



Published in final edited form as:

*Curr Opin Struct Biol.* 2017 April ; 43: 1–9. doi:10.1016/j.sbi.2016.09.011.

## Structural characterization of mammalian bHLH-PAS transcription factors

Dalei Wu<sup>1,2</sup> and Fraydoon Rastinejad<sup>1,\*</sup>

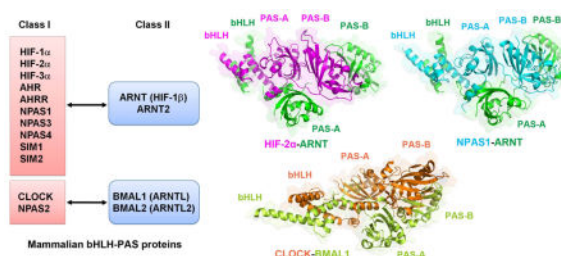
<sup>1</sup>Integrative Metabolism Program, Sanford Burnham Prebys Medical Discovery Institute, Orlando, FL 32827, USA

<sup>2</sup>State Key Laboratory of Microbial Technology, School of Life Sciences, Shandong University, Qingdao, 266237, China

### Abstract

The mammalian basic helix-loop-helix-PER-ARNT-SIM (bHLH-PAS) transcription factors share common architectural features that include a bHLH DNA-binding domain and tandemly positioned PAS domains. The sixteen members of this family include the hypoxia-inducible factors (HIF-1 $\alpha$  and HIF-2 $\alpha$ ), ARNT (also known as HIF-1 $\beta$ ), CLOCK and BMAL1. Most bHLH-PAS proteins have been genetically linked to variety of diseases in humans, including cancers, metabolic syndromes and psychiatric conditions. To function as transcription factors, the bHLH-PAS proteins must form heterodimeric complexes. Recent crystallographic studies of HIF- $\alpha$ -ARNT and CLOCK-BMAL1 complexes have unveiled the organization of their multi-domain bHLH-PAS-A-PAS-B segments, revealing how these architectures can give rise to unique patterns of heterodimerization. As our structural understanding becomes better integrated with ligand-discovery and target gene identification, a more comprehensive picture of their architectural and functional properties will emerge.

### Graphical Abstract



\*Corresponding author: Rastinejad, Fraydoon (frastinejad@sbdisccovery.org).

#### Conflict of interest

Nothing declared.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Introduction

The basic helix-loop-helix-PER-ARNT-SIM (bHLH-PAS) family of proteins are dimeric transcription factors with a variety of distinct functions. These proteins share a common domain arrangement, including a bHLH DNA binding domain, tandemly positioned PAS domains (PAS-A and PAS-B) and a variable transactivation or transrepression domains (TAD or TRD) (Figure 1a). Based on their distinct heterodimerization patterns, the mammalian bHLH-PAS proteins can be divided into two separate classes [1]. Class II (class  $\beta$ ) proteins act as common heterodimerization partners for each of class I (class  $\alpha$ ) members (Figure 1b). Some class II members, such as aryl hydrocarbon receptor nuclear translocator (ARNT) [2,3], may also have the ability to form homodimers, although it is not yet clear if these homodimers are physiologically functional.

As shown in Figure 1b, the class I members of the family includes three hypoxia-inducible factors (HIF-1 $\alpha$ , HIF-2 $\alpha$  and HIF-3 $\alpha$ ), aryl hydrocarbon receptor (AHR), aryl hydrocarbon receptor repressor (AHRR), four neuronal PAS proteins (NPAS1, NPAS2, NPAS3, NPAS4), two single-minded proteins (SIM1, SIM2), and clock circadian regulator (CLOCK). The class II members only consists of four members, ARNT (also known as HIF-1 $\beta$ ), ARNT2, brain and muscle ARNT-like protein 1 (BMAL1, also called ARNTL), and BMAL2 (ARNTL2). Two non-overlapping groups of heterodimers arise (the upper ARNT group and bottom BMAL group in Figure 1b). Through dimerization, the bHLH-PAS proteins form transcriptionally active complexes capable of binding with high affinity to DNA response elements for regulation of their target genes.

Members of the bHLH-PAS family are known to participate in a wide variety of biological processes. These proteins appear to have the ability to register diverse physiological states and/or environmental signals. The HIFs are sensitive to low oxygen stress, and respond to this stress by up-regulating transcriptional programs involving erythropoiesis, angiogenesis, and metabolism that together allow the cells to cope and counteract their low oxygen stress [4]. AHR binds directly to xenobiotics and suggested endogenous ligands, and can induce the expression of detoxifying enzymes within the liver and modulate immune cell differentiation to counteract the cell's xenobiotic and toxic stresses [5]. AHRR, itself a target gene of AHR, lacks the functional transactivation domain that is contained in AHR, and represses AHR activity by sequestering their common dimerization subunit ARNT. It is not known if AHRR can bind to xenobiotic agents, but like the AHR-ARNT heterodimer, the AHRR-ARNT heterodimer directly binds to xenobiotic response element (XRE) DNA [6]. The NPAS1 and NPAS3 proteins are highly expressed in the nervous system and within inhibitory interneurons, and their genetic deficiencies have been linked to a variety of human psychiatric diseases [7]. NPAS4 is an activity-dependent transcription factor required for contextual memory formation [8]. SIM1 and SIM2 have important functions in neurogenesis and embryogenesis [9]. CLOCK and its paralog NPAS2 are the major transcriptional regulators of circadian rhythm, and also control some metabolic gene programs [10].

Previous reviews of the mammalian bHLH-PAS family largely focused on their functions and potential roles in human diseases (e.g. cancer) [1,11,12]. Here we review our understanding of the bHLH-PAS proteins in their three-dimensional architectural contexts

and functional assemblies, highlighting how their domains predispose them for dimerization, ligand-binding, DNA-binding, and protein-protein interactions. We begin by reviewing the current knowledges about how several Class I and Class II members physically assemble into distinct patterns of heterodimers. We then focus more specifically on the HIF-1 $\alpha$  protein to show how DNA-binding and interactions with regulatory proteins are formed.

## Overall structures of bHLH-PAS heterodimers

The N-terminal regions of bHLH-PAS proteins contain three well-defined domains (bHLH, PAS-A and PAS-B), all of which are relatively conserved within their protein family (Figure 1a and b). By contrast, the length and composition at their C-terminal regions are highly variable, and also likely to be structurally flexible. Accordingly, the majority of reported protein structures of this family derive from their N-terminal domains; while the structural information about C-terminal regions is currently represented only by short peptides which are in bound states with other proteins.

