with Arnt and binds hypoxia response elements (HREs; ACGTG core sequence) on target genes. There is a large body of evidence [11•] indicating that HIF-1α mediates the physiological response to both hypoglycemia and hypoxia by upregulating genes that encode glycolytic enzymes, EPO (erythropoietin), and VEGF (vascular endothelial growth factor) [11•]. Recent work has also demonstrated developmental roles for HIFs.

Mouse knockouts of *HIF-1α* [20••–22••] and *Arnt* [23,24] revealed similar but not identical developmental defects. *HIF-1α* null mutants are embryonic lethal. By embryonic day 9 (E9), there was a reduction in vascularization, particularly in the yolk sac and cephalic region. Blood vessel formation was properly initiated but later the normally observed dense vascular structures were replaced by an enlarged and much less branched vasculature, suggesting that the defects are angiogenic. Early mammalian embryonic development proceeds under reduced oxygen levels. The *HIF-1α* mutant developmental defects are consistent with an inability of the embryo to establish an elaborate blood vessel network in response to hypoxia.

Since HIF-1α forms heterodimers with Arnt, it is not surprising that mutants of *Arnt* are embryonic lethal and show defects in vascular structures [23,24]. Abnormalities were observed in placental, yolk sac, and branchial vascularization, as well as formation of the neural tube. The two *Arnt* mutant reports [23,24] differed in yolk sac defects that were observed, likely because of differences in genetic background. The *Arnt* defects overlap with *HIF-1α* defects (e.g. yolk sac vascularization) but are generally less severe. In particular, massive cell death of cephalic mesenchymal cells occurs by E10 [21••,25•] and causes the failure of neural tube closure in *HIF-1α* mutants but not in *Arnt* mutants. Another phenotype observed only in *HIF-1α* mutant mice was myocardial hypertrophy. One possibility is that *Arnt2* and/or *Bmal1* compensates for *Arnt* at these sites. However, mice carrying deletions of *Arnt2* survive up to birth [26•] with characterized thymic defects only. Whether any of the multiply mutant combinations of *Arnt*, *Arnt2* and *Bmal1* will fully recapitulate the *HIF-1α* phenotype remains to be determined.

The key feature of both *HIF-1α* and *Arnt* mutants is that they affect angiogenesis. This is strikingly similar to what is found in *VEGF* mutants, which are defective in both vasculogenesis and angiogenesis [27,28]. There is considerable evidence from tissue culture experiments that HIF-1α:Arnt regulates *VEGF* expression [20••–22••,23]. Correspondingly, *Arnt* mutant mice show reduced *VEGF* expression in the yolk sac and other embryonic sites [23]. In addition, cephalic vascular defects are observed in both *HIF-1α* and *VEGF* mutant mice. Surprisingly, *HIF-1α* mutant embryos express higher levels of *VEGF* mRNA than wild-type embryos [25•], though it is not clear from which tissue the *VEGF* RNA is derived nor if there is compensation by other non-hypoxia-dependent *VEGF* activation mechanisms in *HIF-1α* mutants. Nevertheless, this result questions the unique role of HIF-1α:Arnt in *VEGF* gene regulation. How, then, do *HIF-1α* and *VEGF* mutant mice have near identical vascular defects? One possibility is that cephalic cell death results in depletion of the mesenchymal support cells (pericytes) that are required to support the endothelial cells, thus leading to vascular regression. The other possibility is that endothelial cells mutant for *HIF-1α* are unable to mediate the cellular response to hypoxia even though *VEGF* is overproduced. This model postulates that vascular regression occurs because of a cell-autonomous function of *HIF-1α*.

Expression of *EPAS1*, the second HIF, revealed that it is predominantly expressed in endothelial cells [14], suggesting that it plays a role in hypoxia regulation distinct from *HIF-1α*. Analysis of *EPAS1* mutant embryos [29••] revealed that although morphologically normal (including the vascular system) they do not survive past E15.5, well after the stage when *HIF-1α* mutant embryos die. This suggested that lethality of *EPAS1* mutant embryos occurs as a result of physiological rather than morphological defects. Other than endothelial cells, it was noted that *EPAS1* was prominently expressed in the organ of Zuckerkandl (OZ), which lies near the kidney and testis. The OZ is an embryonic site of oxygen-regulated release of catecholamines. The catecholamines control heart rate. In an elegant series of experiments, *EPAS1* mutants were shown to have decreased heart rate and reduced noradrenaline, the most significant catecholamine in the fetus. Lethality of *EPAS1* mutant mice could be rescued by administration of D,L-threo-3,4-dihydroxyphenylserine, a substrate for noradrenaline. Target genes of *EPAS1* have not been identified, but potential candidates are genes encoding catecholamine biosynthetic enzymes and proteins involved in catecholamine release or uptake. The OZ degenerates during embryogenesis, and the carotid body assumes the function of controlling post-natal respiration and cardiovascular output via O_2 levels and catecholamine synthesis and release. *EPAS1* shows prominent expression in the carotid body suggesting a role in O_2 control during embryogenesis and in the adult.

