bHLH-PAS proteins: functional specification through modular domain architecture

AK Michael¹, CL Partch¹,²*

Abstract
Introduction
The basic helix-loop-helix PER-ARNT-SIM (bHLH-PAS) family of transcription factors respond to a wide range of external stimuli such as oxygen (hypoxia), harmful chemicals (xenobiotic), and light (circadian)⁷,⁸. This review will examine how the modularity of this protein family enables diversity in recognition of cellular signalling cues and subsequent regulation of transcriptional activity. Pairwise interactions between the DNA-binding bHLH and PAS domains establish the general architecture of the heterodimeric transcription factors, while C-terminal regulatory motifs control activity of the complexes. The function of bHLH-PAS proteins has been elaborated throughout evolution through domain shuffling among the regulatory motifs to confer new functionality. This review will also discuss how the activity of bHLH-PAS transcription factors is differentially attenuated in a tissue-specific manner by evolutionarily related proteins to regulate development and the cellular response to environmental stimulus.

Discussion
The authors have referenced some of their own studies in this review. The protocols of these studies have been approved by the relevant ethics committees related to the institution in which they were performed.

The PAS domain
The ability of the bHLH-PAS family to sense and respond to diverse cellular cues can be attributed in many ways to their PAS domain. The PAS fold is the distinguishing feature of bHLH-PAS proteins, showing the highest degree of evolutionary conservation, even between distantly related members of the family. The canonical PAS domain consists of a wide range of external stimuli such as oxygen (hypoxia), harmful chemicals (xenobiotic), and light (circadian)⁷,⁸. This review will examine how the modularity of this protein family enables diversity in recognition of cellular signalling cues and subsequent regulation of transcriptional activity. Pairwise interactions between the DNA-binding bHLH and PAS domains establish the general architecture of the heterodimeric transcription factors, while C-terminal regulatory motifs control activity of the complexes. The function of bHLH-PAS proteins has been elaborated throughout evolution through domain shuffling among the regulatory motifs to confer new functionality. This review will also discuss how the activity of bHLH-PAS transcription factors is differentially attenuated in a tissue-specific manner by evolutionarily related proteins to regulate development and the cellular response to environmental stimulus.

Licensee OA Publishing London 2013. Creative Commons Attribution License (CC-BY)

ing of PAS domains from HIF-2α and its obligate heterodimeric partner ARNT, has led to the discovery of selective, high affinity exogenous small molecules that bind within the PAS domains to regulate protein interaction, suggesting that they may have potential applications for therapeutic modulation of bHLH-PAS function. The ability to target specific bHLH-PAS transcription factors with PAS-binding small molecules and elicit tight control over gene expression pathways highlights the importance of gaining additional mechanistic information into this family of proteins.

PAS domains also mediate protein–protein interactions that occur through several distinct interfaces on the small, globular fold. PAS domains typically use their exposed β-sheet to dimerize with other PAS domains, but recently the opposing α-helical surface has been shown to mediate heterotypic interactions with other PAS domains and regulatory proteins. Proteins in the bHLH-PAS family predominantly use specific interactions between tandem PAS domains (denoted PAS-A and PAS-B) to form stable, heterodimeric transcription factor complexes. Mutation of PAS domains decreases complex formation to regulate transcriptional activity and can even alter selectivity of subunits for their heterodimeric partners. Therefore, PAS-mediated dimerization not only dictates selectivity for bHLH-PAS subunits for one another, but also facilitates DNA binding by forming the dimeric bHLH domain.