The first isolated PAS-B domain structures of HIF-2 $\alpha$  [13] and ARNT [14] were reported more than a decade ago using solution NMR (Figure 1c), which gave rise to a suggested model of heterodimerization between the PAS-B domains of HIF-2 $\alpha$  and ARNT mediated through their respective  $\beta$ -sheet surfaces [14]. This model subsequently guided the authors to mutate surface amino-acid residues on each subunit (HIF-2 $\alpha$  R247E and ARNT E362R), so as to create an artificial salt bridge that could further enhance the binding affinity of two PAS-B domains and facilitate their co-crystallization (Figure 1d) [15]. This artificial PAS-B heterodimer structure not only correlated with the original NMR-based model that two domains can form an interface via their  $\beta$ -sheets in an anti-parallel fashion, but more importantly revealed a large cavity ( $\sim 300 \text{ \AA}^3$ ) with the HIF-2 $\alpha$  PAS-B domain [16–19].

While the studies outlined above utilized only isolated PAS-B domains of bHLH-PAS proteins, they did not address how the multi-domain proteins are architecturally organized. The CLOCK-BMAL1 heterodimer crystal (containing the bHLH-PAS-A-PAS-B segments of both proteins) was reported in 2012, and became the first multi-domain structure obtained for this family [20]. While that study did not show how DNA could bind within the complex, a subsequent crystal structure was reported involving just the isolated bHLH segments of CLOCK-BMAL1 in complex with E-box DNA [21]. In 2013, we reported the structure of the isolated AHR PAS-A domain, which forms a helix-swapped homodimer in the crystal and in solution (Figure 1e) [22]. More recently, we reported the multi-domain (bHLH-PAS-A-PAS-B) crystal structures for both HIF-1 $\alpha$ -ARNT and HIF-2 $\alpha$ -ARNT heterodimers, including these heterodimers captured in a variety of functional states that included bound small molecules and the hypoxia response element (HRE) DNA [23].

The multi-domain heterodimeric structures revealed some surprising lessons about how bHLH-PAS proteins form functional heterodimers through the interactions of their N-terminal bHLH, PAS-A and PAS-B domains. As shown in Figure 1f and 1g, the CLOCK-BMAL1 and HIF-2 $\alpha$ -ARNT structures are each highly intertwined and asymmetric. This finding showed that the PAS domains within each of these heterodimers do not form equivalent (symmetry) interactions. Furthermore, in comparing the two heterodimers, one

observes very distinctive overall quaternary arrangements, whereby two separate groups of bHLH-PAS proteins (those that heterodimerize with ARNT and those that heterodimerize with BMAL1) are architecturally distinct in their overall structures (Figure 1b). The three adjacent domains (bHLH, PAS-A and PAS-B) in each of CLOCK and BMAL1 are spatially close within their individual polypeptides (Figure 1f). Yet in HIF-2 $\alpha$ -ARNT, the three domains of ARNT have no intra-molecular contacts among them, allowing ARNT spirals around the outside of its partner HIF-2 $\alpha$  (Figure 1g). We have more recently discovered, through additional X-ray crystallographic studies, that the domain arrangements and quaternary architectures of NPAS1-ARNT (Figure 1h) and NPAS3-ARNT structures are essentially identical to those of HIF-1 $\alpha$ -ARNT and HIF-2 $\alpha$ -ARNT (Wu *et al.*, unpublished data). Together these findings suggest that many ARNT heterodimers give rise to a similar quaternary architecture, while the BMAL1 heterodimers appear to form an altogether different type of quaternary architecture.

## Communication between Domains

In more closely examining the structures of HIF-2 $\alpha$ -ARNT and CLOCK-BMAL1 complexes, we highlight the amino-acids responsible for aligning the domain-domain interfaces in each heterodimer. As shown in Figure 2a, there are six different domain-domain interfaces in HIF-2 $\alpha$ -ARNT, including four heterodimer domain interfaces and two intra-HIF-2 $\alpha$  interfaces (detailed in Figure 2b). However in CLOCK-BMAL1, only four significant domain-domain interfaces can be found (Figure 2a and b). The total heterodimer interface areas in HIF-2 $\alpha$ -ARNT and CLOCK-BMAL1 are 3550 Å<sup>2</sup> and 4800 Å<sup>2</sup>, respectively (as calculated by the PISA program [24] in Figure 2b). The much larger buried surface area of CLOCK-BMAL1 is mainly contributed by its well-ordered linker regions connecting the domains, while the linkers between the ARNT domains are completely disordered in the crystal of HIF-2 $\alpha$ -ARNT.

Although the overall quaternary structures of these two heterodimers are clearly different (Figure 1f, 1g and 2a), one still finds many shared architectural features between them. The bHLH, PAS-A and PAS-B domains all act as dimerization modules in both heterodimers. Furthermore, some points of contact between corresponding domains (i.e. Interfaces 1, 2 and 4) are somewhat similar in the HIF-2 $\alpha$ -ARNT and CLOCK-BMAL1 (Figure 2c). The buried surface areas of these interfaces are also comparable in size (Figure 2b). Interfaces 1 (bHLH/bHLH) and 2 (PAS-A/PAS-A) are roughly symmetric and mediated mainly by hydrophobic residues (Figure 2c). Interestingly, interface 2 is also similar in structure to the PAS-A homodimer of AHR (Figure 1e), highlighting that the PAS-A domains in all bHLH-PAS proteins may form a common dimerization surfaces that involve swapped N-terminal helices.

By contrast the parallel-faced Interface 4 (PAS-B/PAS-B) in the multi-domain HIF-2 $\alpha$ -ARNT complex (Figure 2c) proved to be dramatically different from the originally proposed model based on early NMR experiments with the isolated HIF-2 $\alpha$  and ARNT PAS-B domains, which had predicted to involve a  $\beta$ -sheet mediated anti-parallel interface (Figure 1d). The inconsistency of the earlier NMR and our crystallographic findings highlights the importance of the entire protein context when studying PAS domains. In further comparing

HIF-2 $\alpha$ -ARNT and CLOCK-BMAL1, one finds there is a small difference at Interface 4 within these complexes. As shown in Figure 2c, the C-terminal helix of HIF-2 $\alpha$  PAS-B domain is wrapped towards the interface, while that of CLOCK PAS-B faces the opposite direction. As a result, the buried surface area of Interface 4 is larger in HIF-2 $\alpha$ -ARNT than that in CLOCK-BMAL1 (Figure 2b). Furthermore, this difference appears to contribute to the overall divergent quaternary organizations of these heterodimers.