Thus, *HIF-1α* is involved in angiogenesis and cell survival and *EPAS1* functions to regulate catecholamine levels during development. This apparent difference in phenotype and function provides evidence for separable hypoxic adaptation pathways during embryogenesis: *HIF-1α* acts at a 'local' cellular level and *EPAS1* acts at a 'systemic' level. Many issues remain including the role of *EPAS1* in endothelial cells and the postembryonic roles of HIFs. Some members of the HIF subfamily may partially compensate for each other. The use of multiply mutant mice and tissue/temporal-specific knockouts can address these issues. Another important issue concerns how HIFs sense and respond to varying O_2 levels and whether all HIFs use the same sensing pathway. It is

provocative that the bacterial FixL protein binds and senses O_2 via heme binding to a PAS domain. Will this prokaryotic mechanism for sensing O_2 be conserved in eukaryotic HIFs?

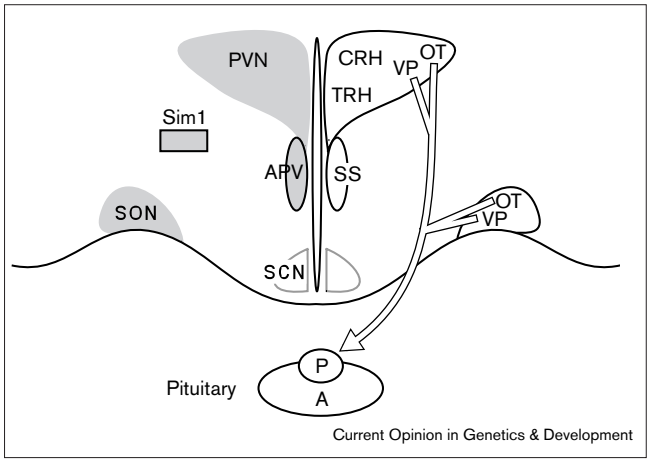
Mammalian bHLH–PAS proteins control the development and physiology of oxygen homeostasis. Is this the case for invertebrates? Insects are relatively resistant to hypoxia but have the ability to respond physiologically [30], developmentally [31], and transcriptionally [32]. The *Drosophila* Trachealess bHLH–PAS protein [33,34] forms a heterodimer with Tango, the *Drosophila* Arnt ortholog [35,36•,37], to direct the development of the trachea, the insect respiratory organ. There is no evidence, however, that Trachaeless:Tango responds to different levels of O_2. Work performed over 40 years ago demonstrated that terminal tracheal branching is regulated by the oxygenation of cells [12,31]. In one plausible model, low O_2 levels within a cell triggers activation of a HIF bHLH–PAS protein that signals the tracheoles to further branch and deliver additional O_2 [12]. Direct evidence for this model has not yet been found. However, there is a hypoxia-inducible HRE-binding activity present in *Drosophila* tissue culture cells [38] and the *Drosophila* Similar bHLH–PAS protein [39] dimerizes with Tango [35], and possesses an O_2-regulated domain found in mammalian HIFs [40•]. Although preliminary, these studies suggest that insects respond to O_2 levels in a manner similar to vertebrates, and work in *Drosophila* has the potential to provide a useful genetic and molecular tool for understanding general mechanisms of O_2 sensing and transcriptional pathways.

Murine *Sim1* and *Drosophila sim*: homologs with similar functions in controlling development of neuronal cell lineages

The *Drosophila single-minded* (*sim*) gene controls transcription and development of the CNS midline cell lineage [1•]. Sim forms heterodimers with Tango, and the complex binds midline enhancer elements (ACGTG core sequence) that reside on target genes [35]. Mammals have two *sim* genes — *Sim1* and *Sim2* [41]. Mutant analyses have demonstrated that *Sim1* functions to control development of specific cell types within the central nervous system like its *Drosophila* counterpart.

Sim1 is expressed in the precursor cells of the paraventricular nucleus (PVN), the anterior periventricular nucleus (aPV), and the supraoptic nucleus (SON) of the mouse hypothalamus [42,43••] (Figure 2). In *Sim1* null mutant mice, the PVN, aPV, and SON cells do not terminally differentiate to express the neuropeptides TRH (thyrotropin-releasing hormone), SS (somatostatin), CRH (corticotropin-releasing hormone), VP (vasopressin), and OT (oxytocin), which define at least five major neuroendocrine cell types within the PVN/SON. As the corresponding PVN/SON region appears hypocellular in newborn mutant mice, it was suggested that these cells died. The PVN/SON has been implicated in providing a trophic factor(s) for the pituicytes located in the posterior pituitary. Consistently, the number of pituicytes in *Sim1* mutants was reduced whereas the anterior pituitary developed normally. It is presumed that the lack of PVN/SON and posterior pituitary function causes the perinatal lethality of *Sim1* mutant mice. Intriguingly, insect brain midline neurons, the formation of which are likely governed by *sim*, include a number of neurosecretory and neuropeptide-producing cells including CRH-like diuretic peptides [44].

