COMMENTARY

Assembly and function of bHLH–PAS complexes

Jennifer L. Fribourgh[a] and Carrie L. Partch[a,b,1]

The basic helix–loop–helix-PER-ARNT-SIM (bHLH–PAS) family of transcription factors coordinates the expression of distinct transcriptional programs to control processes from development to the hypoxia response and beyond. Despite differences in their target genes and modes of regulation, these transcription factors share a common domain architecture, consisting of a bHLH DNA-binding domain followed by tandem PAS domains and intrinsically disordered C-terminal regulatory domains. In PNAS, Seok et al. present the structure of the core bHLH-PAS dimer of the aryl hydrocarbon receptor (AHR)–aryl hydrocarbon nuclear receptor translocator (ARNT) transcription factor bound to DNA (1). This study provides a foundation for understanding how AHR–ARNT specifically recognizes its consensus DNA motif and highlights how changes in interdomain contacts may communicate information about ligand binding to regulate subcellular localization and transcriptional activation.

The bHLH–PAS family is defined by formation of heterodimers comprising class I and class II subunits. Class I proteins are typically regulated by tissue or environmental-specific factors, whereas class II proteins are expressed ubiquitously. Examples of class I proteins include AHR (regulated by xenobiotics), hypoxia-inducible factor-α (HIF-α, regulated by hypoxia), neuronal PAS domain proteins (NPAS, developmentally regulated), and circadian locomotor output cycles protein kaput (CLOCK, circadian rhythms). ARNT is the predominant class II subunit found in bHLH–PAS complexes, whereas the related protein brain and muscle ARNT-like 1 (BMAL1) appears to be primarily dedicated to CLOCK to regulate circadian rhythms. Although cellular studies have largely outlined the different regulatory mechanisms that control heterodimerization, subcellular localization, and activity of these complexes, the structural basis for their assembly and diverse functions has been unclear.

The structure of AHR–ARNT by Seok et al. (1) adds to a recent bounty of bHLH–PAS structures, providing several key insights that address these fundamental questions. As the newest representative of ARNT-containing bHLH–PAS complexes (2–4), this structure solidifies earlier observations noting a similar global architecture among heterodimers that share an ARNT subunit (Fig. 1). Notably, the PAS-A domains of AHR, NPAS3, and HIF1-α all make direct contacts with their own bHLH domains that tether the N-terminal PAS domain in close proximity to DNA. It appears that ARNT and BMAL1 confer distinct differences in the global architecture of the bHLH–PAS heterodimers via the spatial arrangement of their PAS domains (Fig. 1). Specifically, the PAS-A and PAS-B domains of ARNT are completely separated in space from one other, whereas the PAS domains of BMAL1 make numerous interdomain contacts that orient the CLOCK–BMAL1 heterodimer more linearly away from DNA (2, 5). Although the structure of AHR–ARNT by Seok et al. (1) lacks the PAS-B domains, its similarity in packing of the AHR and ARNT PAS-A domains suggests that the overall architecture of the heterodimer will be similar to NPAS1/3–ARNT and HIF1/2-α–ARNT complexes (2–4). Therefore, it appears that differences in both class I and class II subunits lead to two distinct global architectures for the bHLH–PAS family. Given the recent discovery of cross-talk between AHR (6) and HIF-α (7–9) with BMAL1, it will be interesting to see which structural cues predominate in the architecture of these "mixed" bHLH–PAS heterodimers, or whether they have an entirely new global architecture.

Understanding subtle structural differences in the modes of DNA recognition by members of the bHLH–PAS family is crucial because of the vastly different transcriptional programs they regulate. In general terms, a bHLH domain contains four to six basic amino acids that are used to bind DNA, located within a HLH dimerization domain. How is specificity conferred by these transcription factors that have DNA-binding domains that are essentially structurally identical? Presumably, subtle sequence variations in the bHLH domain allow for selection of specific DNA consensus sequences. The study by Seok et al. (1) provides important insight into how AHR–ARNT selectively recognizes the dioxin-response element (DRE) over the closely related hypoxia-response element (HRE). The authors observed that residue R39 of AHR makes numerous contacts with the GC sequence in the canonical DRE, TTGCGTG, whereas ARNT recognizes the

[a]Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 96054; and [b]Center for Circadian Biology, University of California, San Diego, CA 92161

Author contributions: J.L.F. and C.L.P. wrote the paper.

