The aryl hydrocarbon receptor as a model PAS sensor

Emmanuel Vazquez-Rivera\textsuperscript{a, b}, Brenda Rojas\textsuperscript{a}, Jessica C. Parrott\textsuperscript{a}, Anna L. Shen\textsuperscript{a, b}, Yongna Xing\textsuperscript{a, b}, Patrick R. Carney\textsuperscript{a, b}, Christopher A. Bradfield\textsuperscript{a, b, *}

\textsuperscript{a} Molecular and Environmental Toxicology Center, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, United States
\textsuperscript{b} McArdle Laboratory for Cancer Research, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706, United States

\section{A R T I C L E  I N F O}

Handling Editor: Dr. Aristides M. Tsatsakis

\textbf{Keywords:}

Aryl hydrocarbon receptor
PAS proteins
Environmental sensors

\section{ABSTRACT}

Proteins containing PER-ARNT-SIM (PAS) domains are commonly associated with environmental adaptation in a variety of organisms. The PAS domain is found in proteins throughout Archaea, Bacteria, and Eukarya and often binds small-molecules, supports protein-protein interactions, and transduces input signals to mediate an adaptive physiological response. Signaling events mediated by PAS sensors can occur through induced phosphorelays or genomic events that are often dependent upon PAS domain interactions. In this perspective, we briefly discuss the diversity of PAS domain containing proteins, with particular emphasis on the prototype member, the aryl hydrocarbon receptor (AHR). This ligand-activated transcription factor acts as a sensor of the chemical environment in humans and many chordates. We conclude with the idea that since mammalian PAS proteins often act through PAS-PAS dimers, undocumented interactions of this type may link biological processes that we currently think of as independent. To support this idea, we present a framework to guide future experiments aimed at fully elucidating the spectrum of PAS-PAS interactions with an eye towards understanding how they might influence environmental sensing in human and wildlife populations.

\section{1. Introduction}

The Ah receptor (AHR) is a prototype PER-ARNT-SIM (PAS) domain containing protein that is well known for its role in the adaptive metabolism and physiological consequences of a variety of structurally related trace extended aromatic compounds (TEACOPS) and halogenated-dibenzo-p-dioxins (“dioxins”). In this perspective, we provide an overview of AHR signal transduction, presenting this protein as a prototype small molecule sensor with roles in both environmental adaptation and normal physiology. We begin with a brief discussion of the larger family of PAS domain proteins to emphasize the concept that these proteins display similar functions in a variety of microorganisms, animals, and plants. We then move to a description of AHR signaling as a representation of the idea that PAS domains display multiple characteristics; as a site for homotypic dimerization between two PAS proteins, as a site for heterotypic interactions with distinct output proteins and cellular chaperones, and as a site for small molecule binding. We then provide evidence for the idea that the AHR senses both environmental and endogenous signals and then close this perspective with the possibility that additional PAS dimers and higher order PAS-PAS interactions are still undiscovered and that these interactions may explain the pleiotropy of many environmental sensing pathways.

\section{2. The PAS domain}

\subsection{2.1. PAS diversity}

Organisms require mechanisms to adapt to environmental change. The PAS, steroid/nuclear receptor, Laci, Gal, and Mark protein families are examples of ligand-responsive sensors that play essential roles in metabolic adaptation [1–7]. In these systems, sensor proteins often recognize an environmental or metabolic stimulus and initiate signal transduction events to induce physiological change, often through alterations in gene expression. In many organisms, this occurs through an environmental signal that induces a conformational change and/or post-translational modification of the sensor protein. In turn, this activated sensor influences gene expression through a variety of mechanisms, including increased concentration of the sensor protein at genomic regulatory sites or through phosphorelays that influence the levels of downstream transcription factors at genomic elements. Through such pathways, sensing systems can orchestrate the expression of multiple genomic targets involved in environmental adaptation.

\footnotesize{\textsuperscript{*} Corresponding author at: McArdle Laboratory for Cancer Research, 1400 University Ave., Madison, WI 53706, United States.
E-mail address: bradfield@oncology.wisc.edu (C.A. Bradfield).}

https://doi.org/10.1016/j.toxrep.2021.11.017
Received 3 June 2021; Received in revised form 19 November 2021; Accepted 24 November 2021
Available online 26 November 2021
2214-7500/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
The PAS domain was first defined based upon sequence alignments of the fruit fly Per and Sim, as well as the human ARNT gene products, with its name arising from the acronym of the first letter of each founding member (PER-ARNT-SIM) [8–10]. This initial definition of the PAS domain encompassed approximately 250–300 amino acids and harbored two degenerate internal repeats, known as PAS-A and PAS-B (Fig. 1). In more recent publications, these internal PAS-A, PAS-B repeats of approximately one hundred amino acids in length each are now also commonly referred to as “PAS domains.” By either definition, the PAS domain is found in Archaea, Bacteria and Eukarya, commonly in tandem repeats, and often enclosed within a protein that plays a role in the adaptive responses to environmental change [5]. In plants, the PAS domain is often referred to as “Light, Oxygen, Voltage” (LOV) domain. In part, this LOV nomenclature has arisen from the observation that in some plants, the PAS/LOV domain has been shown to bind flavin and act as a photosensor to mediate blue light-induced phototropism [5,11].

