Modeling PAS sensors through *in silico* and genetic approaches

By

Brenda Liset Rojas

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Molecular and Environmental Toxicology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: January 14, 2022

The dissertation is approved by the following members of the Final Oral Committee:

Christopher Bradfield, Professor, Oncology

Sean Ronnekleiv-Kelly, Assistant Professor, Department of Surgery

Jae-Hyuk Yu, Professor, Bacteriology and Genetics

John Hogenesch, Professor, Pediatrics at Cincinnati Children's Hospital Medical Center

ACKNOWLEDGEMENTS

None of this would be possible without the help of many people. In this very long list, I must start by thanking my husband, Alejandro Herrera, who has always believed in me as person and a scientist. Thank you for all the sacrifices you've made to help my dream come true. Thank you for always sticking by me even when I'm driving you crazy. Thank you for our wonderful little family and for being the strength when I can't be.

I need to thank my mentor and advisor, Dr. Christopher Bradfield, whom I affectionately call jefe (translates to boss in Spanish). Thank you for listening to my ideas, for letting me fail and learn on my own throughout this process. I will be grateful to you always for believing in me and never letting me quit when my life was in shambles. Most importantly, thank you for giving me the time I needed to be with my son and grieve his loss. Such a simple gesture allowed me to get back up and keep moving forward.

Everyone in the Bradfield Lab also deserves acknowledgement. Individually, I thank Dr. Emmanuel Vazquez-Rivera, Dr. Morgan Walcheck, Dr. Rachel Wilson, Jessica Parrot, and Dr. Anna Shen. Thank you all for everything you do every day – from helping me order materials to watching my dog, Chula and everything in between. I would not be here without your support and friendship. To Dr. Amarilys Gonzalez-Vazquez – thank you for your unconditional love. I am forever grateful to and for you. To Mark Marohl, I appreciate your guidance, help, and support through this process and during the difficulties of the past year. I consider you all my Madison family.

To my family and friends in Los Angeles, I appreciate your help, love, and support during this process but also when we lost Lucas. To my mom, Elizabeth Garduno – I know you still think I majored in nutrition, but I promise you I didn't. I still can't tell you what to eat. This year was as difficult for you as it was for me, but we will get through this together. To my siblings,

Xavier, Iris, Yesenia, and Leslie – thank you! Special thank you to Juan Vazquez, Gilbert Espino, and Salvador Medina for their continued support. To my nieces and nephews – Arianna, Sebastian, Vincent, Alyssa, Adrian, Isabella, and Lillard (Lily) – I want you to know that you can make any dream possible, and I will always be there to make you snacks or help you with your homework. To my in-laws: Leyla, Jaime, Patty, Marcos, Janett, and Alberto: thank you for always supporting my dream.

In addition to my Madison family, I also have my Cal State LA family. Thank you to Catalina Guasco, Grace Perez, Jessica Martinez, Giovanna Agredano, Dr. Sunil Mangalassary, and my academic mother – Vicki Kubo-Anderson. Thank you for supporting my dream to go to UW-Madison, recognizing my potential, and believing in me.

This list would not be complete without my best friends Evelyn Noyola-Lugo and Rebecca Horta. We've been friends for over 25 years and even when we go months without talking, our conversations pick right back up as though time has not passed. I love you girls and I celebrate this achievement with you. I'm grateful for you and your families, who also supported this crazy dream, and who love our fat-fat as much as we do. All that's left to say is: Go DODGERS!

DEDICATION

This accomplishment is dedicated to my angel son in heaven, Lucas C. Herrera, and my abuelita, Quintila Vela Bello. I hope you are as proud of me as I am of you. Losing you has been the most difficult part of this journey but wherever you are, Mommy and Daddy love you eternally. I miss you every day, fat-fat. Te extrano abuelita.

ABSTRACT

The PAS family of transcription factors play vital roles in important physiological processes. Within this family, PAS sensors, with their characteristic domain, are of interest because they modulate circadian rhythms, xenobiotic metabolism, and oxygen homeostasis, in addition to other functions. The PAS domain was named from the three founding members – Period, Aryl hydrocarbon nuclear translocator, and Sim. Structurally, the domain has been labeled in many ways, however we stick to the PAS A and B nomenclature Our central hypothesis is that PAS proteins are part of a network that coordinate physiological responses to stressors through obscure interactions. To test our hypothesis, we worked on the following aims: (1) To develop a lexicon for the structures partaking in PAS-PAS dimerization, we mined the Protein Data Bank for mammalian proteins. This information helped us refine the terminology and showcase differences between PAS A-PAS A and PAS B-PAS B repeat interactions. Moreover, it taught us that PAS proteins interact with other regions outside their PAS counterpart, thus PAS A-PAS B repeat interconnectivity is common. (2) To determine the ability of our yeast system, based on *Saccharomyces cerevisiae* strain L40, to accommodate steroid receptor signaling, we introduced and activated the mineralocorticoid, androgen, glucocorticoid, and estrogen α receptors individually and exposed them to various ligands. Results from these experiments argue for the possibility of turning our L40 system into a “one-cell system” that can accommodate multiple pathways for complex receptor mixtures. (3) To determine dimerization between PAS B repeats in the prototypical AhR:ARNT complex, we again used our L40 system. We found that PAS B repeats are necessary for successful interconnectivity. Additionally, we took advantage of the new AlphaFold and SwarmDock algorithms to predict PAS B interactions between the AhR and ARA9 and ARA3. This work is significant because it also has the potential to detect biomarkers of disease progression that can function as targets for therapeutic treatments and drugs

CONTENTS

Ch. 1: Introduction- Building a meta detection system for biologically active compounds using the yeast *Saccharomyces cerevisiae* L40

Receptors exploit dimerization to partner with members of their own families, or other families, as homo or heterodimers that form cellular-response networks. Dimerization permits: (1) increased protein stability, (2) increased complexity to avoid unwanted interactions, and (3) increased binding affinity and specificity for DNA sequences that encode pertinent target genes. To fully understand these cellular-response networks, we began focusing on interactions within and between two major families: PAS sensors and the steroid receptors. The PAS sensors arise from sequences that are of ancient origin and are found in all kingdoms of life. In animals, PAS domains are found a variety of structures regulating muscular tissue (myogenesis) and neuron (neurogenesis) formation during embryonic development, production of red blood cells and platelets in the bone marrow (hematopoiesis), and sleep-wake cycles (circadian rhythms). Similarly, steroid receptors regulate keen processes related to metabolism, sex differentiation, and growth. This family includes estrogen α, glucocorticoid, mineralocorticoid, and androgen receptors, among others.

Our motivation, to better understand the interaction within and between these protein families, was driven by an interest in employing these sensors and receptors to detect and quantify important biologically active compounds found in several important matrices. In the PAS sensor family, only the ligand binding pathway of the aryl hydrocarbon receptor (AhR) has been studied as a biosensor. While some steroid receptors, such as ER, have been proposed as biosensors, their application has also not yet been realized. We define biologically active compounds as chemicals that exert a physiological effect beyond their nutritional value. The presence of unwanted biologically active compounds is a concern in foods, personal care products, drugs, and supplements. In this regard, endocrine disrupting chemicals are the largest and best-known class because they target sensitive steroid receptor pathways thus interfering with important biological processes. The endocrine disruptor class includes intentional and incidental chemicals such as preservatives (butylated hydroxyanisole, parabens), fragrances, and plasticizers (phthalates), often present in small amounts.

Although effects on steroid receptor pathways are extensively studied, chemicals that interact with the AhR are also contaminants of concern and have been linked to pathologies and/or cancers. For example, less refined mineral oils and paraffins may harbor polyaromatic hydrocarbons (PAHs) that activate the AhR pathway. Furthermore, there is growing evidence that some drugs may affect sleep/wake cycles, regulated by the *CLOCK* gene, another remember of the PAS sensor family. Moreover, plants and microbes harbor additional PAS sensor proteins with a wide spectrum of stimuli that may harbor potential as regulatory concern or as potentially commercially valuable compounds[1]. Therefore, there is a need to identify the multiplicity of PAS pathways with potential as valuable biosensors and better understand the mechanisms by which they carry out the physiological response.

The goal for this dissertation is to provide the background information that can ultimately lead to the development of a one-cell biosensor system that can accommodate multiple PAS stimuli and nuclear receptor ligands. The foundation of this system was pioneered almost twenty year ago, and is based on a yeast expression system, built on sensor chimeras with the DNA binding domain of the LexA protein, and read-outs resulting from sensor dependent activation of a reporter gene (beta galactosidase, β-gal) controlled by the Lex operator (LexO)[2,3]. Because of the ambitious nature of this goal, our initial focus was on elucidating the potential of such a system and identifying its limitations. Therefore, in this thesis, we focused on pathways we perceived as holding the greatest utility as biosensors, i.e., the PAS sensor, AhR, and the steroid receptors: estrogen α (ER), mineralocorticoid (MR), glucocorticoid (GR), and androgen (AR). These pathways were chosen based upon their perceived utility and the preliminary data from this laboratory that they are amenable to a single cell biosensor system.

Our initial objectives were multiple and included: 1) Gaining insight into the breadth of bioactive molecules that could be detected with a spectrum of sensors and that could be functional in a yeast/LexA dependent system; and 2) Understanding the potential for pathway interference by understanding the domain structure and interaction surfaces of PAS proteins.

We propose that this understanding will guide us through future attempts to develop a meta-analytical system that can detect multiple classes of biologically active molecules in one tube systems. Such efforts will require an informed decision as to which strategy is best: to develop specific yeast strains for a range of individual pathways or to use a singular yeast with all the pathways of interest co-expressed in a single clone. Not surprisingly, while this scientific decision provided initial guidance to our experiments, our investigations took some interesting turns which led to several exciting observations regarding PAS protein domain structure and their interactions.

In **Chapter 2**, we ask a basic question: how many distinct steroid receptor pathways can be successfully employed in our LexA-yeast system? Put another way, what is the potential of this system to interrogate multiple biologically active classes? To answer this question, we employ our yeast based AhR ligand detection system as a positive control and use it as a model to expand the concept to four important members of the steroid receptor superfamily. The rationale for this chapter was to demonstrate that a broad range of biologically active compounds, endocrine disrupting chemicals, could be detected using our yeast-based system. To this end, we successfully developed a system where the full-length steroid receptors: ERα, MR, GR, and AR were each fused to an *Escherichia coli* domain analogous to a DNA binding domain known as LexA. Each fused protein was exposed to various ligands and 3 out 4 were able to detect compounds successfully. In addition, we used AlphaFold, the recently developed protein prediction algorithm from Deep Mind, to predict the structure of our fused AhR and steroid receptors. The data from this chapter shows that our yeast system can accommodate steroid receptors and PAS protein signaling while the predicted structures help us think about how the system may be modified to improve future results.

In **Chapter 3**, we attempt to better understand the PAS domain to understand biosensor design and thus reduce potential crosstalk across PAS signaling molecules that might be co-expressed in a single cell. To accomplish this task, we take advantage of experimental PAS

structures deposited into the Protein Data Bank (PDB) to develop a lexicon and clarify domain interactions that have been observed over the past decades. To reduce confusion in the field, we develop formal definitions for common features of PAS domains, such as PAS domain and PAS repeat. We also demonstrate that the PAS family of proteins showcase at least 4 different dimerization modes, some of which are supported by structures outside the PAS domain. We provide evidence to support a novel finding, that PAS A repeat interactions are different than those observed in PAS B repeats: for example a loop in the PAS A repeat in the HIF2α:ARNT dimer establishes contacts with DNA, while PAS A and B repeats within HIF2α interconnect. This experimental evidence challenges data that supports a “like interacts with like” ideology whereby certain parts of PAS proteins interact exclusively with their counterparts in other PAS proteins (i.e. PAS A only interacts with PAS A).

In **Chapter 4**, we continue developing our models of PAS domain interactions using our yeast system as a model. We focus on the familial PAS prototype dimer, the AhR:ARNT (aryl hydrocarbon nuclear translocator) and determine the sufficiency of the PAS B repeat in dimerization. Using the yeast two-hybrid, a technique that relies on the reconstitution of a transcription factor to determine physical interactions between two proteins, we show for the first time that both the AhR and ARNT use PAS B repeats to dimerize. We further investigate the role of PAS B in interconnectivity with other proteins involved in AhR signaling, such as ARA9 and ARA3, by taking AlphaFold’s predicted proteins and docking them using the SwarmDock server. This chapter is important because it serves as a baseline to inform how the AhR uses its PAS B repeat for more than ligand binding. Additionally, we use these ideas to develop a new AhR signaling pathway visual that underscores the importance of domain interactions.

A conclusion of this work is that while considerable additional development is necessary, a one-cell yeast system shows promise as a biosensor for activators of these two major nuclear receptor families: PAS and steroid. Through these investigations, we also provide information that facilitates the prediction of unknown interactions that may guide future optimization studies

and will lead to improved understanding for complex receptor mixture responses – where more than one pathway triggers a physiological response.

REFERENCES

1 Vazquez-Rivera, E. *et al.* The aryl hydrocarbon receptor as a model PAS sensor. *Toxicology Reports* **9**, 1-11, doi:https://doi.org/10.1016/j.toxrep.2021.11.017 (2022).

2 LaPres, J. J., Glover, E., Dunham, E. E., Bunger, M. K. & Bradfield, C. A. ARA9 modifies agonist signaling through an increase in cytosolic aryl hydrocarbon receptor. *Journal of Biological Chemistry* **275**, 6153-6159, doi:10.1074/jbc.275.9.6153 (2000).

3 Hogenesch, J. B. *et al.* 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**, 8581-8593 (1997).

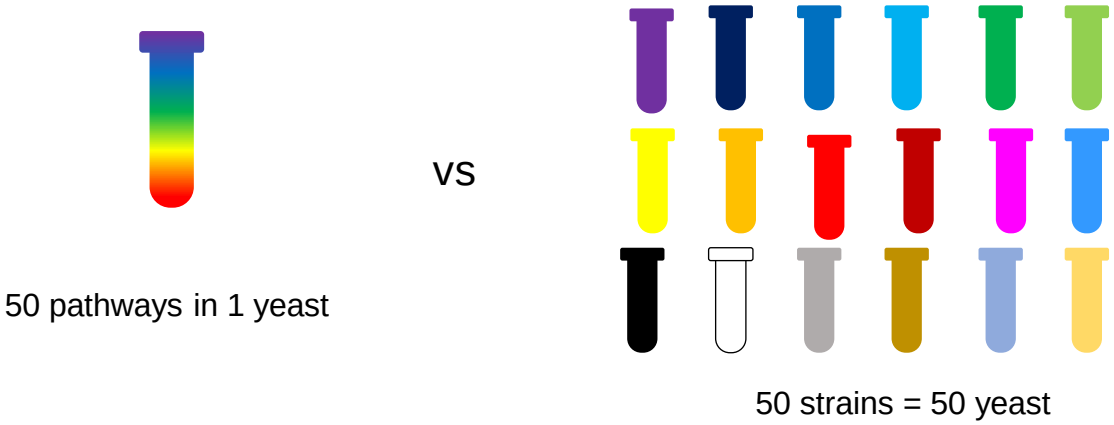


Figure 1: The framework for this work. The purpose of this work is to understand PAS protein interactions to develop a yeast-based system that can accommodate many pathways in 1 yeast or many strains, giving rise to individual pathways. To understand these interactions, we must understand which regions on PAS proteins are interacting and their arrangements.

Ch. 2: Development of yeast-based steroid receptor biosensors

INTRODUCTION

Steroid receptors were first described in the 1970s at the interface for nutrition and medicine, where it was postulated that chemical messengers, referred to as hormones, were responsible for a host of physiological responses[1]. The steroid receptor superfamily is comprised of proteins that detect ligands, such as steroids, drugs, and other lipophilic compounds resulting in signal transduction and playing major roles in sex differentiation, growth, and metabolism. The estrogen receptor alpha (ERα), glucocorticoid (GR), androgen (AR), and mineralocorticoid (MR) receptors are a few of the most well understood members of in this superfamily. As a result of their varied but essential roles in physiology, they are subject to interference from endocrine disrupting chemicals – chemicals that hinder the communication and signaling processes for steroid receptors leading to adverse health outcomes.[2]

Breakthrough publications by the Ptashne group[3] and Fields and Song[4] cemented yeast as a model organism that could be used to study protein domains and protein-protein interactions. Compared to the use of animals and cell lines, yeast was relatively inexpensive to manipulate and easy to grow. Additionally, human genes could also be introduced and modified using homologous recombination and versatile expression and reporter plasmids, as well as yeast chromosomes. Thus, yeast was an important component that led to the birth of synthetic biology, whereby fine-tuning the expression of a protein of interest at a specific step in a specific pathway can be used to optimally control a process that yields a specific product.

Perhaps the best known use of yeast has been in the production of fermented foods including beer and wine[5] where enzymes involved in alcohol biosynthesis have been modified to impart or modify flavors or other sensorial qualities that affect consumer acceptance[6]. In yet another example, yeast has also been employed in the biosynthesis of cannabinoids by introducing a hexanoyl-CoA pathway while simultaneously exploiting the native mevalonate

pathway – a metabolic pathway found in eukaryotes that metabolizes acetyl-CoA into isoprenoids, later converted to cholesterol, steroids, etc.[7].

All these benefits have facilitated the introduction of synthetic transcription factors[8] like steroid and other nuclear receptors, into yeast to study the delicate interplay between gene structure and protein function[9]. Furthermore, yeast holds advantages for the industrial and environmental communities because it can be employed as a biosensor able to detect and quantify compounds through reporter systems and chimeras. Thus, yeast holds great potential as a tool for the development of mechanistic bioassays for endocrine disrupting chemicals[10-12].

Recently, an Expert Consensus statement[2] was published identifying 10 key characteristics that identify endocrine disrupting chemicals. The group highlighted the importance of developing reliable bioassays for each key characteristic. Such consensus statements point to the importance of validation, reproducibility, and sensitivity in bioassays design. The four-parameter logistic curve (4PL) is commonly used to analyze the pharmacological properties of a ligand and the robustness of a system. Derived from classical receptor theory, the 4PL model is a powerful way to characterize nuclear receptor signaling. Upon ligand binding, the receptor complex generates a pharmacological response where the ligand activates or antagonizes the receptor[13]. As the name implies, four parameters are calculated from a quantified dose-response relationship – most commonly a semi-logarithmic plot with a sigmoidal S shape. The four parameters are: (1) “bottom” aka E_0 describes the baseline or zero-dose response, (2) EC_{50} showing the concentration needed to effect 50% of the maximum response, (3) Hillslope which depicts the “steepness” of the most linear portion of the S shape and attempts to account for multiple binding sites, and (4) “top” aka $E_{max,}$ showcasing the maximal effect of the ligand (also known as efficacy)[13].

We set out to survey how readily nuclear receptors could be converted into yeast-based biosensors. Therefore, we chose four model receptors and introduced the GR, ERα, AR, and MR into the yeast *Saccharomyces cerevisiae* L40 and exposed them to various ligands. We

hypothesized that we could generate dose-response relationships that were comparable in utility and sensitivity to the LexA- mouse AhR^{b1} fusion that is ligand inducible and reproducible.

MATERIALS AND METHODS

Plasmids: Mouse ERα (BAJ65337.1), mouse GR (NP_001191190.1), human MR (NP_000892.2), and mouse AR (NP_038504.1) were synthesized by GeneUniversal (Newark, DE) and VectorBuilder (Chicago, IL) and cloned into plasmid 535 (pBTM116)[14] giving rise to yeast expression plasmids 2275 (LexA-mERα), 2293 (LexA-mGR), 2294 (LexA-huMR), and 2295 (LexA-mAR) (Figure 1) . The plasmid pBTM116 is approximately 5685 bp, contains the LexA (1-202) domain and is followed by a multiple cloning site. It also harbors a *TRP1* gene and a high-copy 2μ origin of replication[14]. Mouse ERα, mouse GR, human MR, and mouse AR were each inserted into restriction sites *EcoRI* and *BamHI*, to ensure that each was in-frame with LexA (Figure 1). The positive control for these experiments, plasmid 703 (LexA-NΔ166AhR), has been described previously[15].

L40: The yeast *Saccharomyces cerevisiae* reporter strain L40[14] ATCC MYA-3332 *(MATa ade2 his3 leu2 trp1 LYS::lexA-HIS3 URA3::lexA-LacZ)* was streaked and grown for 3 days in YPD agar (Takara Bio, San Jose, CA) plates with ampicillin, a complete medium, at 30 ˚C. One colony was selected and grown overnight in 10 mL, pH adjusted to 5.8, YPD broth with ampicillin (Takara Bio, San Jose, CA) at 30 ˚C and 220 RPM. After the allotted time, each expression plasmid was transformed into competent L40 cells using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA). To make competent yeast cells: the cells were centrifuged, and the supernatant was discarded. The pellet was washed with 10 mL of the Frozen-EZ Yeast Solution-1 and subjected to centrifugation. The supernatant was discarded and the pellet was resuspended in 1 mL of the Frozen-EZ Yeast Solution-2. To transform yeast: 50 µL of competent cells were mixed with 0.2-1 µg yeast expression plasmids 2275, 2294 or 2295 and 500 µL of Frozen-EZ Yeast Solution-3 was added and mixed thoroughly. The mixture was incubated in a water bath at 30 ˚C for 45 minutes and mixed every 15 minutes. One hundred microliters of the transformation mixture was spread on to synthetic medium lacking

tryptophan (Takara Bio, San Jose, CA) and containing ampicillin. The plates were incubated for up to 4 days at 30 ˚C to allow transformant growth.

Exposure to compounds: Ten micromolar concentrations in dimethylsulfoxide (DMSO) (Sigma Aldrich, Burlington, MA) served as the stock solution for most ligands: 17β-estradiol (Sigma Aldrich, Burlington, MA), genistein (Thermo Fisher Scientific, Waltham, MA), bisphenol A (Thermo Fisher Scientific, Waltham, MA), dihydrotestosterone (Sigma Aldrich, Burlington, MA), aldosterone (Sigma Aldrich, Burlington, MA), and β-napthoflavone (Sigma Aldrich, Burlington, MA). One transformant colony, representative of each transformation, was grown overnight in synthetic medium lacking tryptophan broth (Takara Bio, San Jose, CA) and containing ampicillin at 30 ˚C and 210 RPM. One hundred microliters were pipetted into clear 96 well plates and the OD_{600} was measured and adjusted to ~0.50 using the ClarioStar Plus plate reader (BMG Labtech, Cary, NC). Stock solutions for each compound were then used to create dose-response curves. Five microliters of each solution at the appropriate concentrations were dispensed into white 96 well plates using the Tecan Liquid Handler (Tecan, Morrisville, NC). Ninety-five microliters of OD_{600} adjusted cells were pipetted on to the same white plate, covered with a clear lid, wrapped in aluminum foil, and incubated at 30 ˚C and 210 RPM for 2 hours.

Measuring β-galactosidase activity: After the allotted time, yeast cells were lysed and chemiluminescent activity from the β-galactosidase enzyme was measured using the Gal-Screen assay (Thermo Fisher Scientific, Waltham, MA). The Gal-Screen Reaction Buffer was made as follows: the Gal-Screen substrate was diluted 1:25 with the Gal-Screen Buffer-B. One hundred microliters of the reaction buffer were dispensed to each well. The plate was covered in foil and incubated at 26-28 ˚C and 210 RPM for ~ 60-90 minutes. The plate was then placed in the ClarioStar Plus plate reader and values were measured using the luminometer reader function and the MARS data analysis software (BMG Labtech, Cary, NC).

Data Analysis: All protocols were performed with at least 3 biological and 3 technical replicates. The data from MARS was analyzed using PRISM GraphPad v9.2.0 (San Diego, CA) using the four-parameter logistic curve (4PL) and a Levy-Jennings quality control graph to measure precision throughout runs. The limit of detection (LOD) was calculated from the 4PL as has been described previously[16].

For normalized results, the highest raw value was designated as the top, while the lowest raw value was designated as the bottom. From there, all responses are calculated as a percentage of the highest value and described as % β-galactosidase relative light units (RLU). For parallel-line analysis (PLA) for each environmental estrogen, the 4PL Hillslope was compared. GraphPad assigns a p-value: the lower the p-value the more different the slopes are.

3-D protein predictions: Amino acid sequences for LexA-steroid receptor fusions were predicted using the AlphaFold server[17]. Since full-length steroid receptors have not been resolved, we exploited the deep learning properties of the AlphaFold algorithm which uses homologous templates and multiple sequence alignments. Models were then visualized using PyMOL v2.5 (Schrodinger, New York, NY).

RESULTS AND DISCUSSION

A major objective of this study was to examine the potential of a yeast based chimeric system to detect a broad range of biologically active compounds –in parallel- using mammalian sensor proteins of the steroid receptor family, as a model. As outlined in Chapter 1, to develop a meta-analysis system capable of detecting dozens of classes of biologically active agents, we envision two basic assay scenarios. The first, is an approach where we establish multiple assay systems based on unique sensors for specific chemical targets and run them in parallel (i.e., one sample is interrogated with ER, AR, MR, AhR, etc. systems independently). While such an approach requires greater costs and labor due to the multiplicity of strains employed and independence of each assay, it could allow immediate determination of the class of target being activated (e.g., ER vs AR). The second approach is one where multiple targets are interrogated at once. Here we imagine multiple co-expressed sensor chimeras as LexA-fusions and a single reporter driven by Lex operators (LexO). In this scenario, a variety of biologically active classes of compounds could be rapidly detected in a single strain. The primary shortcoming of this system being that all positive samples could require a follow-up with a second assay that discriminates the identity of the specific pathway at play. Another potential shortcoming of this co-expression approach is that pathways may interact and confound results. In this regard, the AhR (the Ah receptor) and ERα have been shown to physically interact and such contacts may interfere with assay information gain. While interactions between sensor families will ultimately require identification empirically, Chapter 3 begins our approach down that experimental path.

These ideas are based on studies from the early expression studies of the PAS protein sensor, AhR, in yeast; a chimera of AhR and LexA was shown to robustly reproduce the pharmacology of the mammalian receptor[18]. Thus, the positive control for these experiments is a LexA fused-N-terminal deletion of mouse AhR^{b1} (NΔ166-mAhR) exposed to βNF at various concentrations (**Figure 3**). This construct was chosen because it also reflects a potential

direction for sensor development, as it is an attempt at a minimal domain construct, employed to reduce the number of domains present in the sensor while also minimizing any interactions or cross talk across systems. Simply, this construct was engineered several years ago with the purpose of understanding AhR signaling independent of heterodimerization with other bHLH-PAS proteins through its PAS A repeat or basic helix loop helix (bHLH) domains[15]. At the time of its construction, we thought of interactions between PAS proteins as requiring HLH and PAS domains, yet we now know most PAS proteins employ an A'α helix – a region N-terminus to PAS A – to stabilize PAS A-PAS A interactions[19-22]. This domain is also absent from the NΔ166-mAhR clone employed as a positive control here. Thus, this AhR chimera signals in the absence of its primary ARNT dimerization domains (HLH, A' alpha and PAS A repeat), implying that the domains required for AhR signaling in yeast reside in the ligand binding domains, which map to the C-terminal half of the PAS domain (PAS B repeat), included in this clone.