Besides the interfaces formed between corresponding domains in these heterodimers, there are also several interfaces formed by distinct domains in each subunit. Among the two PAS-A/PAS-B domain interfaces in HIF-2 $\alpha$ -ARNT (Figure 2d), one (Interface 3) is formed between ARNT PAS-A and HIF-2 $\alpha$  PAS-B, and the other (Interface 5) is formed between the PAS-A and PAS-B domains within HIF-2 $\alpha$ . By contrast, only a single PAS-A/PAS-B domain interface is found in BMAL1 (Interface 3). When all these PAS-A/PAS-B interfaces above are compared through superimposition (Figure 2d), interestingly, one finds the same conserved “patch” (mainly four residues on two neighboring  $\beta$ -strands) on each PAS-A domains is responsible for forming interactions with a variety of different surfaces from the PAS-B domains. In addition, it has been shown that the PAS-A/PAS-B interface of BMAL1 is very similar to that of PER2 protein [20]. PER has both PAS-A and PAS-B domains but lacks a bHLH domain within its polypeptide [25]. Overall these observations echo a previous suggestion that PAS proteins are particularly capable of making use of their  $\beta$ -strands in forming dimerization interfaces [26]. Interface 6 of HIF-2 $\alpha$ -ARNT is formed by the bHLH and PAS-A domains of HIF-2 $\alpha$  (Figure 2e). This type of interface was not observed in the CLOCK-BMAL1 complex, but we observed an identical interface in the HIF-1 $\alpha$ -ARNT complex [23], as well as in NPAS1-ARNT and NPAS3-ARNT complexes (Wu *et al.*, unpublished data).

## Interactions between HIF-1 $\alpha$ and other macromolecules

In addition to the heterodimerization with their partners, bHLH-PAS proteins must interact with other macromolecules (e.g. DNA and proteins) to both regulate and mediate their distinct physiological functions as transcription factors. Here, with an emphasis on the most thoroughly-studied example HIF-1 $\alpha$ , we summarize the current structural understandings of these macromolecular interactions (Figure 3a). As discussed above and shown in Figure 3b, the first step in forming a functional transcription factors is for HIF-1 $\alpha$  to physically dimerize with ARNT using their contiguous N-terminal bHLH, PAS-A and PAS-B domains. Once the productive heterodimer is formed, HIF-1 $\alpha$ -ARNT can bind efficiently to the core HRE sequence (5'-NRCGTG-3', where R is A or G) associated with its target genes [27], and subsequently recruit transcriptional coactivators such as CBP/p300 via HIF's transactivation domains (TAD) located within its C-terminal segment [28]. Our crystal structures of HIF-1 $\alpha$ -ARNT-DNA (Figure 3c) and HIF-2 $\alpha$ -ARNT-DNA revealed how the bHLH domains cooperate to recognize the major groove face of HRE sequence (5'-TACGTG-3'), and how specific interactions between base pairs and residues are formed in the complex [23].

As a sensor of the oxygen levels, the HIF-1 $\alpha$  protein is only stable and able to accumulate inside cells under low-oxygen stress (hypoxia). When the oxygen level is normal

(normoxia), two specific proline residues (P402 and P564, either or both) in the oxygen-dependent degradation domain (ODD) of HIF-1 $\alpha$  are hydroxylated by prolyl hydroxylase domain (PHD) proteins (Figure 3a and b). The PHD proteins utilize both oxygen and 2-oxoglutarate (2-OG) as co-substrates. The PHD proteins are among the most important of known HIF-1 $\alpha$  regulators. As shown in Figure 3d, the crystal structure of PHD2 catalytic domain, in complex with a 17-residue peptide (including P564) of the HIF-1 $\alpha$  ODD, reveals the interactions involved within the hydroxylation step [29]. Once the proline residues are hydroxylated, the von Hippel-Lindau tumor-suppressor protein (pVHL) recognizes and binds to the HIF-1 $\alpha$  protein and further recruits an E3 ubiquitin-protein ligase complex that targets HIF-1 $\alpha$  for degradation (Figure 3a). The reported structure of HIF-1 $\alpha$  ODD peptide bound to the pVHL-Elongin B-Elongin C complex has also shown how the hydroxyl-P564 is recognized by pVHL through highly specific molecular interactions (Figure 3e) [30,31].

Factor-inhibiting HIF (FIH) also regulates the HIF-1 $\alpha$  subunit, by hydroxylating an asparagine residue (N803) near the C-terminus of HIF-1 $\alpha$  in an oxygen-dependent manner (Figure 3a and b). This asparagine hydroxylation subsequently blocks the interactions between HIF-1 $\alpha$  and CBP, and represses the transcriptional activity. The crystal structure of FIH complexed with a C-TAD fragment of HIF-1 $\alpha$  and 2-OG has also been reported, showing that the TAD peptide adopts a loop conformation near the N803 hydroxylation site, and sends N803 precisely into the active site of FIH (Figure 3f) [32]. The NMR solution structure of the TAZ1 motif of CBP and a similar C-TAD peptide of HIF-1 $\alpha$  has shown that the N803 residue is located on a short helix with its sidechain buried in the interface with CBP (Figure 3g), which can explain why hydroxylated N803 leads to reduced interactions between HIF-1 $\alpha$  and CBP/p300 [33,34].

Since the HIF pathway is clearly involved in the cancer progression (especially solid tumors) [35], inhibitors of both HIF-1 $\alpha$  and HIF-2 $\alpha$  have long been pursued in the anti-cancer drug discovery [17,36,37]. Our crystal structures of HIF-1 $\alpha$ -ARNT and HIF-2 $\alpha$ -ARNT heterodimers led to the discovery of multiple hydrophobic ligand-binding pockets, including those positioned directly inside the PAS domains [23]. This finding highlights the importance of PAS domains beyond their dimerization functions as potential ligand-binding domains, and further opens the door for more aggressively targeting HIF proteins with drug-like molecules for therapeutic benefit in the cancer setting [4]. Interestingly, the ARNT protein also harbors open cavities within its tandem PAS domains, suggesting that the wider mammalian bHLH-PAS proteins may all be endowed with ligand-sensing pockets within their PAS-A and PAS-B domains. In addition to the critical subunit dimerization between HIF- $\alpha$  and ARNT, the direct interactions of HIF- $\alpha$  with other regulatory proteins (e.g. those shown in Figure 3) have also been targeted to modulate HIF activities in different therapeutic settings [38].