Modularity within the bHLH-PAS family
If structural conservation within the PAS domains allows heterodimerization within the bHLH-PAS family, the C-terminal regulatory domains contain transactivation (TAD) or repressor domains that control their activity. Although members of the family share a relatively high degree of sequence similarity within the bHLH and PAS domains, a sequence alignment clearly segregates genes together into clusters that regulate the same pathway (Figure 1). These findings highlight the importance of letting conservation guide identification of functional subclasses within bHLH-PAS proteins and provide insight into potential mechanisms of regulation. While the N-terminus of this protein family is structured by the presence of bHLH and PAS domains, the C-terminus tends to be intrinsically disordered with only short regions of predicted secondary structure, a common feature of transcription factors that serve as scaffolds for assembly of the eukaryotic transcriptional machinery. The C-terminal regulatory domains also show a higher degree of divergence among family members and confer functional constraints by interacting with pathway-specific transcriptional regulators and controlling stability of the complex. For example, the stability of HIF-1α is regulated by oxygen concentration through conserved prolines in the oxygen-dependent degradation domain and an asparagine in the TAD. Under normoxic conditions, proline hydroxylation association of HIF1-α with the von Hippel-Lindau E3 ubiquitin ligase, targets HIF-1α for proteosome-mediated degradation.

In mammals, two bHLH-PAS transcription factors within the family, ARNT and BMAL1 (Brain and Muscle ARNT-Like 1) act as general factors that heterodimerize with multiple partners to control all of bHLH-PAS signalling. The functional specificity of ARNT or BMAL1 is conferred by the dimerisation partner with which they interact, allowing for tissue- or pathway-specific regulation of target genes. The non-redundant roles of ARNT and BMAL1 proteins in mediating interactions with diverse partners suggest that complex mechanisms may exist for cross-talk between signalling pathways. The extent to which inter-pathway regulation plays a role in transcriptional regulation is still not well understood; however, there is some evidence that HIF-1α can compete with AhR for recruitment of ARNT to interfere with the dioxin signalling pathway. The overlap between

---

**Figure 1:** bHLH-PAS transcription factors are evolutionarily related and have homologous repressors with internal deletions. Many of the important bHLH-PAS signalling pathways are regulated by paralog repressors that share high sequence homology (line length within cladogram represents evolutionary distance) with the activator, but lack a domain that the positive element possesses. SIM2 represses the neurogenesis pathway and lacks a transactivation domain. IPAS1 attenuates the hypoxia response, but lacks a transactivation domain and cannot form a DNA-bound complex. AhRR cannot bind xenobiotics as AhR can (denoted with the chemical symbol) and therefore represses AhR:ARNT activity by competing for ARNT binding. Yellow dashed line indicates that ARNT1 acts as the obligate heterodimeric partner in these pathways.

---

Licensee OA Publishing London 2013. Creative Commons Attribution License (CC-BY)

pathways and potential crosstalk presents the intriguing potential for another level of transcriptional control to be driven by local protein concentration, stoichiometry and affinity of these complexes, and should be explored further.

**Evolutionarily related repressors of bHLH-PAS transcription factors**

When comparing bHLH-PAS proteins across signalling pathways, one interesting trend emerges: within each pathway, there is often a homologous repressor that shares modular domain architecture with an activator\(^{23-25}\). The repressors either lack domain(s) necessary for activation or possess an additional domain with repressive activity. Examples of how the bHLH-PAS protein family exploits gene duplication and domain shuffling to add additional layers of transcriptional regulation are discussed below.

The mammalian SIM1/2 proteins are homologs of the *Drosophila melanogaster* gene single-minded (sim) and regulate gene expression during development\(^{26}\). SIM1 and SIM2 share a high degree of sequence similarity (~90%) within the PAS domains; however, they have highly divergent C-termini\(^{27,28}\). Both proteins heterodimerize through similar PAS-mediated interactions with their cognate partner, ARNT; however, due to differences in their C-termini, the heterodimers have opposing activities on transcriptional regulation. The SIM1:ARNT complex activates target genes, while SIM2, with its distinct C-terminal sequence that contains two repressive domains, quenches the transactivation domain of ARNT to inhibit gene expression (Figure 2)\(^{29}\).