Figure 2

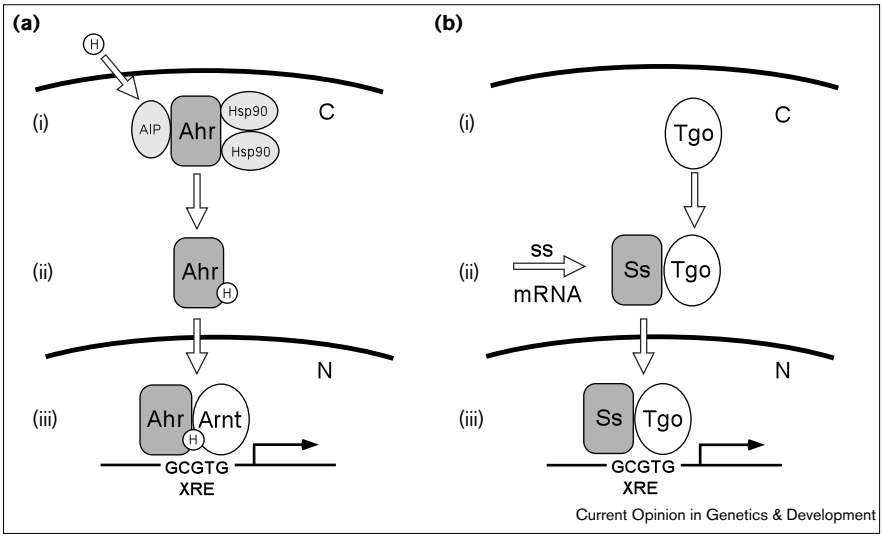


Sim1 expression and function in the hypothalamus–pituitary axis. *Sim1* is expressed in the PNV, APV, and SON of the hypothalamus (depicted by shading on the left side). The corresponding right side shows the five major neuropeptide-secreting cell types, CRH, TRH, SS, OT, and VP, located within the three nuclei. OT and VP neurons project (arrow) to the posterior (P) pituitary. In *Sim1* mutant mice, all five cell types fail to terminally differentiate. In addition, the posterior pituitary is reduced in size because of the lack of the SON/PVN, which presumably secretes trophic factors required for the growth of the posterior pituitary. A, anterior pituitary.

The phenotype described for *Sim1* mutant mice is similar to that of mice mutant for *Brain-2*, a POU domain transcription factor gene, which also lacks CRH, OT, and VP neurons in the PVN/SON and has a reduced number of pituicytes. As *Brain-2* expression in the prospective PVN/SON is missing in the *Sim1* mutant, *Sim1* most likely utilizes *Brain-2* to relay its function in CRH, OT and VP neuron specification. Interestingly, the *Drosophila Drifter* gene [45] is a homolog of *Brain-2* and functions downstream of *sim* in CNS midline cell development.

Although genetically *Sim1* acts upstream of *Brain-2* in specifying the CRH/VP/OT neurons, this does not preclude its direct involvement in regulating CRH/VP/OT transcription. The recent observation [46] that Clock:Bmal1 heterodimers activate the VP promoter in the SCN raises the possibility that Sim1 together with Arnt or Arnt2 may directly regulate the expression of VP, and possibly OT, CRH, TRH, or SS in the PVN/SON both during development as well as after birth. It remains an intriguing possibility that *Sim1* mediates the

Figure 3



Different modes of Ahr/Spineless function. **(a)** In vertebrates, Ahr acts as a cytoplasmic receptor that binds aryl hydrocarbons and directs transcription of genes that encode hydrocarbon-metabolizing enzymes. (i) Initially, it is maintained in a ligand-responsive state as a complex with Hsp90 and Ahr-interacting protein (AIP). (ii) Upon binding to a hydrocarbon ligand (H), Ahr dissociates from AIP and Hsp90, translocates to the nucleus (N), and dimerizes with Arnt. (iii) The Arh:Arnt complex binds to the xenobiotic response element (XRE) on target genes. **(b)** The *Drosophila spineless* gene (*ss*) is a developmental regulator that behaves differently from vertebrate Ahr. (i) In the absence of a bHLH–PAS partner protein, Tango protein (Tgo) resides in the cytoplasm (C). (ii) When the *ss* gene is activated (e.g. in the distal antenna), *ss* mRNA and Ss protein are synthesized. Within the cytoplasm, Ss and Tgo proteins dimerize. (iii) The Ss:Tgo protein complex enters cell nuclei, and binds XRE sequences on target genes.

homeostatic capacity of the hypothalamus by responding to the physiological status of the organism (perhaps by binding effector molecules directly), and controls stress by regulating hormonal gene expression. In this regard, although *Drosophila sim* function is required for *Drifter* expression, Sim also likely functions later to control midline glial transcription in combination with Drifter and other transcription factors [47,48•].