## Closing remarks

The mammalian bHLH-PAS proteins are an important class of transcription factors that require subunit dimerization among family members. The first set of multi-domain structures reported, of CLOCK-BMAL1, HIF-1 $\alpha$ -ARNT and HIF-2 $\alpha$ -ARNT complexes helped explain how bHLH-PAS proteins can recognize their heterodimerization partners,



further showing the existence of two distinct architectural classes distinguished at the level of quaternary structure (Figure 1) [20,23]. A larger and more comprehensive picture of macromolecular complexes involving other bHLH-PAS proteins is still needed to reveal if other modes of quaternary architecture are possible with this class of heterodimers. Our broad understanding of how such structures are linked to distinct functions in each case still require a great deal of biochemical and cell-based studies, particularly those involving other classes of interacting proteins and the potential endogenous ligands utilized by each bHLH-PAS member.

The basic physiology and molecular architectures of this group have been best characterized so far for HIF- $\alpha$ -ARNT and CLOCK-BMAL1 complexes. Only in the case of AHR and HIF- $\alpha$  proteins, there has been work to show their small-molecule binding potentials. Given that the PAS domains within this family are highly conserved in their basic core three-dimensional architecture, but still vary significantly in the residues within their putative ligand-binding pockets, the remaining bHLH-PAS members also present as candidates for future drug-discovery efforts. These efforts will be particularly valuable since genetic defects in members of this family have already been linked with significant disease states in humans, including cancers, metabolic diseases, autoimmune diseases and psychiatric diseases (Table 1).

Beyond the identification of natural ligands and drug candidates that can be directed specifically at each member, much consideration must be given to how small-molecule binding at the PAS domains can mechanistically alter the physiological responses of these factors in their roles as regulators of genetic programs. Another class of mammalian transcription factors, the nuclear receptors, have been far better studied in that both endogenous and synthetic small-molecule modulators are clearly ascribed now to most family members, and a clear mechanistic understanding developed for how ligand binding is transmitted to regulation of nuclear receptor function [39]. If our hypothesis proves true that the bHLH-PAS factors all have endogenous ligands and function as ligand-binding transcription factors, these ligands will also then provide the optimal tools for exploring the downstream target gene programs for each individual member. As a class of transcription factors with great therapeutic promise, the bHLH-PAS proteins are now emerging as exciting candidates for in-depth explorations using integrated structural, chemical and cell-based approaches.

## Acknowledgments

This work was supported by the National Institutes of Health [grant number NIGMS 1R01GM117013] and U.S. ARMY Medical Research [grant number W81XWH-16-1-0322].

## References

1. Kewley RJ, Whitelaw ML, Chapman-Smith A. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol.* 2004; 36:189–204. [PubMed: 14643885]
2. Sogawa K, Nakano R, Kobayashi A, Kikuchi Y, Ohe N, Matsushita N, Fujii-Kuriyama Y. Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation. *Proc Natl Acad Sci U S A.* 1995; 92:1936–1940. [PubMed: 7892203]

3. Wang F, Shi S, Zhang R, Hankinson O. Identifying target genes of the aryl hydrocarbon receptor nuclear translocator (Arnt) using DNA microarray analysis. *Biol Chem*. 2006; 387:1215–1218. [PubMed: 16972789]
- 4. Keith B, Johnson RS, Simon MC. HIF1alpha and HIF2alpha: sibling rivalry in hypoxic tumour growth and progression. *Nat Rev Cancer*. 2012; 12:9–22. This review thoroughly discussed both overlapping and unique (even opposite) roles of HIF-1 $\alpha$  and HIF-2 $\alpha$ , in many aspects of tumor growth and potential therapeutics.
5. Moura-Alves P, Fae K, Houthuys E, Dorhoi A, Kreuchwig A, Furkert J, Barison N, Diehl A, Munder A, Constant P, et al. AhR sensing of bacterial pigments regulates antibacterial defence. *Nature*. 2014; 512:387–392. [PubMed: 25119038]
6. Mimura J, Ema M, Sogawa K, Fujii-Kuriyama Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes Dev*. 1999; 13:20–25. [PubMed: 9887096]
7. Erbel-Sieler C, Dudley C, Zhou Y, Wu X, Estill SJ, Han T, Diaz-Arrastia R, Brunskill EW, Potter SS, McKnight SL. Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc Natl Acad Sci U S A*. 2004; 101:13648–13653. [PubMed: 15347806]
8. Ramamoorthi K, Fropf R, Belfort GM, Fitzmaurice HL, McKinney RM, Neve RL, Otto T, Lin Y. Npas4 regulates a transcriptional program in CA3 required for contextual memory formation. *Science*. 2011; 334:1669–1675. [PubMed: 22194569]
9. Ema M, Morita M, Ikawa S, Tanaka M, Matsuda Y, Gotoh O, Saijoh Y, Fujii H, Hamada H, Kikuchi Y, et al. Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis. *Mol Cell Biol*. 1996; 16:5865–5875. [PubMed: 8927054]
10. DeBruyne JP, Weaver DR, Reppert SM. CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nat Neurosci*. 2007; 10:543–545. [PubMed: 17417633]
- 11. McIntosh BE, Hogenesch JB, Bradfield CA. Mammalian Per-Arnt-Sim proteins in environmental adaptation. *Annu Rev Physiol*. 2010; 72:625–645. This review summarized the functions of bHLH-PAS proteins and some other mammalian proteins containing the PAS domain, in a unique environmental responsive angle with the comprehensive analysis of several major pathways involved. [PubMed: 20148691]
- 12. Bersten DC, Sullivan AE, Peet DJ, Whitelaw ML. bHLH-PAS proteins in cancer. *Nat Rev Cancer*. 2013; 13:827–841. This review detailed the different while connected roles of many mammalian bHLH-PAS proteins and their related proteins in cancer. [PubMed: 24263188]
13. Erbel PJ, Card PB, Karakuzu O, Bruick RK, Gardner KH. Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor. *Proc Natl Acad Sci U S A*. 2003; 100:15504–15509. [PubMed: 14668441]
14. Card PB, Erbel PJ, Gardner KH. Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. *J Mol Biol*. 2005; 353:664–677. [PubMed: 16181639]
15. Scheuermann TH, Tomchick DR, Machius M, Guo Y, Bruick RK, Gardner KH. Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor. *Proc Natl Acad Sci U S A*. 2009; 106:450–455. [PubMed: 19129502]
16. Key J, Scheuermann TH, Anderson PC, Daggett V, Gardner KH. Principles of ligand binding within a completely buried cavity in HIF2alpha PAS-B. *J Am Chem Soc*. 2009; 131:17647–17654. [PubMed: 19950993]
17. Rogers JL, Bayeh L, Scheuermann TH, Longgood J, Key J, Naidoo J, Melito L, Shokri C, Frantz DE, Bruick RK, et al. Development of inhibitors of the PAS-B domain of the HIF-2alpha transcription factor. *J Med Chem*. 2013; 56:1739–1747. [PubMed: 23363003]
18. Scheuermann TH, Li Q, Ma HW, Key J, Zhang L, Chen R, Garcia JA, Naidoo J, Longgood J, Frantz DE, et al. Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nat Chem Biol*. 2013; 9:271–276. [PubMed: 23434853]
19. Scheuermann TH, Stroud D, Sleet CE, Bayeh L, Shokri C, Wang H, Caldwell CG, Longgood J, MacMillan JB, Bruick RK, et al. Isoform-Selective and Stereoselective Inhibition of Hypoxia Inducible Factor-2. *J Med Chem*. 2015; 58:5930–5941. [PubMed: 26226049]