To develop our chimera sensor system, we first carefully examined the LexA-AhR-chimera as a prototype. The dose-response (D-R) follows the expected shape until ~7 μM, where an unexpected drop appears in the response. This drop coincides with the upper solubility limit of the ligand as βNF as ~2.51 μM in water (determined visually). However, in this work, the solubility is probably different, as ligand is dissolved in media that contains salts and carbon sources that allow yeast growth. Additionally, yeast excrete metabolites into the medium that presumably raise solubility for βNF. The upper response limit value is relatively high compared to the other receptors in this study (as will be discussed later) but is considered of value as it allows an estimate of the top value and allows generation of data that fits 4PL analysis. Based on the 4PL, the EC_{50} for βNF in our system is ~0.43 μM, slightly higher than other reported studies, particularly in contrast to mammalian cell lines, indicating a lower potency in this system (Vazquez-Rivera, in review). One group reports that the AhR ligand TCDD may stick to plastic when stored due to its hydrophobic nature thus with less ligand in

solution, exposure concentrations are lower than they appear[23]. This is also a possibility with βNF as it also possesses similar hydrophobic qualities. In our lab, βNF solutions were always stored in glass vials to limit this confounding factor. Yet other groups suggest yeast detoxification mechanisms through efflux pumps as possible reasons for reduced potency as shown through increased βNF EC_{50} values. In our lab, we have been able to successful employ a yeast-two hybrid system that results in a lower EC_{50} for βNF, therefore efflux pumps are an unlikely explanation. Moreover, the co-expression of modifiers (i.e., ARA9 and ARA3) has also been shown to improve sensitivity by almost two-orders of magnitude (Vazquez-Rivera, in review).

Since potency is affected by receptor binding affinity, defined as the strength of the interaction between the drug and the receptor[13], these results led us to postulate that the PAS A repeat may also play a role in strengthening the interaction between the PAS B repeat of AhR and βNF. When considering crystallized 3D structures of PAS protein heterodimers, the PAS A and B repeats on class α proteins interact in such a way that the β-sheets for both partners appear to almost back into each other[21,22]. It is possible that this is a native configuration even in the absence of a dimerization partner like ARNT.

The steepness of the linear portion of the D-R is estimated at ~2, indicating a somewhat steep curve. With AhR, this result is unexpected since it is generally accepted that AhR binds one ligand at a time; therefore, the Hillslope was anticipated to yield a value closer to 1. In the last parameter, the top value, a high efficacy is observed. If classic receptor theory is to be accepted then E_{max} describes that "the maximum binding is proportional to the maximum number of receptors" and this value is an extension of the top value[13]. As with the Hillslope, the efficacy is poorly understood as a parameter especially in heterologous yeast systems. To quantitate reproducibility, the Levy-Jennings analysis was employed (**Figure 2**). Overall, we can

state that all four parameters for this assay in our lab have working coefficients of variance (CV) ~20%.

Our primary objective is to develop a singular yeast strain (e.g., L40) and a singular reporter (LexO driven β-galactosidase) and integrate it with a wide variety of sensor/receptor systems in parallel (see Introduction and Chapter 1). After evaluating the response of a prototype AhR sensor system, we then asked whether this same LexA-LexO based reporter system could be readily applied to members of the steroid receptor superfamily. The rationale is: if LexA-LexO in L40 was broadly applicable, then it could be the centerpiece of a meta system development path. Therefore, our initial objective was to ask two questions. (a) By choosing sensor systems at random, how many would work in our yeast system solely based on design and dependent upon on published structural domains? (b) Could most sensor systems perform in accordance with pharmacology comparable to mammalian cell culture systems of much greater cost? To these ends, four full-length steroid receptors (GR, ER, AR, and MR) were fused in-frame to LexA (**Figure 1**) and their pharmacology was examined following exposure to various ligands. In the case of the AR (**Figure 3B**), MR **(Figure 3C**) and ERα (**Figure 3D**), exposure to prototype ligands generated a full D-R with similar bottom values (~1700-3000 RLU). Top values were different between AR and ERα, 10^6 vs 10^5 RLU, respectively. In our hands, MR did not exhibit the top values necessary to determine the 4 parameters and GR did not provide a graded response at all. The EC_{50} for AR and ERα were: 3.67 nM, and 0.80 nM, respectively. The Hillslope for each was: 1.272, 1.781, respectively. The LOD was calculated to determine our assay's sensitivity to low concentrations for each ligand. The LOD is lowest for the ERα and highest for the AhR (**Figure 4**). This pattern is expected based on the bottom RLU values observed; few studies specify this information; however, they are difficult to compare because the units aren't always congruent.

Interestingly, in this study, GR did not respond to prototype ligands, even after exposure to dexamethasone, a ligand with high affinity (**Figure 3A**). This was surprising because our

laboratory and others have successfully expressed the GR in yeast and we have functionally expressed it as a Gal4 chimera, previously[18,24]. Although this construct was sequenced, obvious mutations in synthesis were not revealed; we have not investigated its failure at this time. Rather, we emphasize the fact that of three out of four steroid receptors were immediately successful as yeast-based sensors in a L40 Lex system. This suggests considerable potential for this approach, although some individual sensors such as GR, may require additional study for their implementation.

The bottom values observed with the steroid receptor chimeras were not surprising because we anticipated that use of full-length proteins in our assay would yield lower baseline values compared to constructs with N-terminal deletions. This is also the case the case between the AhR construct, NΔ166-mAhR, described here, and a LexA-full-length mAhR[b1] (Vazquez-Rivera, in review). These same differences are also observed in top values for these constructs. However, differences between AR and ERα are not yet understood—they may arise from different efficiencies of the system because of protein expression levels or naturally existing yeast modifiers that are pathway specific. For example there are studies that postulate the existence of protein sites capable of binding estradiol and corticosteroids in yeast, though their consequence on 4PL analysis has also not been elucidated[25]. While it can be hypothesized that the E_{max} (top values) may increase due to differences in receptor concentration, our results do not give any indication for such an increase. Lower EC_{50} values have been reported in other yeast-based biosensors including the yeast estrogen assay (YES) and the yeast androgen assay (YAS)[26]. The complete protocol for these assays including their interlaboratory EC_{50} were not readily available, however a short version was accessible online and shows a protocol that can take up to 7 days to execute. Our protocol can be performed in as little as one day if yeast is available, saving time and money.

In yet another assay, exploiting 3 different engineered yeast strains, the EC_{50} for AR, MR, and ERα were between: 3-27 nM for DHT, 14-100 nM for aldosterone, and 0.08-0.77 nM

for estradiol, respectively[27,28]. Although our assay has comparable EC_{50} values, our assay relies solely on the transcription factor while the strains in the Ito-Harashima et. al (2015, 2020) rely on transformation of the yeast with the transcription factor and the nuclear coactivator, NCOA1. These systems also are dependent upon the cognate response element and thus cannot be co-expressed without the addition of additional target promoters. Moreover, the NCOA1 is a member of the bHLH-PAS family of proteins and its overexpression is known to potentiate activity for steroid receptors by changing the delicate balance between coactivator and corepressor concentrations[29,30]. The authors describe their data in fold change therefore it is hard to determine what their bottom and top values were for further comparison. Additionally, the effect of an interactor on the EC_{50} or the top values is not fully understood. Some interactors have been shown to shift the D-R curve to the left and/or increase the top value[18,31,32]. Another effect to consider is the possibility of homodimerization between 2 $LexA_{DBD}$ and 2 full-length steroid receptors, so that 2 or more fused proteins may interact. The consequences of said interactions on any of the 4 parameters have also not been explored.

Although all endocrine disrupting chemicals are of concern, there is an emphasis on understanding the role and effect of xenoestrogens in the human body. This is because these compounds can have agonistic and antagonistic effects. They can also bioaccumulate in organisms, especially humans; because we have not evolved the mechanisms for adaptation[33]. In women, endogenous estrogen is found in the range between 36.7 pM to 1.47 nM, where the highest values are observed in premenopausal women, so any alterations to this delicate balance can be hazardous to human health[2] (**Figure 5**). In our system, our reproducibility for the ERα showed a coefficient of variance of up to 29% (**Figure 5**). This CV is higher than the AhR control and would need to be improved for increased reproducibility and use outside the lab; nonetheless, our system was robust enough to test for the presence of other types of environmental estrogens, including xenoestrogens and one phytoestrogen.

We exposed our yeast to DMSO-solvated genistein, bisphenol A (BPA) and tamoxifen. Genistein is a soy-derived compound with a structure similar to estradiol thought to confer protection from breast cancers; this hypothesis stems from the low prevalence of the disease in women of Asian-descent[34]. The compound BPA is man-made, used to harden plastics, and used as a coating inside canned goods to prevent corrosion from acidic foods. Consequently, BPA is regulated as an indirect additive or food contact surface. Tamoxifen is a drug and selective estrogen receptor modulator: It can act as an agonist for estrogen receptors in breast cells, and an antagonist in uterine and bone cells. This property is exploited to treat premenopausal women diagnosed with breast cancer[35].

In general, the D-R curves for each compound are quite different (**Figure 6**). The estradiol curve is the same one observed in **Figure 4D** and is the reference compound because it is the endogenous ligand for the ERα. The second compound on the graph is genistein, with an EC_{50} of 7.44 μM and a top value ~17% lower than estradiol. Additionally, the Hillslope is much steeper with a value closer 4. The third compound is BPA with an EC_{50} of 70.87 μM, a top value ~60% lower than estradiol, a Hillslope of 2.077. The fourth compound, tamoxifen does not fit the 4PL model, thus none of the parameters are provided with any confidence. Based on the R squared values, the 4PL model was a good fit for the data for the rest of the ligands.
From these data, we can observe that our system effectively recapitulates the mammalian potency rank for ERα ligands. Additionally, we can see that in our system, these ligands deviate from parallelism (*p-value <0.001*) – a phenomenon where ligands that behave the same way should yield statistically similar slopes allowing the calculation of a relative potency. This result makes sense because the assumption within parallelism is that the reference and test compounds have the same biologically active agent. Although our test ligands have similar structures, it is not clear, at this time, that they bind or produce the same allosteric changes in ERα; however, if we take into account other receptors that have been crystallized, such as HIF2α[22], it is likely that different amino acids are used to bind each ligand effectively and may

change the conformation of the protein. Furthermore, in this system we cannot calculate relative potencies for the ERα but can screen for estrogenic contaminants.

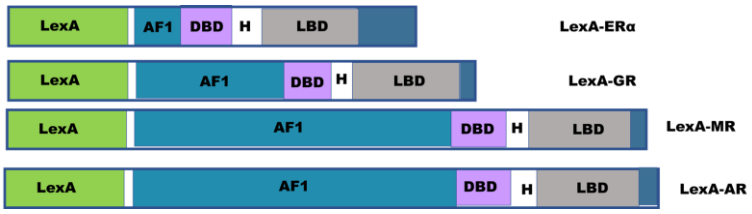
To understand the structural determinants that could play a role in the responses observed, the sequences were introduced into Deep Mind's AlphaFold structure prediction. In this system, the full LexA functions as a DNA binding domain and binds LexA operators in the L40 genome. The LexA closely resembles the helix-turn-helix DNA binding motif and is 202 amino acids long. Our LexA-AhRΔ166 construct shows the remnants of the PAS A repeat separated from the PAS B repeat by the transactivation domain **(Figure 7**). This conformation is different from predicted heterodimers having both PAS repeats and ligand in PAS B[36]. For comparison, we also pulled the full-length AhR predicted structure from the AlphaFold database (AF-P30561-F1) and the PAS repeats continue to be separated by the TAD. However, the full-length prediction also shows a portion of the TAD near the bHLH (**Figure 8**). It's possible that the AlphaFold prediction reflects the inactivated and untransformed AhR, thus the conformation in other predicted models reflects the liganded, activated, and transformed AhR.

Full-length steroid receptors have yet to be fully resolved structurally. However, the zinc finger (ZnF) DBD and a total of 12 helices – that makeup the hydrophobic ligand binding domain – have been crystallized separately for various steroid receptors. A switch between active and inactive conformations relies heavily on helix 12 (H12), found C-terminus[37]. The active form allows recruitment of coactivators by H3, H4, and H12; together they form a surface where coactivators can bind to LxxLL sequences (**Figure 9**)[37]. On the LexA-MR protein, AlphaFold predictions show the LexA separated whereby a portion is closer to the LBD and the second part is near the ZnF DBD. Furthermore, LBD helices cluster near the DBD. The prediction for LexA-AR is different than what is observed with LexA-MR. The LexA is packed closer together, while the LBD helices sit atop the LexA (**Figure 10**) and the DBD sits above the LBD. For the final prediction, LexA-ERα, the configuration is like LexA-MR, however the ZnF sit atop the N-terminus region of LexA, while the LBD helices are observed over the LexA C-terminus (**Figure**

11). When looking at these predictions, it is important to note that none have been modeled with ligands nor their homodimeric partners thus these conformations may change upon ligand and/or partner binding.

CONCLUSION

Yeast offers a cheap alternative to mammalian cell-based biosensors. For this biosensor, we took advantage of the reporter strain L40 and developed a fused-steroid receptor system that was functionable for most proteins tested. In our system, the ERα showed the lowest limit of detection and was used for further ligand activation studies; estradiol, BPA, tamoxifen, and genistein were successfully detected. By taking advantage of the new AlphaFold technology to predict what these fused proteins could look like, we hope to understand fused proteins and the mechanisms by which they may function in model organisms and systems.



Plasmid #	Protein Accession #	Plasmid Name	Restriction sites	Top Primer Bottom Primer
2275	BAJ65337.1	LexA-mERα	*EcoRI-BamHI*	5'-GAATTCATGACCATGACCCTTCACACCAA-3' 5'-GGATCCTCAGATCGTGTTGGGGAAGC-3'
2293	NP_001191190.1	LexA-mGRα	*EcoRI-BamHI*	5'-GAATTCATGGGAAATGACCTGGGATTC-3' 5'-GGATCCTCACTTTTGATGAAACAGAAG-3'
2294	NP_000892.2	LexA-huMR	*EcoRI-BamHI*	5'-GAATTCATGGAGACCAAAGGCTACCACAG-3' 5'-GGATCCTCACTTCCGGTGGAAGTAGAGC-3'
2295	NP_038504.1	LexA-mAR	*EcoRI-BamHI*	5'-GAATTCATGGAGGTGCAGTTAGGGCTGGG-3' 5'-GGATCCTCACTGTGTGTGGAAATAGATGG-3'

Figure 1: Constructs developed for this study. Full-length steroid receptors were inserted into pBTM116. The resulting protein is an N-terminus LexA-fusion. Constructs were transformed in yeast strain L40 and exposed to various concentrations of DMSO-solvated estradiol, dexamethasone, aldosterone and dihydrotestosterone to generate dose-response curves.

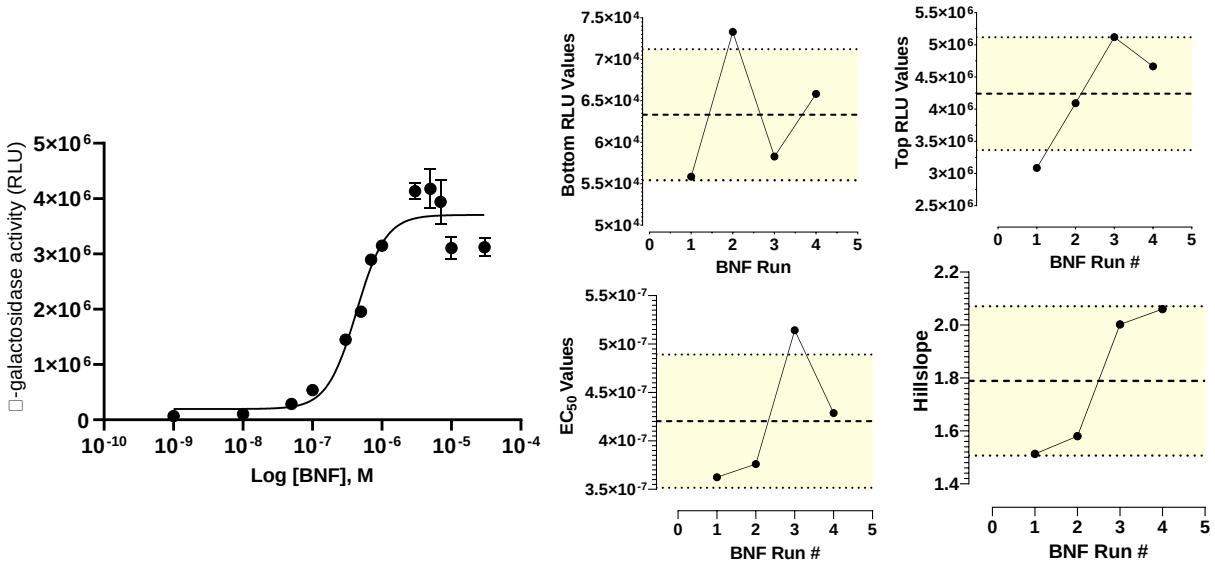


Figure 2: LexA-NΔ166mAhR^{b1} served as the control for our experiments. Statistical analyses show that the average ± CV (%) for 5 runs with 4 technical replicates each were: bottom: 66331 ± 13; top: 424071 ± 21; EC_{50}: 0.42 µM ± 16.40; and Hillslope: 1.789 ± 15.770.

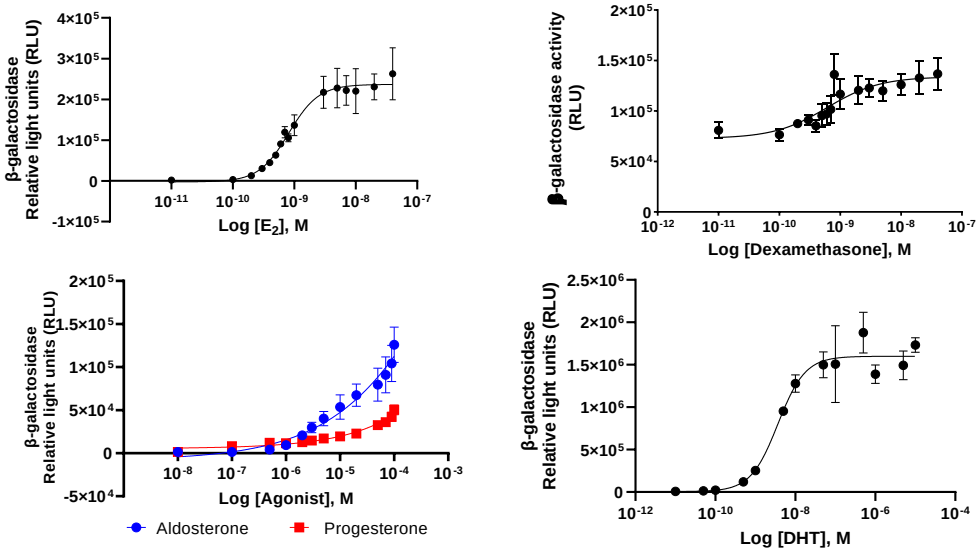


Figure 3: Dose-response curves for each LexA-fused steroid receptor. (A) The ERα exposed to the endogenous ligand estradiol. (B) GR did not signal properly and was not used for further studies (C) MR exposed to aldosterone and progesterone (D) AR exposed to dihydrotestosterone.

Table: 1. Yeast-based steroid receptor biosensors reported in the literature.

Steroid receptor	Ligand	EC_{50} (nM)	EC_{50} (nM) this study
ERα	Estradiol	0.531, 0.100, 0.200	0.80
	Genistein	356	7446
	Bisphenol A	1640, 4000	78710
	Tamoxifen	3400	n/a
GR	Dexamethasone	16000[25] 280[25]	n/a
MR	Aldosterone	6500[25] 650[25] 790[25] 320[25]	n/a
AR	Dihydrotestosterone	28[26], 2, 0.400, 3.50 11[26] 23[26]	3.67

Gaido KW, Leonard LS, Lovell S, Gould JC, Babaï D, Portier CJ, McDonnell DP. Evaluation of chemicals with endocrine modulating activity in a yeast-based steroid hormone receptor gene transcription assay. Toxicol Appl Pharmacol. 1997 Mar;143(1):205-12. doi: 10.1006/taap.1996.8069. PMID: 9073609

Chu WL, Shiizaki K, Kawanishi M, Kondo M, Yagi T. Validation of a new yeast-based reporter assay consisting of human estrogen receptors alpha/beta and coactivator SRC-1: application for detection of estrogenic activity in environmental samples. Environ Toxicol. 2009 Oct;24(5):513-21. doi: 10.1002/tox.20473. PMID: 19161236.

Mertl J, Kirchnawy C, Osorio V, Grininger A, Richter A, Bergmair J, Pyerin M, Washüttl M, Tacker M. Characterization of estrogen and androgen activity of food contact materials by different in vitro bioassays (YES, YAS, ERα and AR CALUX) and chromatographic analysis (GC-MS, HPLC-MS). PLoS One. 2014 Jul 7;9(7):e100952. doi: 10.1371/journal.pone.0100952. PMID: 25000404;

Table 2: Limit of detection (LOD) for each receptor plus ligand. The LOD was calculated to identify the sensitivity of our assay to ligand in the presence of yeast cells. The values were calculated using the 4PL and are within the 95% confidence interval.

Receptor	Ligand	Bottom RLU	Top RLU	EC_{50} (nM)	Hillslope	LOD (nM)
AhR	β-napthoflavone	65824	3.70E+06	4200	2.060	27.7
GR	dexamethasone	n.d.	n.d.	n.d.	n.d.	n.d
ERα	estradiol	1777	2.37E+05	0.80	1.781	4.90 x E-05
MR	aldosterone					
AR	dihydrotestosterone	4299	1.60E+06	3.67	1.272	7.80 x E-04

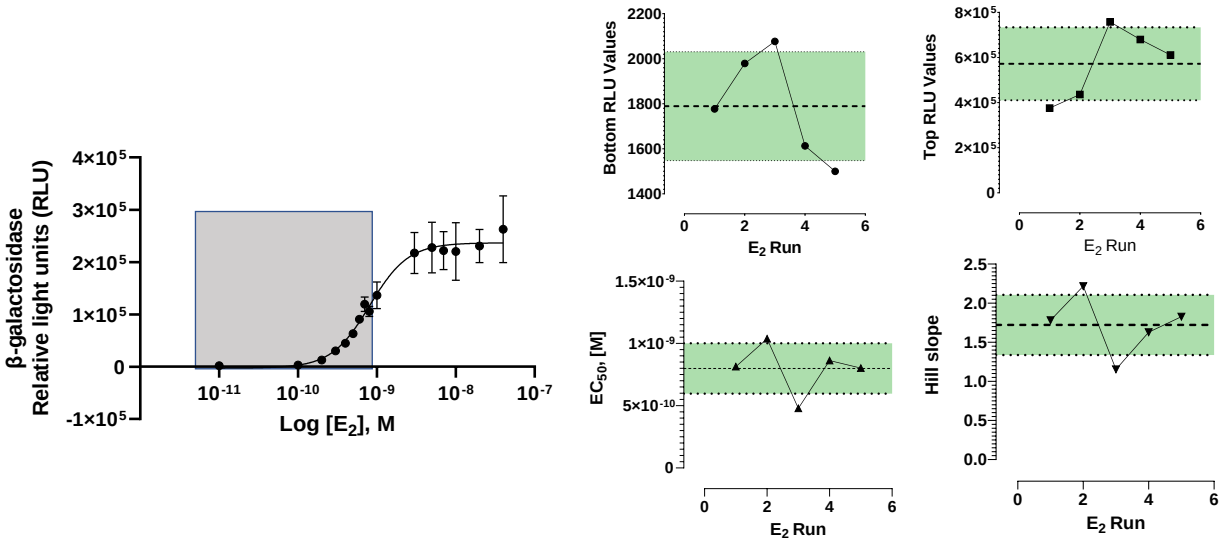


Figure 5: With a low LOD, the ERα was exposed to estradiol. Endogenous estradiol in pre-menopausal women is found in the range of 36.7 pM to 1.47 nM. Postmenopausal women tend to have lower values circulating. Statistical analyses show that average ± CV (%) for 5 runs with 4 technical replicates each were: bottom: 1789 ± 14; top: 571919 ± 29; EC_{50}: 0.80 µM ± 25; and Hillslope: 1.721 ± 22.

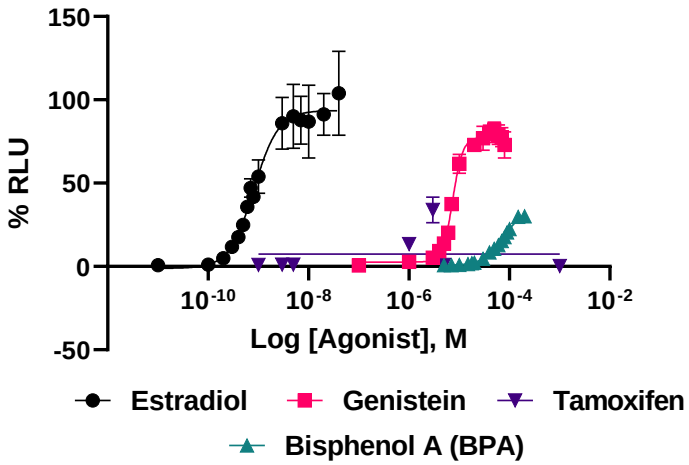


Figure 6: The ERα exposed to various ligands. In our assay, the ERα recapitulates mammalian potency rank, whereby estradiol has the lowest EC_{50} followed by genistein, and BPA. Tamoxifen values did not fit the 4PL analysis thus none of the parameters can be included with any confidence. The highest top values also follow the pattern observed with EC_{50}. Interestingly, the Hillslopes deviate from parallelism as confirmed through PLA with a *p-value <0.001*. Normalized data shown here for an easy comparison.

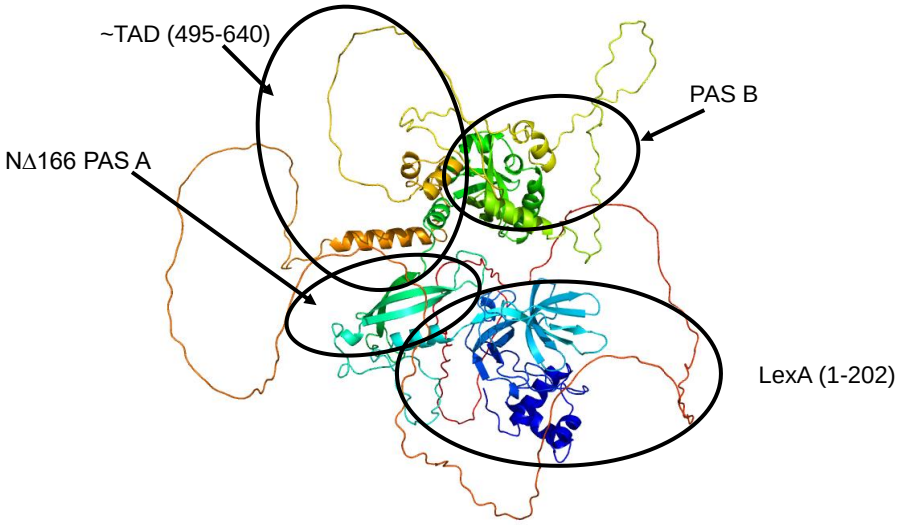


Figure 7: AlphaFold prediction for LexA-AhRΔ166. Fused protein sequence was entered into AlphaFold for prediction. The LexA serves as the DNA binding domain and showcases a helix-turn-helix structure. The prediction appears to separate the PAS A from PAS B repeats. Image is colored for direction, blue to red, N-terminus to C-terminus.