In line with evolutionary conservation, one might expect the mammalian *Sim* genes to be expressed in the ventralmost portion of the spinal cord, the floorplate, which shares similar functions with the insect CNS midline cells, including guiding axons across the midline — however, neither *Sim1* nor *Sim2* are expressed there. *Sim1* is expressed in the V3 neurons *adjacent* to the floorplate (alas, so close but so far!) [42,49•]. The insect CNS midline cells not only function in midline-directed axon guidance but also consist of functional neurons and specialized glial cells. In contrast, the vertebrate floorplate does not consist of neurons but is comprised of a specialized cell type. It could be that the multiple functions of the insect midline cells are subdivided in the vertebrate CNS into more specialized cells, and that *Sim1* functions only in the V3 subset and not the floorplate. Interestingly, *Sim1* expression in the V3 cells is abolished in *Nkx2.2* mutant mice [49•]. This mutant results in transformation of V3 cells to a more dorsal, motoneuron fate. *Nkx2.2* is the homolog of the *Drosophila ventral nervous system defective (vnd)* gene, which controls the development of the ventral CNS, but is repressed in the CNS midline cells by *sim* [50,51]. Although *Nkx2.2* and *vnd* are conserved in patterning ventral neural tissue, their regulatory hierarchy and mode of regulation (activation vs. repression) with respect to *Sim1* and *sim* appears to be reversed in the two organisms.

Spineless and Ahr: homologs with different functions in vertebrates and insects

Ahr and Arnt proteins form a heterodimeric DNA-binding complex that controls the physiological response to toxin metabolism [52]. They are highly conserved in the vertebrate species examined, and broadly expressed. Ahr acts as a cytoplasmic receptor for aryl hydrocarbons (e.g. dioxin). Upon binding ligand, Ahr translocates to the nucleus, and forms a DNA-binding activation complex with Arnt (Figure 3a). The physiological response to aryl hydrocarbons has not been observed in insects and other invertebrates [53]. This is despite the fact that well-conserved Ahr and Arnt genes are found in both *Drosophila* [54•] and *Caenorhabditis elegans* [55•].

The *Ahr* homologs in *Drosophila* is the *spineless* gene [54•]. Null mutants of *spineless* are viable and show a number of adult phenotypes. The distal region of the leg (tarsus) is deleted and sensory bristles are reduced in size. The most striking mutant phenotype is a transformation of distal antenna into distal leg. *spineless* is specifically expressed in the affected cells — the precursors to antenna, sensory cells, and tarsus. Spineless forms a DNA-binding heterodimeric complex with Tango (Figure 3b) [56•]. Although localized in the cytoplasm of most cells, Tango accumulates in the nucleus at all embryonic sites with ectopic *spineless* expression, further supporting the notion that Spineless and Tango form a heterodimeric complex *in vivo*. This result argues against Spineless acting as a cytoplasmic receptor for a diffusible ligand that controls its nuclear transport and function, Instead, it suggests that

spineless function is directly associated with its expression, and if there is a ligand for Spineless it is ubiquitously present and most likely not regulatory.

In all aspects, Spineless functions in a similar way to the Sim and Trachealess developmental regulatory proteins that also interact with Tango [36•]. One important technical development is the ability to use *spineless* heterozygous flies to genetically screen for interacting genes [56•]. Heterozygous deficiencies of *spineless* have a weak antennal defect. This feature was employed to screen for dominant enhancers of the *spineless* antennal phenotype. One of the *Enhancer of spineless* complementation groups was *tango*, which validates the ability of the screen to identify relevant interacting genes. This type of approach, coupled with the *Drosophila* genome project, promises to be a powerful tool to understand how bHLH–PAS genes control development.

Drosophila spineless has developmental roles but no known role(s) in toxin metabolism. Vertebrate *Ahr* primarily functions in a physiological role, although *Ahr* knockout mice show modest developmental defects of the liver [57,58]. It has been argued that ancestral *Ahr/spineless* was involved in chemosensation [56•]. In vertebrates, the Ahr protein took on a role as a toxin sensor, and in arthropods, Ss functions to control development of the antenna, a chemosensory organ. It will be particularly interesting to determine the function of *Ahr* from *C. elegans* as well as other organisms.

Conclusions and perspectives

bHLH–PAS proteins represent an important link between environment, physiology, and development, and progress in understanding their function has greatly accelerated in recent years. The genome projects will undoubtedly identify additional PAS proteins. It will be particularly interesting to identify new animal PAS–Plus proteins (to date they have been mainly identified in plants and bacteria) and determine whether they act as 'direct' sensing molecules. Study of bHLH–PAS protein function needs to determine the extent that these proteins mediate environmental influences on development and physiology. With the notable exception of mutants of the rhythm and *Ahr/spineless* PAS genes, which are viable, most bHLH–PAS gene mutants are embryonic lethal. The full spectrum of bHLH–PAS gene functions, both developmental and physiological, will require an examination of their roles later in development and in adults. Different bHLH–PAS proteins often bind the same DNA element; this may allow for convenient cross-regulation and the integration of multiple environmental and cellular signals. It will be important to identify additional *cis* and *trans*-acting factors that dictate transcriptional specificity and cross-regulation of these bHLH–PAS proteins. *Drosophila* will prove to be particularly useful, as a result of the use of speedy and sophisticated genetic and transgenic approaches.