The hypoxia-inducible factor-1 (HIF-1) is an oxygen-labile protein that mediates adaptive responses to reduced oxygen availability by heterodimerizing with ARNT to bind DNA and activate genes involved in glycolysis and angiogenesis, promoting survival in low-oxygen conditions\(^{30,31}\). Beside HIF-1\(\alpha\), there are two additional members of the HIF bHLH-PAS superfamily: HIF-2\(\alpha\), also known as endothelial PAS domain protein 1 (EPAS1), and HIF-3\(\alpha\)\(^{32,33}\). A novel transcriptional repressor of this pathway has been identified, which is a splicing variant of the HIF-3\(\alpha\) locus, referred to as inhibitory PAS domain protein 1 (IPAS1). This alternative splicing event produces a truncated protein product that does not contain a transactivation domain. IPAS1 competes with ARNT to form a complex with HIF-1\(\alpha\), forming an abortive IPAS1:HIF-1\(\alpha\) complex that cannot bind DNA at hypoxia response elements. In this way, splicing variants can generate repressors to attenuate the cellular response to hypoxia\(^{34}\). In healthy tissue, the HIF1\(\alpha\)-ARNT complex plays a beneficial role in human physiology and development; however, this adaptation also increases oxygen supply to hypoxic microenvironments in tumours to permit proliferation and tumourigenesis\(^{35}\).

Understanding how the activity HIF1\(\alpha\)-ARNT complex is modulated endogenously by the repressor IPAS1 will further our understanding of the mechanisms in place that regulate the escape from oxygen sensitivity in cancer progression.

Exposure to harmful chemicals that humans encounter in the environment, such as dioxin, results in the expression of xenobiotic

---

**Figure 2**: Tissue-specific expression of bHLH-PAS transcriptional repressors provides cell-type specific responses to external stimuli and cell fate control. a) In almost all hypoxic mammalian tissues, HIF-1\(\alpha\) forms a heterodimer with ARNT to activate hypoxia response genes, including genes involved in neovascularisation. Selective upregulation of IPAS1 in the cornea inhibits the transcriptional activity of HIF-1\(\alpha\)-ARNT by binding to HIF-1\(\alpha\) and sequestering the protein. This tissue-specific regulation prevents detrimental neovascularisation within the cornea in mammals. b) During mouse embryonic development early muscle progenitor cells show high expression of SIM1, which can bind ARNT and activate genes. As development progresses, SIM1 expression is downregulated and SIM2 expression dominates to form an inactive transcriptional complex with ARNT to attenuate expression of SIM2 target genes.
response genes to break down these toxic compounds. This response is mediated in part at the transcriptional level by the AhR:ARNT complex. Dioxin and other halogenated aromatic hydrocarbons bind to the PAS-B domain of the transcription factor AhR, breaking up the interaction with AhR and the chaperone protein, Hsp90 that retains AhR in the cytoplasm. After this, AhR is free to enter the nucleus and form a heterodimer with ARNT to bind DNA, recruit transcriptional coactivators and activate xenobiotic response genes. The activity of AhR:ARNT is adjusted by the aryl hydrocarbon receptor repressor (AhRR) that binds to ARNT in a ligand-independent manner and sequesters the protein away from AhR. AhRR shares a high degree of similarity with AhR (PAS-A, 59.7%; C-terminus, 19.7%) and uses this molecular mimicry to form an inactive transcriptional complex that downregulates the xenobiotic pathway. While studies suggest that AhRR may act as a tumour suppressor, the functional significance of this negative regulation remains to be fully elucidated. However, it has been postulated that AhRR is up-regulated in delicate tissues such as the germline, to protect against toxic metabolites generated by the breakdown of these polycyclic aromatic hydrocarbons.

**Tissue-specific regulation of gene expression**

Each cell type demands the ability to differentially regulate gene expression and cellular responsiveness to external cues, suggesting that mechanisms exist to modulate the output of environmentally sensitive bHLH-PAS signalling pathways in a tissue-specific manner. This is accomplished at the molecular level by distinct expression patterns of transcriptional activators and repressors, which allows for fine-tuning of the cellular response to global inputs, such as low oxygen concentration or xenobiotic exposure, depending upon the identity and needs of the cell.