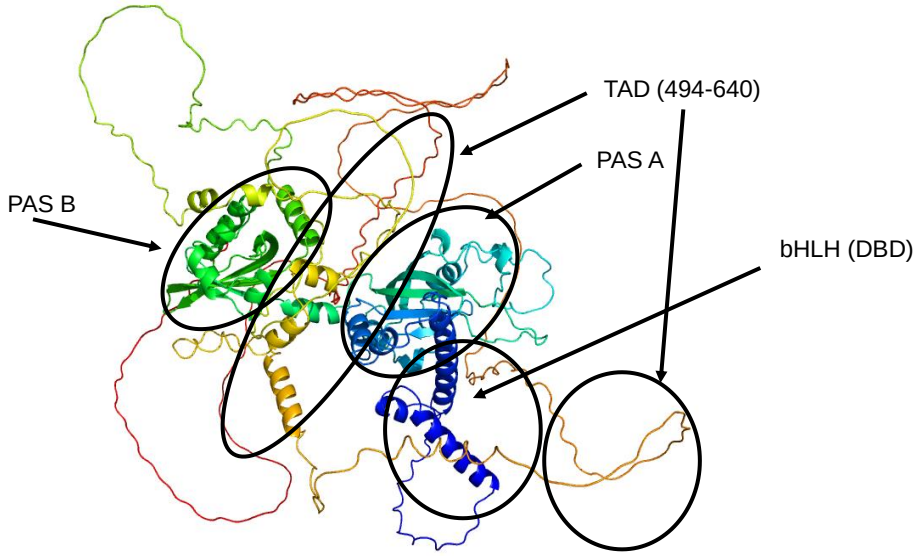


Figure 8: AlphaFold prediction for AhR. Much like the LexA-fused AhR, the full-length AhR is predicted to separate the PAS A from PAS B by interjecting a portion of the TAD between the repeats. Additionally, another portion of the TAD is observed near the bHLH. Image is colored for direction, blue to red, N-terminus to C-terminus (*AF-P30561-F1*).

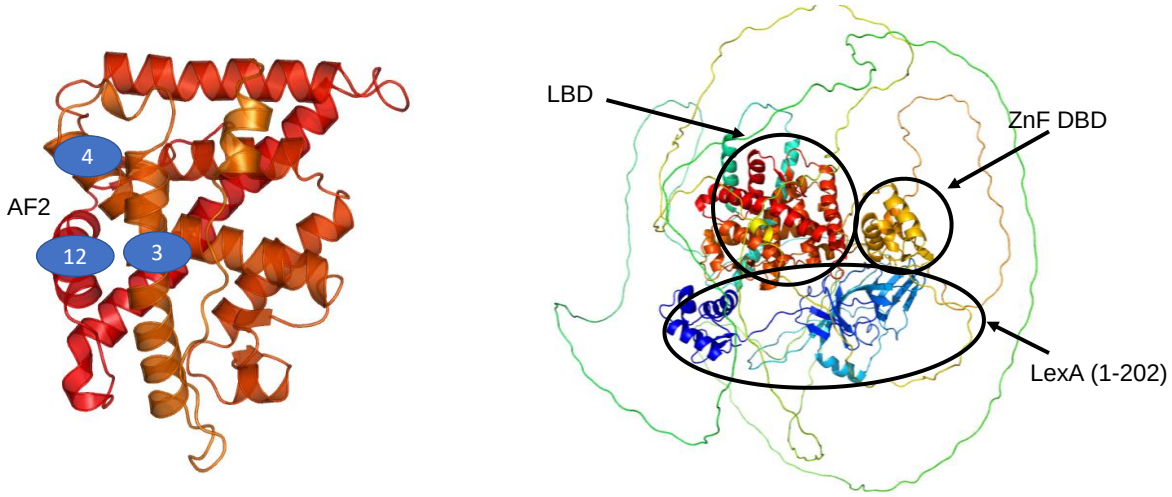


Figure 9: AlphaFold prediction for LexA-MR. The ligand binding domain for steroid receptors is composed of 12 helices. (*Left*) H3, H4, and H12 change conformation to provide the surface for AF2. (*Right*) Predicted fused LexA-MR is oriented to show the beginning of the LexA (*dark blue*) on the bottom left, while the end of the LexA is shown on the bottom right (*light blue*). Image is colored for direction, blue to red, N-terminus to C-terminus.

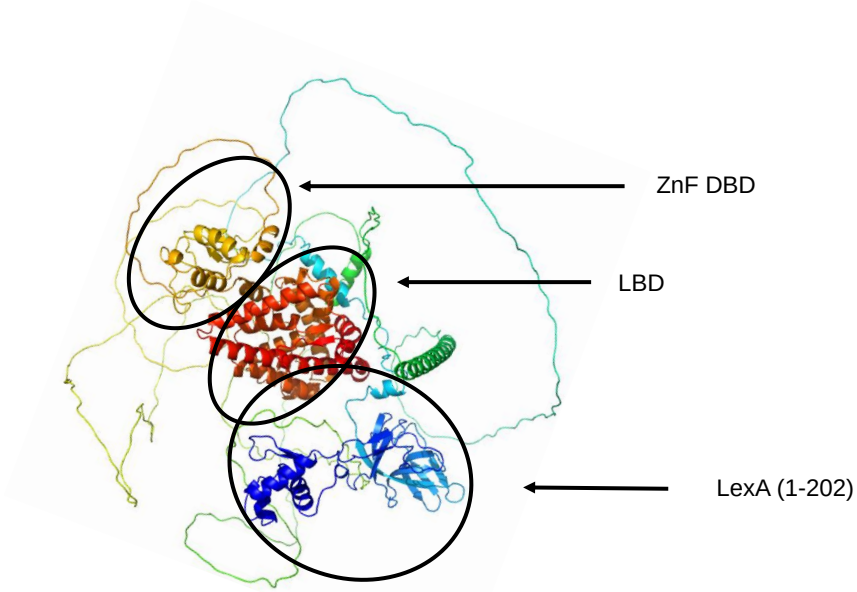


Figure 10: AlphaFold prediction for LexA-AR. The predicted configuration shows the LexA domain packed tightly, while the LBD helices sit on top of it. The DBD is situated on top of the LBD. This is different from what is observed with other fused steroid receptors used in this study. Image is colored for direction, blue to red, N-terminus to C-terminus.

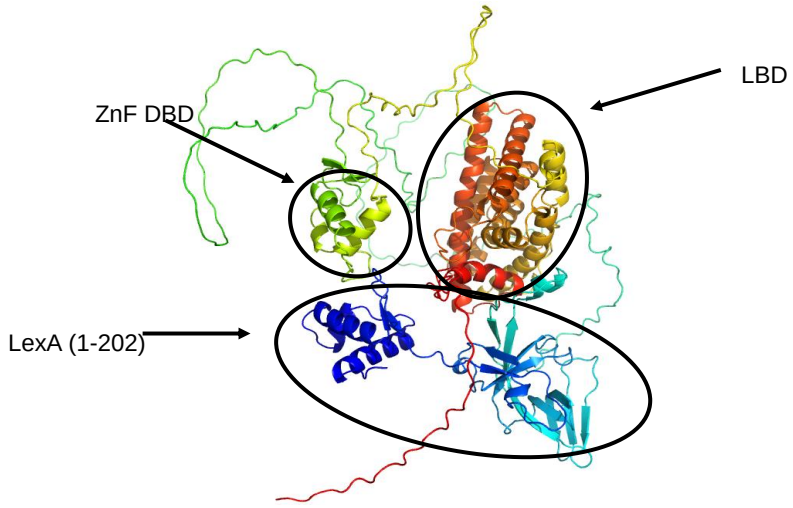


Figure 11: AlphaFold prediction for LexA-ERα. Much like LexA-MR, the prediction shows LexA N-terminus is separated from the C-terminus. The LBD domain sits atop the C-terminus part of LexA, while the DBD sits atop the N-terminus part of LexA. Image is colored for direction, blue to red, N-terminus to C-terminus.

REFERENCES

1 Evans, R. M. The Nuclear Receptor Superfamily: A Rosetta Stone for Physiology. *Molecular Endocrinology* **19**, 1429-1438, doi:10.1210/me.2005-0046 (2005).

2 La Merrill, M. A. *et al.* Consensus on the key characteristics of endocrine-disrupting chemicals as a basis for hazard identification. *Nature Reviews Endocrinology*, 1-13, doi:10.1038/s41574-019-0273-8 (2019).

3 Brent, R. & Ptashne, M. A EUKARYOTIC TRANSCRIPTIONAL ACTIVATOR BEARING THE DNA SPECIFICITY OF A PROKARYOTIC REPRESSOR. *Cell* **43**, 729-736, doi:10.1016/0092-8674(85)90246-6 (1985).

4 Fields, S. & Song, O. K. A NOVEL GENETIC SYSTEM TO DETECT PROTEIN PROTEIN INTERACTIONS. *Nature* **340**, 245-246, doi:10.1038/340245a0 (1989).

5 Zhang, C. Y. *et al.* Increased esters and decreased higher alcohols production by engineered brewer's yeast strains. *European Food Research and Technology* **236**, 1009-1014, doi:10.1007/s00217-013-1966-1 (2013).

6 Denby, C. M. *et al.* Industrial brewing yeast engineered for the production of primary flavor determinants in hopped beer. *Nature Communications* **9**, doi:10.1038/s41467-018-03293-x (2018).

7 Luo, X. *et al.* Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123-126, doi:10.1038/s41586-019-0978-9 (2019).

8 Redden, H., Morse, N. & Alper, H. S. The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Research* **15**, 1-10, doi:10.1111/1567-1364.12188 (2015).

9 Botstein, D. & Fink, G. R. Yeast: an experimental organism for 21st Century biology. *Genetics* **189**, 695-704, doi:10.1534/genetics.111.130765 (2011).

10 Godowski, P., Picard, D. & Yamamoto, K. Signal transduction and transcriptional regulation by glucocorticoid receptor-LexA fusion proteins. *Science* **241**, 812-816, doi:10.1126/science.3043662 (1988).

11 Metzger, D., White, J. H. & Chambon, P. The human oestrogen receptor functions in yeast. *Nature* **334**, 31-36, doi:10.1038/334031a0 (1988).

12 Carver, L. A., Jackiw, V. & Bradfield, C. A. THE 90-KDA HEAT-SHOCK PROTEIN IS ESSENTIAL FOR AH RECEPTOR SIGNALING IN A YEAST EXPRESSION SYSTEM. *Journal of Biological Chemistry* **269**, 30109-30112 (1994).

13 Salahudeen, M. S. & Nishtala, P. S. An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice. *Saudi Pharmaceutical Journal* **25**, 165-175, doi:10.1016/j.jsps.2016.07.002 (2017).

14 Hollenberg, S. M., Sternglanz, R., Cheng, P. F. & Weintraub, H. Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Molecular and Cellular Biology* **15**, 3813-3822, doi:10.1128/mcb.15.7.3813 (1995).

15 Carver, L. A., LaPres, J. J., Jain, S., Dunham, E. E. & Bradfield, C. A. Characterization of the Ah receptor-associated protein, ARA9. *Journal of Biological Chemistry* **273**, 33580-33587, doi:10.1074/jbc.273.50.33580 (1998).

16 Masdor, N. A.

17 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

18 LaPres, J. J., Glover, E., Dunham, E. E., Bunger, M. K. & Bradfield, C. A. ARA9 modifies agonist signaling through an increase in cytosolic aryl hydrocarbon receptor. *Journal of Biological Chemistry* **275**, 6153-6159, doi:10.1074/jbc.275.9.6153 (2000).

19 Seok, S. H. *et al.* Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 5431-5436, doi:10.1073/pnas.1617035114 (2017).

20 Huang, N. *et al.* Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337**, 189-194, doi:10.1126/science.1222804 (2012).

21 Wu, D., Su, X., Potluri, N., Kim, Y. & Rastinejad, F. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. 1-15, doi:10.7554/eLife.18790 (2016).

22 Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **524**, 303-308, doi:10.1038/nature14883 (2015).

23 Miller, C. A. A human aryl hydrocarbon receptor signaling pathway constructed in yeast displays additive responses to ligand mixtures. *Toxicology and Applied Pharmacology* **160**, 297-303, doi:10.1006/taap.1999.8769 (1999).

24 Miesfeld, R., Godowski, P. J., Maler, B. A. & Yamamoto, K. R. Glucocorticoid receptor mutants that define a small region sufficient for enhancer activation. *Science (New York, N.Y.)* **236**, 423-427, doi:10.1126/science.3563519 (1987).

25 Feldman, D., Do, Y., Burshell, A., Stathis, P. & Loose, D. S. An Estrogen-Binding Protein and Endogenous Ligand in Saccharomyces cerevisiae: Possible Hormone Receptor System. *Science* **218**, 297-298, doi:doi:10.1126/science.6289434 (1982).

26 Arnold, S. F., Robinson, M. K., Notides, A. C., Guillette, L. J. & McLachlan, J. A. A yeast estrogen screen for examining the relative exposure of cells to natural and xenoestrogens. *Environ Health Perspect* **104**, 544-548, doi:10.1289/ehp.96104544 (1996).

27 Ito-Harashima, S. *et al.* Construction of reporter gene assays using. *Genes Environ* **42**, 20, doi:10.1186/s41021-020-00159-x (2020).

28 Ito-Harashima, S. *et al.* Construction of sensitive reporter assay yeasts for comprehensive detection of ligand activities of human corticosteroid receptors through inactivation of CWP and PDR genes. *J Pharmacol Toxicol Methods* **74**, 41-52, doi:10.1016/j.vascn.2015.06.001 (2015).

29 Glass, C. K., Rose, D. W. & Rosenfeld, M. G. Nuclear receptor coactivators. *Current Opinion in Cell Biology* **9**, 222-232, doi:10.1016/S0955-0674(97)80066-X (1997).

30 Szapary, D., Huang, Y. & Simons, S. S. Opposing Effects of Corepressor and Coactivators in Determining the Dose-Response Curve of Agonists, and Residual Agonist Activity of Antagonists, for Glucocorticoid Receptor-Regulated Gene Expression. *Molecular Endocrinology* **13**, 2108-2121, doi:10.1210/mend.13.12.0384 (1999).

31 Dunham, E. E., Stevens, E. A., Glover, E. & Bradfield, C. A. The aryl hydrocarbon receptor signaling pathway is modified through interactions with a Kelch protein. *Mol Pharmacol* **70**, 8-15, doi:10.1124/mol.106.024380 (2006).

32 Kaul, S., Blackford, J. A., Cho, S. & Stoney Simons, S. Ubc9 is a novel modulator of the induction properties of glucocorticoid receptors. *Journal of Biological Chemistry* **277**, 12541-12549, doi:10.1074/jbc.M112330200 (2002).

33 Nilsson, R. Endocrine Modulators in the Food Chain and Environment. *Toxicologic Pathology* **28**, 420-431, doi:10.1177/019262330002800311 (2000).

34 Miodini, P., Fioravanti, L., Di Fronzo, G. & Cappelletti, V. The two phyto-oestrogens genistein and quercetin exert different effects on oestrogen receptor function. *Br J Cancer* **80**, 1150-1155, doi:10.1038/sj.bjc.6690479 (1999).

35 Metzger, D., Ali, S., Bornert, J. M. & Chambon, P. Characterization of the amino-terminal transcriptional activation function of the human estrogen receptor in animal and yeast cells. *J Biol Chem* **270**, 9535-9542, doi:10.1074/jbc.270.16.9535 (1995).

36 Corrada, D., Soshilov, A. A., Denison, M. S. & Bonati, L. Deciphering Dimerization Modes of PAS Domains: Computational and Experimental Analyses of the AhR:ARNT Complex Reveal New Insights Into the Mechanisms of AhR Transformation. *PLoS Computational Biology* **12**, doi:10.1371/journal.pcbi.1004981 (2016).

37 Delfosse, V. *et al.* Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes. *Proc Natl Acad Sci U S A* **109**, 14930-14935, doi:10.1073/pnas.1203574109 (2012).

Chapter 3: Mammalian PAS family crystal structures for the non-structural biologist

INTRODUCTION

Organisms detect rapidly changing extracellular environments to regulate the proper physiological response for survival. This adaptation often depends on sensors that employ protein-protein interactions to generate dimers and higher order complexes that influence cellular physiology through complex networks or interactomes. In chordates, “PAS proteins” are an important family of biological sensors that detect changes in oxygen levels, light, membrane voltage, and the presence of certain small molecules, to influence adaptive physiological responses, often through changes in gene expression that are the result of PAS protein dimers that bind to genomic enhancer elements. This binding to enhancers is often mediated through N-terminal basic helix-loop helix motifs (bHLH) which are common in many PAS proteins.

Phylogenetic analyses demonstrates that PAS domains are found in all major kingdoms of life, suggesting that this structure evolved from a common ancestor and quickly adapted to the needs of multicellular organisms[1]. The PAS domain has many functions in biology. While PAS domains can serve to bind small molecules, they also appear to play a major role in dimerization with other PAS family members (we refer this to as homo-family interaction). In this review we differentiate interactions as follows: PAS-PAS interactions that take place between the exact same protein are homotypic, while interactions between two distinct PAS proteins are heterotypic. Alternatively, PAS domains may confer additional specificity for members of other protein families, such as heat shock proteins or other transcription factors (we refer to these interactions as hetero-family) [2-4]

In our laboratory, mammalian PAS proteins are divided into six classes: (i) alpha class which commonly act as sensors of environmental cues, (ii) beta class which act as dimeric partners for the alpha class to regulate responsive gene expression, (ii) gamma class which include a number of transcriptional coactivators, (iv) delta class which class which we view as repressors of the alpha beta dimers, (v) epsilon class which describes a pair of PAS domain

containing enzymes, and (vi) kappa class which include a series of ion channel protein subunits **(Figure 1)**. This chapter focuses on the nomenclature and rules that govern alpha, beta, and delta class PAS interactions. For additional information on the underlying biology, the reader is referred to a recent review[5].

The term "PAS domain" was originally identified through sequence homology analysis of the three founding proteins assigned to the family: *Drosophila* Per[6], human Arnt[7], and *Drosophila* Sim[8] (**Figure 2**). This ~270 amino acid region was shown to be comprised of two internal, highly degenerate repeats of ~50 amino acids each, deemed PAS A and PAS B[8]. Early definition of the larger PAS domain was also based on sequence-function studies, and structure prediction methodology (helices, hydrophobicity etc.). Due to the low level of amino acid sequence conservation in the PAS domain (e.g., as low as 15% identity across family members, **Figure 2**), the definition of the PAS domain has been subject to several proposed changes over the years and can be quite confusing to many in the field. For example, some have redefined the PAS domain, using the term "PAS/PAC domain[9]," to emphasize a ~45 amino acids motif (PAC domain) found near the C-terminal end of the PAS B repeat. To add to the complexity, the term "PAS-fold" has been proposed to represent this PAS-PAC region. Thus, the "PAS fold" definition describes PAS B repeats but does not represent PAS A domains which do not appear to harbor a PAC region[10]. These definitions become particularly important in mammalian PAS biology because several PAS proteins encoded by the mammalian genome only harbor a single PAS repeat/fold. Without the PAS A and PAS B organization required by the earliest definition of PAS domain, we must reexamine this domain to include these newfound members (**Figure 1**).

For clarity, throughout this chapter we will use the term "PAS domain" to define a protein region that harbors any number of PAS repeats in a row. In mammalian biology, we observe that the PAS domain harbors only one or two PAS repeats (although this number can be higher in other kingdoms of life). These repeats are named in alphabetical order from the N-terminal

side (i.e., PAS A, PAS B, etc.). It is important to note that we consider the PAS repeats as the fundamental unit of the PAS domain and our definitions of PAS domains and PAS repeats include linker sequences that may not be resolvable/observable by current structural methodologies. Finally, we will not use the PAS-PAC terminology as we consider the PAC domain to be a structural component of the PAS B repeat.

Describing this field of science can be confusing due to the imperfect, conflicting and changing nomenclature that has been based on amino acid sequence and structure prediction, rather than experimentally observed three-dimensional structure. This series of *in silico* experiments is an effort to address this confusion and build more robust definitions of domains, as well as more robust models of interactions and signaling. The methodology employed in this chapter integrates our understanding of PAS protein cell biology with the recent achievements in the structural biology of PAS proteins that have arisen from both X-ray crystallographic and NMR spectroscopic methods.

METHODS

As a learning set, we focused on sequences and crystal structures of PAS proteins of the alpha, beta and delta classes and have chosen to examine only those structures and protein sequences where dimerization between two PAS proteins is represented in the elucidated structure (**Figure 3 and Table 1**). To accomplish our tasks, several computational approaches were employed, these include, PyMol Molecular Graphics System v1.8.2.1, Phyre2[11], Alpha-Fold[12], and JalView[13]

RESULTS AND DISCUSSION

Definitions based on structure

In this chapter, we discuss the similarities and differences observed in crystallized PAS homo-family dimers: (a) CLOCK:ARNTL, (b) PER1:PER1, (c) PER2:PER2, (d) PER3:PER3 (e) HIF2α:ARNT, (f) HIF1α:ARNT, (g) NPAS1:ARNT, (h) NPAS3:ARNT, (h) AHR:ARNT and (i) AHRR:ARNT. We hypothesize that the PAS family of transcription factors has various modes of heterodimerization to facilitate the proper cellular response. We argue that PAS A and PAS B repeats have different but equally important functions in the transcription process. The goal of this work is to explain crystal structures to the non-structural biologist in a way that can move this field forward.

Defining the PAS Repeat: Using visual inspection, Phyre2, and Jalview analyses to identify beta strands and alpha helices in PAS domains from all alpha-beta, alpha-gamma, gamma-gamma dimers that have been elucidated structurally (**Table 1 and Figure 3**), we observed a simple pattern of structural motifs that is common across all PAS repeats. As a representation of this analysis: a sequence alignment of the bHLH and PAS domains of proteins from the learning set was generated, and the alignments were performed only on those regions which could be structurally elucidated (i.e., any unstructured loops were not included) (**Figure 3**). These analyses support a singular definition of PAS repeats conforming to an earlier described concept where each PAS repeat is composed of 5 anti-parallel β-strands and 4 α-helices[10], each PAS repeat begins with a β-strand and ends with a β-strand (**Figure 4, 5, and 6**). The β-strands (Aβ, Bβ, Gβ, Hβ, Iβ) form β-sheets that conform to cavities capable of binding small ligands. We define the posterior side of the β-sheet as the “back” of the structure and cavity as the “face” of the structure[14], just like other seminal findings. Importantly, an additional alpha helix is consistently observed on the amino terminal side of all PAS A domains from alpha and beta class proteins (and some delta class). This alpha helix, denoted here as A’ helix

(**Figures 4, 5, and 6**) can be seen in corresponding crystal structures as the central dimerization driver between most PAS A repeats. Contrary to most simple models of PAS protein dimerization employed in a number of recent reviews, it appears that interaction of PAS domains is not driven by PAS-A-PAS A interactions but rather by the A' helix. In conclusion, we identify all PAS repeats by this structural rule (e.g., starting with a beta strand and ending with a beta strand) and describe the A' helix as an accessory motif commonly found on the amino terminal side of the PAS A repeat and on the C-terminal side of the bHLH motif, supporting both dimerization and positioning of the basic region within the major groove of DNA.

The motif driven definition of the PAS A repeat is marked improvement over identification of this region via sequence alignment. Although PAS tertiary architecture is conserved, primary sequences are not and can be difficult to recognize by homology searches at the amino acid level. For example, between mouse and human, PAS A motifs share about 35% homology, while PAS B motifs show about 31% homology[15]. Interestingly, between PAS A and B, there is less than 20% homology[15]. Despite low sequence homology, all PAS proteins examined in this study (100%) conform to the following motif organization: A-beta strand, B beta strand**,** C alpha-Helix, D alpha helix, E alpha helix, F alpha helix, G beta strand, H beta strand, I beta strand (denoted as Aβ, Bβ,Cα,Dα,Eα,Fα,Gβ,Hβ,Iβ). Biologically, the presence of more than one PAS repeat is thought to provide selective advantages through diversification of PAS function, such as increased dimerization specificity or avidity, capacity to bind small molecules and an increased ability to transduce cellular signals.

Rules of dimerization

With formal definitions and understanding of the PAS domain and its repeats in hand, we turned our attention to the ways in which these domains interact, focusing on homo-family interactions. To accomplish this, we analyzed each dimeric pair from our learning set visually and informatically using PyMol, JalView, DNAStar (MegAlign®. Version 17.2. DNASTAR.

Madison, WI), and Phyre2. Our objective was two-fold: first, to identify common modalities of PAS protein interactions. Second, to address the idea commonly used in the PAS field of research (including by our own laboratory) that depicts the signal transduction pathway as a symmetrical pairing of bHLH-PAS A and PAS B repeats associated with the active state of the dimer (**Figure 7**). This common representation of PAS signal transduction has its roots in the symmetry of known bHLH protein and other nuclear receptors such as glucocorticoid receptor and estrogen receptor[16], as well as domain analyses studies focusing on the aryl hydrocarbon receptor (AhR) and its obligate partner, the aryl hydrocarbon receptor nuclear translocator (ARNT)[17], but is being contradicted by initial data from structural studies from PAS field. Therefore, we systematically analyzed our learning set for the mechanisms by which PAS domains interact, searching for similarities, symmetry, and differences.

The CLOCK:ARNTL Dimer: The first dimer to be structurally resolved from the mammalian PAS family, containing both bHLH, and PAS domains (but not C-terminal variable region), was CLOCK:ARNTL (aka, BMAL, MOP3)[18]. This dimer is a transcriptional activator found in a variety of cell types, including the suprachiasmatic nucleus (SCN) of the hypothalamus[15]. It regulates daily physiological and behavioral activities for the 24-hour circadian cycle. In a day, circadian inputs such as light, food, or hormones upregulate levels of CLOCK, permitting dimerization with ARNTL, and subsequent binding of the dimer to the promoter of various outputs genes and for the repressor of the pathway, PER[15]. The CLOCK, ARNTL, and PER mRNA and proteins oscillate throughout the day; when CLOCK levels are up, PER levels are down and vice versa. CLOCK:ARNTL levels are highest during the day, while PER levels increase in the late afternoon or evening. This increase in PER facilitates translocation into the nucleus where it disrupts the CLOCK:ARNTL dimer and represses its own transcription. The following day, CLOCK is once again activated and the process re-starts. The SCN is not the only region with a molecular clock of this kind; all cells express clock genes that create oscillations in a similar manner. In some cells, a paralogue of CLOCK, known as

NPAS2[19] is an important regulator of this rhythmic pathway. This CLOCK:ARNTL dimer is a central transcriptional engine that synchronizes clocks in almost all cell types and is referred to as the master pacemaker.