2.2. PAS domain functions

The PAS domain is encoded in sensor proteins found in all kingdoms of life. This domain functions through multiple mechanisms, as a site for small molecule binding and as a site for homotypic (i.e., PAS-PAS) or heterotypic (PAS-other) protein-protein interactions. One highly conserved role of a PAS domain is as a site for small molecule binding and sensory function (Fig. 2) [5,12]. In microbial systems, the PAS domain is found in sensor proteins such as the photoactive yellow protein (PYP) and FixL. [13–16]. In the PYP protein, the PAS domain covalently binds 4-hydroxycinnamic acid as a chromophore, while in FixL the PAS domain holds heme as part of an oxygen-binding site. In these two examples, the PAS domain directly functions as a sensor for either light or oxygen to regulate phototropism or nitrogen fixation, respectively. Importantly, modelling of the PAS-B domain of the ARNT protein suggests this domain binds its prototype ligand, 2,3,7,8-tetrachlorodibenzo-p-dioxin within this same PAS repeat fold (Fig. 2) [17–19].

In addition to binding chromophores, PAS domains can also serve as a protein-protein interaction surface. One example of a homotypic interaction (where two PAS domains interact) is the PAS transcription factor Aureochrome 1a from Phaeodactylum tricornatum (PtAu1a), where the PAS domains of two PtAu1a proteins homodimerize in the presence of its substrate PAHs, was commonly referred to as “PAS domains” [38,39]. Reports of inducibility were followed by the proposal that a collection of cytochromes-P450 dependent monoxygenases (P450 s) named the Aryl hydrocarbon receptor (AhR), also participates in this light sensing/circadian adaptation pathway [23]. In the dark VVD exists as a homodimer with each monomer binding a flavin. When VVD absorbs light, a conformational change exposes the PAS domain of VVD at the same time as the PAS domain of WC-1 is opened up [24]. This transient unpairing of dimers allows a new dimerization in which VVD replaces WC-2 as a partner of WC-1, creating a transcriptionally inert complex [25]. Through this feedback mechanism the amount of active WCC and VVD modulates the light response allowing the organism to respond to incremental levels of light rather than only light versus dark [26].

Interactions of PAS with different domains (i.e., PAS-nonPAS) also serve to influence signal transduction in response to environmental stimuli. Examples of such heterotypic interactions include; 1) The interaction of the PASB domain of the mammalian NCOA1 transcriptional coactivator with the LXXLL motif within the STAT6 transcription factor and 2) The interaction of the PASB domain of the mammalian CLOCK protein with its CRY regulator [27,28]. There are several important examples of heterotypic interactions of PAS domains occur intramolecularly in sensor biology. One example of such an interaction is the PAS domain of the light activated transcription factor El1222 and its covalently linked DNA binding motif [29].

3. The ARYL hydrocarbon receptor as a model mammalian PAS sensor

3.1. History

The adaptive metabolism of polycyclic aromatic hydrocarbons (PAHs) has long been an area of scientific investigation due to the carcinogenicity of many congeners and their widespread dispersion in the environment [30]. Sources of PAHs from human activity include combustion of wood and fossil fuels for energy and heat, tobacco smoking, use of coal tar sealants and asphalt, coal liquefying plants, coke and aluminum production, barbecuing, and smoking or charring of food over fire. Natural emissions of PAHs are also significant and include, wildfires, petroleum seepage, coal deposits, and volcanic activities [31,32]. Similarly, structurally related halogenated-dioxins, -dibenzofurans and biphenyls are also environmental contaminants that commonly arise from human activity and sometimes natural sources [33,34]. Thus, organisms have been in contact with environmentally ubiquitous TEA-COPs for millions of years, with the use of fire and human industrial activity only serving to increase exposure of certain populations in recent millennia.

Given their widespread occurrence, toxicity and carcinogenicity, PAHs like benzo[a]pyrene (BAP) served as early prototypes in studies of xenobiotic metabolism, bioactivation and detoxification. This research revealed that a collection of cytochromes-P450 dependent monoxygenases (P450 s) named “aryl hydrocarbon hydroxylase” (AHH), had a major influence on the biological half-lives of PAHs in many species [35,36]. Approximately fifty years ago, it was observed that this PAH metabolic system, was inducible by its substrates [37]. Due to parallels with the LacI system, this “appearance” of an AHH enzymatic activity in the presence of its substrate PAHs, was commonly referred to as “induction” [38,39]. Reports of inducibility were followed by the proposition that a “binding species” or a “receptor” for PAHs mediates the upregulation of AHH activity and led to the introduction of the idea that this system may have arisen as a protection against the toxicity of PAHs,
as well as structurally related phytoalexins, and TEACOPS in the environment \[33, 40–43\].

The study of a genetic polymorphism in mice that influenced a strain’s inducibility upon response to PAHs led to the identification of the \(\text{Ah}^b\) locus and the designation that strains were either “responsive” (prototype strain C57BL/6 or B6) or “nonresponsive” (prototype strain DBA/2 or D2) \[44, 45\]. Genetic studies identified a single autosomal locus as primarily responsible for this differential responsiveness, giving rise to the nomenclature: \(\text{Ah}^b\) \((b \text{ from B6})\) to define the responsive allele and \(\text{Ah}^d\) \((d \text{ from D2})\) to define the nonresponsive allele. More recent genomic nomenclature alters this locus designation to \(\text{Ahr}\) instead of \(\text{Ah}\) and it is this terminology we will use through the remainder of this review (i.e., \(\text{Ahr}^b\) vs \(\text{Ah}^b\)).

Our understanding of metabolic adaptation to PAHs was aided by the development of radioligands of the AHR \[42, 46\]. These reagents, derived from the radiolabelling of PAHs or the more potent dioxin structure, led to the discovery of a binding protein or “receptor” within target cells and tissues \[47\]. These studies also revealed that “nonresponsive” \(\text{Ahr}^d\) mice are more appropriately described as “hyporesponsive” as they mount an inductive response, although much less robustly, than the \(\text{Ahr}^b\) strains. Biochemical characterization of this binding species revealed that it was a soluble protein associated with chaperones such as the 90 kDa heat shock protein (HSP90), with later studies adding the smaller co-chaperones such as ARA9 (aka XAP2 or AIP), p23, and possibly ARA3 (aka NS1BP) to this complex (Figs. 3 and 4, see below) \[48–55\].