One of the most well understood examples of tissue-specific regulation of bHLH-PAS signalling pathways occurs in the cornea. During sleep, closure of the eyelid creates a hypoxic environment within cells of the cornea. Despite continual exposure to low oxygen levels that would normally stimulate HIF-1α-dependent gene expression to increase neovascularisation, the cornea remains completely avascular, which is necessary for vision. Selective down-regulation of the hypoxia pathway in the cornea can be attributed to the unique expression pattern of the HIF-1α-related repressor, IPAS1. Exposure of the corneal epithelium to hypoxia results in a rapid HIF-dependent induction of IPAS1 mRNA and repression of HIF-1α:ARNT activity by sequestering HIF-1α from ARNT in an abortive HIF-1α:IPAS1 complex incapable of binding DNA. Repression by IPAS1 in hypoxic conditions specifically in the cornea is believed to play an important role in preventing neovascularisation in the cornea and blindness.

In addition to their participation in pathways that respond to environmental stimuli, many bHLH-PAS proteins also play a direct role in development and cell fate determination. The *Drosophila* bHLH-PAS protein Tango, the ortholog of the mammalian ARNT protein, regulates specific stages of development. Like ARNT, Tango is broadly expressed throughout tissues and forms a heterodimer with SIM to control development of the central nervous system midline cells. However, Tango also forms a heterodimer with its bHLH-PAS partner Trachealess to control proper development and tubulogenesis within tracheal cells and the salivary duct. Because expression of sim and Trachealess is restricted to the cell lineages that they control, they possess the ability to regulate specific developmental programmes through selective expression of their specific target genes upon binding to their shared partner, Tango. In mammals, SIM proteins have similar functions that control embryonic development and cell differentiation. In situ hybridisation studies in mouse embryos show largely non-overlapping expression patterns of mSIM1 and mSIM2, with mSIM1 being expressed in early limb muscle precursor cells and selective SIM2 expression in cells after migration (Figure 2b).

While many studies have quantified mRNA expression levels of bHLH-PAS genes across different tissues, the mechanisms controlling these differential expression patterns remain to be elucidated and deserve further study. Advent of genome-wide techniques will allow for identification of transcription factor binding sites and chromatin modifications that dictate the epigenetic state of the cell. These studies will illustrate the roles that bHLH-PAS proteins play in controlling cell- and tissue-specific gene expression and how these pathways respond to environmental insults.

**Conclusion**

Nature has tinkered with the bHLH-PAS family, using the PAS domain module as a building block to establish the foundation of a specific transcription factor architecture that is accessorised with regulatory motifs that modulate activity. bHLH-PAS transcription factors exemplify the concept of functional expansion through modular domain shuffling in molecular evolution, using the well-conserved PAS fold to bring together stimulus-responsive proteins with their partners, while conserved domains and motifs outside of the PAS fold have been swapped to activate or repress the pathway. Over 29,000 PAS domain-containing proteins have been identified from archaeabacteria to humans (SMART database), most of which likely sense...
environmental cues to act within transcripcional networks. We are just beginning to understand the intricacies of these proteins and the pathways they act within. Additional studies need to be done to further characterize these proteins using sequence conservation paired with bioinformatic analyses of predicted regulatory motifs that will help inform biological function of paralogs and other related genes. Moreover, the capability to determine protein expression, DNA binding and transcriptional regulation genome-wide by these proteins in response to external stimuli will help us better understand tissue-specific regulation of bHLH-PAS proteins and their biological role in keeping mammals healthy and in tune with their environment.

Abbreviations list

AhR, aryl hydrocarbon receptor; AhRR, aryl hydrocarbon receptor repressor; ARNT, aryl hydrocarbon receptor nuclear translocator; BMAL1, Brain and Muscle ARNT-Like 1; EPAS1, endothelial PAS domain protein 1; HIF-1, hypoxia-inducible factor-1; IPAS1, inhibitory PAS domain protein 1; ODD, oxygen-dependent degradation domain; TAD, transactivation domain.

References