In accordance with our classification scheme, CLOCK is a class α sensor, while ARNTL is a class β dimerization partner[20]. Analysis of the CLOCK:ARNTL dimer crystal structure (pdb:4F3L) reveals an asymmetrical but parallel configuration[18]. Generally, the three dimerization regions – bHLH, PAS A, and PAS B – interact between both proteins, i.e., the CLOCK bHLH interacts with the ARNTL bHLH and so on (**Figure 8**). Yet, in contrast to simplistic symmetrical models as depicted in **Figure 7**, in space, bHLH-bHLH and PAS A-PAS A interactions are lateral, while PAS B-PAS B interactions are medial and stacked on top of PAS A-PAS A. Moreover, the PAS B-PAS B configuration is different than PAS A-PAS A and is observed in anterior-posterior orientation. Additionally, within CLOCK, the bHLH and PAS A repeats interact directly, while the corresponding regions in ARNTL do not. In both CLOCK and ARNTL, PAS A-PAS B contacts are also made so that PAS As interact with both PAS Bs (**Figure 9 and 10**).

In analyzing PAS A-PAS A interactions, the CLOCK:ARNTL structure was the first mammalian PAS pair to reveal a helix N-terminus to PAS A of each protein, deemed A'α-helix, that plays a major role. Based on our definition of PAS repeats, the A'α-helix is not part of the PAS A repeat and is an independent accessory motif. Dividing CLOCK and ARNTL with an imaginary midline, allows one to see that the A'α-helix of one protein crosses over to establish extensive contacts with the "back" of the PAS A β-sheet of the other, while also interfacing with one another (**Figure 9**). Outside of the A'α-helix, the "backs" of PAS A β-sheets point toward this imaginary midline and the "faces," containing the 4 helices, point outward. Therefore, the A'α-helices on both proteins stabilize PAS A-PAS A interactions.

The PAS A and PAS B repeats connect through a linker. The ARNTL linker is solvent accessible, flexible, and has a few interconnections with the "back" of the CLOCK PAS A β-

sheet. On the other hand, the CLOCK linker buries between the dimer establishing contacts with both PAS A and B repeats on both proteins. Interactions through PAS B happen on a frontal plane and rely on an anterior/posterior position on top of the sagittal PAS A-PAS A. Interestingly, PAS B repeats adopt a "face to back" configuration whereby the CLOCK PAS B face assembles against the ARNTL PAS B (**Figure 10**).

The PER1, PER2, and PER3 Homodimers: Repression plays a significant role in gene regulation. PERs exhibit one of the most common repressor mechanisms that exists in nature: the "dominant-negative," where dimerization occurs between a monomer containing a DNA binding domain (DBD) and a monomer lacking a DBD. The authors describing the original PAS domain mention a conversation with Michael Rosbash as the basis for the assumption that *Drosophila* PERs do not have a bHLH or other DBD[8]. The idea of PER as a dominant negative of the CLOCK cycle was also proposed many years ago by our lab[21]. Interestingly, the initial mouse PERs publication used amino acid alignments to show the existence of HLH domains on each PER[22]. It is unclear when the consensus that mouse PERs lack the basic-DBD arose, however repression capabilities are hypothesized to come from their ability to bind transcription factors leading to the presence of 1 DBD, instead of the 2 needed, hindering target gene activation[23]. More information is needed to identify the structures present N-terminus to PAS domains and their roles in repression.

Mammalian PERs are interesting because they have been shown to homo- and heterodimerize with each other possibly to facilitate nuclear translocation, although the effect on repression has not been explored in detail[24,25]. This ability to homo- and heterodimerize make mouse PERs one of the more important, yet poorly understood, dimerization anomalies in the PAS family. *Drosophila* PERs have a helix C-terminal to PAS B and a similarly located helix was found in mouse PERs, dubbed by the authors – helix αE –important for PAS domain interactions[26-28]. To streamline the naming of these accessory motifs outside the PAS domain, we will refer to αE as B2'α helix, using A'α helix as the counterpart example. These non-

canonical structures outside PAS domains and their interactions stabilize all PER homodimers. Interestingly, the B2'α helix on PERs appear to also contain a nuclear export signal (NES)[26,27]. Additionally, where PER1 and PER3 have the A'α helix, PER2 does not, hence 2 additional and different dimerization modes are observed in PERs alone.

Multiple dimers/oligomers of PERS have recently been structurally elucidated using X-ray crystallography. The first dimerization mode comes from PER2 homodimers[26]. In PER2 (pdb: 3GDI) homodimers, there is a lack of interactions between both PAS As; of note, they also lack the A'α helix. Thus without PAS A-PAS A interactions, PER2 homodimers rely mostly on intermolecular PAS A-PAS B and PAS B-PAS B repeat interactions[26] (**Figure 11**). Direct contacts between PAS A on molecule 1 and PAS B on molecule 2 are managed through the loop akin to Dα-Eα in of PAS A and the B'2α helix C-terminus to PAS B; this facilitates direct contact between the PAS A repeat and the NES located on B2'α helix (L460, I464, L467, M469)[26]. This phenomenon is not unusual as NES are typically found in amphipathic α-helices, where non-polar residues are buried within the dimerization interface, and the last residue tends to be solvent exposed for protein export[29]. For PAS B-PAS B interfacing, the Hβ, the Iβ strands, and the Hβ-Iβ loops are needed (**Figure 12**). Phenylalanines on the Hβ and the Iβ strands at positions 415 and 425 on molecule 1 establish stacking interactions with the same amino acids on molecule 2. Phenylalanines are important for hydrophobic stacking interactions and are usually found buried in the protein core. The Hβ-Iβ loop on molecule 1 harbors a tryptophan at position 419 that establishes contact with the Iβ strand at the glycine at position 428 on molecule 2[26]. Additionally, W419 also interfaces with a tryptophan at position 412 on the Iβ strand[26]. Tryptophan residues are energetically expensive; as such they are found at sites that are important for protein folding and function[30] Understanding the role of these specific interactions could provide clues about the role of the PER2 homodimer.

The second dimerization mode within PERs is observed in PER1 (pdb:4DJ2) and PER3 (pdb: 4DJ3) homodimers, whereby PAS B-PAS B contacts are made, as well as PAS A- PAS A

contacts (**Figure 13**). Although PER1 and PER3 proteins show A'α helices, they don't use them for PAS A-PAS A contacts[27]. Instead, PER1 and PER3 homodimers rely on αF helices (referred as αC by authors).[27] Intriguingly, the helices don't cross over to interact with the back of the beta sheets as other dimers do, instead, the helices line up and establish contact among themselves to stabilize dimerization (**Figure 14**). For PAS B-PAS B interactions, the backs of each β-sheet align toward the midline and interact. The helices and a few β-strands are observed on the outer perimeter. Much like PER2, the Hβ, the Iβ strands, and the Hβ-Iβ loops play prominent roles. In PER1 homodimers, the tryptophan at position 448, on the Hβ-Iβ loop of molecule 1 interfaces with glycine (G455) on the Iβ strand on molecule 2. Additionally, W448 on molecule 1 interacts with a tryptophan at position 441 on molecule 2 (**Figure 15**). These positions are analogous in PER3 whereby W359 on the Hβ-Iβ loop of molecule 1 interacts with G368 on the Iβ strand and W352 on the Hβ strand of molecule 2 (**Figure 16**). Much like PER2, these interactions have no known biological consequences at this time and need to be studied for biological relevance.

The HIF2α:ARNT and HIF1α:ARNT Dimers: Oxygen homeostasis is maintained primarily by HIFs1-3α and their partner ARNT. The HIFαs are upregulated under hypoxic conditions and heterodimerize with ARNT to transcribe target genes responsible for survival and adaptation to low oxygen levels. Under normoxic conditions, HIFs are constantly produced and degraded, however during hypoxic conditions, enzymes that hydroxylate HIFs lose activity because they need oxygen. This hydroxylation step is important as it permits recognition and ubiquitination by Vh1, an E3 ubiquitin ligase. Without ubiquitination, HIF concentrations increase in the cell, they translocate to the nucleus, bind ARNT and initiate target gene transcription, such as erythropoietin – a hormone that prompts red blood cell production. The HIF1α and HIF2α proteins are the most studied and most similar in structure. The HIF3α (IPAS, MOP7) proteins are an enigmatic member of this subfamily that may act as a repressor of the pathway and is also referred to as inhibitory PAS. Although HIF1α is found in all tissues, HIF2α and HIF3α expression is restricted to vascular endothelial cells, placenta, lungs, interstitial cells of

the retina, liver parenchymal cells and type II pneumocytes.

The ARNT protein is particularly fascinating because it is promiscuous and has been shown to dimerize with approximately 12 of the 22 PAS proteins, including itself. Outside of PAS proteins, the ARNT is also thought to interact with various coactivators, like TACC3[31], and steroid receptors like the estrogen receptor alpha (ERα)[32]. The HIF1α:ARNT (pdb: 4ZPR): and the HIF2α:ARNT (pdb: 4ZPK) dimers display modes that are similar but distinctive from the CLOCK:ARNTL dimer therefore ARNT dimers represent yet another type of PAS interactions (**Figure 17 and Figure 18**).

The overall structure shows PAS A-PAS A and PAS B-PAS B interactions between HIF1α/HIF2α and ARNT . Just like CLOCK_ARNTL, both HIF1α:ARNT and HIF2α:ARNT dimers use their A'α helices to stabilize PAS A-PAS A interfaces[33]; the helices cross the midline to interact with the back of opposing beta sheets (**Figure 19**). There are some unique aspects present in these two pairs. First, are the distinct PAS A-PAS B configurations within HIF1α/HIF2α not previously observed in the pairs aforementioned (**Figure 21** and **22**). Second, in the HIF2α:ARNT, additional contacts are observed between PAS A and DNA; specifically N184 and K186 on the Gβ-Hβ loop of the HIF2α (**Figure 20**)[33]. Currently, this interaction is unique among crystallized PAS dimers. Third, the ARNT PAS A repeat not only interacts with the HIF1α/HIF2α PAS A, but also interfaces with the HIF1α/HIF2α PAS B repeats (**Figure 21** and **22**). To establish these connections, the ARNT PAS A relies on what appears to be the Fα helix, Gβ, and Hβ strands while HIF1α/HIF2α use a portion of the linker between its PAS A-PAS B repeats and the Hβ and Iβ strands[33].

In the HIF1α:ARNT dimer, the Fα on HIF1α plays an important role for PAS B-PAS B stabilization, as well as the apparently disordered region between AA 247-279[33]. On ARNT, the Hβ, Iβ, and the Hβ-Iβ loop establish the necessary connections with HIF1α (**Figure 23**). Since HIF1α and HIF2α are highly homologous at the PAS domain and repeats, it is expected that the protein interfaces are the same for both proteins. Interestingly, for HIF2α the Aβ, Bβ, Aβ-Bβ

loop, Fα helix, and AA 353-359, disordered and located outside of PAS B, all interact with ARNT PAS B[33]. This is not the case with HIF1α, therefore homology does not appear to be the only requirement necessary for PAS B-PAS B interactions between HIF1α/HIF2α and ARNT.

The NPAS1:ARNT and NPAS3:ARNT Dimers: As their name implies, neuronal PAS (NPAS) proteins reside in the nervous system and mutations have been linked to schizophrenia and autism[34-36]. Little is known about their pathway although some binding partners have been identified. For instance, NPAS4 can dimerize with ARNT, ARNT2, and ARNTL2[15]. NPAS4 shares homology of ~ 30% and 43% with NPAS1, at the bHLH and PAS A domains, respectively[15]. NPAS1 shares similarity of ~76%, 98%, and ~47% in the bHLH, PAS A and PAS B domains.[15] Thus, it is possible that NPAS1 and NPAS3 may also dimerize with ARNT2 and ARNTL2, much like NPAS4.

The NPAS dimers have a structure like HIFα:ARNT dimers where the ARNT PAS A interacts with NPAS PAS A and PAS B repeats, while the ARNT PAS B only interacts with the NPAS PAS B. Also, like most other PAS dimers discussed before, NPAS1:ARNT (pdb:5sy5) (**Figure 24**) and NPAS3:ARNT (pdb: 5sy7) (**Figure 25**) PAS A-PAS A connections are stabilized by their A'α helices and beta sheets [37] (**Figure 26**). In the NPAS1:ARNT dimer, the NPAS1 PAS B- ARNT PAS A interaction relies on the linker between ARNT PAS A-PAS B, the Fα helix, Gβ, Hβ, and Iβ strands; while the NPAS1 PAS B uses the linker between PAS A-PAS B, Gβ, Hβ, Iβ, and a loop near Gβ-Hβ (AA: 366-369)[37]. Although NPAS1 appears to have its PAS A and PAS B repeats in close contact, only the Hβ on PAS A and Gβ on PAS B appear to interconnect (**Figure 27**). For PAS B-PAS B configuration, the NPAS1 uses helices Eα and Fα and the Aβ-Bβ loop as well as AA 406-421, which appear to be a loop followed by a helix outside the PAS B domain[37](**Figure 28**). The ARNT counterpart requires more regions to coordinate interaction: Bβ, Bβ-Cα loop, Cα, Dα, Hβ, Hβ-Iβ loop and the Iβ. Based on these observations, it can be hypothesized that NPAS1 requires a lot of flexibility to interact and possibly carry out its biological function.

In NPAS3:ARNT, the ARNT PAS A employs the Eα, Fα helices, Gβ, Gβ-Hβ loop and Hβ for PAS A to establish contact with the NPAS3 PAS B. The NPAS3 PAS B also utilizes a lot of flexible regions (linker and the loop following PAS B at position 438-443) to establish contact with the ARNT PAS A, as well as some beta strands (Hβ and Iβ)[37]. Interestingly, the NPAS3 PAS A has very little contact with its PAS B (**Figure 28**), however both repeats interact with the ARNT linker. For PAS B-PAS B (**Figure 29**) interconnectivity, NPAS3 relies on the Aβ-Bβ loop, Eα, Eα-Fα loop, Fα and a helix outside of PAS B at positions 449-453; ARNT uses all structures except Eα, Fα and Gβ and the loops between them. This phenomena is similar to what is observed in the NPAS1:ARNT interactions and may happen as result of the high homology between NPAS1 and NPAS3.

The AHR:ARNT and AHRR:ARNT Dimers: The prototype PAS heterodimer, historically, has been the aryl hydrocarbon receptor (AHR) and its partner ARNT. This is because AHR is the only known mammalian PAS protein to be activated by ligand binding. When hydrophobic ligands diffuse through the cell membrane and migrate into the cytosol, they encounter the un-activated AHR. Once activated, the AHR transforms and dimerizes with nuclear ARNT. The AHR:ARNT heterodimer employs its bHLH domains to bind response elements on target gene promoters for various enzymes and its repressor: the aryl hydrocarbon receptor repressor (AHRR). Unsurprisingly, the AHRR bears a striking resemblance to the AHR, however it lacks the PAS B repeat and is unable to bind ligands.

The AHR:ARNT (pdb: 5V0L) activated dimer has been crystallized at the bHLH domain and PAS A repeat[38] (**Figure 30**). The AHR PAS B repeat has proven difficult to crystallize. Hence several *in silico* predictions of the full length AHR:ARNT dimer also exist[39-41]. Much like other PAS dimers, PAS A-PAS A interactions are facilitated through A'α helix (**Figure 31**). *In silico* predicted structures imagine PAS B interactions with ligand and ARNT in configurations similar to HIF2α:ARNT and CLOCK:ARNTL yielding very different models. It is unlikely that AHR will resemble either PAS dimer because unlike its predecessors, it binds hydrophobic ligands. It

seems more plausible that the AHR:ARNT will also have a unique PAS B-PAS B configuration.

The AHRR:ARNT (pdb: 5Y7Y) dimer represents the second major anomaly in the PAS family[42] (**Figure 32**). Just like before, the AHRR:ARNT dimer stabilizes PAS A-PAS interactions by swapping the A'α helix for interactions with the opposite protein (**Figure 33**). However, there are subtle differences between AHR and AHRR PAS A repeats that may account for differences in their binding modes. Firstly, from the crystal structure the AHRR PAS A is approximately 20 amino acids larger than the AHR PAS A. A second important observation is that the Eα helix is longer than the Fα helix; the Eα helix does not appear to have direct contacts with the ARNT. Additionally, an extra helix following the Fα helix on the AHRR, which we refer to as F2α, contributes to ARNT PAS B interactions along with the Fα helix and the Eα-Fα loop. Lastly, the Hβ-Iβ loop is roughly 17 amino acids and, if HIF2α can be taken as an example, may be involved in DNA interactions that are worth future exploration. The ARNT PAS B establishes contacts with the AHRR PAS A through its linker, Aβ, Bβ, Bβ-Cα loop, Fα-Gβ loop, Hβ, Hβ-Iβ loop, and Iβ (**Figure 34**). All the deviations observed on AHRR are of interest because they allow interactions with both the ARNT PAS A and PAS B repeats. As most ARNT bound PAS proteins show the ARNT PAS B completely separated from the partner protein PAS A, understanding the role of this interaction in transcription could also provide more clues about how AHRR carries out its repressor function.

Similarities and Differences among PAS dimers

Many lessons have been learned from crystal structures. In analyzing the overall structures, it is easy to see that all proteins are asymmetrical, tightly wound with some form of crossing over along the two major domains, bHLH and PAS. Across, the PAS A repeat is also bigger than the PAS B repeat (**Table 1, Figure 3, Figure 35**) likely because of bigger loops in the Fα-Gβ, Gβ-Hβ, and the Hβ-Iβ loops. Furthermore, in HIF2α the Gβ-Hβ loop establishes direct contacts with response elements on the DNA. This is an important finding because loops

are often missing from crystal structures because they are hard to resolve and tend to be predicted. It is possible that resolving more loops could help us understand the role they play in transcription. Invariably, the PAS A repeats also use accessory motifs – A'α helices – to stabilize interactions. The mouse PER2 homodimers are interesting because they lack the A'α helix and do not rely on PAS A repeat interactions for dimerizations. However, PER1 and PER3 do have A'α helices and do not use them for PAS A interconnectivity; instead, they rely on their Fα helices for stabilization. These findings together provide additional evidence that PAS A play a universal role in bringing PAS proteins together and facilitating dimerizations.

Perhaps the most captivating revelation have been the interactions observed between PAS B repeats. Roughly 30 amino acids shorter and denser than PAS A repeats, PAS B repeats show overall complex and different configurations in different orientations. The Fα helix, Hβ-Iβ loop play prominent roles in helical "front"- β-sheet "backs" contact for most dimers. As with all things in life, some "rules" are meant to be broken, so it follows that mouse PER homodimers and the AHRR:ARNT heterodimer interact in unexpected ways that use the PAS B repeat.

Although ARNTL and ARNT have been relegated to dimerization partners, crystal structures suggest they may play a more prominent role in the transcription process. For example, ARNT wraps around its partner and the ARNT PAS A interacts with both PAS A and PAS B repeats on the partner protein. In contrast, ARNTL wraps around its partner in a more linear fashion whereby the ARNTL PAS A interacts with the CLOCK PAS A and so on. These major differences may also provide clues about partner selection and adaptability to the needs of the partner protein.

CONCLUSION

The PAS family of transcription factors are an exciting set of proteins that showcase different ways of heterodimerization; this has become apparent from crystal structures resolved in the past decade. Crystal structures have allowed us to see that: (1) PAS domains are composed of two repeats, A and B, and a linker; (2) PAS repeats begin with a beta strand and end with a beta strand; (3) the PAS A repeat tends to be bigger than the PAS B repeat likely because of bigger loops; (4) the PAS A repeat heterodimerizes by using an accessory motif, the A'α helix; and (5) amino acid alignments are misleading for the PAS family. Many of these observations hold true for most PAS proteins, however PER and AHRR proteins break these rules. For this reason, they are listed as a separate class deemed class γ because they differ structurally and in their dimerizations.

Although a lot of the terminology discussed here has been historically used, few publications have explicitly defined it and thusly, many words are employed in the literature to mean the same thing. Through this work, we emphasize the importance of a common terminology for the PAS family with a structural basis thereby continuously moving the field forward. Furthermore, we expect that as more structures are resolved, we continue to add to these observations thereby elucidating the complex interactions that govern PAS protein function and signal transduction.

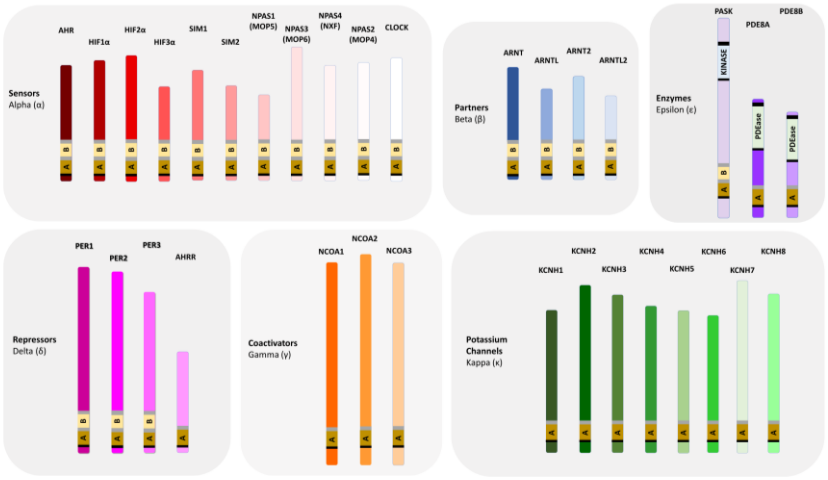


Figure 1: The PAS family of proteins. There are 33 PAS proteins with significant roles. In this proposed classification scheme, we disseminate them into class: (i) alpha sensors, (ii) beta dimerization partners, (iii) epsilon enzymes, (iv) delta repressors, (v) gamma coactivators, and (vi) espsilon potassium voltage-gated channels.

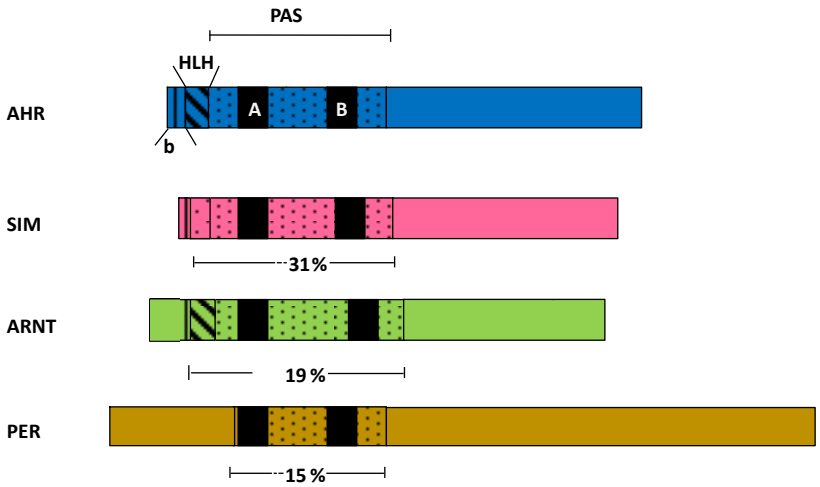


Figure 2: Classical view of PAS family founding members. The PAS domain was first described by Nambu et. al (1991); through amino acid alignments and sequence identity, the group showcased a new family of proteins – PAS. Further studies in our lab found sequence identify among the founding members to range between 15 to 31%; however 3D analysis confirms highly conserved structures at the PAS domain.

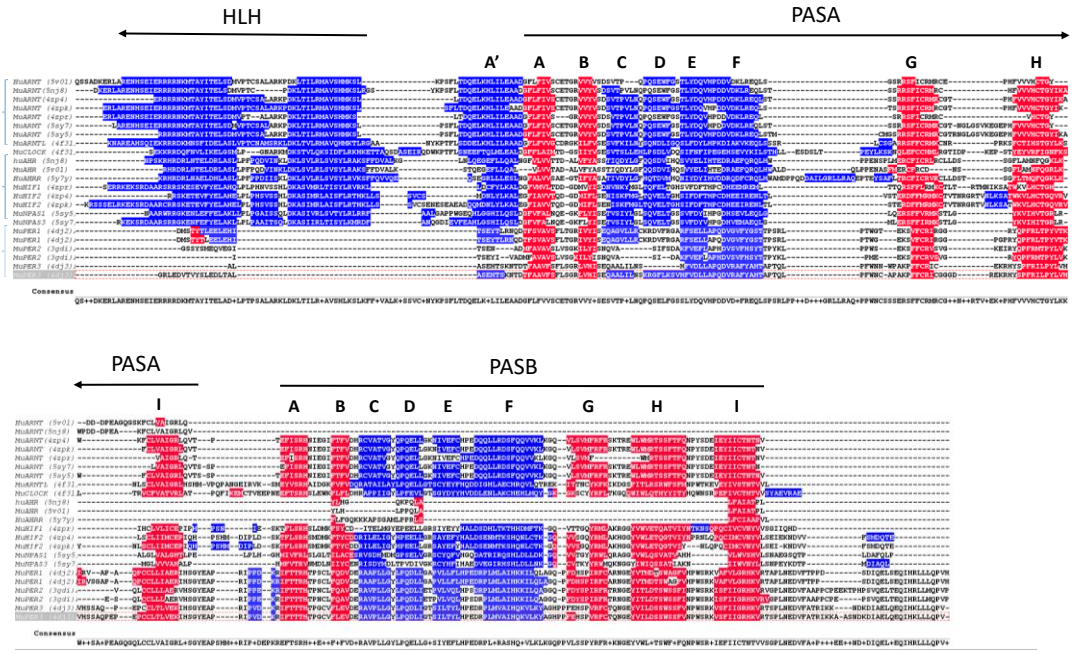


Figure 3: Sequence alignment for PAS proteins used in this study. Using Jalview and crystal structures from the Protein Data Bank, we aligned sequences to determine common structural elements observed. The major domains are the DNA binding domain, helix-loop-helix (HLH), and the PAS domain, composed of PAS A and PAS B repeats. Blue = helix in crystal; red = beta-strand in crystal

Table 1: PAS heterodimer learning set. The PAS proteins from the PDB used in this study and their structural features are outlined in detail. Additionally, the chain ID, the amino acid position, and the length of each structural feature is also represented.