Despite the striking similarities between invertebrate and vertebrate bHLH–PAS proteins, there are some interesting differences. Vertebrate Ahr is a receptor and requires ligand binding to function appropriately, whereas the *Drosophlia* developmental regulatory proteins, Sim, Trachealess, and Spineless, function in a ligand-independent mode. Does the necessity for ligand interactions represent a fundamental distinction between vertebrate and invertebrate bHLH–PAS proteins? It will be interesting to determine whether vertebrate developmental bHLH–PAS proteins, such as Sim1 and Sim2, require a ligand for proper function. Another distinction is that *Drosophila* Arnt/Tango is cytoplasmic in the absence of bHLH–PAS partner proteins, whereas vertebrate Arnt is predominantly nuclear. Does this also represent a fundamental difference between vertebrates and invertebrates in how bHLH-PAS proteins function, particularly in the ability to mediate signaling pathways? Finally, it is important to use the knowledge gained from understanding how these important regulatory proteins function to resolve practical problems in medicine and agriculture. Human genetics and molecular genetic analysis in animal model systems can provide insight into pathologies associated with bHLH–PAS gene polymorphisms, including effects on heterozygous individuals. Structural knowledge of the ligand-binding and interaction properties of the PAS domain now permits novel approaches for designing ligand-responsive proteins.

Acknowledgements

For simplicity, we have usually identified genes described in this review by a single name and reference, even though in numerous cases they were discovered by more than one lab. We would like to thank Chris Doe, Elizabeth Getzoff, Noah May, Jacques Michaud, Tonia Von Ohlen, Jean-Luc Pellequer, Greg Semenza, Celeste Simon, Paul Taghert, and Brenda Temple for helpful advice and assistance. Work in our labs is supported by NICHD and the National Science Foundation (ST Crews) and National Institutes of Health, the John Merck Fund, the Arnold and Mabel Beckman Foundation, the Sloan Foundation, and a Damon Runyon Scholar Award (C-M Fan).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Crews ST: **Control of cell lineage-specific development and transcription by bHLH-PAS proteins.** *Genes Dev* 1998, **12**:607-620.

A general review of bHLH–PAS proteins focusing predominantly on the role of *Drosophila* bHLH–PAS proteins in controlling development.

2. Lagarias DM, Wu S-H, Lagarias JC: **Atypical phytochrome gene structure in the green alga *Mesotaenium caldariorum*.** *Plant Mol Biol* 1995, **29**:1127-1142.

3. Zhulin IB, Taylor BL, Dixon R: **PAS domain S-boxes in Archae, Bacteria and sensors for oxygen and redox.** *Trends Biochem Sci* 1997, **22**:331-333.

4. Ponting CP, Aravind L: **PAS: a multifunctional domain family comes to light.** *Curr Biol* 1997, **7**:674-677.

5. • Dunlap JC: **Molecular bases for circadian clocks.** *Cell* 1999, **96**:271-290.

This article reviews the molecular control of circadian rhythms in fungi, plants, insects, and mammals, and describes the roles of bHLH-PAS proteins in rhythm gene regulation.

6. Borgstahl GEO, Williams DR, Getzoff ED: **1.4 Å structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore.** *Biochemistry* 1995, **34**:6278-6287.

7. Gong W, Hao B, Mansy SS, Gonzalez G, Gilles-Gonzalez MA,
• Chan MK: **Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction.** *Proc Natl Acad Sci USA* 1998, **95**:15177-15182.
Like PYP, the structure of FixL illustrates how the heme ligand fits into the PAS domain glove. Will this mechanism for O_2-sensing PAS proteins also be utilized by vertebrate O_2-sensing bHLH–PAS proteins?

8. Cabral JH, Lee A, Cohen SL, Chait BT, Li M, Mackinnon R: **Crystal**
• **structure and functional analysis of the HERG potassium channel N terminus: a eukaryotic PAS domain.** *Cell* 1998, **95**:649-655.
HERG has an intracellular PAS domain. The structure closely resembles PYP. Mutational analysis suggests that the PAS domain is an interaction domain, and that it may react with other regions of HERG. The data presented in this study and two others [6,7•] strongly support the notion that there is a common structure to the PAS domain.

9. Pellequer J-L, Brudler R, Getzoff ED: **Biological sensors: more than**
• **one way to sense oxygen.** *Curr Biol* 1999, **9**:416-4180
A review of PAS protein structure with an emphasis on the oxygen-sensing FixL PAS protein.