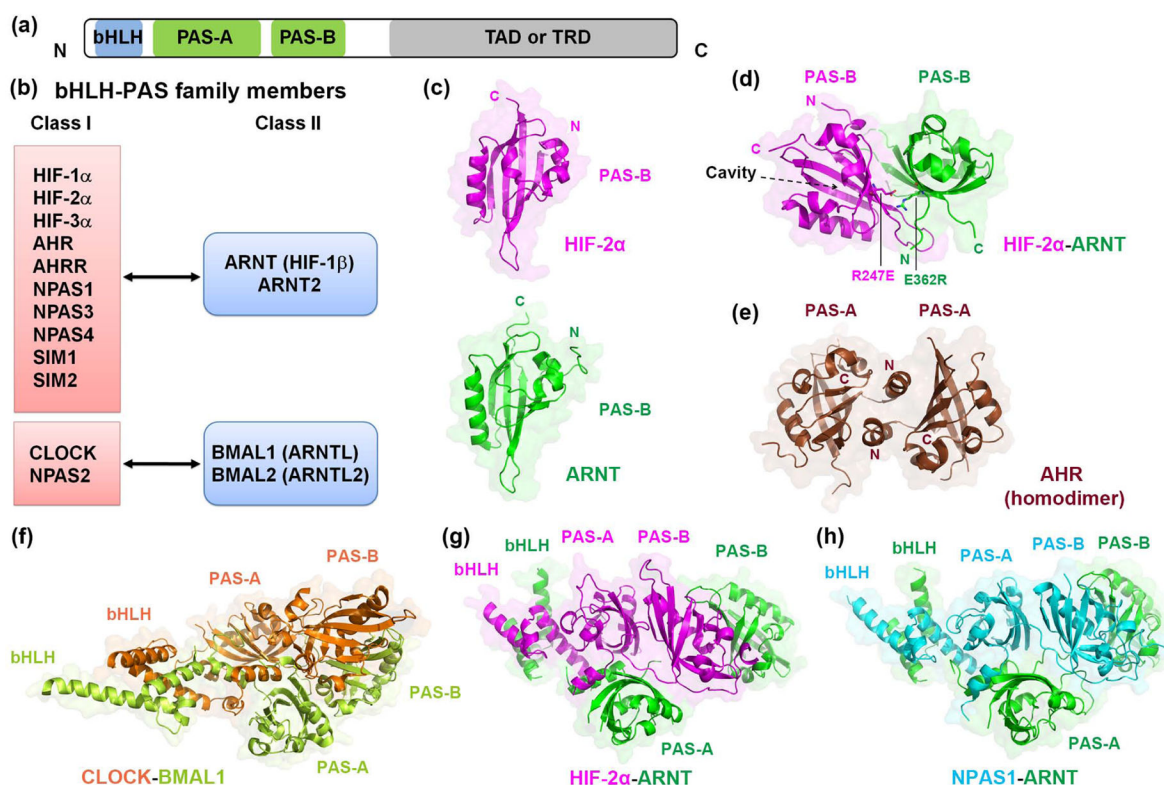


- 20. Huang N, Chelliah Y, Shan Y, Taylor CA, Yoo SH, Partch C, Green CB, Zhang H, Takahashi JS. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science*. 2012; 337:189–194. This study described the first multi-domain crystal structure of bHLH-PAS complex, opening the door to understand how the bHLH, PAS-A and PAS-B domains collaborate in the heterodimerization. [PubMed: 22653727]
- 21. Wang Z, Wu Y, Li L, Su XD. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. *Cell Res*. 2013; 23:213–224. [PubMed: 23229515]
- 22. Wu D, Potluri N, Kim Y, Rastinejad F. Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain. *Mol Cell Biol*. 2013; 33:4346–4356. [PubMed: 24001774]
- 23. Wu D, Potluri N, Lu J, Kim Y, Rastinejad F. Structural integration in hypoxia-inducible factors. *Nature*. 2015; 524:303–308. This study reported several multi-domain structures of HIF-1 $\alpha$ -ARNT and HIF-2 $\alpha$ -ARNT, including those bound to DNA and small-molecule ligands. These structures reveal how the HIF complexes recognize HRE sequence, and also provide useful information for future drug discovery targeting the HIFs. [PubMed: 26245371]
- 24. Krissinel E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res*. 2015; 43:W314–319. [PubMed: 25908787]
- 25. Hennig S, Strauss HM, Vanselow K, Yildiz O, Schulze S, Arens J, Kramer A, Wolf E. Structural and functional analyses of PAS domain interactions of the clock proteins *Drosophila* PERIOD and mouse PERIOD2. *PLoS Biol*. 2009; 7:e94. [PubMed: 19402751]
- 26. Moglich A, Ayers RA, Moffat K. Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure*. 2009; 17:1282–1294. This review systemically discussed the structural properties of PAS domains, especially the plastic interfaces used in their homo- or hetero-dimerization. [PubMed: 19836329]
- 27. Wenger RH, Stiehl DP, Camenisch G. Integration of oxygen signaling at the consensus HRE. *Sci STKE*. 2005; 2005:re12. [PubMed: 16234508]
- 28. Semenza GL. Hypoxia-inducible factors in physiology and medicine. *Cell*. 2012; 148:399–408. [PubMed: 22304911]
- 29. Chowdhury R, McDonough MA, Mecinovic J, Loenarz C, Flashman E, Hewitson KS, Domene C, Schofield CJ. Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure*. 2009; 17:981–989. [PubMed: 19604478]
- 30. Hon WC, Wilson MI, Harlos K, Claridge TD, Schofield CJ, Pugh CW, Maxwell PH, Ratcliffe PJ, Stuart DI, Jones EY. Structural basis for the recognition of hydroxyproline in HIF-1  $\alpha$  by pVHL. *Nature*. 2002; 417:975–978. [PubMed: 12050673]
- 31. Min JH, Yang H, Ivan M, Gertler F, Kaelin WG Jr, Pavletich NP. Structure of an HIF-1  $\alpha$  - pVHL complex: hydroxyproline recognition in signaling. *Science*. 2002; 296:1886–1889. [PubMed: 12004076]
- 32. Elkins JM, Hewitson KS, McNeill LA, Seibel JF, Schlemminger I, Pugh CW, Ratcliffe PJ, Schofield CJ. Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1  $\alpha$ . *J Biol Chem*. 2003; 278:1802–1806. [PubMed: 12446723]
- 33. Dames SA, Martinez-Yamout M, De Guzman RN, Dyson HJ, Wright PE. Structural basis for Hif-1  $\alpha$ /CBP recognition in the cellular hypoxic response. *Proc Natl Acad Sci U S A*. 2002; 99:5271–5276. [PubMed: 11959977]
- 34. Freedman SJ, Sun ZY, Poy F, Kung AL, Livingston DM, Wagner G, Eck MJ. Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1  $\alpha$ . *Proc Natl Acad Sci U S A*. 2002; 99:5367–5372. [PubMed: 11959990]
- 35. LaGory EL, Giaccia AJ. The ever-expanding role of HIF in tumour and stromal biology. *Nat Cell Biol*. 2016; 18:356–365. [PubMed: 27027486]
- 36. Semenza GL. Hypoxia-inducible factors: mediators of cancer progression and targets for cancer therapy. *Trends Pharmacol Sci*. 2012; 33:207–214. [PubMed: 22398146]