								PAS A			PAS B		
PDB #	PAS Protein	Chain ID	Uniprot #	bHLH	Length	A'α helix	Length	Beginning	End	Length	Beginning	End	Length
	mCLOCK	B	O08785	71-126	**55**	143-154	**11**	239	319	**80**	339	440	**101**
4F3L	mARNTL	A	Q9WTL8	43-89	**46**	106-118	**12**	120	256	**136**	275	375	**100**
	mHIF2a	B	P97481	29-75	**46**	87-93	**6**	98	228	**130**	243	343	**100**
4ZPK	mARNT	A	P53762	100-142	**41**	160-172	**12**	174	343	**169**	362	463	**101**
	mHIF1a	B	Q61221	16-73	**57**	90-96	**6**	99	225	**126**	242	341	**99**
4ZPR	mARNT	A	P53762	87-140	**53**	161-173	**12**	175	342	**167**	362	462	**100**
	mNPAS1	B	P97459	53-103	**50**	137-145	**8**	148	290	**142**	305	405	**100**
5SY5	mARNT	A	P53762	99-140	**41**	160-173	**13**	174	342	**168**	362	462	**100**
	mNPAS3	B	Q9QZQ0	60-115	**55**	149-163	**14**	165	320	**155**	337	438	**101**
5SY7	mARNT	A	P53762	90-140	**49**	160-173	**13**	175	343	**168**	363	462	**99**
	mAHR	B	P30561	37-80	**43**	114-119	**5**	125	264	**139**	N/A	N/A	**N/A**
5V0L	hARNT	A	P27540	90-142	**52**	160-172	**12**	177	339	**162**	N/A	N/A	**N/A**
	hAHRR	A	A9YTQ3	39-90	**51**	116-123	**7**	125	276	**151**	N/A	N/A	**N/A**
5Y7Y	bARNT	B	Q9BE97	98-144	**46**	160-172	**12**	174	342	**168**	362	463	**101**
4DJ2	mPER1	A,B	O35973	N/A	**N/A**	201-213	**12**	219	343	**124**	359	461	**102**
3GDI	mPER2	A,B	O54943	N/A	**N/A**	N/A	**N/A**	190	314	**124**	330	434	**104**
4DJ3	mPER3	A,B	O70361	N/A	**N/A**	121-126	**5**	131	253	**122**	270	372	**102**

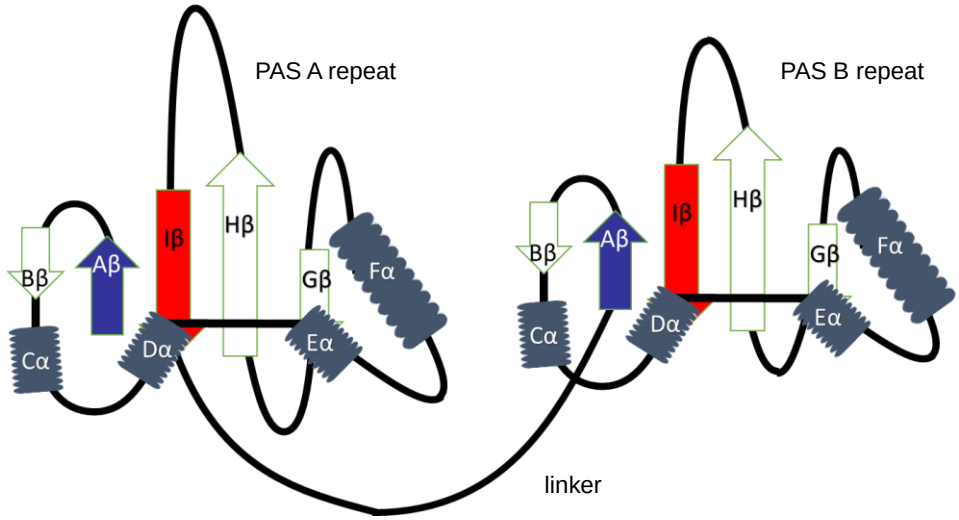


Figure 4: Generic PAS domain defined. Cartoon representation of a PAS domain. Each PAS domain is composed of any number of PAS repeats however it is common to see 1 or 2 repeats. In such an arrangement the PAS A is ~150 amino acids, followed by a linker ~20 amino acids,, and a PAS B repeat of ~100 amino acids. Blue indicates the start of the repeat and red is the end of the repeat.

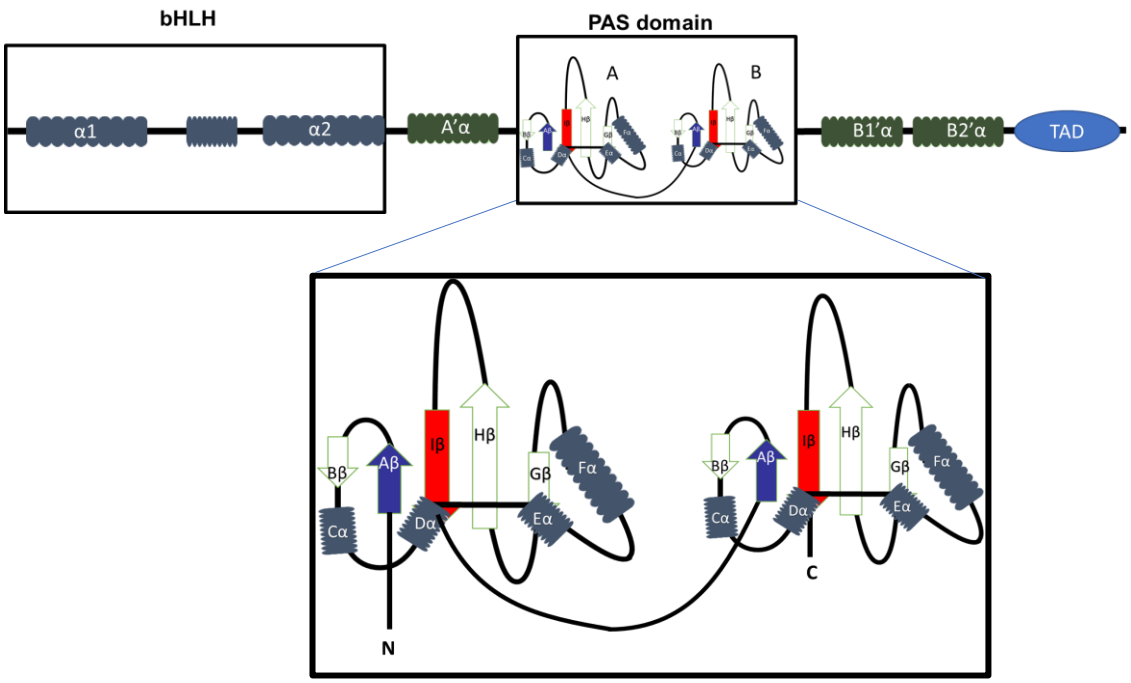


Figure 5: Linear representation of a generic PAS protein. In this chapter, most of the proteins studied contain both PAS repeats. From right to left (N- to C-terminus) the bHLH DBD is apparent, followed by the accessory motif A'α helix. This helix is involved in PAS A-PAS A repeat interactions. Immediately following the accessory motif is the PAS domain, ending with additional helices and the TAD. Image is not drawn to scale; red indicates beginning of repeat, while royal blue is the end of the repeat.

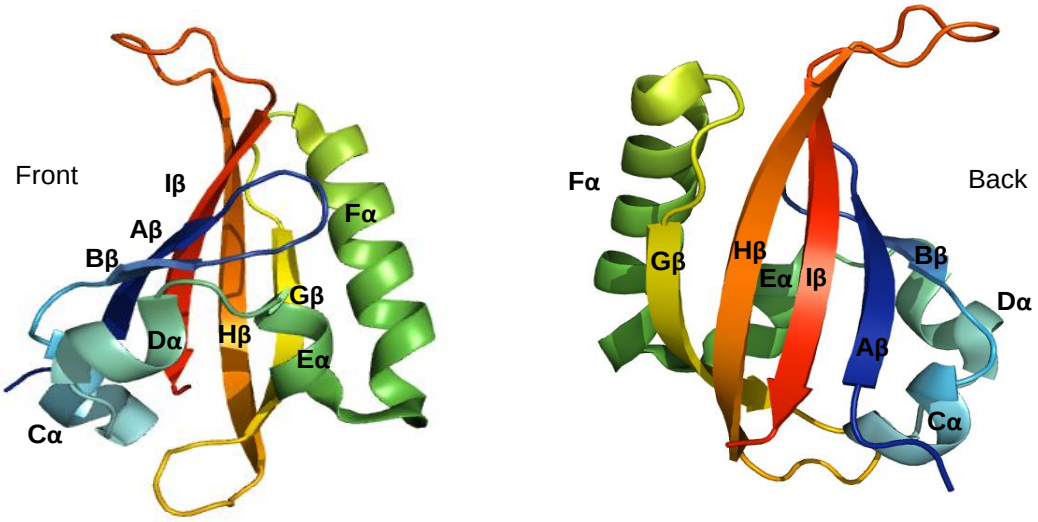


Figure 6: Representation of mammalian PAS domains. Using the ARNT PAS repeat, the 5 anti-parallel β-strands and 4 helices are prominent. Image adapted from (pdb:5y7y) and colored to show direction. (Left) The “front” of the PAS domain forms a cavity and contains ~4 helices. (Right) The “back” of the PAS domain forms a sheet.

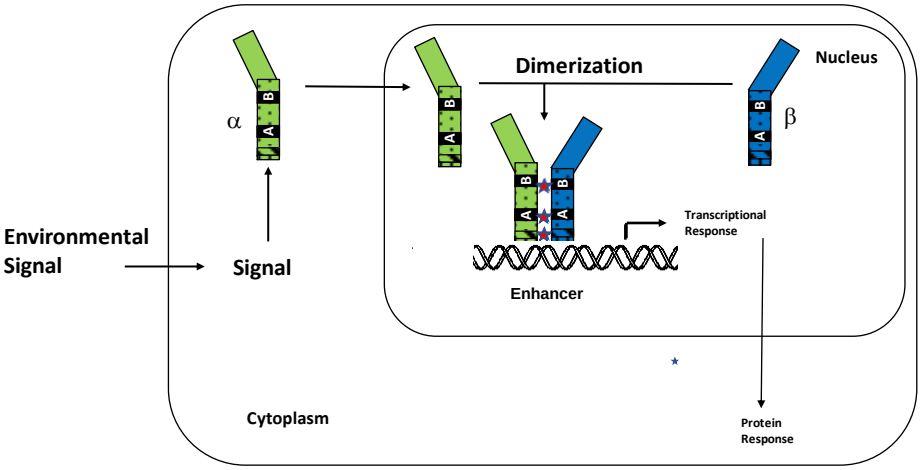


Figure 7: Classical schematic for class α-β dimerization. As sensors, the class α must detect an environmental signal to begin its transcriptional pathway. In this classical representation,, the inactivated sensor is found in the cytoplasm; upon detection of the environmental signal, the sensor moves into the nucleus where it dimerizes, using bHLH, PAS A and PAS B repeats symmetrically with a class β partner. Together, they subsequently bind enhancer elements on the DNA to initiate the proper physiological response.

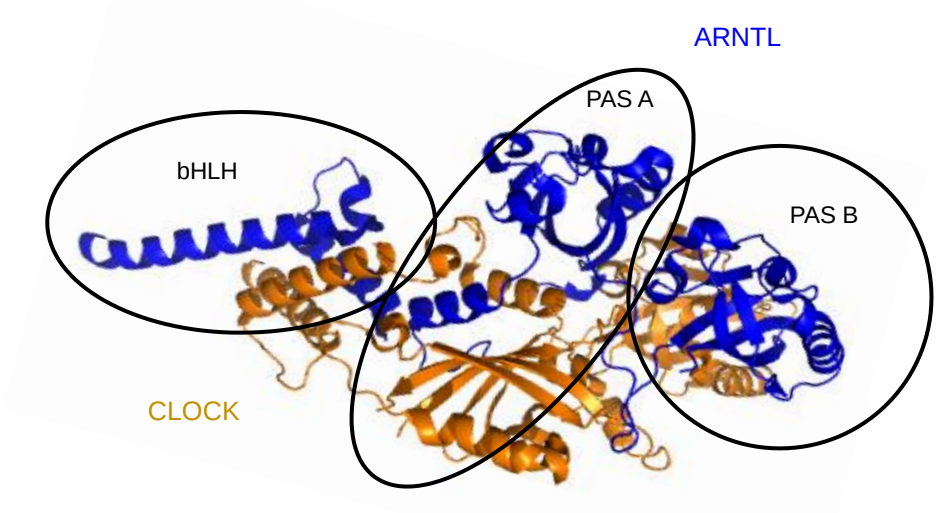


Figure 8: The Mammalian PAS dimer CLOCK:ARNTL (pdb:4F3L). Within the current classification scheme, CLOCK is a class α sensor and ARNTL is a class β dimerization partner. The two major domains on both proteins were shown to interact $bHLH_{CLOCK}$-$bHLH_{ARNTL}$, and PAS_{CLOCK}-PAS_{ARNTL}. Within the PAS domain, there are interactions between PAS A_{CLOCK}-PAS A_{ARNTL} and PAS B_{CLOCK}-PAS B_{ARNTL}

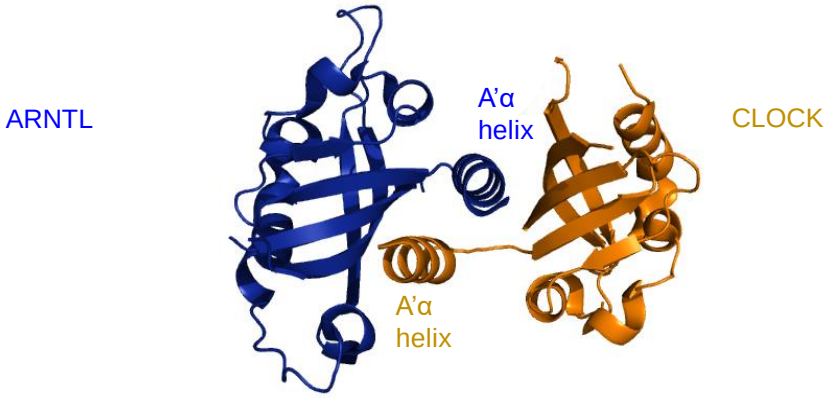


Figure 9: PAS A interactions stabilized by A'alpha helix. As the first resolved mammalian heterodimer with both PAS A and B repeats, it pioneered the idea that the A'α helices cross an imaginary midline to interact with the back of the beta sheets. In contrast, the "face" containing the cavity points away from this midline.

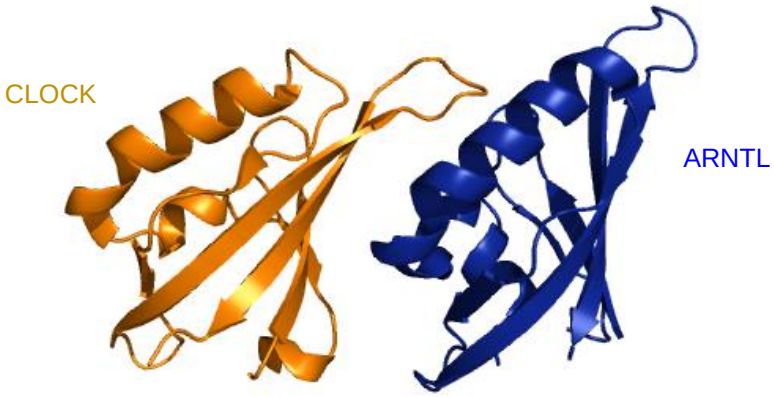


Figure 10: CLOCK-ARNTL PAS B interactions. The CLOCK protein is known to interact exclusively with ARNTL. The PAS B repeat dimer shows a “back to face” configuration where the back of the CLOCK beta sheet and the Hβ-Iβ loop interacts with the Fα helix on the “face” of the ARNTL PAS B repeat.

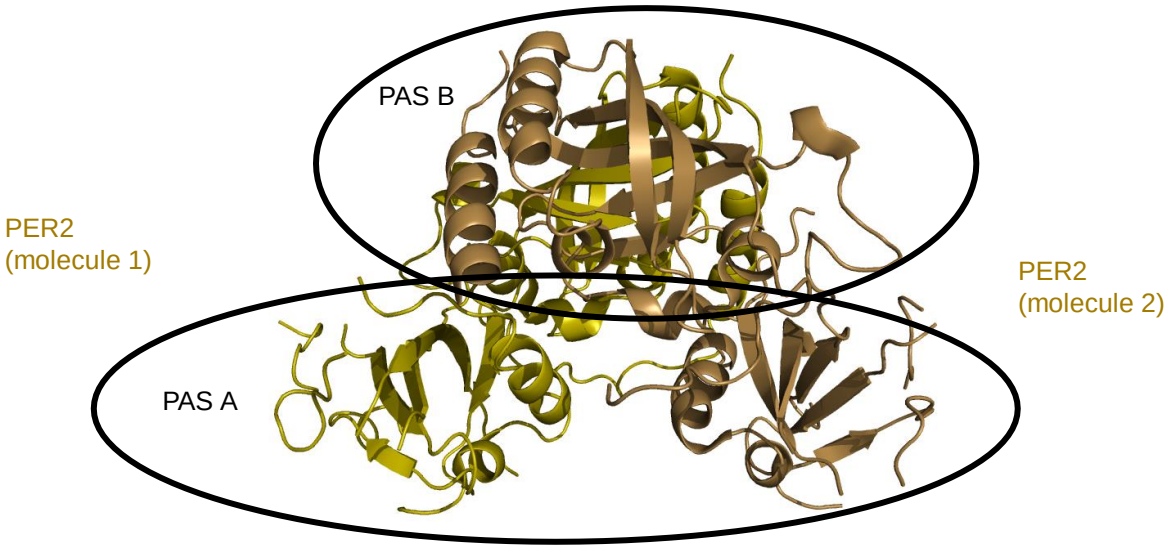


Figure 11: The PER2 homodimer is the anomaly in the anomaly group (pdb: 3gdi). Within the PAS family, PERs are unusual because they can homo- and heterodimerize with each other. The PER2 homodimer lacks the A'α helix and does not use PAS A repeats to dimerize, relying more heavily on PAS B repeat interactions.

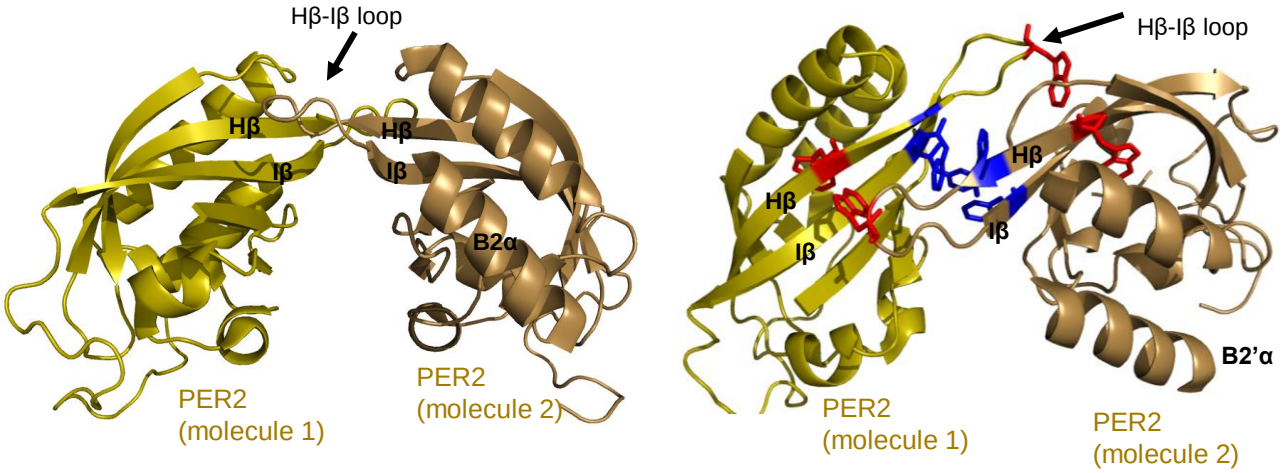


Figure 12: Reliance on PAS B repeats for PER2 homodimer. Mammalian PER2 homodimers use Hβ, Iβ and the Hβ-Iβ loops to stabilize PAS B repeat interactions (left). Stacking interactions through phenylalanine (blue) and tryptophan (red) residues on Hβ, Iβ, and Hβ-Iβ loop are observed (right). The crystal structure also shows B2'α – a helix C-terminus of the PAS B repeat that harbors a nuclear export signal (NES) **.**

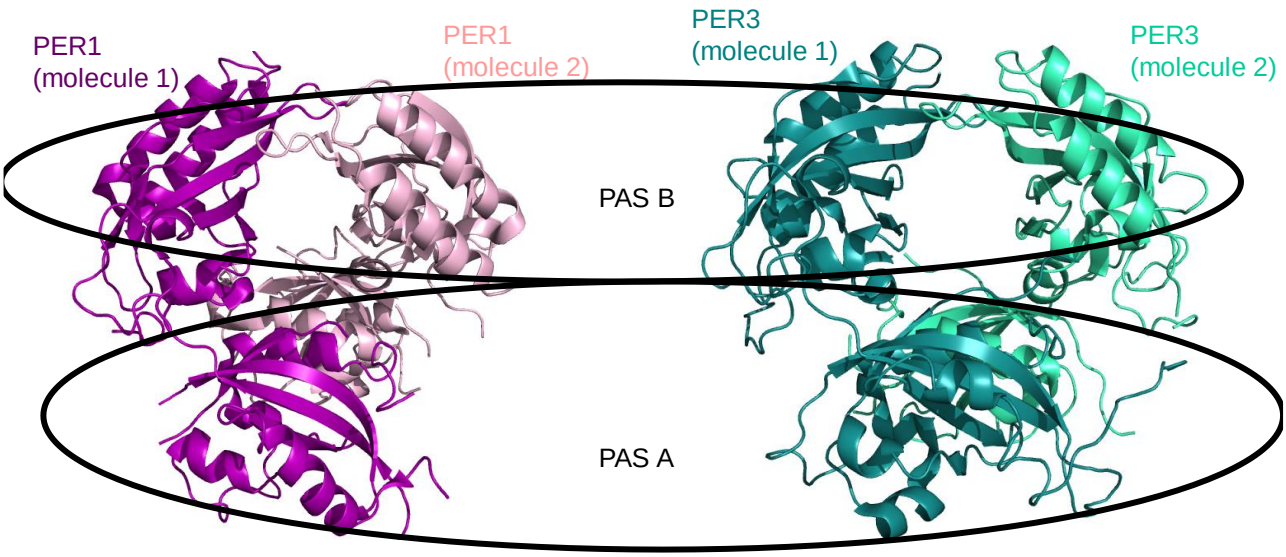


Figure 13: Mammalian PER1 and PER3 homodimers. In these homodimers, PAS A-PAS A and PAS B-PAS B interactions are observed. Unlike PER2, both PER1 and PER3 contain A'α helices, however they are not used in PAS A-PAS A interactions.

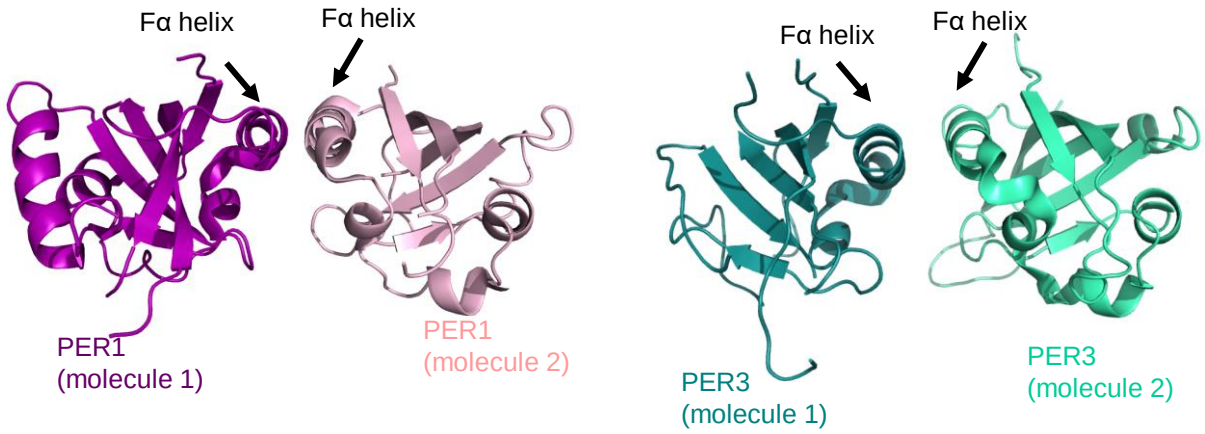


Figure 14: Mammalian PER1 and PER3 PAS A repeat dimers. Both PER1 and PER3 contain the A'α helix but don't rely on them for PAS A-PAS A interactions. Instead, each molecule relies heavily on the Fα helix to interact; interestingly, these helices do not cross the "midline" and do not interact with the back of the beta sheet of the other.