Radioligands also led to the development of competitive binding assays that allowed structure-activity studies to explain agonist potency. These studies provided evidence that dioxin ligands that bind with higher affinity to the receptor site are more potent inducers of the target P450s \[56\]. The subsequent observation that this binding affinity segregated with the \(\text{Ahr}^d\) and \(\text{Ahr}^b\) polymorphism in mice provided the final formal evidence for the existence of an AHR \[57\].

The cloning of the P450 encoding genes that comprised AHH activity, i.e., \(\text{CYP1A1, CYP1A2, and possibly CYP1B1}\), as well as the identification of the genomic cis-elements that controlled their expression in response to PAH exposure, led to the discovery of the regulatory enhancers linked to the induction phenomenon \[36, 58–60\]. Specifically, early experiments demonstrated that the consensus sequence, GGCGTGC (sometimes also defined as TNGCGTG), was bound by a heterodimeric complex and controls the upregulation genes like \(\text{CYP1A1}\) in response to PAH and dioxin ligands \[59, 61–63\]. This genomic enhancer element goes by many names: xenobiotic response element (XRE, which we will use here), dioxin response element (DRE), and Ah response element (AHRE).

### 3.2. Discovery of mammalian PAS proteins

A watershed moment in our understanding of adaptive metabolism, and PAS proteins writ large, arose from a somatic-cell genetics approach aimed at identification of gene products essential for the ligand-induced, AHR-dependent, induction of \(\text{CYP1A1}\) in cell culture. This work resulted...
in the molecular cloning of a molecule called ARNT ("aryl hydrocarbon receptor nuclear targeter") that was essential for the AHR to gain a high affinity for the nuclear compartment and chromatin upon ligand binding [8]. Sequence analysis of ARNT provided two important observations. First, as noted above, the ARNT protein displays homology with the Sim and Per gene products of the fruit fly (D. melanogaster) allowing the initial definition of the PAS homology domains (Fig. 1) [64,65]. Second, homology searches revealed ARNT (as well as SIM, but not PER) harbor a bHLH domain immediately N-terminal to PAS. Such bHLH domains had been previously observed in partners of dimeric transcription factors such as MyoD and Myc [66]. This concept led to the proof that ARNT was a part of a protein complex with that directly bound enhancer elements within the genome [62,67,68].

The use of the photoaffinity ligand, 2-azido-3-iodo-7,8-dibromodibenzo-p-dioxin allowed the purification of the AHR from B6 cytosol, as well as its primary amino acid sequence and antibodies. In turn, this led to the ultimate cloning of the corresponding cDNA and structural gene [46,67,69-72]. Of central importance was that the cDNA encoded the second mammalian member of the PAS family and harbored an adjacent bHLH domain (Fig. 4). Taken in sum, these observations suggested a dimeric partnership between the AHR and ARNT lead to the signal transduction model described in Fig. 3 and discussed more below. If we add the ARNT as an additional founding member in chordates, then of the original four PAS proteins described, all harbor two PAS repeats (denoted PAS-A and PAS-B), three out of four (SIM, ARNT and AHR) harbor a bHLH domain immediately N-terminal to the PAS repeats, yet, the C-terminal halves display much lower sequence homology (Fig. 4) [65,73].

4. Molecular biology of the aryl hydrocarbon receptor

4.1. Protein-protein interactions

The observation that two bHLH-PAS proteins, the AHR and ARNT, are essential for CYP1A1 induction led to the demonstration that signaling is dependent upon an AHR-ARNT heterodimer with specificity for the XRE sequence [62,68,74]. With this concept in hand, functional domain maps of the AHR and ARNT were rapidly developed (Fig. 4) [74-80]. Based upon the recognition that bHLH-domains commonly act in dimeric pairs of transcription factors to position basic-alpha helices within the major groove of enhancer DNA, it was quickly demonstrated that the basic region helix found in the bHLH of both the AHR and ARNT would be required for binding to the XRE [67,74,77,80-82] (Figs. 3 and 4).

The PAS-A domain in both the AHR and ARNT, was found to play an essential role in signaling. Molecular and crystallographic studies support the idea that cooperation between the HLH and PAS-A domains is the primary support for dimerization between these two proteins. In turn, these interactions position the basic alpha helix within the major groove of DNA for sequence specific contacts at XREs. Thus, a clear role for this domain is as a PAS-PAS dimerization surface, contributing to AHR-ARNT interaction and formation of a competent transcriptional complex [74,77,83-85].

While some data exists to support the idea that the PAS-B domain is important in AHR-ARNT dimerization, a greater body of evidence documents its role in heterotypic interactions by the AHR. In support of a role for such protein-protein interactions is the early observation that this region of the AHR represses AHR-activity [74,76,97], and the parallel mapping studies demonstrating Hsp90 binds to this same repressing region of PAS-B [77,79,98]. A common interpretation of this data is that a chaperone complex (possibly a dimer of Hsp90, and the cochaperones ARA9, p23 and perhaps ARA3) is holding the PAS-B domain in a conformation that can accept ligand and prevent dimerization with ARNT or inappropriate contacts with other proteins in the cytosol that might lead to its aggregation and inactivation [51,53,89-91]. Once ligand binds, conformational changes are induced which weakens or reorganizes the chaperone complex, simultaneously revealing, nuclear localization and possibly ARNT-dimerization motifs within the AHR.