10. Pellequer J-L, Wager-Smith KA, Kay SA, Getzoff ED: **Photoactive**
• **yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily.** *Proc Natl Acad Sci USA* 1998, **95**:5884-5890.
The authors of this report used the structure of PYP to model a PAS domain of the Arnt bHLH–PAS protein. The conclusion is that the structure of the Arnt PAS-B domain strongly resembles PYP and it is likely that all PAS domains will have a similar structure. This prediction has, so far, concurred with the structures of FixL and HERG. The authors make testable predictions regarding sequences potentially used for bHLH–PAS protein heterodimerization and other ligand and protein interactions.

11. Semenza GL: **Hypoxia-inducible factor 1: master regulator of O_2**
• **homeostasis.** *Curr Opin Genet Dev* 1998, **8**:588-594.
This article and [12] are general reviews regarding recent work on the physiological and developmental roles of mammalian hypoxia-inducible transcription factors.

12. Guillemin K, Krasnow MA: **The hypoxic response: huffing and HIFing.** *Cell* 1997, **89**:9-12.

13. Wang GL, Jiang B-H, Rue E, Semenza GL: **Hypoxia-inducible factor 1 is a basic helix-loop-helix-PAS heterodimer regulated by cellular O_2 tension.** *Proc Natl Acad Sci USA* 1995, **92**:5510-5514.

14. Tian H, McKnight SL, Russell DW: **Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells.** *Genes Dev* 1997, **11**:72-82.

15. Gu Y-Z, Moran SM, Hogenesch JB, Wartman L, Bradfield CA: **Molecular characterization and chromosomal localization of a third a-class hypoxia inducible factor subunit, HIF3α.** *Gene Expression* 1998, **7**:205-213.

16. Zhou Y-D, Barnard M, Tian H, Ring H, Francke U, Shelton J, Richardson J, Russell DW, McKnight SL: **Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system.** *Proc Natl Acad Sci USA* 1997, **94**:713-718.

17. Hogenesch JB, Chan WK, Kackiw VH, Brown RC, Gu Y-Z, Pray-Grant M, Perdew GH, Bradfield CA: **Characterization of a subset of the basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway.** *J Biol Chem* 1997, **272**:8581-8593.

18. Takahata S, Sogawa K, Kobayashi A, Ema M, Mimura J, Ozaki N, Fujii-Kuriyama Y: **Transcriptionally active heterodimer formation of an Arnt-like PAS protein, Arnt3, with HIF-1α, HLF, and Clock.** *Biochem Biophys Res Commun* 1998, **248**:789-794.

19. Hogenesch JB, Gu Y-Z, Jain S, Bradfield CA: **The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors.** *Proc Natl Acad Sci USA* 1998, **95**:5474-5479.

20. Iyer NV, Kotch LE, Agani F, Leung SW, Laughner E, Wenger RH,
•• Gassmann M, Gerhart JD, Lawler AM, Yu AY, Semenza GL: **Cellular and developmental control of O_2 homeostasis by hypoxia-inducible factor 1α.** *Genes Dev* 1998, **12**:149-162.
See annotation [22••].

21. Ryan HE, Lo J, Johnson RS: **HIF-1α is required for solid tumor**
•• **formation and embryonic vascularization.** *EMBO J* 1998, **17**:3005-3015.
See annotation [22••].

22. Carmeliet P, Dor Y, Herbert JM, Fukumura D, Brusselmans K,
•• Dewerchin M, Neeman M, Bono F, Abramovitch R, Maxwell P *et al.*: **Role of HIF-1α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis.** *Nature* 1998, **394**:485-490.
This paper and [20••,21••] describe the initial phenotypic analyses of mouse knockouts of the HIF-1α gene and its developmental roles.

23. Maltepe E, Schmidt JV, Baunoch D, Bradfield CA, Simon MC: **Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT.** *Nature* 1997, **386**:403-407.

24. Kozak KR, Abbott B, Hankinson O: **ARNT-deficient mice and placental differentiation.** *Dev Biol* 1997, **191**:297-305.

25. Kotch LE, Iyer NV, Laughner E, Semenza GL: **Defective**
• **vascularization of HIF-1α-null embryos is not associated with VEGF deficiency but with mesenchymal cell death.** *Dev Biol* 1999, **209**:254-267.
This report provide a detailed examination of mesenchymal cell death in *HIF-1α* mouse knockouts. The surprising conclusion is that defective vascularization associated with mesenchymal cell death is not caused by an absence of VEGF. In contrast, the authors report increased levels of VEGF in *HIF-1α* mutant mice, although levels in the mesenchymal cells were not directly examined.

26. Wines ME, Tiffany AM, Holdener BC: **Physical localization of the**
• **mouse *aryl hydrocarbon receptor nuclear translocator-2* (*Arnt2*) gene with the *c^{112K}* deletion.** *Genomics* 1998, **51**:223-232.
This report demonstrates that a deletion of the *Arnt2* locus is viable. The only phenotypes observed were defects in the thymus.