- 37. Cardoso R, Love R, Nilsson CL, Bergqvist S, Nowlin D, Yan J, Liu KK, Zhu J, Chen P, Deng YL, et al. Identification of Cys255 in HIF-1 $\alpha$  as a novel site for development of covalent inhibitors of HIF-1 $\alpha$ /ARNT PasB domain protein-protein interaction. *Protein Sci*. 2012; 21:1885–1896. [PubMed: 23033253]

- 38. Wilkins SE, Abboud MI, Hancock RL, Schofield CJ. Targeting Protein-Protein Interactions in the HIF System. *ChemMedChem*. 2016; 11:773–786. This review nicely summarized the known structures of HIF- $\alpha$  complexed with its interacting proteins, and the current drug discovery progress targeting these interactions for various therapeutic purposes. [PubMed: 26997519]
39. Huang P, Chandra V, Rastinejad F. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu Rev Physiol*. 2010; 72:247–272. [PubMed: 20148675]
40. Patard JJ, Rioux-Leclercq N, Fergelot P. Understanding the importance of smart drugs in renal cell carcinoma. *Eur Urol*. 2006; 49:633–643. [PubMed: 16481093]
41. Eltzschig HK, Bratton DL, Colgan SP. Targeting hypoxia signalling for the treatment of ischaemic and inflammatory diseases. *Nat Rev Drug Discov*. 2014; 13:852–869. [PubMed: 25359381]
42. Semenza GL. Hypoxia-inducible factor 1 and cardiovascular disease. *Annu Rev Physiol*. 2014; 76:39–56. [PubMed: 23988176]
43. Denison MS, Soshilov AA, He G, DeGroot DE, Zhao B. Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol Sci*. 2011; 124:1–22. [PubMed: 21908767]
44. Wu D, Nishimura N, Kuo V, Fiehn O, Shahbaz S, Van Winkle L, Matsumura F, Vogel CF. Activation of aryl hydrocarbon receptor induces vascular inflammation and promotes atherosclerosis in apolipoprotein E $^{-/-}$  mice. *Arterioscler Thromb Vasc Biol*. 2011; 31:1260–1267. [PubMed: 21441140]
45. Quintana FJ. The aryl hydrocarbon receptor: a molecular pathway for the environmental control of the immune response. *Immunology*. 2013; 138:183–189. [PubMed: 23190340]
46. Murray IA, Patterson AD, Perdew GH. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nat Rev Cancer*. 2014; 14:801–814. [PubMed: 25568920]
47. Kamnarsan D, Muir WJ, Ferguson-Smith MA, Cox DW. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. *J Med Genet*. 2003; 40:325–332. [PubMed: 12746393]
48. Pickard BS, Christoforou A, Thomson PA, Fawkes A, Evans KL, Morris SW, Porteous DJ, Blackwood DH, Muir WJ. Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. *Mol Psychiatry*. 2009; 14:874–884. [PubMed: 18317462]
49. Stanco A, Pla R, Vogt D, Chen Y, Mandal S, Walker J, Hunt RF, Lindtner S, Erdman CA, Pieper AA, et al. NPAS1 represses the generation of specific subtypes of cortical interneurons. *Neuron*. 2014; 84:940–953. [PubMed: 25467980]
50. Sun X, Lin Y. Npas4: Linking Neuronal Activity to Memory. *Trends Neurosci*. 2016; 39:264–275. [PubMed: 26987258]
51. Bonnefond A, Raimondo A, Stutzmann F, Ghoussaini M, Ramachandrapa S, Bersten DC, Durand E, Vatin V, Balkau B, Lantieri O, et al. Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. *J Clin Invest*. 2013; 123:3037–3041. [PubMed: 23778136]
52. Partonen T, Treutlein J, Alpmann A, Frank J, Johansson C, Depner M, Aron L, Rietschel M, Wellek S, Soronen P, et al. Three circadian clock genes Per2, Arntl, and Npas2 contribute to winter depression. *Ann Med*. 2007; 39:229–238. [PubMed: 17457720]
53. Gamble KL, Berry R, Frank SJ, Young ME. Circadian clock control of endocrine factors. *Nat Rev Endocrinol*. 2014; 10:466–475. [PubMed: 24863387]
54. Kelleher FC, Rao A, Maguire A. Circadian molecular clocks and cancer. *Cancer Lett*. 2014; 342:9–18. [PubMed: 24099911]