4.2. Ligand binding

Despite extensive information available regarding the pharmacology of AHR ligands, there exists a gap in our understanding of the structure of the AHR ligand-binding domain (LBD) due to the lack of crystal structure for its PAS-B. The LBD of the AHR was initially mapped by covalently binding an [[125I]-labeled photoaffinity ligand followed by CNBr cleavage and micro-sequencing which revealed the labeled peptide fragment was coincident with PAS-B [67]. This overlap of the LBD with the PAS B domain was supported by a number of subsequent molecular studies [74,77,92]. More recent mutagenesis, and homology model-derived structures, also indicate PAS-B as being important for ligand binding and provide a preliminary view of this PAS-B-ligand-bound structure and provide evidence that a V375A polymorphism within this region explains much of the variability in ligand responsiveness observed across the mouse Ahr* and Ahr* mouse strains (Fig. 2) [17-19,93-96].

While the N-terminal halves of the AHR and ARNT appear to play central roles in chaperone interactions, dimerization, DNA binding and ligand induced transformation, the C-terminal halves of these factors appear to harbor regions that influence expression of genomic targets once the complex is bound to chromatin [77,78,86,88,97]. In this regard, multiple subdomains in the C-terminus of the AHR are often reported, with the idea that each confers weak transactivation potency alone, but act synergistically. One possible consequence of this multiplicity is that it may enable the AHR to interact with a variety of transcription factors and activate transcription from a variety of promoters [61,86].

4.3. Variation and polymorphism

Molecular examination of the AHR open reading frame from the human, mouse and rat AHR explains the receptor molecular weight variation observed both within and across species. In large part, this size difference, as documented by western blot analysis and photoaffinity labelling of receptors, is due to altered termination codon usage in the final exon (exon 11) of the structural gene [43,71]. While still a debated concept, the marked difference in molecular weight of the AHR observed across species was initially thought of as evidence that AHR structure may be more evolutionarily responsive to the environmental niche of a given species, as compared to genes encoding receptors for endogenously generated hormones [43,98]. While such a model is difficult to prove, it is an intriguing proposition, with early evolutionary analyses providing some support and arguments against this idea [40,99].

The observation that Ahr polymorphisms in rodent models (e.g., the V375A polymorphism in mice, described above, or the splice junction polymorphism observed in rats, described in detail elsewhere [100]) can lead to alterations in ligand response, spurred considerable interest in whether polymorphisms leading to hyper- or hypomorphic Ahr alleles might be common in human populations [101,102]. While investigations of SNPs (single nucleotide polymorphisms) can harbor significant biases reflecting the geography or population of focus, interrogation of the public database, dbSNP [103], indicates that common SNPs (defined here as greater than 1% in a population) appear in humans at frequencies similar to loci encoding most other nuclear receptors [101].

Of the recorded SNPs resulting in nonsynonymous (missense) alterations within the Ahr gene, only a few, e.g., P517S, R554 K and V570I, are commonly reported in various human populations, while non- synonymous and intronic SNPs are much more common [104-106]. In one study of a Japanese population, a number of novel rare genetic variants were detected (K17 T, K401R, N487D, and I514T) in addition to the more common R554 K allele [107]. Although this study did not
include a functional analysis, it provides one example of the diversity of rare, potentially important, missense variants in unique populations. In a second study, haplotypes of the three more common SNPs noted above were examined for their functional impact on AHR signaling in vitro [104]. Interestingly, the initial studies indicated that haplotypes corresponding to I570 and K554 yield a hypomorphic receptor, perhaps due to an attenuated transactivation domain or decreased stability in vivo. In contrast to the missense SNPs, known synonymous SNPs and SNPs within introns and other noncoding regions, number over a hundred in human populations. While less obvious, nonsynonymous and intronic SNPs have the potential to influence AHR signaling through impacts on splicing, mRNA stability or even codon usage. While the importance of intronic SNPs leading to hypomorphic Ahr alleles in the rat is well documented, a parallel in humans has not yet been reported [100]. A summary of functionally important polymorphisms in the AHR open reading frames across species is presented in Fig. 5.

5. Aryl hydrocarbon receptor pathways

5.1. The AHR adaptive pathway

It is now possible to describe the classical pathway for the adaptive metabolism of PAHs in humans and most mammals (Fig. 3). Upon entry of PAHs into cells, these ligands bind to the AHR in the cytosol inducing a conformational change in the AHR that loosens associations with its chaperone complex and exposes a nuclear localization sequence (NLS). The ligand-activated receptor complex then translocates into the nucleus, where the AHR sheds or rearranges its cellular chaperones and binds to the constitutively nuclear ARNT proteins through PASA, PASB and HLH domains. The dimerization of the AHR-ARNT heterodimer positions the basic alpha helices to recognize the XREs within the major groove of DNA and this is associated with the recruitment of coactivators and chromatin rearrangements, leading to the transcriptional activation of target genes, such as CYP1A1, CYP1A2, and CYP1B1.

The classic adaptive pathway depicted in Fig. 3 is dependent upon a variety of homotypic and heterotypic protein interactions (e.g., PAS-PAS and PAS-chaperone, respectively). Yet, a considerable body of evidence suggests that the role of protein interactions in AHR biology is more complex, and outputs extend beyond genes encoding xenobiotic metabolizing enzymes as the mechanism depicts in Fig. 3 [108]. While we commonly associate the adaptive pathway with the upregulation of target genes through the dimeric AHR-ARNT binding to XREs, a large body of evidence indicates that the AHR is also involved in the up- and down-regulation of a variety of additional genomic targets, most of which are not typically associated with xenobiotic metabolism [109–111]. Importantly, the mechanisms underlying these ‘nonclassical’ signaling pathways appears to occur by mechanisms that are distinct from that depicted in Fig. 3. While not entirely elucidated, such mechanisms commonly employ heterotypic protein interactions that link the AHR to other cellular signaling molecules, some examples include the estrogen receptor, EZF and RelA [112–115].