27. Carmeliet P, Ferreira V, Breier G, Pollefeyt S, Kieckens L, Gertsenstein M, Fahrig M, Vandenhoeck A, Harpal K, Eberhardt C *et al.*: **Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele.** *Nature* 1996, **380**:435-439.

28. Ferrara N, Carver-Moore K, Chen H, Dowd M, Lu L, O'Shea KS, Powell-Braxton L, Hillan KJ, Moore MW: **Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene.** *Nature* 1996, **380**:439-442.

29. Tian H, Hammer RE, Matsumoto AM, Russell DW, McKnight SL: **The**
•• **hypoxia-responsive transcription factor EPAS1 is essential for catecholamine homeostasis and protection against heart failure during embryonic development.** *Genes Dev* 1998, **12**:3320-3324.
An elegant paper describing mouse knockout of the *EPAS1* gene. *EPAS1* mice die late in embryogenesis from cardiac failure. *EPAS1* functions during embryogenesis in the Organ of Zuckerkandl by controlling the hypoxia-regulated release of catecholamines. This neurotransmitter controls fetal heart beat.

30. Krishnan SN, Sun Y, Mohsenin A, Wyman RJ, Haddad GG: **Behavioral and electrophysiologic responses of *Drosophila melanogaster* to prolonged periods of anoxia.** *J Insect Physiol* 1997, **43**:203-210.

31. Wigglesworth VB: **Growth and regeneration in the tracheal system of an insect, *Rhodnius prolixus* (Hemiptera).** *Q J Microsc Sci* 1954, **95**:115-137.

32. Ma E, Haddad GG: **Anoxia regulates gene expression in the central nervous system of *Drosophila melanogaster*.** *Mol Brain Res* 1997, **46**:325-328.

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

34. Isaac DD, Andrew DJ: **Tubulogenesis in *Drosophila*: a requirement for the *trachealess* gene product.** *Genes Dev* 1996, **10**:103-117.

35. Sonnenfeld M, Ward M, Nystrom G, Mosher J, Stahl S, Crews S: **The *Drosophila tango* gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development.** *Development* 1997, **124**:4583-4594.

36. Ward MP, Mosher JT, Crews ST: **Regulation of *Drosophila* bHLH**
• **PAS protein cellular localization during embryogenesis.** *Development* 1998, **125**:1599-1608.
In the absence of other bHLH–PAS partner proteins, Tango is cytoplasmic; however, it is nuclear in cells that express Sim and Trh, which are dimerization partners of Tango. When either *sim* or *trh* are ectopically expressed, their products and Tango are nuclear. This suggests a model in which Sim/Trh forms a heterodimer with Tango in the cytoplasm and the complex translocates into the nucleus. The paper argues against ligand-driven nuclear translocation of Sim and Trh.

37. Ohshiro T, Saigo K: **Transcriptional regulation of *breathless* FGF receptor gene by binding of TRACHEALESS/dARNT heterodimers to three central midline elements in *Drosophila* developing trachea.** *Development* 1997, **124**:3975-3986.

38. Nagao M, Ebert BL, Ratcliffe PJ, Pugh CW: ***Drosophila melanogaster* SL2 cells contain a hypoxically inducible DNA binding complex which recognises mammalian HIF-1 binding sites.** *FEBS Lett* 1996, **387**:161-166.

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

40. • Bacon NCM, Wappner P, O'Rourke JF, Bartlett SM, Shilo B, Pugh CW, Ratcliffe PJ: **Regulation of the *Drosophila* bHLH-PAS protein Sima by hypoxia: functional evidence for homology with mammalian HIF-1α.** *Biochem Biophys Res Comm* 1998, **249**:811-816.
Demonstrates that the *Drosophila* Similar bHLH–PAS protein [39] has a carboxy-terminal region (not in the PAS domain) that, when fused to Gal4, can confer O_2-sensitive regulation. This region is functionally analogous to the mammalian HIF protein hypoxia responsive domain-1 and suggests that *Drosophila* Similar is an insect HIF.

41. Michaud J, Fan C-M: ***Single-minded* — two genes, three chromosomes.** *Genome Research* 1997, **7**:569-571.

42. Fan C-M, Kuwana E, Bulfone A, Fletcher CF, Copeland NG, Jenkins NA, Crews S, Martinez S, Puelles L, Rubenstein JLR, Tessier-Lavigne M: **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 Neuro* 1996, **7**:1-16.

43. •• Michaud JL, Rosenquist T, May NR, Fan C-M: **Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1.** *Genes Dev* 1998, **12**:3264-3275.
In this study, the authors generated murine knockouts of the *Sim1* gene, demonstrating that the PVN, SON, and aPV nuclei of the hypothalamus failed to form and that expression of genes in those lineages was altered. Evidence is provided for transcriptional regulatory hierarchy between *Sim1* and *Brain-2*, which is conserved between their fly homologs *sim* and *drifter* [45], respectively.