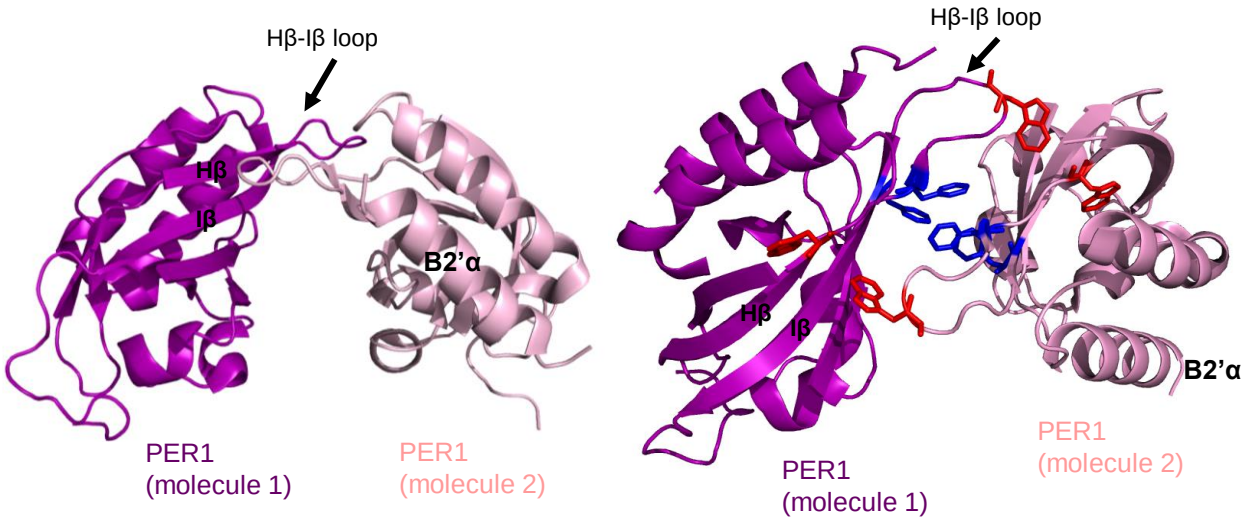


Figure 15: PAS B repeat interactions in PER1 homodimers. Mammalian PER1 homodimers use Hβ, Iβ and the Hβ-Iβ loops to stabilize PAS B repeat interactions (left). Stacking interactions through phenylalanine (blue) and tryptophan (red) residues on Hβ, Iβ, and Hβ-Iβ loop are observed (right). The crystal structure also shows B2'α – a helix C-terminus of the PAS B repeat that harbors a nuclear export signal (NES) **.**

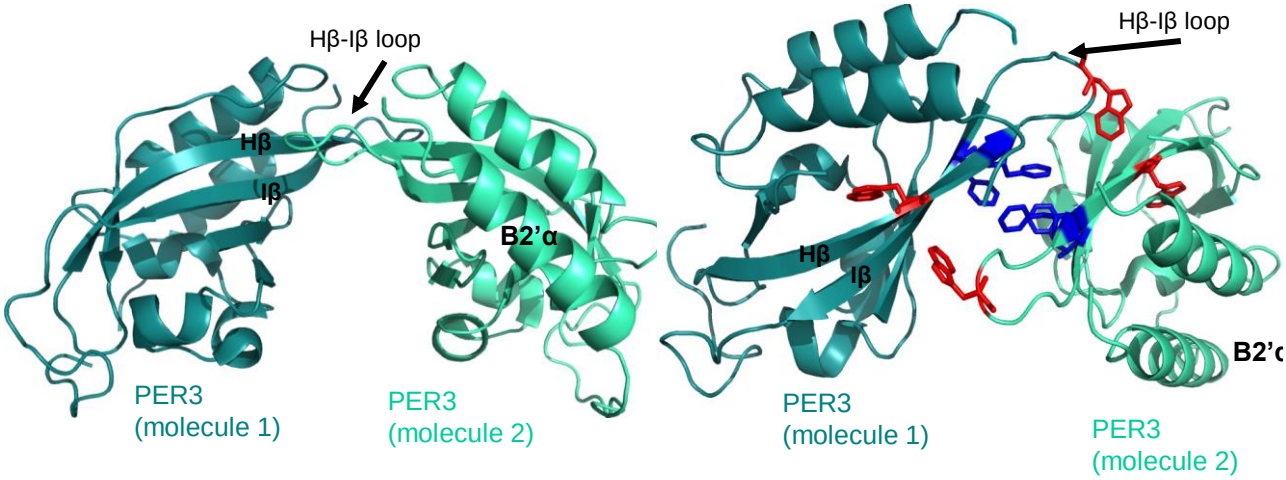


Figure 16: PAS B repeat interactions in PER3 homodimers. Mammalian PER3 homodimers use Hβ, Iβ and the Hβ-Iβ loops to stabilize PAS B repeat interactions (left). Stacking interactions through phenylalanine (blue) and tryptophan (red) residues on Hβ, Iβ, and Hβ-Iβ loop are observed (right). The crystal structure also shows B2'α – a helix C-terminus of the PAS B repeat that harbors a nuclear export signal (NES) **.**

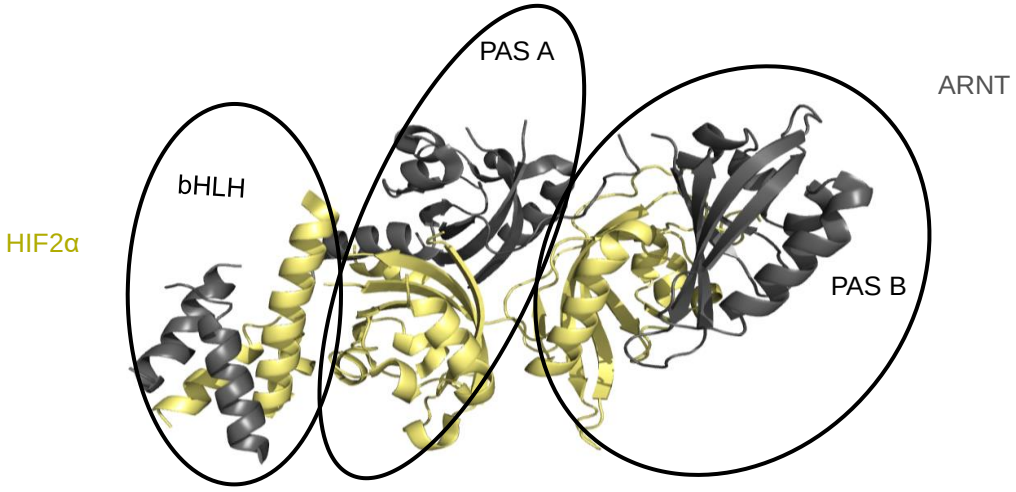


Figure 17: Mammalian HIF2α-ARNT (pdb:4zpk). This is a different configuration from CLOCK:ARNTL but some elements remain the same: In this dimer, HIF2α is the class α sensor and ARNT is the class β dimerization partner. Overall, the $bHLH_{HIF2α}$-$bHLH_{ARNT}$ and the $PAS_{HIF2α}$-PAS_{ARNT} domains interact. Unique to ARNT dimerizations: the PAS $A_{HIF2α}$-and PAS B_{ARNT} have no direct interactions, PAS $A_{HIF2α}$-PAS $B_{HIF2α}$ interact, and PAS A_{ARNT} interfaces with both PAS $A_{HIF2α}$ and PAS $B_{HIF2α}$.

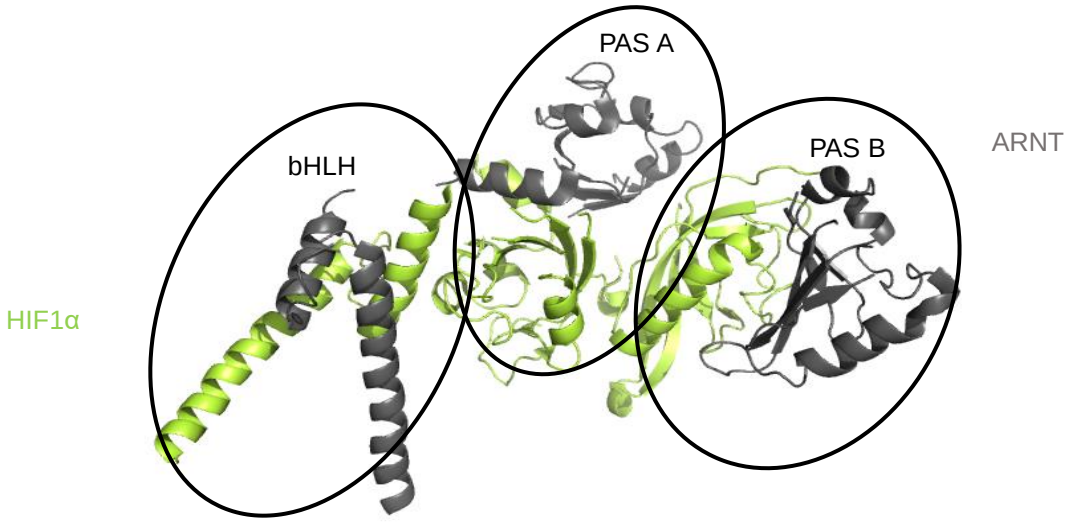


Figure 18: Mammalian HIF1α-ARNT (pdb:4zpr). In this dimer, HIF1α is the class α sensor and ARNT is the class β dimerization partner. The configurations is like that observed in HIF2α:ARNT. Overall, the $bHLH_{HIF1\alpha}$-$bHLH_{ARNT}$ and the $PAS_{HIF1\alpha}$-PAS_{ARNT} domains interact.

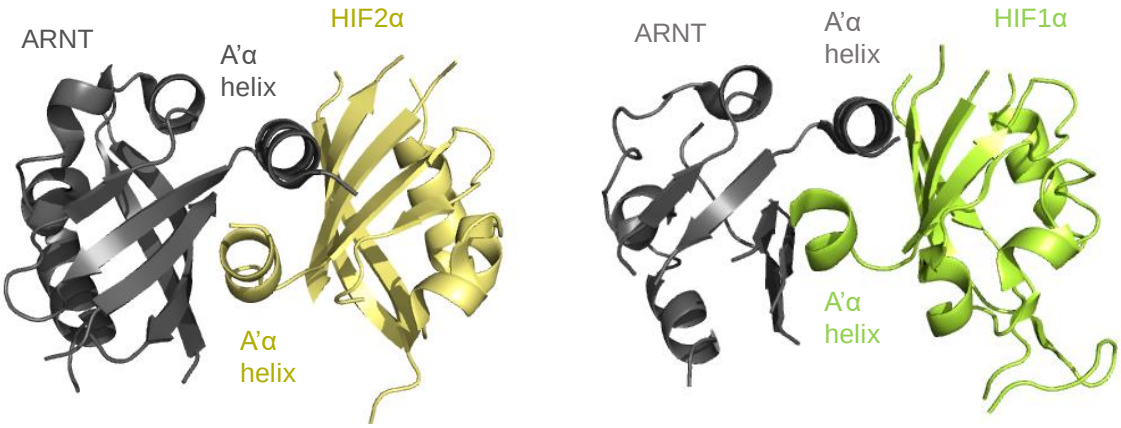


Figure 19: PAS A interactions for HIF2α:ARNT and HIF1α:ARNT. The HIF2α (pdb: 4zpk) and HIF1α (pdb:4zpr) proteins each dimerize with ARNT and rely on A'α helices to stabilize interactions between the PAS A repeats. In the case of HIF2α, interactions with DNA are also established using amino acids N184 and K186.

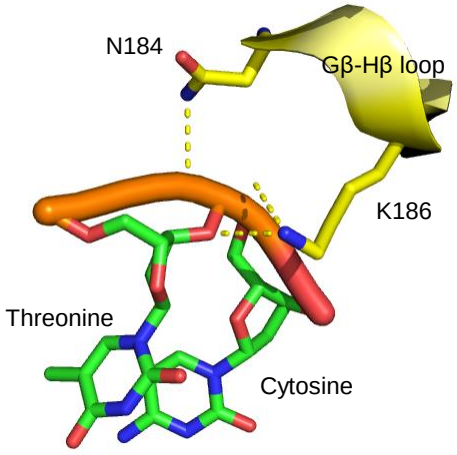


Figure 20: HIF2α: ARNT DNA contacts. Although PAS A is mostly known for dimerization, in HIF2α:ARNT, the N184 and K186 residues on the Gβ-Hβ loop interact with threonine and cytosine in DNA.

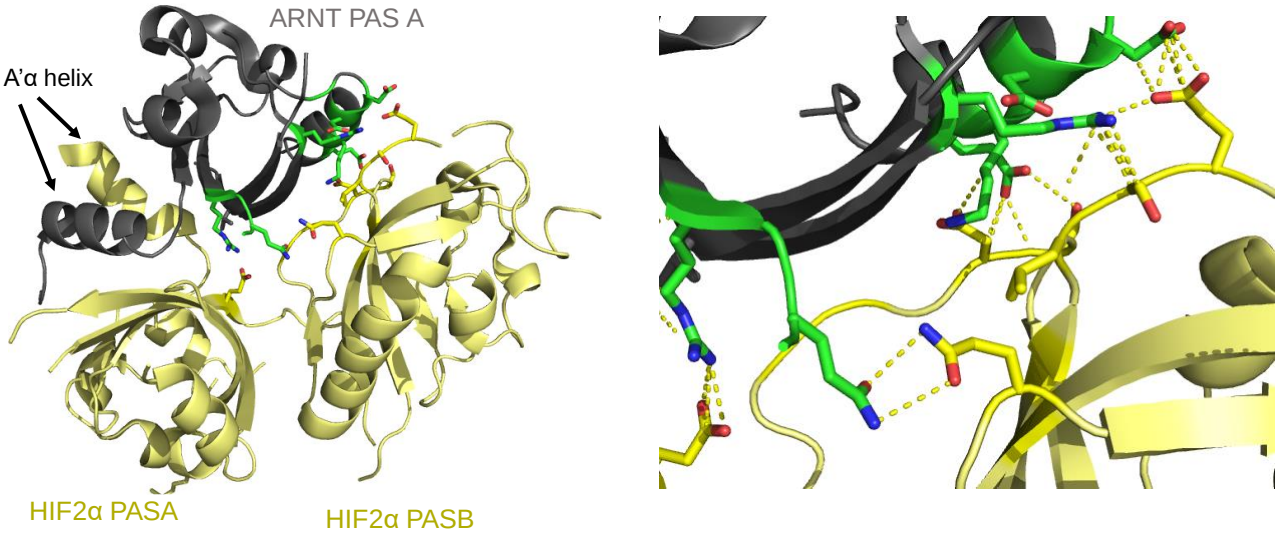


Figure 21: HIF2α PAS A and PAS B interactions with ARNT PAS A. *(Left)* The ARNT dimers have distinct conformation that relies on ARNT PAS A interacting with both repeats in HIF2α. *(Right)* Most of the interactions apparently require a portion of the disordered region. Connections are not obvious between the HIF2α PAS A and B repeats. Green = ARNT residues, yellow = HIF2α residues

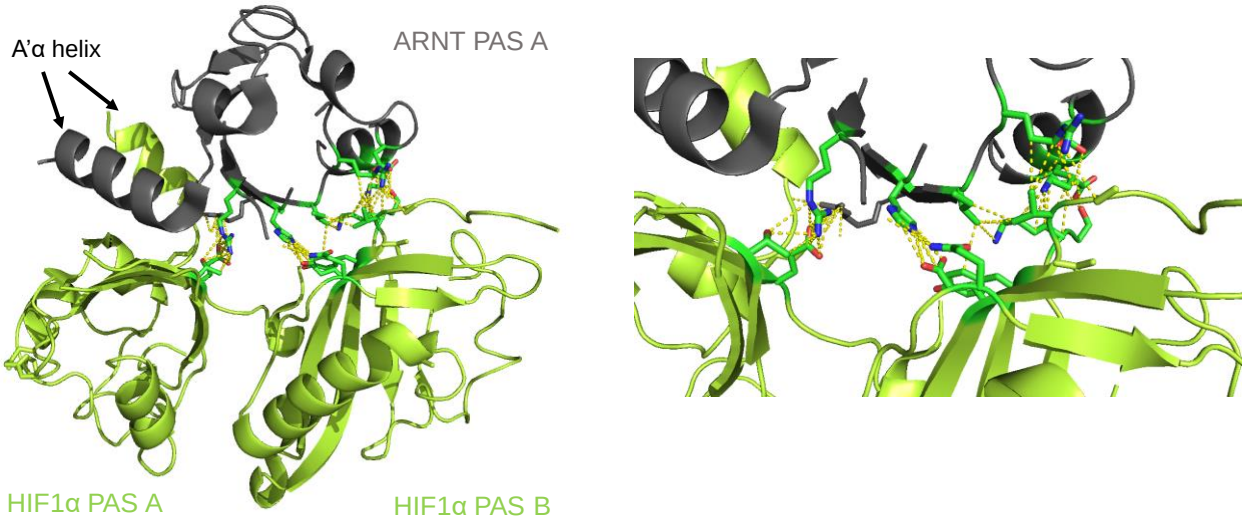


Figure 22: Intramolecular PAS A and PAS B interactions within HIF1α. *(Left)* The ARNT dimers have distinct conformation that relies on ARNT PAS A interacting with both repeats in HIF1α. *(Right)* Most of the interactions apparently require a portion of the disordered region. Connections are not obvious between the HIF1α PAS A and B repeats. Green = ARNT residues, lime green = HIF1α residues

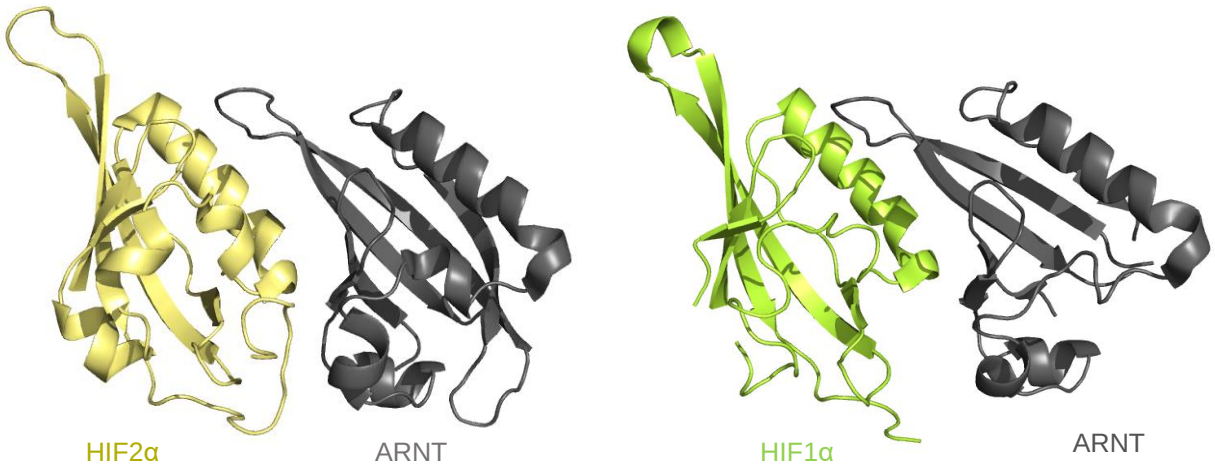


Figure 23: HIF2a: ARNT and HIF1α:ARNT PAS B interactions. The ARNT PAS B interacts exclusively with HIF2α and HIF1α PAS B repeats. Interaction is like CLOCK:ARNTL PAS B repeats.

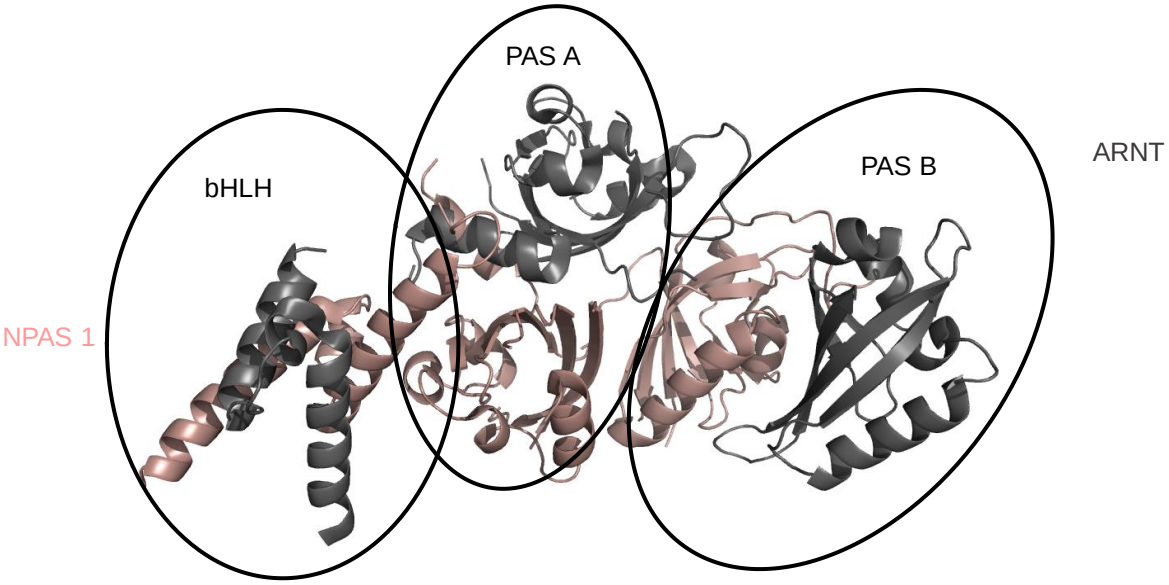


Figure 24: Mammalian NPAS1:ARNT dimer (pdb: 5SY5). In this dimer, NPAS1 is the class α sensor and ARNT is the class β dimerization partner. The configuration is like that observed in HIF2α:ARNT. Overall, the $bHLH_{NPAS1}$-$bHLH_{ARNT}$ and the PAS_{NPAS1}-PAS_{ARNT} domains interact.

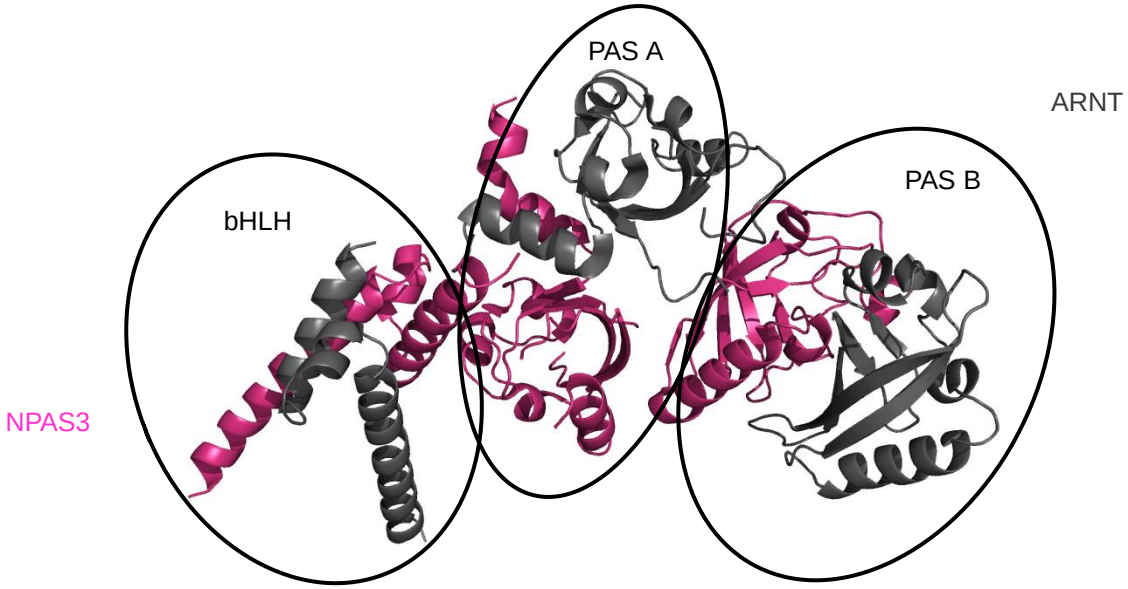


Figure 25: Mammalian NPAS3: ARNT. The NPAS3 is the class α sensor and ARNT is the class β dimerization partner. The configuration is like that observed in NPAS1:ARNT. Overall, the $bHLH_{NPAS3}$-$bHLH_{ARNT}$ and the PAS_{NPAS3}-PAS_{ARNT} domains interact.

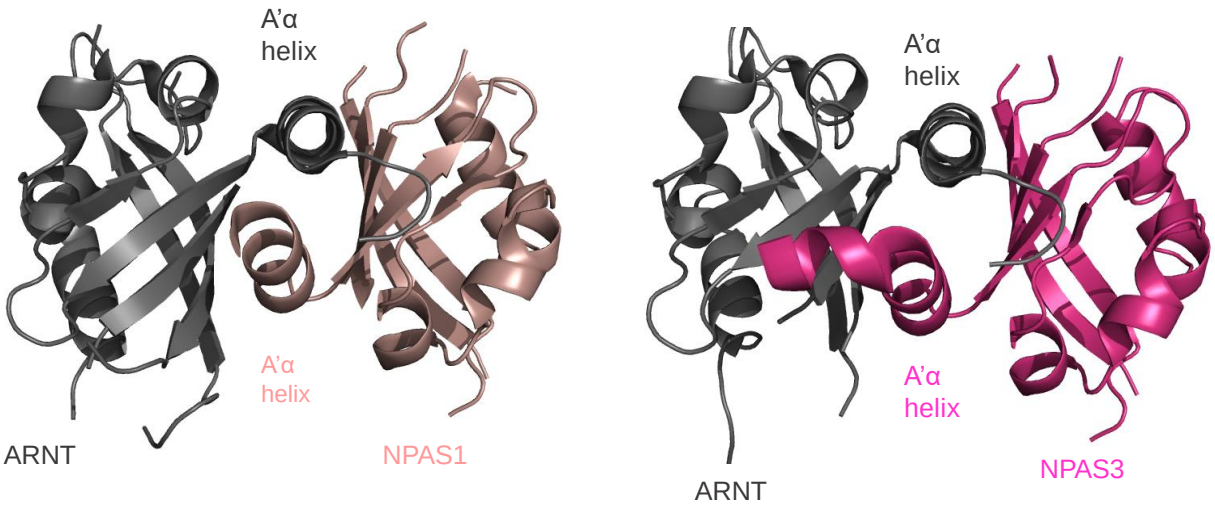


Figure 26: PAS A repeat interactions between NPAS1:ARNT and NPAS3:ARNT

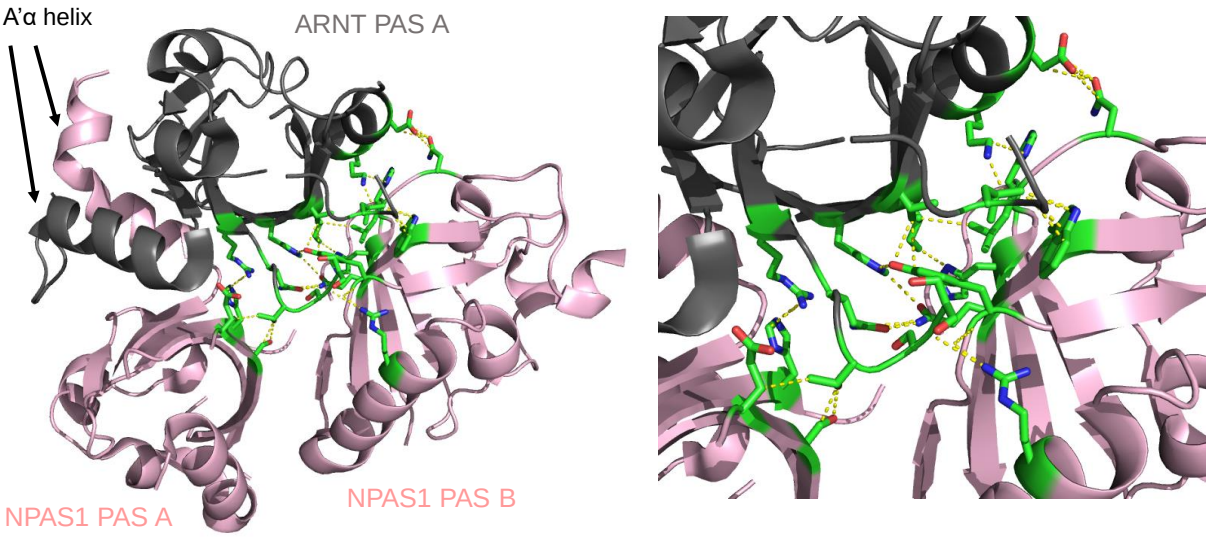


Figure 27: NPAS1 PAS A and B repeat interactions with ARNT PAS A. *(Left)* The ARNT PAS A interacting with both repeats on NPAS1. *(Right)* Most of the interactions apparently require a portion of the disordered region as well as the beta sheet. Some connections also exist between the NPAS1 PAS A and B repeats.