The adaptive pathway described above, appears to be suppressed through a number of mechanisms [116,117]. While there are numerous reported mechanisms, the one most relevant to this perspective on PAH proteins is mediated by the “AHR-Repressor” or AHRR [118]. This protein, which has been shown to dimerize with ARNT, is a close structural homolog of the AHR but is missing a region corresponding to part of its PAS-B repeat (Fig. 5). One idea is that the AHRR competes for dimerization with ARNT and thereby competes with AHR-ARNT for XRE occupancy [118]. Although not completely understood, the AHRR appears to provide feedback inhibition through three potential mechanisms: competition for ARNT, competition for the XRE, and transcriptional repression of target genes upon XRE binding through direct repression on their promoters [116–119].

5.2. The AHR cognate pathway

While the AHR does influence metabolic adaptation in response to environmental chemicals, this may not be the sole evolutionary driver for its conservation in animal species. In support of an additional “cognate pathway,” observations from animal models null for AHR expression describe a variety phenotypes [120–125]. While early reports of AHR-null phenotypes were subtle and often differed across laboratories and species, a limited list includes alterations in peripheral lymphocyte populations, alterations in gastrointestinal immunity, a patent ducus venosus (DV), reduced litter sizes, and age-dependent cardiac hypertrophy [124,126–129]. Given that the immunological aspects of AHR physiology have garnered considerable recent attention, the reader is referred to some excellent reviews for more insight [130,131].

A variety of studies in recombinant mouse models support the idea that the cognate signaling pathways of the AHR have similarities to the adaptive pathway described above. A few experimental observations in support of this idea include the observation that mutant alleles of two proteins important in the adaptive pathway, the ARNT and ARA9, are also required for normal DV closure [132,133]. Moreover, knock-in mutations at the Ahr locus that ablate the XRE binding by the AHR-ARNT dimer disrupt many aspects of the cognate pathway such as DV closure and lymphocyte development [134,135]. Importantly, the target genes of the cognate pathway are still unclear and may or may not be distinct form the adaptive pathways. In this regard, conditional null alleles indicate that AHR signaling to regulate DV closure or barrier immunity are occurring in cellular compartments not traditionally associated with the adaptive response [131,136,137].

6. Ligands

6.1. Endogenous and cognate ligands of the AHR

While initial attention to the AHR arose from an interest in understanding the response to xenobiotics, there has long been a search for endogenous or cognate ligands. Evidence that AHR-null animal models display such a wide variety of phenotypes gives import to such efforts but does not rule out the possibility that the AHR acts in a ligand-independent manner [128]. Along this line of thought, it is possible that the binding of xenobiotic ligands is an independent role for the AHR, selected for independently throughout evolution. For this and future discussion of this topic, we propose two criteria to define a “cognate ligand.” The first is that the ligand can be found in animal tissues naturally and in an evolutionarily consistent manner. The second is that the ligand must be shown to be functionally linked with the
receptor for some essential biological process. To date, we have many interesting ligands that appear to meet the first criteria. Where we have less data is in the proof of a link to a biological function. While, to date, we have no physiologically proven cognate ligand, a variety have been reported and summarized in a number of excellent reviews on the topic [127,131,138-141].

6.2. Proligands

In accordance with numerous structure-activity studies performed over the last decades, an early heuristic to define AHR ligands is that they must fit a planar binding pocket with dimensions of a flat hydrophobic rectangle with dimensions of approximately 14 Å × 12 Å × 5 Å [142-147]. Recent structural predictions based on homology models provide additional insights, although to be confident, we must ultimately await structural elucidation of this binding through approaches such as nuclear magnetic resonance or crystallography [17,19,146,149].

The repeated observation that the AHR displays binding affinity for larger planar aromatic structures lead some to speculate that smaller one- or two-ring aromatic compounds are probably not ligands, even though they might lead to receptor activation in vivo [127]. An important early example of this phenomenon was the natural product indole-3-carbinol (I3C), a plant auxin derived from tryptophan in edible Brassica family plants [150]. Early on, this compound was highly touted as a dietary anticarcinogen and determined to be a powerful dietary activator of AHR-regulated xenobiotic metabolizing enzymes such as the CYP1 monoxygenases. Indole-3-carbinol is now emerging as a paradigm of the prol gland concept in AHR biology, which postis that many small aromatic compounds are metabolically, chemically, or spontaneously converted into extended aromatic structures by natural processes [127,151,152]. In the case of I3C, the conversion occurs in the low pH environment of the stomach, where the acid-catalyzed condensation of I3C generates the potent AHR agonist, indololo[3,2,b]carbazole (IC2) [151,153].