The authors declare no conflict of interest.

See companion article on page 5431.

[1]To whom correspondence should be addressed. Email: cpartch@ucsc.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1705408114

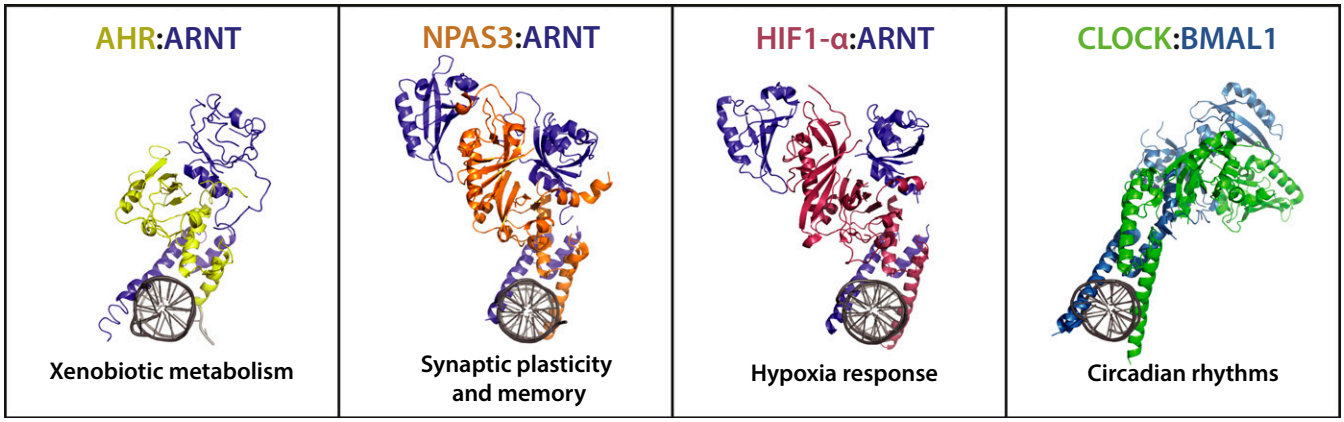

Fig. 1. The global architecture of bHLH–PAS heterodimers. From *Left* to *Right*, the structure of ARNT-containing complexes: AHR–ARNT bound to DNA encoding the DRE (PDB ID code 5V0L), NPAS3–ARNT bound to DNA encoding the HRE (PDB ID code 5SY7), and HIF1-α–ARNT bound to the HRE (PDB ID code HZPR). On the far right, the CLOCK–BMAL1 bHLH–PAS structure (PDB ID code 4F3L) is aligned to the bHLH domain of the CLOCK–BMAL1 bHLH–E-box DNA structure (PDB ID code 4H10).

GTG half-site (1). Correspondingly, mutation of R39 eliminated high-affinity recognition of the DRE, highlighting subtle changes in the bHLH domain that encode for specific recognition of consensus motifs. Sequences outside the bHLH may also play a role in DNA recognition for some bHLH–PAS complexes: the structure of HIF2-α–ARNT bound to DNA depicts contacts between the PAS-A domain of HIF2-α that appear to expand its recognition of DNA beyond the consensus HRE motif (3). Therefore, changes in the global architecture of bHLH and PAS domains across the family could play an underappreciated role in the recognition of DNA and gene regulation. Coming to a better understanding of this will require more DNA-bound structures at higher resolution to resolve potential functional roles for the PAS-A domain in DNA recognition.

AHR has a special role within the bHLH–PAS family as a sensor of chemically diverse xenobiotics and endogenous ligands (10). The AHR PAS-B binds these compounds directly to regulate nuclear entry and association with ARNT, but the partially structured nature of this domain in its unliganded state has precluded its structure determination thus far. Perhaps coexpression of AHR and ARNT in the presence of xenobiotic analogs might allow determination of this holy grail of bHLH–PAS complex structures. Notably, ligand binding by AHR leads to release of its molecular chaperones and exposure of a nuclear localization signal in the bHLH domain; upon nuclear entry, AHR heterodimerizes with ARNT to fulfill its transcriptional obligation. To explore the structural basis for this potential communication between the PAS and bHLH domains, Seok et al. (1) explored the importance of interdomain interactions observed in their structure. Notably, perturbation of the bHLH–PAS-A interface in AHR appears to disconnect ligand binding from exposure of the nuclear localization signal, allowing for constitutive nuclear localization independent of ligand binding in vivo (1). This finding suggests that allosteric channels between these domains in AHR are important for ligand sensing and nuclear localization.