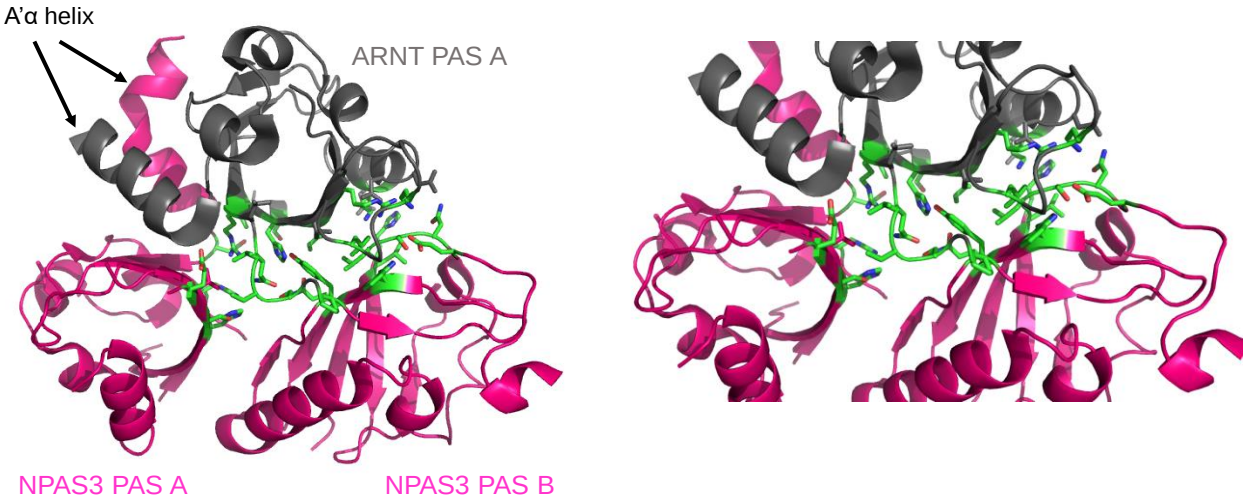


Figure 28: NPAS3 PAS A and B repeat interactions with ARNT PAS A. *(Left)* The ARNT PAS A interacting with both repeats on NPAS3. *(Right)* Most of the interactions apparently require a portion of the disordered region as well as the beta sheet. Some connections also exist between the NPAS3 PAS A and B repeats.

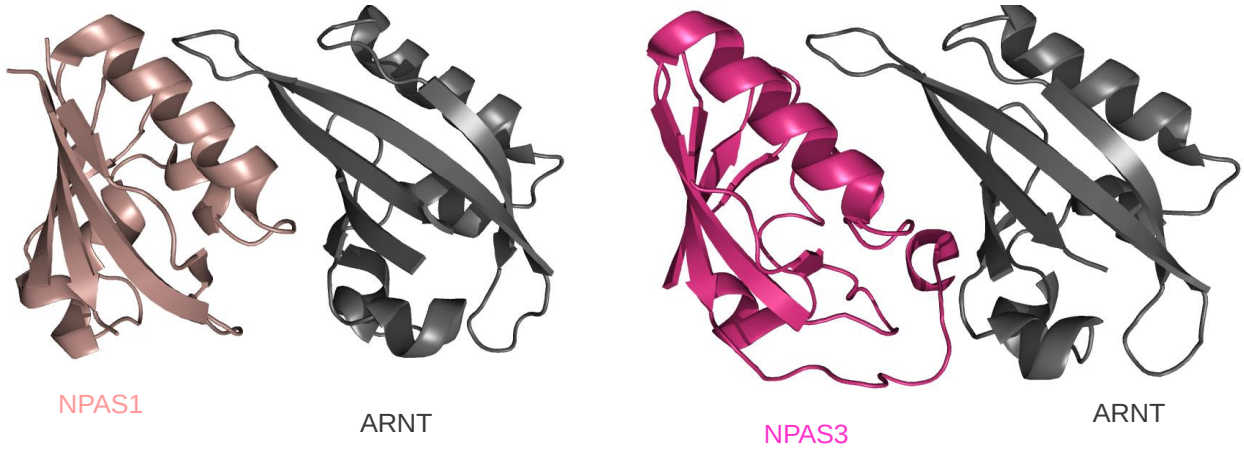


Figure 29: PAS B repeat interactions between NPAS1:ARNT and NPAS3:ARNT. The ARNT PAS B interacts exclusively with NPAS1 and NPAS3PAS B repeats. Interaction is like CLOCK:ARNTL PAS B repeats.

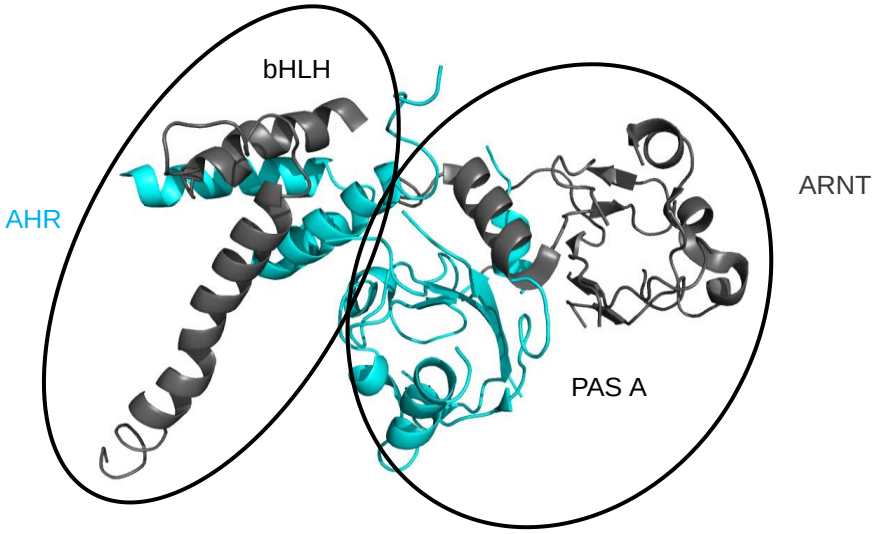


Figure 30: AHR:ARNT PAS A repeat interactions (pdb: 5V0L). The full AhR protein has not been crystallized, however the resolution of its bHLH and PAS A in dimer form with ARNT has been carried out.

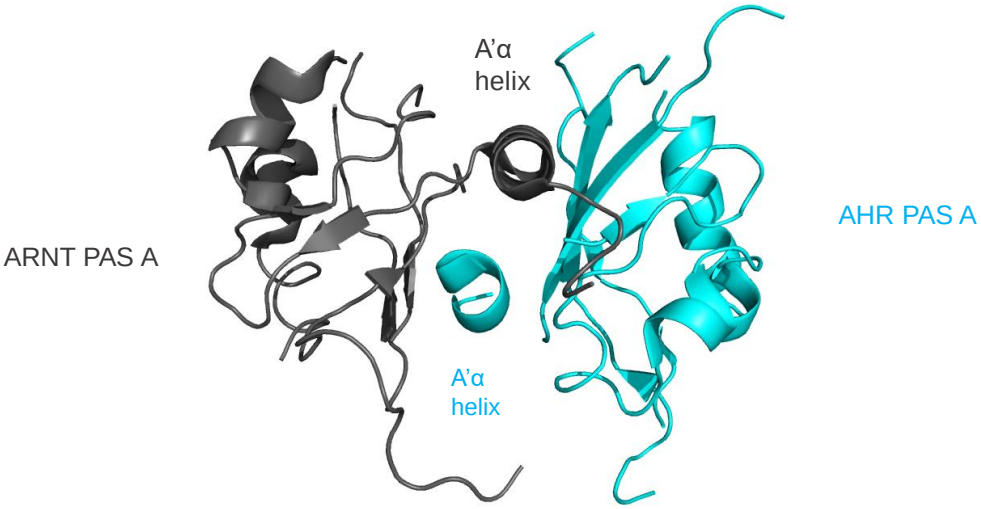


Figure 31: AHR PAS A- ARNT PAS A. As has been the pattern with most of the resolved dimers, the A'α helix cross the midline to interact with the back of the beta sheet, comprised of the 5 anti-parallel beta strands.

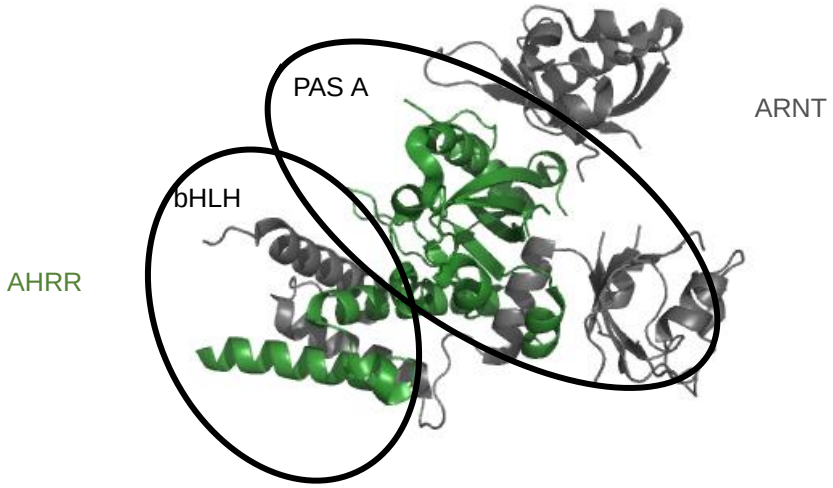


Figure 32: AHRR PAS A ARNT PAS A and B (pdb:5Y7Y). The AHRR belongs to the class δ and is the only δ-β dimer resolved. Since AHRR lacks PAS B, only the bHLH and PAS A for AHRR are present, while both ARNT PAS repeats are visible.

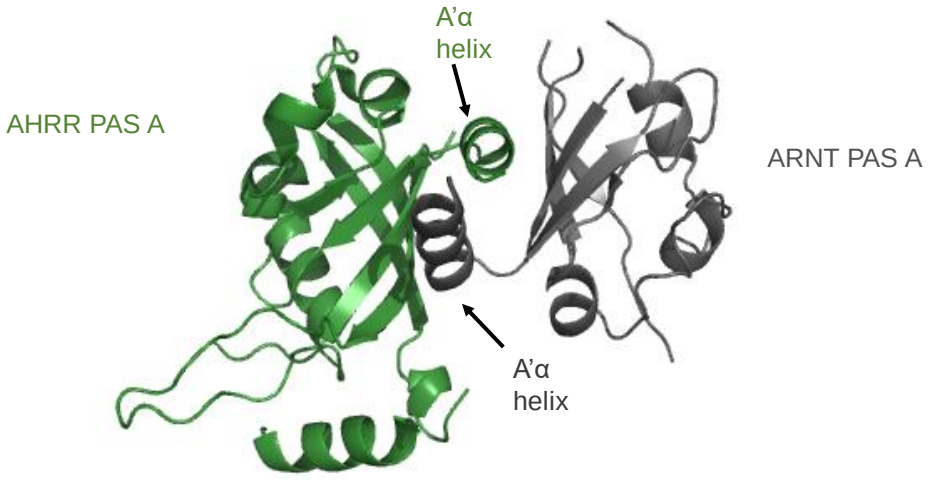


Figure 33: AHRR PAS A ARNT PAS A. Although this is a class δ-β dimer, the organization between PAS A-PAS A repeats is the same as class α-β dimers. Like other dimers before it, the A'α helices cross over to stabilize and facilitate interconnectivity.

Figure 34: AHRR PAS A ARNT PAS B. Without a PAS B repeat in the AHRR, the ARNT PAS B sits atop the AHRR PAS A**.** This configuration has not been observed in ARNT dimers thus it appears the ARNT partner may accommodate its partners to garner the appropriate biological response.

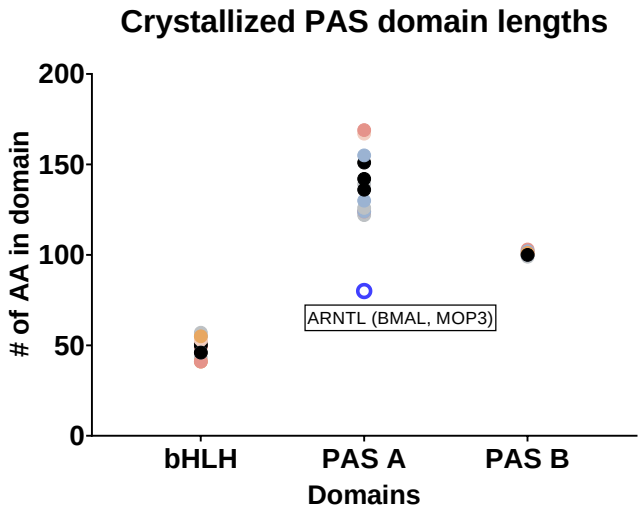


Figure 35: Domain and repeat lengths observed in crystallized heterodimers. Lengths from Table 1 for bHLH, PAS A and PAS B repeats were graphed. Since ARNT has been crystallized numerous times, it was only used once to avoid overrepresentation. The bHLH is ~50 AA, the PAS A varies between 80 and 170 AA, while PAS B is ~100 AA. The most obvious outlier is the ARNTL PAS A at about 80 AA .

REFERENCES

1 Mei, Q. & Dvornyk, V. Evolution of PAS domains and PAS-containing genes in eukaryotes. 385-405, doi:10.1007/s00412-014-0457-x (2014).

2 Carver, L. A., Jackiw, V. & Bradfield, C. A. THE 90-KDA HEAT-SHOCK PROTEIN IS ESSENTIAL FOR AH RECEPTOR SIGNALING IN A YEAST EXPRESSION SYSTEM. *Journal of Biological Chemistry* **269**, 30109-30112 (1994).

3 Carver, L. A., LaPres, J. J., Jain, S., Dunham, E. E. & Bradfield, C. A. Characterization of the Ah receptor-associated protein, ARA9. *Journal of Biological Chemistry* **273**, 33580-33587, doi:10.1074/jbc.273.50.33580 (1998).

4 Beischlag, T. V. *et al.* Recruitment of the NCoA/SRC-1/p160 Family of Transcriptional Coactivators by the Aryl Hydrocarbon Receptor/Aryl Hydrocarbon Receptor Nuclear Translocator Complex. *Molecular and Cellular Biology* **22**, 4319-4333, doi:10.1128/mcb.22.12.4319-4333.2002 (2002).

5 Vazquez-Rivera, E. *et al.* The aryl hydrocarbon receptor as a model PAS sensor. *Toxicology Reports* **9**, 1-11, doi:https://doi.org/10.1016/j.toxrep.2021.11.017 (2022).

6 Crews, S. T., Thomas, J. B. & Goodman, C. S. The Drosophila single-minded gene encodes a nuclear protein with sequence similarity to the per gene product. *Cell* **52**, 143-151, doi:10.1016/0092-8674(88)90538-7 (1988).

7 Reyes, H., Reisz-Porszasz, S. & Hankinson, O. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193-1195, doi:10.1126/science.256.5060.1193 (1992).

8 Nambu, J. R., Franks, R. G., Hu, S. & Crews, S. T. The single-minded gene of Drosophila is required for the expression of genes important for the development of CNS midline cells. *Cell* **63**, 63-75, doi:10.1016/0092-8674(90)90288-P (1990).

9 Ponting, C. P. & Aravind, L. Vol. 7 R674-R677 (Elsevier, 1997).

10 Hefti, M. H., Franc, K.-j., Vries, S. C. D., Dixon, R. & Vervoort, J. The PAS fold A redefinition of the PAS domain based upon structural prediction. **1208**, 1198-1208, doi:10.1111/j.1432-1033.2004.04023.x (2004).

11 Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858, doi:10.1038/nprot.2015.053 (2015).

12 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

13 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).

14 Henry, J. T. & Crosson, S. Ligand-Binding PAS Domains in a Genomic , Cellular , and Structural Context. doi:10.1146/annurev-micro-121809-151631 (2011).

15 McIntosh, B. E., Hogenesch, J. B. & Bradfield, C. A. Mammalian Per-Arnt-Sim Proteins in Environmental Adaptation. *Annual Review of Physiology* **72**, 625-645, doi:10.1146/annurev-physiol-021909-135922 (2010).

16 Evans, R. M. The Nuclear Receptor Superfamily: A Rosetta Stone for Physiology. *Molecular Endocrinology* **19**, 1429-1438, doi:10.1210/me.2005-0046 (2005).

17 Reisz-Porszasz, S., Probst, M. R., Fukunaga, B. N. & Hankinson, O. Identification of functional domains of the aryl hydrocarbon receptor nuclear translocator protein (ARNT). *Molecular and Cellular Biology* **14**, 6075-6086, doi:10.1128/mcb.14.9.6075 (1994).

18 Huang, N. *et al.* Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337**, 189-194, doi:10.1126/science.1222804 (2012).

19 Reick, M., Garcia, J. A., Dudley, C., Science, S. L. M. & undefined. NPAS2: an analog of clock operative in the mammalian forebrain. *science.sciencemag.org*.

20 Gu, Y.-Z., Hogenesch, J. B. & Bradfield, C. A. The PAS Superfamily: Sensors of Environmental and Developmental Signals. *Annual Review of Pharmacology and Toxicology* **40**, 519-561, doi:10.1146/annurev.pharmtox.40.1.519 (2000).

21 Hogenesch, J. B., Gu, Y. Z., Jain, S. & Bradfield, C. A. The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 5474-5479, doi:10.1073/pnas.95.10.5474 (1998).

22 Sun, Z. S. *et al.* RIGUI, a Putative Mammalian Ortholog of the Drosophila period Gene. *Cell* **90**, 1003-1011, doi:10.1016/S0092-8674(00)80366-9 (1997).

23 Amoutzias, G. D., Robertson, D. L., Peer, Y. V. D. & Oliver, S. G. Choose your partners : dimerization in eukaryotic transcription factors. doi:10.1016/j.tibs.2008.02.002 (2008).

24 Zylka, M. J. *et al.* Molecular Analysis of Mammalian Timeless. *Neuron* **21**, 1115-1122, doi:10.1016/S0896-6273(00)80628-5 (1998).

25 Yagita, K. *et al.* Dimerization and nuclear entry of mPER proteins in mammalian cells. *Genes Dev* **14**, 1353-1363 (2000).

26 Hennig, S. *et al.* Structural and functional analyses of PAS domain interactions of the clock proteins Drosophila PERIOD and mouse period2. *PLoS Biology* **7**, 0836-0853, doi:10.1371/journal.pbio.1000094 (2009).

27 Kucera, N. *et al.* Unwinding the differences of the mammalian PERIOD clock proteins from crystal structure to cellular function. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3311-3316, doi:10.1073/pnas.1113280109 (2012).

28 Yildiz, O. *et al.* Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD. *Mol Cell* **17**, 69-82, doi:10.1016/j.molcel.2004.11.022 (2005).

29 la Cour, T. *et al.* Analysis and prediction of leucine-rich nuclear export signals. *Protein Eng Des Sel* **17**, 527-536, doi:10.1093/protein/gzh062 (2004).

30 Barik, S. The Uniqueness of Tryptophan in Biology: Properties, Metabolism, Interactions and Localization in Proteins. *Int J Mol Sci* **21**, doi:10.3390/ijms21228776 (2020).

31 Sadek, C. M. *et al.* Isolation and characterization of AINT: a novel ARNT interacting protein expressed during murine embryonic development. *Mech Dev* **97**, 13-26, doi:10.1016/s0925-4773(00)00415-9 (2000).

32 Matthews, J., Wihlén, B., Thomsen, J. & Gustafsson, J.-A. Aryl hydrocarbon receptor-mediated transcription: ligand-dependent recruitment of estrogen receptor alpha to 2,3,7,8-tetrachlorodibenzo-p-dioxin-responsive promoters. *Molecular and cellular biology* **25**, 5317-5328, doi:10.1128/MCB.25.13.5317-5328.2005 (2005).

33 Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **524**, 303-308, doi:10.1038/nature14883 (2015).

34 Zhou, Y. D. *et al.* Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 713-718, doi:10.1073/pnas.94.2.713 (1997).

35 Brunskill, E. W., Witte, D. P., Shreiner, A. B. & Potter, S. S. Characterization of Npas3 , a novel basic helix-loop-helix PAS gene expressed in the developing mouse nervous system. *Mechanisms of Development* **88**, 237-241, doi:10.1016/S0925-4773(99)00182-3 (1999).

36 Erbel-Sieler, C. *et al.* Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13648-13653, doi:10.1073/pnas.0405310101 (2004).

37 Wu, D., Su, X., Potluri, N., Kim, Y. & Rastinejad, F. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. 1-15, doi:10.7554/eLife.18790 (2016).

38 Seok, S. H. *et al.* Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 5431-5436, doi:10.1073/pnas.1617035114 (2017).

39 Xing, Y. *et al.* Identification of the Ah-receptor structural determinants for ligand preferences. *Toxicological Sciences* **129**, 86-97, doi:10.1093/toxsci/kfs194 (2012).

40 Giani Tagliabue, S., Faber, S. C., Motta, S., Denison, M. S. & Bonati, L. Modeling the binding of diverse ligands within the Ah receptor ligand binding domain. doi:10.1038/s41598-019-47138-z.

41 Szöllösi, D., Erdei, Á., Gyimesi, G., Magyar, C. & Hegedüs, T. Access path to the ligand binding pocket may play a role in xenobiotics selection by AhR. *PLoS ONE* **11**, doi:10.1371/journal.pone.0146066 (2016).

42 Sakurai, S., Shimizu, T. & Ohto, U. cro The crystal structure of the AhRR – ARNT heterodimer reveals the structural basis of the repression of AhR-mediated. **292**, 17609-17616, doi:10.1074/jbc.M117.812974 (2017).

Chapter 4: Evidence for asymmetrical dimerization and PAS B dependency on the AhR-ARNT dimer.

INTRODUCTION

The aryl hydrocarbon receptor (AhR)[1] is an enigmatic ligand activated transcription factor that binds a variety of compounds and regulates an adaptive metabolic response[2,3]. In its inactivated form, the protein is found in the soluble fraction of the cell bound to chaperones: HSP90[4], ARA3[5], ARA9[6,7] and p23. Upon binding a ligand, the AhR changes conformation, reveals a nuclear localization sequence (NLS), translocates to the nucleus, and swaps its chaperones for its obligate partner, ARNT (**Figure 1, Left**). The resultant dimer binds to genomic enhancers and upregulates the expression of a battery of genes encoding xenobiotic metabolizing enzymes, as well as the AhR repressor (AhRR) which appears to downregulate the induced pathway[8].

The AhR, ARNT and the AhRR belong to the PAS family of proteins and serve as a signaling prototype for the family of 22 mammalian proteins. The three family members of the AHR pathway each harbor a bHLH DNA binding domain, a PAS domain comprised of repeats, known as PAS A and B, as well as a C-terminal “Transcriptional Regulatory Domain” (AD) (Chapter 2). Each of these motifs are separated by flexible linkers with little sequence homology across family members. Early data generated in cell culture models, yeast, or *in vitro*, supports a role for the bHLH domains in dimerization and DNA recognition, the PAS A repeat in dimerization, the PAS B repeat in ligand binding and chaperone association, and the AD in recruitment of coactivators or corepressors[3].

The three-dimensional structures of the full-length AhR protein, the full length ARNT protein, the full-length AhR-ARNT dimer, the AhR complex with chaperones, or the AhR complex with coactivators have not been resolved. Recently, the bHLH and PAS A domains of the AhR complex have been crystallized by two laboratories [9] where they largely agree and display a degree of symmetry. Interestingly, these structures also supported a model where an A’ alpha helix is the primary driver of PAS-A-PAS-A interactions, indicating that the helix and not

PAS A repeats, have direct contacts with each other[9-11] (Chapter 2). Nonetheless, these helices facilitate proximity of PAS A repeats thereby permitting the correct configuration for dimerization. Perhaps most importantly, the structures and molecular biology to date, provide no real information on the mechanisms by which the PAS B repeat of the AhR interacts with either chaperones or ARNT. In particular, the current data does not inform us about the symmetry of the PAS B interactions, whether PAS-B-PAS-B interactions occur at all, or whether they are similar in type to PAS-A-PAS-A interactions (**Figure 1, Right**).

To better understand the role of the PAS B repeat in AhR biology, we have employed the interaction trap in the yeast *S cerevisiae* and coupled these results from our modelling approach using computational and visualization tools such as Alpha-Fold, PyMol etc. (Chapter 2). We begin these experiments by presenting our interpretation of the simple model of dimerization that has driven the molecular biology of the AhR field for the last twenty years (**Figure 1, Left**). This model assumes all domains interact in a linear fashion, whereby specific regions interact exclusively with their counterparts on other PAS proteins[12,13]. Put another way, the HLH of the AhR interacts with the HLH of ARNT, the PAS A repeats interact, and PAS B repeats interact (**Figure 1, Left**) thus the mode in which the PAS A-PAS A interaction occur is assumed to be identical to the mode in which PAS B domains interact. While PAS A -PAS A repeat interactions has been documented by us and others, there is no compelling data to demonstrate the PAS B-PAS B domains dimerization occurs. Advances in crystallization techniques and protein modelling have changed the way we think about PAS protein interactions. Recent studies have provided compelling evidence that PAS A-PAS B interactions may be central to dimer formation and that PAS A repeats interact with the partner PAS A and the B repeats (CLOCK:ARNTL[14], PER1:PER1[15], PER2:PER2[16], PER3:PER3[15], HIF1α:ARNT[10], HIF2α:ARNT[10], NPAS1:ARNT[11], and NPAS3:ARNT[11]) (Chapter 2). Furthermore, predicted[17] AhR:ARNT heterodimers suggest the pair may also share PAS B contacts.

To shed light on this biology, we have focused on the structure function of the PAS B repeat of the AhR, with particular attention to its modes of protein interaction. Our primary hypothesis is that PAS B interactions are sufficient for AhR:ARNT dimerization. To test this idea, we exploited the yeast two-hybrid screening using a LexA-AhR bait and an ARNT prey construct. Additionally, to understand the effect of structural features on dimerization and ligand binding, we performed random mutagenesis of the AhR PAS B repeat and visualized these changes. We found that some amino acid substitutions resulted in loss of function because of global disturbances for both dimerization and ligand binding. To tie these concepts together, we also propose a new AhR:ARNT pathway schematic to visualize PAS B interactions.

MATERIALS AND METHODS

Yeast two-hybrid:

Protein interactions were determined using the two-hybrid approach in the yeast Saccharomyces cerevisiae. Our modified yeast interaction trap was employed to identify PAS domains interactions between AHR and ARNT *in vivo*. The LexA chimeras were constructed to fuse AHR and ARNT domains with the DNA binding domain of the bacterial protein LexA amino acids 1-202).

Bait AhR Plasmids: Mouse AhR^{b1} (AAA02896.1)[1] was synthesized and cloned into plasmid 535 (pBTM116)[18] giving rise to yeast expression plasmids 703 (LexA-NΔ166AhR^{b1})[19] and 739 (LexA-m$AhR^{b1}\Delta$AD)[20] (**Figure 2**). The plasmid pBTM116 is approximately 5685 bp, contains the LexA (1-202) domain, followed by a multiple cloning site, harbors a *TRP1* gene, and a high-copy 2µ origin of replication[18]. In plasmid 739, the signature glutamine rich AD region was deleted: mapping to amino acids 495-640.

Prey ARNT Plasmids: Human ARNT (AAA51777.1)[21] was cloned into plasmid pSGBCU giving rise to full-length ARNT for expression plasmid 574[20] (FL-ARNT bait) and has been described before (Figure 2). Plasmid 574 is low copy number origin of replication (CEN/ARS) and contains the auxotrophic marker *LEU2*. The ARNT was also designed *in silico* using DNAStar (DNAStar, Inc, Madison, WI); an N-terminus deletion of up to amino acid position 345 was synthesized and cloned into pYX242 (Millipore, Bulington, MA) to make plasmid 2353 (PAS B-ARNT bait) (Figure 2). Plasmid 2353 is high copy number (2µ) with the *LEU2* marker.