Perhaps more important than the identity of I3C as a prol gland is the idea that this process may be a common route of ligand production in normal physiology. In this regard, tryptophan (TRP) is metabolized to the alpha-keto acid indole-3-pyruvic acid (I3P) by at least two enzymes (D-amino acid oxidase and aspartate aminotransferase), and I3P spontaneously condenses into TEACOPS with high AHR binding affinity in vivo [154-156]. The aromatic amino acid TRP is a particularly interesting source of endogenous ligands. This substrate can be metabolically and chemically converted into a variety of polyaromatic structures. For example, the TRP condensation product, 6-formyindolo[3,2,b]carbazole (FICZ), is often considered a top candidate as a cognate ligand of the AHR. This compound is generated by UV irradiation of TRP in the skin and displays an AHR binding affinity in the 10⁻¹¹ molar range (Kₐ) [152,157]. Other TRP metabolites that show potential as physiological endogenous ligands include the TEACOP products produced spontaneously from the TRP metabolite kynurenine (Kyn) [158,159]. These KYN derived TEACOPs have been shown to display potent agonist activity and may be responsible for aspects of reported immunological activity of Kyn on T-cells [160,161]. Interestingly, other TRP metabolites and condensation products are also candidates as ligands; these include simple indoles, the indigoids and cinnabarinic acid [140,162,163].

6.3. Evidence the AHR requires endogenous activation

Multiple experimental lines provide evidence that the cognate pathway requires endogenous ligand activation. First, more than one lab has reported physiological conditions where XRE-driven reporters are activated under conditions consistent with the presence of an endogenous ligand [111,164]. Second, mouse models hypomorphic for AHR expression can be rescued with pre/perinatal exposure to a remarkably potent agonists such as 2,3,7,8-tetrachloridibenzo-p-dioxin or SU5416 (i.e., DV closure) [132,165]. Third, overexpression of the CYP1A1 gene in the mouse leads to phenotypes similar to AHR-null animals. This last observation being consistent with a model where adaptive metabolic clearance of an endogenous ligand is required for normal barrier immunity in the gut [128].

7. The PAS protein family in mammals

Following the cloning and characterization of ARNT and the AHR from mice, rats and humans, a number of additional important PAS proteins were identified in mammals [73,166]. Among these are the HIF-alphas which dimerize with ARNT (also known as HIF-beta) to regulate the hypoxia response and the CLOCK, ARNTL and PER proteins which play central roles in the maintenance of circadian rhythms [73,167-171]. In each of these cases, like the AHR-ARNT, a dimeric pairing of two distinct PAS proteins drives a central transcriptional response through a cognate enhancer element. Also, like the AHR-ARNT system, each of these systems regulates environmental adaptation and each is also important for essential physiological developmental processes. Importantly, the CLOCK and HIF stories are recorded in detail as the result of the recent Nobel Prizes associated with their discovery [170,172].

In early attempts at discovery of novel mammalian PAS factors, a variety of nomenclature schemes have been used. For example, in our laboratory, when each newly discovered factor was identified through searches of expressed sequence tags, the unique clones were designated Member of PAS-1, 2, 3... (MOP1 etc.) [173]. In parallel, the another laboratory employed a similar strategy and their nomenclature denoted the site of expression and number of predicted (e.g., NPAS1, for neuronal PAS 1 etc.) [174]. Additionally, novel bHLH-PAS members have been named based on similarity to existing genes or predicted function [118,175-179]. A recent survey of the human genome reveals this family is quite large, with 33 PAS-domain encoding genes found by homology search or functional cloning (Fig. 6) [73]. Twenty-two of these proteins are documented or presumed to play roles in transcriptional regulation. Nineteen of these twenty-two transcriptional regulators harbor a (bHLH) domain immediately N-terminal to their PAS-A domain. Of the remaining eleven PAS proteins, eight are potassium channels, two are phosphodiesterases, and one is a serine-threonine kinase [73].

Based upon our knowledge of AHR-ARNT mediated signal transduction, sequence alignment homology, as well current understanding of the biological function of each PAS family member, our laboratory classifies mammalian PAS members as either alpha-class, beta-class, gamma-class, delta class, or kappa-class (Fig. 6). The alpha-class includes the eleven PAS proteins that commonly display restricted cell type expression patterns, that are most commonly associated with the sensing environmental stimuli or cellular state, and that transduce those signals through an induced pairing with one of four promiscuous beta-class PAS protein partners [168,180-182]. We propose Beta-class to denote PAS proteins that are more widely expressed and often act as essential partners for a broad spectrum of family partners from the alpha-class. In this proposed PAS protein nomenclature, we add the designation of the three known human coactivators (NCOA1, NCOA2 and NCOA3) as Gamma-class because of their unique roles in modulating transcription and because little is currently known about their PAS-PAS interactions. Additionally, we designate as Delta-class, those PAS proteins involved in transcriptional signaling pathways, but that are missing one of the hallmark domains of this family. For example, the AHR is missing PAS-B and the three PERs are missing a canonical bHLH domain. These structures have led to a general suspicion that Delta-class can be thought of as repressors. Two final classes of mammalian PAS proteins are the Kappa class (for potassium channels) and the Epsilon class (enzymes).

To date, the vast majority of known PAS protein interactions in mammalian systems are homotypic ones, that is, through PAS-PAS domain pairing and HLH domain pairing. This pairing then positions...
the basic region of the bHLH domain to produce a competent DNA binding structure capable of recognizing cognate enhancers, with each member recognizing one half site of the element (Fig. 3). Of these dimers that directly influence gene expression through cognate enhancer elements, the best understood are alpha-beta class pairings like the AHR-ARNT, HIF1-alpha-ARNT (aka HIF1-beta) or CLOCK-ARNTL (aka bMAL1 or MOP3). While homodimers such as ARNT-ARNT and PER-PER homodimers have been reported, the biological consequences of such interactions through recognition of genomic enhancers is unclear (e.g., ARNT and PER homodimers) [82,183]. Similarly, other non-alpha-beta class interactions have also been reported, but it is currently unclear if such interactions require either partner’s PAS domain for this binding (e.g., NCOA1 AND ARNT) [184].