Currently, AHR is the only mammalian bHLH–PAS transcription factor known to be regulated by binding of endogenous ligands, although at least one bHLH–PAS protein in insects binds developmentally regulated hormones (11). Does AHR-like ligand binding by PAS domains regulate other members of this protein family? Leveraging the existence of buried, hydrophilic cavities in PAS-B domains, Gardner and colleagues identified a series of small molecules that selectively target the HIF2-α or ARNT PAS-B domain to regulate heterodimer formation (12–14) or recruitment of coactivator proteins (15, 16). A careful analysis of PAS domains in CLOCK, BMAL1, NPAS1, NPAS3, ARNT, HIF1-α, and HIF2-α identified potential ligand-binding pockets in both PAS-A and PAS-B domains, suggesting the broad potential for ligand binding in these transcription factors (4). Advances in the structural biology of a native, ligand-binding bHLH–PAS complex, AHR–ARNT, by Seok et al. (1) brings us one step closer to understanding native ligand binding by PAS domains and highlights the potential for discovery of ligands that similarly regulate other bHLH–PAS transcription factors.

Acknowledgments

This work was funded by National Institutes of Health Grants GM107069 and GM121507 (to C.L.P.), and a University of California, Santa Cruz Chancellor's Postdoctoral fellowship (to J.L.F.).

1 Seok S-H, et al. (2017) Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proc Natl Acad Sci USA* 114:5431–5436.
2 Huang N, et al. (2012) Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* 337:189–194.
3 Wu D, Potluri N, Lu J, Kim Y, Rastinejad F (2015) Structural integration in hypoxia-inducible factors. *Nature* 524:303–308.
4 Wu D, Su X, Potluri N, Kim Y, Rastinejad F (2016) NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. *eLife* 5:5.
5 Wang Z, Wu Y, Li L, Su XD (2013) Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. *Cell Res* 23:213–224.
6 Xu CX, Krager SL, Liao DF, Tischkau SA (2010) Disruption of CLOCK-BMAL1 transcriptional activity is responsible for aryl hydrocarbon receptor-mediated regulation of Period1 gene. *Toxicol Sci* 115:98–108.
7 Adamovich Y, Ladeuix B, Golik M, Koeners MP, Asher G (2017) Rhythmic oxygen levels reset circadian clocks through HIF1α. *Cell Metab* 25:93–101.
8 Peek CB, et al. (2017) Circadian clock interaction with HIF1α mediates oxygenic metabolism and anaerobic glycolysis in skeletal muscle. *Cell Metab* 25:86–92.
9 Wu Y, et al. (2017) Reciprocal regulation between the circadian clock and hypoxia signaling at the genome level in mammals. *Cell Metab* 25:73–85.

10 Stockinger B, Di Meglio P, Gialitakis M, Duarte JH (2014) The aryl hydrocarbon receptor: Multitasking in the immune system. *Annu Rev Immunol* 32:403–432.
11 Charles JP, et al. (2011) Ligand-binding properties of a juvenile hormone receptor, Methoprene-tolerant. *Proc Natl Acad Sci USA* 108:21128–21133.
12 Chen W, et al. (2016) Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* 539:112–117.
13 Key J, Scheuermann TH, Anderson PC, Daggett V, Gardner KH (2009) Principles of ligand binding within a completely buried cavity in HIF2alpha PAS-B. *J Am Chem Soc* 131:17647–17654.
14 Scheuermann TH, et al. (2009) Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor. *Proc Natl Acad Sci USA* 106:450–455.
15 Guo Y, et al. (2013) Regulating the ARNT/TACC3 axis: Multiple approaches to manipulating protein/protein interactions with small molecules. *ACS Chem Biol* 8:626–635.
16 Partch CL, Gardner KH (2011) Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proc Natl Acad Sci USA* 108:7739–7744.