44. Patel M, Chung J-S, Kay I, Mallet AI, Gibbon CR, Thompson KSJ, Bacon JP, Coast GM: **Localization of *Locusta*-DP in locust CNS and hemolymph satisfies initial hormonal criteria.** *Peptides* 1994, **15**:591-602.

45. Anderson MG, Perkins GL, Chittick P, Shrigley RJ, Johnson WA: ***drifter*, a *Drosophila* POU-domain transcription factor, is required for correct differentiation and migration of tracheal cells and midline glia.** *Genes Dev* 1995, **9**:123-137.

46. Jin X, Shearman LP, Weaver DR, Zylka MJ, de Vries GJ, Reppert SM: **A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock.** *Cell* 1999, **96**:57-68.

47. Wharton JKA, Franks RG, Kasai Y, Crews ST: **Control of CNS midline transcription by asymmetric E-box elements: similarity to xenobiotic responsive regulation.** *Development* 1994, **120**:3563-3569.

48. • Sanchez-Soriano N, Russell S: **The *Drosophila* SOX-domain protein Dichaete is required for the development of the central nervous system midline.** *Development* 1998, **125**:3989-3996.
Provides genetic evidence that the *Drosophila* Dichaete SOX domain transcription factor interacts genetically with the Drifter POU domain protein to control midline glial expression. In combination with [47], these results suggest that *slit* midline glial transcription is regulated combinatorially by Sim, Tango, Dichaete, and Drifter.

49. • Briscoe J, Sussel L, Hartigan-O'Connor D, Jessell TM, Rubenstein JLR, Ericson J: **Homeobox gene *Nkx2.2* and specification of neuronal identity by graded Sonic hedgehog signalling.** *Nature* 1999, **398**:622-627.
A demonstration that vertebrate *Nkx2.2* regulates neuronal identity in the spinal cord and controls *Sim1* expression.

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

51. Xiao H, Hrdlicka LA, Nambu JR: **Alternate function of the *single-minded* and *rhomboid* genes in development of the *Drosophila* ventral neuroectoderm.** *Mech Dev* 1996, **58**:65-74.

52. Rowlands JC, Gustafsson J-A: **Aryl hydrocarbon receptor-mediated signal transduction.** *Crit Rev Toxicol* 1997, **27**:109-134.

53. Hahn ME, Poland A, Glover E, Stegement JJ: **Photoaffinity labeling of the Ah receptor: phylogenetic survey of diverse vertebrate and invertebrate species.** *Arch Biochem Biophys* 1994, **310**:218-228.

54. • Duncan DM, Burgess EA, Duncan I: **Control of distal antennal identity and tarsal development in *Drosophila* by *spineless-aristapedia*, a homolog of the mammalian dioxin receptor.** *Genes Dev* 1998, **12**:1290-1303.
Comprehensive molecular genetic analysis of the *Drosophila spineless* gene demonstrates that it is homologous to vertebrate *Ahr*. Although highly related by sequence identity, insect Ss is a developmental regulatory protein and vertebrate Ahr is a physiological regulator.

55. • Powell-Coffman JA, Bradfield CA, Wood WB: ***Caenorhabditis elegans* orthologs of the aryl hydrocarbon receptor and its heterodimerization partner the aryl hydrocarbon receptor nuclear translocator.** *Proc Natl Acad Sci USA* 1998, **95**:2844-2849.
The authors identify *Ahr* and *Arnt* homologs in *C. elegans*, indicating that bHLH–PAS proteins are highly conserved from worm to human. They further show that dimerization, interaction with Hsp90, and DNA binding properties of C. *elegans* AHR are similar to mammalian AHR. Studies in yeast suggest that a post-translational process is required for nuclear localization or transcriptional activation, but that activation does not involve dioxin binding.

56. • Emmons R, Duncan D, Estes P, Mosher J, Sonnenfeld, Ward M, Duncan I, Crews S: **The *Drosophila* Spineless-Aristapedia and Tango bHLH-PAS proteins interact to control antennal and tarsal development.** *Development* 1999,**126**:3937-3945.
Demonstrates that *Drosophila* Ss and Tango function as heterodimers *in vivo*. Provides evidence that Ss function is not controlled by binding of an exogenous ligand, but is controlled by cell-specific *ss* expression. This differs from vertebrate Ahr, the function funtion which is ligand-dependent.

57. Fernandez-Salguero P, Pineau T, Hilbert DM, McPhail T, Lee SST, Kimura S, Nebert DW, Rudikoff S, Ward JM, Gonzalez FJ: **Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor.** *Science* 1995, **268**:722-726.

58. Schmidt JV, Su GH-T, Reddy JK, Simon MC, Bradfield CA: **Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development.** *Proc Natl Acad Sci USA* 1996, **93**:6731-6736.