Given that a complete description of the interactome of all mammalian PAS proteins has not yet been completed, it may still be of values to consider the possibility that more complex networks exist among PAS family members. In this regard, one early idea developed to explain toxicity of highly potent dioxin analogs was that the hyper-activated AHR may induce dimerization with additional bHLH-PAS partners other than ARNT and influence gene expression through unique enhancer elements, or through competition for limiting PAS-partners such as ARNT [72,171,185]. If, as presented above, PAS-PAS interaction surfaces exist in potassium channels, phosphodiesterases, regulators of development, circadian factors, then each of these areas of biology may be influencing each other in unappreciated ways. Put another way, the unknown partnering of PAS domains could represent paths of communication between physiological processes that have not been previously recognized.

8. Summary

The AHR has served as a paradigm for the PAS sensor superfamily and the roles of these proteins in environmental adaptation. When we look back at the research on this ancient domain found in all kingdoms of life, a few common themes emerge. For example, this domain is often associated with environmental sensing mechanisms that may have its origin in the adaptation to ‘Light, Oxygen and Voltage’, and in many chordates has evolved to mediate adaptive responses to certain xenobiotics, play important roles in barrier immunity, orchestrate midline development, and regulate an organismal response to circadian time. We also see how the PAS domain harbors multiple properties: First as a pocket for the binding of ligands as distinct as heme, 4-hydroxycinnamic acid, and 2,3,7,8-tetrachlorodibenzo-p-dioxin (Fig. 2). Second as a protein-protein interaction surface that dictates both homotypic and heterotypic interactions; With homotypic interactions often generating dimers capable of recognizing target genomic elements, and heterotypic interactions that control levels of signaling and provide a potential link to heterologous output pathways.

Data availability

No data was used for the research described in the article. Data will be made available on request.

Declaration of Competing Interest

Chris Bradfield reports financial support was provided by National Institutes of Health. Chris Bradfield reports a relationship with National Institutes of Health that includes: funding grants. Chris Bradfield has patent none pending to none.

Acknowledgments

This work was supported by the National Institutes of Health Grants R35ES028377, T32ES007015, T32GM008692, T32CA009135 and P30CA14520. The UW-Madison Science & Medicine Graduate Research Scholars Program, and the National Science Foundation Graduate Research Fellowship under the grant no. DGE-1256259. Thanks to Igor Zhulin, Carrie Partch, Alex Veith, and Robert Bourret for helpful discussions.

References

[1] S.E. Bondos, L. Swint-Kruse, K.S. Matthews, Flexibility and disorder in gene regulation: LacI/GalR and hox proteins, J. Biol. Chem. 290 (41) (2015) 24669–24677, https://doi.org/10.1074/jbc.R115.685032.
[2] A. Grove, Regulation of metabolic pathways by MarR family transcription factors, Comput. Struct. Biotechnol. J. 15 (2017) 366–371, https://doi.org/10.1016/j. csbj.2017.06.001.
[3] D.J. Mangelsdorf, C. Thummel, M. Beato, P. Herrlich, G. Schütz, K. Umesono, B. Blumberg, P. Kastner, M. Mark, P. Chambon, R.M. Evans, The nuclear receptor superfamily: the second decade, Cell 83 (6) (1995) 835–839.
[4] A. Moglich, R.A. Ayers, K. Moffit, Structure and signaling mechanism of Per-ARNT-sim domains, Structure 17 (10) (2009) 1282–1294, https://doi.org/10.1016/j.str.2009.08.011.
[5] B.L. Taylor, I.B. Zhulin, PAS domains: internal sensors of oxygen, redox potential, and light, Microbiol. Mol. Biol. Rev. 63 (2) (1999), 479–.,

Fig. 6. A proposed functional classification scheme of PAS protein interactions. The alpha class (reds) includes those proteins that often have restricted cell specific expression and bind with beta class proteins, often in response to an environmental stimulus or cellular state. The beta class (blues), are more widely expressed and act as binding partners for proteins from multiple PAS families, especially alpha class. The delta class (pinks) includes proteins that are often considered repressors and harbor deletions of classical domains such as PAS-B or bHLH. The gamma class (orange), include the mammalian coactivators (NCOAs) and are known for their unique role in transcriptional modulation. Finally, we show two PAS protein families that were not discussed in detail in this review; The epsilon class (purple), includes enzymes with PAS domains, while the kappa class (green) includes potassium channels. The bHLH, PAS A and PAS B domains are depicted as A and B respectively in the proteins in which they adhere to motifs as designated by UNIPROT [191].
[151] C.A. Bradfield, L.F. Bjelkes, Structure—activity relationships of dietary indoles: a proposed mechanism of action as modifiers of xenobiotic metabolism, J. Toxicol. Environ. Health 21 (3) (1987) 311–323. https://doi.org/10.1080/15287960709713011.

[152] A. Rannug, U. Rannug, H. Rosenkranz, L. Winquist, R. Westerholm, E. Agrell, A. Graefstrom, Certain photooxidized derivatives of tryptophan bind with very high affinity to the aryl receptor and are likely to be endogenous signal substances, J. Biol. Chem. 262 (32) (1987) 15422–15427.

[153] L.F. Bjelkes, J.Y. Kim, K.R. Grove, J.C. Bartholomew, C.A. Bradfield, Aromatic hydrocarbon responsiveness-receptor agonists generated from indole-3-carbinol in vitro and in vivo: comparisons with 2,3,7,8-tetrachlorodibenzop-456-dioxin, Proc. Natl. Acad. Sci. 88 (21) (1991) 9543, https://doi.org/10.1073/pnas.88.21.9543.