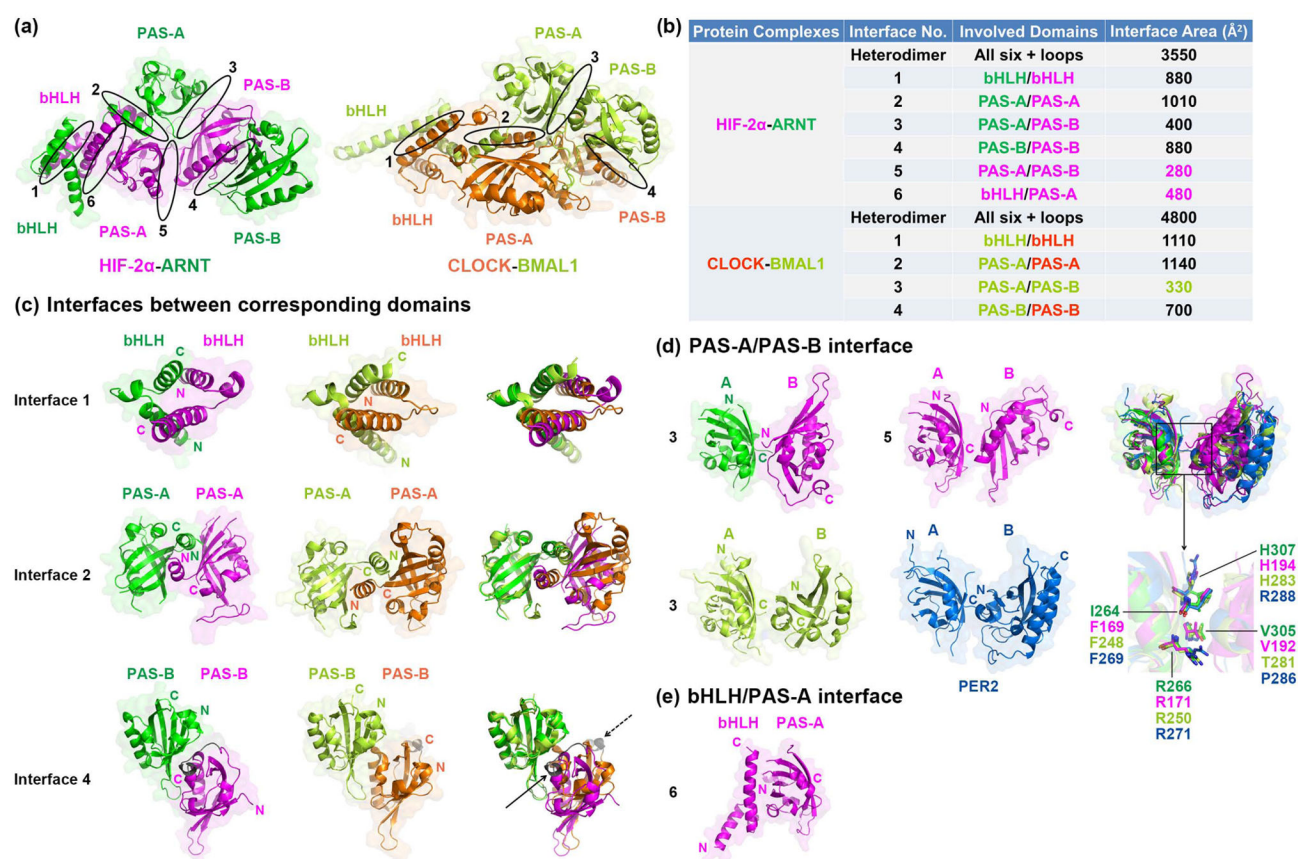
**Highlights**

- Mammalian bHLH-PAS proteins are heterodimeric transcription factors.
- Dimerization involves the bHLH, PAS-A and PAS-B domains.
- Crystal structures of heterodimeric complexes reveal asymmetric architectures.
- These proteins may also bind to distinct ligands through PAS domains.
- Their variable C-terminal regions form interactions with regulatory protein



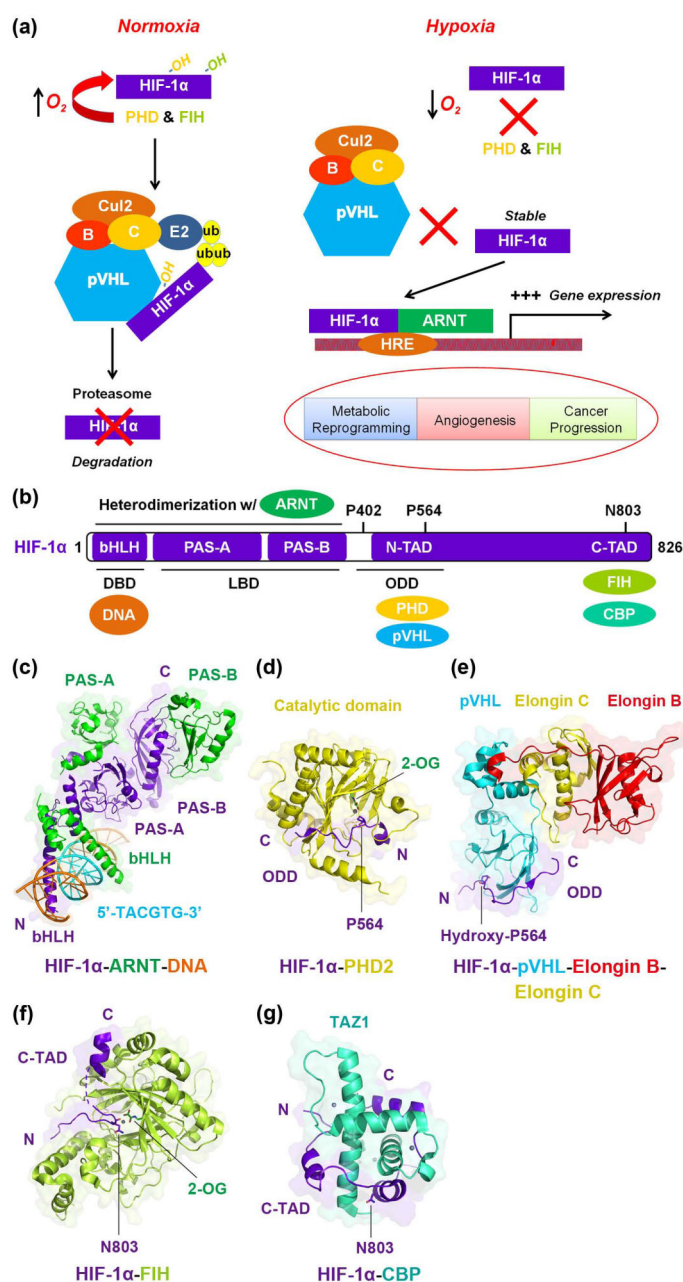
**Figure 1.**

The bHLH-PAS family of transcription factors. (a) Domain organization within the bHLH-PAS polypeptides. bHLH, basic helix-loop-helix; PAS, PER-ARNT-SIM; TAD, transactivation domain; TRD, transrepression domain. (b) Members of the mammalian bHLH-PAS family and their selective heterodimerization partners. (c) NMR solution structures of the single PAS-B domains of HIF-2 $\alpha$  (magenta, PDB: 1P97) and ARNT (green, PDB: 1x00). (d) Crystal structure of the PAS-B heterodimer formed by mutated HIF-2 $\alpha$  and ARNT (PDB: 3F1P). The two mutations (R247E and E362R) are shown and labeled in the figure, as well as the location of HIF-2 $\alpha$  cavity. (e) Crystal structure of the AHR PAS-A domain homodimer (brown, PDB: 4M4X). (f-h) Overall crystal structures of three complexes CLOCK-BMAL1 (PDB: 4F3L) (f), HIF-2 $\alpha$ -ARNT (PDB: 4ZP4) (g) and NPAS1-ARNT (h), with each of bHLH, PAS-A and PAS-B domains labeled. The colors used for CLOCK, BMAL1, HIF-2 $\alpha$ , ARNT and NPAS1 are orange, lemon, magenta, green and cyan, respectively.

**Figure 2.**

Analysis of domain-domain interfaces. (a) Overall structures of HIF-2 $\alpha$ -ARNT and CLOCK-BMAL1 complexes, with each of domains labelled and interfaces numbered. The colors used for HIF-2 $\alpha$ , ARNT, CLOCK and BMAL1 are magenta, green, orange and lemon, respectively. (b) A table summarizing the information for each domain interface about the composition and calculated interface area. The numbers in black color are for the heterodimer interfaces; and those in magenta and lemon are for the intramolecular domain interfaces of HIF-2 $\alpha$  and BMAL1, respectively. (c) Comparisons of the interfaces between corresponding domains (Interface 1 for bHLH/bHLH, 2 for PAS-A/PAS-A, and 4 for PAS-B/PAS-B) within two heterodimers, by the side-by-side displaying (left) and ARNT/BMAL1 alignment (right). The solid arrow shows position of the ARNT PAS-B C-terminal helix (colored in black), and the dashed one shows that of the CLOCK helix (gray). (d) Comparison of the PAS-A/PAS-B interfaces revealing a conserved position at two  $\beta$ -strands of the PAS-A domain ("patch") interacting with different positions at the PAS-B domains. ARNT, HIF-2 $\alpha$ , BMAL1 and PER2 (PDB: 3GDI) are colored in green, magenta, lemon and blue, respectively. The alignment of interfaces by the PAS-A domains is displayed on the right, with the enlarged figure below showing the four conserved residues on the "patch". (e) The unique bHLH/PAS-A interface of HIF-2 $\alpha$  within the HIF-2 $\alpha$ -ARNT complex.



**Figure 3.**

Functional interactions formed between HIF-1 $\alpha$  and other macromolecules. **(a)** The oxygen-dependent regulation of HIF-1 $\alpha$ -ARNT transactivation by pVHL-mediated HIF-1 $\alpha$  protein degradation. PHD, prolyl hydroxylase domain protein; FIH, Factor-inhibiting HIF; pVHL, the von Hippel-Lindau protein; ub, ubiquitin. **(b)** A diagram of HIF-1 $\alpha$  protein showing the function and interacting macromolecules (DNA or proteins) for each domain. DBD, DNA binding domain; LBD, ligand binding domain; ODD, oxygen-dependent degradation domain; TAD, transactivation domain. **(c)** Crystal structure of HIF-1 $\alpha$ -ARNT-DNA complex showing the binding and recognition of HRE site (5'-TACGTG-3', colored in cyan) on the



DNA by bHLH domains (PDB: 4ZPR). **(d)** Crystal structure of the catalytic domain of PHD2 complexed with the ODD region of HIF-1 $\alpha$  including the P564 residue (DPB: 3HQR). **(e)** Crystal structure of P564-hydroxylated HIF-1 $\alpha$  ODD peptide binding to the pVHL-Elongin B-Elongin C complex (PDB: 1LQB). **(f)** Crystal structure of FIH in complex with the C-terminal TAD fragment of HIF-1 $\alpha$  including the N803 residue (PDB: 1H2M). **(g)** NMR solution structure of the TAZ1 motif of CBP binding to the C-TAD peptide of HIF-1 $\alpha$  including the N803 residue (PDB: 1L8C). Colors used in the labels match those used in figures for the same components.

**Table 1**

The bHLH-PAS transcription factors and their links to human diseases

Proteins	Diseases
HIF-1 $\alpha$ , HIF-2 $\alpha$	cancer [28], VHL disease [40], anemia [41], heart failure [42]
AHR	toxicity response [43], inflammation [44], autoimmune diseases [45], cancer [46]
NPAS1, NPAS3	psychiatric disorders (schizophrenia [47], bipolar [48], autism [49])
NPAS4	contextual memory formation impairment [50]
SIM1	severe early-onset obesity [51]
SIM2	cancer [12]
CLOCK, NPAS2	seasonal depression [52], metabolic diseases [53], cancer [54]