[154] L.P. Nguyen, E.L. Hsu, G. Chowdhury, M. Dostalek, G.L. Wang, B.H. Jiang, E.A. Rue, G.L. Semenza, Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS superfamily that interacts with components of the dioxin signaling pathway, J. Biol. Chem. 272 (13) (1997) 8581–8593, https://doi.org/10.1074/jbc.272.13.8581.

[155] Y.D. Zhou, M. Barnard, H.Tian, X.Li, H.Z. Ring, U. Francke, J. Shelton, J. Richardson, D.W. Russell, S.L. McKnight, Molecular characterization of two mammalian BHLH-PAS domain proteins selectively expressed in the central nervous system, Proc. Natl. Acad. Sci. U. S. A. 94 (2) (1997) 713–718, https://doi.org/10.1073/pnas.94.2.713.

[156] Z.S. Sun, U. Albrecht, O. Zbuchenko, J. Bailey, G. Eichelle, C.C. Lee, RIGUL, a putative mammalian ortholog of the drosophila period gene, Cell 90 (6) (1997) 1003–1011, https://doi.org/10.1016/S0092-8674(00)80636-9.

[157] Z.P. Shearman, M.J. Zylka, D.R. Weaver, L.F. Kolakowski, S.M. Reppert, Two period homologs: circadian expression and photic regulation in the suprachiasmatic nuclei, Neuron 19 (6) (1997) 1261–1269, https://doi.org/10.1016/S0896-6273(00)80417-1.

[158] C.M. Fan, E. Kivawana, A. Bullone, C.F. Fletcher, N.G. Copeland, N.A. Jenkins, S. Crews, S. Martinez, L. Puritis, J.L.R. Rubenstein, M. Tessier-Lavigne, Expression patterns of two murine homologs of drosophila period-single-minded suggest possible roles in embryonic patterning and in the pathogenesis of down syndrome, Mol. Cell. Neurosci. 7 (1) (1996) 1–16, https://doi.org/10.1006/mcne.1996.0001.

[159] M. Ema, M. Morita, S. Ikawa, Y. Matsuda, O. Gotay, S. Sajish, H. Fujii, H. Hamada, Y. Kikuchi, Y. Fujii-Kuriyama, Two new members of the murine sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis, Mol. Cell. Biol. 16 (10) (1996) 5805–5817.

[160] M.R. Probst, C.-M. Fan, M. Tessier-Lavigne, O. Hankinson, Two murine homologs of the drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein, J. Biol. Chem. 272 (7) (1997) 4451–4457, https://doi.org/10.1074/jbc.272.7.4451.

[161] D.C. Bersten, A.J. Peet, M.L. Whitelaw, BHLH-PAS proteins in cancer, Nat. Rev. Cancer 13 (12) (2013) 827.

[162] Y.Z. Gu, J.B. Hogenesch, C.A. Bradfield, The PAS superfamily: sensors of environmental and developmental signals, Annu. Rev. Pharmacol. Toxicol. 40 (2000) 519–561, https://doi.org/10.1146/annurev.pharmtox.40.1.519.

[163] B.E. McIntosh, J.B. Hogenesch, C.A. Bradfield, Mammalian per-antm-sim proteins in environmental adaptation, Annu. Rev. Physiol. 72 (2010) 625–645, https://doi.org/10.1146/annurev.physiol.012809.155030.

[164] M.J. Zylka, L.P. Shearman, J.D. Levine, X. Jin, D.R. Weaver, S.M. Reppert, Molecular analysis of mammalian timeless, Neuron 21 (5) (1998) 1115–1122, https://doi.org/10.1016/S0896-6273(00)00580-9.

[165] T.V. Beschlag, S. Wang, D.W. Rose, J. Torisch, S. Reisz-Porszaz, K. Muhammad, W.E. Nelson, M.R. Probst, M.G. Rosenfeld, O. Hankinson, Recruitment of the NCoA/SRC-1/P160 family of transcriptional coactivators by the aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator complex, Mol. Cell. Biol. 22 (12) (2002) 4319–4333, https://doi.org/10.1128/MCB.22.12.4319-4333.2002.

[166] K. Gradin, J. McGuire, R.H. Wender, I. Kvetkoiva, M.L. thielew, R. Toftgård, L. Dorner, M. Gasman, M. Fjeld-Larsen, E. Eriksen, and the human circadian clock mutants to

[167] S. Sakurai, T. Shimizu, U. Ohto, The crystal structure of the AhRR heterodimer reveals the structural basis of the repression of AhR-mediated transcription, J. Biol. Chem. 279 (11) (2004) 12044–12051, https://doi.org/10.1074/jbc.M308128200.

[168] H. Miyatake, M. Mukai, S.-Y. Park, S. Adachi, K. Tamura, H. Nakamura, Sensory mechanism of oxygen and export signals of the human aryl hydrocarbon receptor, J. Biol. Chem. 273 (5) (1998) 2895–2899, https://doi.org/10.1074/jbc.273.5.2895.

[169] M.W. Young, Time travels: a 40-year journey from drosophila clock mutants to human circadian disorders (nobel lecture), Angew. Chem. Int. Ed. 57 (36) (2018) 11532–11539, https://doi.org/10.1002/anie.201803337.

[170] J.B. Hogenesch, W.K. Chan, V.H. Jackiw, R.C. Brown, Y.Z. Gu, M. Praygant, G. H. Perdew, C.A. Bradfield, 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 (13) (1997) 8581–8593, https://doi.org/10.1074/jbc.272.13.8581.

[171] J.W. Young, Travel: a 40-year journey from drosophila’s clock mutants to human circadian disorders (nobel lecture), Angew. Chem. Int. Ed. 57 (36) (2018) 11532–11539, https://doi.org/10.1002/anie.201803337.