L40: The yeast *Saccharomyces cerevisiae* reporter strain L40[18] ATCC MYA-3332 *(MATa ade2 his3 leu2 trp1 LYS::lexA-HIS3 URA3::lexA-LacZ)* was streaked and grown for 3 days in YPD agar (Takara Bio, San Jose, CA) plates with ampicillin at 30˚C. The YPD medium is a complete medium. One colony was selected and grown overnight in 10 mL, pH adjusted to 5.8, YPD broth with ampicillin (Takara Bio, San Jose, CA) at 30 ˚C and 220 RPM. After the allotted

time, bait-prey pair expression plasmids combinations were transformed into competent L40 cells using the Frozen-EZ Yeast Transformation II Kit (Zymo Research, Irvine, CA).

To make competent yeast cells: the cells were centrifuged, and the supernatant was discarded. The pellet was washed with 10 mL of the Frozen-EZ Yeast Solution 1 and centrifuged. The supernatant was discarded. Finally, the pellet was resuspended in 1 mL of the Frozen-EZ Yeast Solution 2. To transform yeast: 50 µL of competent cells were mixed with 0.2-1 µg yeast expression plasmid combinations: (a) controls 535 + 609, 739 + 368, & 739 + 609 and (b) 739 + 574 & 739 + 2353. Each combination was thoroughly mixed with 500 µL of Frozen-EZ Yeast Solution 3. The mixture was incubated in a water bath at 30 ˚C for 45 minutes and mixed every 15 minutes. 150 µL of the transformation mixture was spread on to synthetic medium lacking tryptophan and leucine (Takara Bio, San Jose, CA) and containing ampicillin. The plates were incubated for up to 4 days at 30 ˚C to allow transformant growth.

Exposure to compounds: Ten micromolar β-napthoflavone (βNF) were prepared in dimethylsulfoxide (DMSO) (Sigma Aldrich, Burlington, MA). One transformant colony, representative of each transformation, was grown overnight in synthetic medium lacking tryptophan and leucine broth (Takara Bio, San Jose, CA) and containing ampicillin at 30 ˚C and 210 RPM. One hundred microliters were pipetted into clear 96 well plates and the OD_{600} was measured and adjusted to ~0.5 using the ClarioStar Plus plate reader (BMG Labtech, Cary, NC). The βNF stock solution was then used to create dose-response curves. Five microliters at the appropriate concentrations were dispensed into white 96 well plates using the Tecan Liquid Handler (Tecan, Morrisville, NC). Ninety-five microliters of OD_{600} adjusted cells were pipetted on to the same white plate, covered with a clear lid, wrapped in aluminum foil, and incubated at 30 ˚C and 210 RPM for 2 hours.

Measuring β-galactosidase activity: After the allotted time, yeast cells were lysed and chemiluminescent activity from the β-galactosidase enzyme was measured using the Gal-

Screen assay (Thermo Fisher Scientific, Waltham, MA). The Gal-Screen Reaction Buffer was made as follows: the Gal-Screen substrate was diluted 1:25 with the Gal-Screen Buffer B. One hundred microliters of the reaction buffer were dispensed to each well. The plate was covered in foil and incubated at 26-28 ˚C and 210 RPM for ~ 60-90 minutes. The plate was then placed in the ClarioStar Plus plate reader and values were measured using the luminometer reader function and the MARS data analysis software (BMG Labtech, Cary, NC).

Data Analysis: All protocols were performed with at least 3 biological and 3 technical replicates. The data from MARS was analyzed using PRISM GraphPad v9.2.0 (San Diego, CA) using the four-parameter logistic curve (4PL)[22]. For normalized results, the highest raw value was designated as the maximum, while the lowest raw value was designated as the minimum. We then employed GraphPad software to calculate variables via the 4PL model and all data is reported as a percentage of the highest and lowest values. The data is described as % β-galactosidase relative light units (RLU).

AhR PAS B random mutagenesis

Random Mutagenesis: The full-length open reading from of the AhR ^{b1} receptor was amplified by PCR and cloned into the pGEMT easy vector (designated plasmid 1060). Ten micrograms of this plasmid was subjected to hydroxylamine mutagenesis in 0.5. ml volume, for one hour, at 75 degrees centigrade (1M hydroxylamine, 1M potassium phosphate , pH 6.0). After mutagenesis, the DNA was purified on a Sephadex G50 column, precipitated in ethanol and resuspended in 10 ul (10 mM Tris, 1 mM EDTA, Ph 8.0). One microliter of this plasmid library was transformed into JM1089 E coli, generating 200,000 independent colonies that were remixed and used to generate purified DNA by standard methods. Plasmid 883 was constructed in a Cen/ARS vector such that the open reading frame contained the LexA DNA binding domain fused to amino acids 167-805 of the murine AhR^{b1}. Using EcoRI/BamHI digestion, the library insert was moved into the corresponding region of plasmid 883 to generate the point mutant

library. This generated more than 12,000 independent clones and these were mixed for DNA preparation and yeast transformation (**Figure 3**).

Eight thousand yeast colonies were screened using the X-Gal overlay assay in the presence of βNF ligand, as described previously[20]. Negative clones drove the LexO reporter, beta-galactosidase, and were blue in color after one hour. Loss of function (LOF) mutants were white or pale blue at this same time point. All possible LOF mutants were picked and grown in 15 ml media for two days and then subjected to western blotting for AhR expression Any truncation mutants were eliminated from further consideration, and those showing expression of full-length protein were sequenced by Sanger methodology to identify point mutation location. This procedure yielded a total of 12 independent LOF clones (**Figure, 3 bottom**). All LOF clones were confirmed in duplicate using a liquid β-galactosidase assay, as described previously, and those results are presented in **Table 2**. Data are presented as a percentage of the wildtype Ahr ^{b1} allele, and the Ahr ^{d} allele, and a vector containing no insert were generated as a reference.

Complementation with homo- and hetero-family associated proteins: ARNT, ARA9 and ARA3. To determine if any associated proteins could rescue the signaling phenotypes of LOF mutations, we expressed each mutant with version of ARNT (pL 791), ARA9 (pL 792), and ARA3 (pL 810) via genetic complementation. To this end, the plasmids 791, 792, and 810 were transformed into the mating strain AMR70 (MATα*, his3, lys2, trp1, leu 2*) and plated as a thin film, on agar plates, in either the presence or absence of βNF. Each mutant (along with positive and negative control plasmids) were transformed into the L40 yeast strain (MATa, *his3delta200, trp1-901, leu2-2, LYS::lexop4-HIS, URA3::lexop8lacZ*) and each was streaked onto the lawn of the strains harboring the potential complementing proteins. Interactions were measured via the β-galactosidase overlay assay and are depicted in **Figure 6**.

Modelling protein-protein interactions: The AhR (AF-P30561-F1), ARA9 (AF-O00170-F1), ARA3 (AF-Q9Y6Y0-F1), and ARNT (AF-P53762-F1) predicted structures were retrieved from the AlphaFold server by DeepMind[23] and uploaded onto SwarmDock (https://bmm.crick.ac.uk/~svc-bmm-swarmdock/index.html). The AlphaFold AhR predicted structure was docked with ARA9 and ARNT. For ARA9, we are assuming that the interaction takes place before dimerization with ARNT. For interactions between the predicted AhR:ARNT dimer[17] and ARA3, we used the predicted AhR structure kindly provided by the Bonati group because we are assuming that the interaction takes place after dimerization with ARNT. All predictions were visualized using PyMOL v4.6.0 (Schrodinger Inc., New York, NY).

RESULTS AND DISCUSSION

The PAS family of proteins are an important class of biosensors that detect environmental stressors and mediate adaptive responses of vital functions[24]. Well known examples of mammalian PAS sensors include those that regulate the responses to low oxygen conditions, circadian responses to photic cues, as well as exposure to certain planar aromatic xenobiotics. In our model, these proteins carry out their adaptive roles through dimerization, where a class α PAS protein senses the cue and binds with its cognate class β PAS partner. Together, they position HLH domains to provide specific contacts within the mammalian genome to regulate specific targets (**Figure 1, Left**).

Several early observations suggested that α-β proteins heterodimerize with symmetry, i.e., where the two proteins are aligned in a parallel manner. First, all α- and β-class PAS proteins have similarly organized domain maps, depicted from the N-terminus as bHLH, PAS A, PAS B and hypervariable transactivation domain (AD). Second, bHLH domains have been crystallized from various dimeric transcription factors and present as a symmetrical feature where each basic region lies within the major groove of DNA to provide sequence binding specificity[25]. Third, initial descriptions of the structures of another sensor superfamily of transcription factors, the steroid receptors, are indicative of symmetry at almost all levels of structure[26].

The simplest illustration of symmetry between two PAS protein partners is where “like domains” interact in series: *i.e*., the HLH DNA binding domains interact, the PAS A repeats interact, and the PAS B repeats interact[12,13]. Moreover, such a simplistic model predicts that the PAS A repeats will interact with each other the same way PAS B domains interact; however there is little, if any, direct supporting evidence[12,13]. This poses the possibility of an alternate model for AhR signaling, where the PAS B domains of the AhR and ARNT do not interact at all; thus, the PAS B of the AhR can be viewed simply as a subunit that can be transformed when

binding ligands and releasing chaperones which then, allow the PAS A domain to interact **(Figure 1, Right bottom**). It has long been known that PAS B is an important region for binding chaperones, like Hsp90[4], p23, ARA9[6,7], and ARA3[5].

This provocative idea and crystallized structures for other PAS heterodimers led us to hypothesize that PAS B was sufficient for dimerization in AhR:ARNT. We employed the yeast two-hybrid to test this hypothesis and mutagenized amino acids in the AhR PAS B region to understand the effect on the AhR-ARNT dimerization, but also on ARA3 and ARA9 interactions. These data were then interpreted using algorithms and models developed in Chapter 2.

To develop our interaction assay system, we first tested our LexA-AhR bait for autoactivation (signaling in the absence of ligand) (**Figure 4, Left**). In this experiment, we rely on the ARNT AD to facilitate and carry out the response after dimerizing with the AD-lacking AhR. To determine the responsiveness of the LexA-Δ166 construct at various ligand concentrations, we generated dose-response curves using the ligand βNF[19]. To demonstrate the utility of the interaction trap, we examined its ligand inducibility when we transformed yeast with LexA-AhR bait and a FL-ARNT prey (**Figure 4, Right**). The LexA-AhR bait and the PAS B-AD of ARNT were also co-transformed to test the idea that the ARNT PAS B dimerizes with the AhR PAS B, independent of the PAS A repeat (PAS B sufficiency) (**Figure 5**). Overall, the traditional yeast two-hybrid pair had a lower EC_{50} (26.3 vs 105 nM) and maxima value when compared to LexA-AhR bait + PAS B ARNT prey; however, the Hillslopes were similar (**Figure 5**).

These results are intriguing because while they provide the first evidence that the PAS B domains of the AhR and ARNT interact, changes to the ARNT protein concentration should not impart any changes to the ligand binding function and shape of the dose-response curve, as reflected in the EC_{50}. Taking this idea even further, if classic receptor theory is at play here, then the maximal value and the EC_{50} are constants for a first-order Hill dose response curve; thus the amount of gene product is proportional to the ligand-bound receptor[27]. It is possible that

although the AhR and ARNT PAS As may not be involved in the AhR PAS B-ligand binding kinetics, the PAS A-PAS A interaction may play a role in transient DNA interactions downstream that could affect the DNA state resulting in changes in the maximal value and the EC_{50}. A similar phenomenon has been proposed with steroid receptor homodimers [27], whereby changes in the top and EC_{50} values have been observed in the presence of cofactors.

Mathematically, changes in these two parameters of the dose-response equation for steroid homodimers have arisen from one biological change – a decrease in downstream complexes because of transient cofactor binding on DNA that: (a) changes the DNA state, (b) allows the binding of another cofactor, and/or (c) changes mRNA state during translation[27]. Although the AhR:ARNT complex is a heterodimer, there is evidence that another PAS member, the HIF2α in dimer form with ARNT, uses a loop in its PAS A to interact with DNA[10]. Biological consequences of this interaction and whether this extends to other PAS members are still a mystery and need further investigation. However the AhRR:ARNT crystallized complex shows long loops in the PAS A repeat[28] that may translate to the AhR:ARNT heterodimer because it serves as the main repressor of the AhR pathway.

In addition to PAS B sufficiency, we also asked: which amino acids in the AhR PAS B play a significant role in PAS B interactions specifically with ARA9, ARA3, and ARNT? We postulated this information could find utility in identifying the faces of the PAS domains that were interacting homo- and hetero-family. To accomplish this, the AhR PAS B in LexA-Δ166 was subjected to chemical mutagenesis via hydroxylamine, transformed with ARA9, ARA3 or ARNT, exposed to βNF, and observed for growth/color in media containing β-galactosidase and/or DMSO or βNF. The resulting mutants are outlined in Figure 6 (Left) along with the location of the mutations based on the predicted AhR:ARNT heterodimer[17]. The AhR PAS B repeat is predicted to span between amino acids 284-384, roughly 100 amino acids. A total of 14 mutations were observed between amino acid 291-381, indicating success in targeting the AhR

PAS B. Additionally, some mutants were observed more than once indicating near mutagenesis saturation.

In **Figure 6,** LexA-Δ166 was plated along with the mutants to show their response in the presence and absence of βNF. Mutants 381, 362, and 331 still responded slightly to the ligand while they were considered to have completely lost function despite normal expression levels across all mutants (data not shown). To rescue function, each mutant was transformed with ARA9, ARA3, and ARNT (**Figure 6, Right**). The ARA9 protein is a known chaperone that modifies AhR signaling, in part, by increasing the receptor levels, though the exact mechanism is not completely understood[6,7,29]. As a chaperone protein, it is expected that ARA9 will bind the AhR even without ligand as observed in mutant proteins (mutations at positions 294, 303, 315, 331, 362, and 381). However, in the presence of ligand, the ARA9 appears to rescue even more LOF mutants (291, 294, 295, 303, 305, 315, 331, 362, 368, and 381). These results allow us to see the immunophilin power of ARA9; thus, we posit that these LOF mutants might still bind heat shock proteins in yeast and in the absence of ligand only a few amino acids are available for ARA9 interactions. In the presence of ligand, the AhR activates and changes conformation, possibly releasing energy in the form of ATP. The ATP is hydrolyzed by heat shock proteins, providing the energy needed to stabilize and facilitate interactions between AhR and ARA9[30,31]. Moreover, this change in conformation exposes additional regions that also permit AhR:ARA9 connections. In turn, these connections with ARA9 increase the AhR levels because it decreases the pool that binds to ARNT therefore decreasing the pool that is degraded after transactivation and gene AhR-mediated induction[31]. This may also explain the high background observed for all control, but specifically the AhR:ARA9 plates as well as the dose-response curve observed in **Figure 4, Right**.

Interactions between the AhR and ARA9 have been mapped between amino acids 166 and 491; structurally, the last two beta-strands of PAS A, the linker, PAS B and a portion outside of PAS B[6,7,23]. To get a general idea of how ARA9 might dock AhR, we employed the AlphaFold

predicted structures for ARA9 (AF-O00170-F1) and unliganded AhR (AF-P30561-F1) and used the SwarmDock server to predict protein-protein configuration. The TPR 3 helix acts like a bridge between the AhR PAS A and PAS B (**Figure 7**). Importantly, this same region of ARA9 (the N terminus) has been previously identified as the essential domain for AHR interactions[19,29]. Furthermore, the bulk of the ARA9 protein appears to interconnect with the PAS B repeat. These visuals support experimental evidence for interconnectivity regions and is a direction for future experiments.

The function and mechanism of the ARA3 protein in AhR mediated signaling is less clear than either ARA9 or ARNT. The ARA3 was pulled from a B cell library using an AhR bait in the same screen that identified ARA9. Later, it was determined that the ARA3 clone was missing the first 111 amino acids, thus the full-length ARA3 known as FLARA3 (aka NSIBP) was a 642 amino acid protein harboring a BTB, BACK, IVR domains and six Kelch repeats[5]. Domain analysis experiments show that ARA3 improves AhR signaling through the BACK, IVR, and the 6 kelch repeats; it did not require the N terminal BTB domain. In these earlier experiments, the AhR relied on the PAS domain and additional amino acids beyond the PAS B repeat[5]. Data on **Figure 6**, suggests that the AhR doesn't interact well with ARA3 until the receptor binds ligand; possibly because the transformation step is needed. Moreover, amino acids 291, 294, 303, and 368 may interact better with ARA3 after this conformational change. Using the Bonati dimer as a point of reference (**Figure 8, Left**), which assumes that the AhR:ARNT dimer closely resembles the HIF2α:ARNT dimer, it appears the ARA3 might interact with the helical face of the AhR PAS B, opposite of the ARNT protein. This observation could also explain why studies regularly show little to no interactions between ARNT and ARA3 (unpublished observations).

Lastly, we tested the interaction of each mutant with the obligate partner, ARNT. As expected, no interactions were observed without βNF likely because the activated and transformed AhR develops a higher affinity for the ARNT protein. Thus, in the absence of ligand, we don't expect AhR to interact with ARNT. Furthermore, we know from most PAS structural

studies that the PAS A repeat is proven to be important for homo-family dimerization[9-11,14,15]. Interestingly, 1 out of 13 mutations (G368R at the Hβ-Iβ loop) on the AhR PAS B repeat slightly increased AhR mediated response. Using the Bonati AhR:ARNT dimer to identify the location of the mutations, we can see that the Hβ-Iβ loop does not appear to play a role in dimer formation thus it's hard to determine from the model how position 368 impacts ARNT PAS B binding. Three positions (331, 381, and 362) slightly decrease the AhR mediated response in the presence of the ARNT protein. From the predicted AhR:ARNT dimer, at 331 we can see interactions with the ARNT PAS A repeat, while at position 381, we can see interactions with the PAS B repeat (**Figure 9 and Figure 10**). Thus, we can surmise that both positions compromise ARNT binding. At position 362, the model does not predict interactions with ARNT, perhaps other factors play a role, such as steric hindrance between the amino acid at the location and other structures within the AhR PAS B. When we compare these results to data on **Figure 5**, it suggests that in the absence of the PAS A repeat on AhR, dimerization with ARNT would be compromised; however, in the absence of the ARNT PAS A repeat, dimerization would be reliant on PAS B repeats (preliminary data from our lab supports this idea). The strength of these PAS B repeat interactions have been described as weak in other studies[12,13], and even transient (personal communication with Kevin Gardner). Attempts to dock AhR and ARNT (AF-P53762-F1) using SwarmDock were unsuccessful.

To demonstrate findings in this lab and those of others, we propose a new schematic to depict and clarify AhR dimerization (**Figure 11**) and the signaling pathway (**Figure 12**). To showcase the difference between inactivated and transformed/activated AhR, a linear representation with chaperones is best suited to detail domains when inactivated; however, upon ligand binding, the activated AhR is best shown in configuration ready to accept the ARNT protein to form the dimer (**Figure 11*, Bottom Right***). It's important to show the ability of the domains and accessory motifs to interact with other domains. With all these concepts in mind,

the improved pathway gives a reader a better understanding of the crossover interactions necessary to carry out AhR signaling (**Figure 12**).

CONCLUSION

Advances in structural biology have helped elucidate the dimerization patterns within the PAS family of biosensors. Accessory motifs, repeats and domains interact more freely than had been previously described challenging previous notions that specific regions could only interact with their counterparts in other proteins, for example PAS A repeats could only interact with other PAS A repeats in other PAS proteins. The AhR is among the most widely studied proteins in biology however the entire structure has not been resolved. For the first time, we show that the PAS B repeat in ARNT is sufficient for dimerization with the PAS B repeat in AhR. Additionally, we used chemical mutagenesis and predicted AhR and AhR:ARNT models to determine the role of the PAS B repeat in interactions with other proteins relevant to the signaling pathway, including ARA9 and the ARA3 proteins. We found that ARA9 uses a helix N - terminal to the TPR domains to bridge interconnectivity between the inactivated AhR PAS A and B. Our models also predict that the ARA3 protein interacts with both the PAS A and B through the helical interfaces. Finally, we found that AhR PAS B mutations mostly decreased interactions with ARNT. From this knowledge, we propose the use of a new schematic that differentiates between the inactivated and the activated/transformed AhR in the signaling pathway to help readers understand this crucial difference.

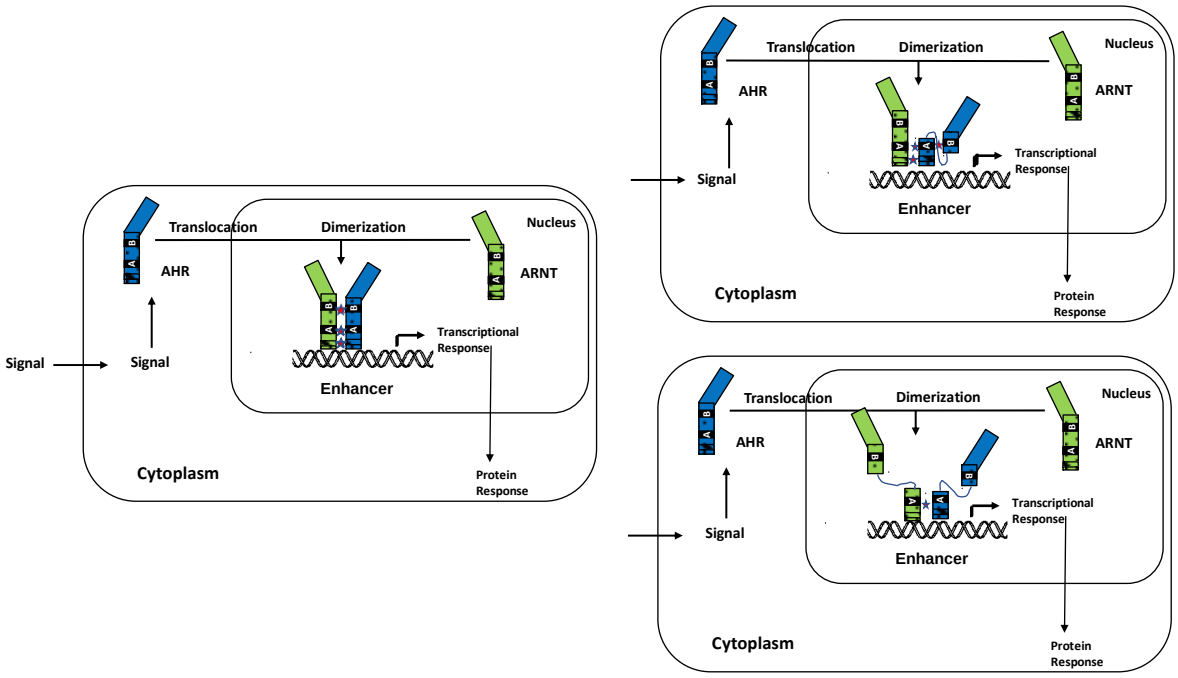


FIGURE 1: Models of AhR-ARNT PAS B Interactions. (*Left*) The most assumed model of AhR-ARNT signaling. Note interactions of HLH, PAS A, and PAS B domains, symmetry, and parallelism of the two proteins. (*Right Upper*): In this imagined model, PAS B domains do not interact with partner protein, instead they interact with PAS A and other chaperone etc. (not shown). *(Bottom left)*, in this model the interactions are parallel, but B domains do not interact with each other.

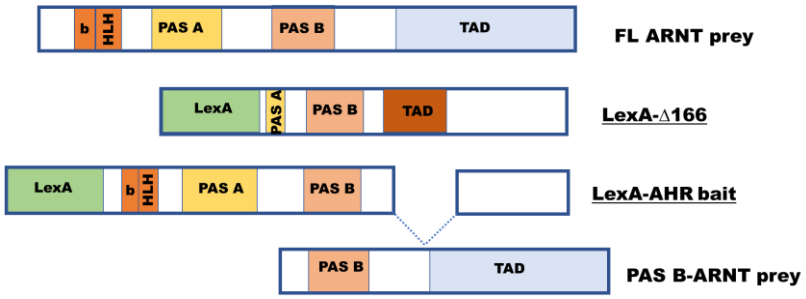
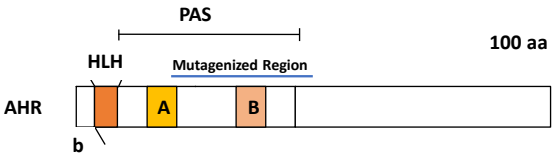


Figure 2: Constructs used to study AhR:ARNT PAS B sufficiency. (Top) The yeast two hybrid was employed to identify dimerization between AhR and ARNT PAS B repeats. The LexA-AHR functions as the bait for the FL-ARNT and PAS B-ARNT preys. To understand the effect of PAS B repeat mutations on AhR, LexA-Δ166 was used.




```
167
EFQRQLHWALNPDSAQGVDEAHGPPQAAVYYTPDQLPPENASFM

ERCFRCRLRCLLDNSSGFLAMNFQGRLKYLHGNKKGKDGALLPP

QLALFAIATPLQPPSILEIRTKNFIFRTKHKLDFTPIGCDAKGL
                                      L  YN
ILGYTEVELCTRGSGYQFIHAADILHCAESHIRMIKTGESGMTV
  D I         E               Y        N
FRLLAKHSRWRWVQSNARLIYRNGRPDYIIATQRPLTDEEGREHL
                 C     R L    V     M
QKRSTSLPFMFATGEAVLYEISSPFSPIM 418
```

Figure 3: Random Mutagenesis of the PAS B repeat region. (*Top*), see text for details. Region mutagenized by hydroxylamine is denoted by blue line. (*Bottom*), the amino acid sequence, and the resultant loss of function mutants that were obtained. Inserted residue is below the mutagenized residue

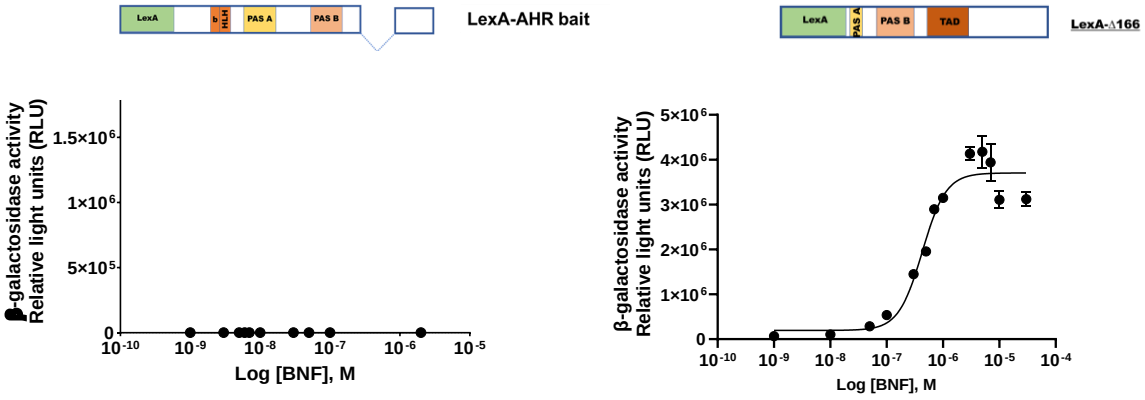
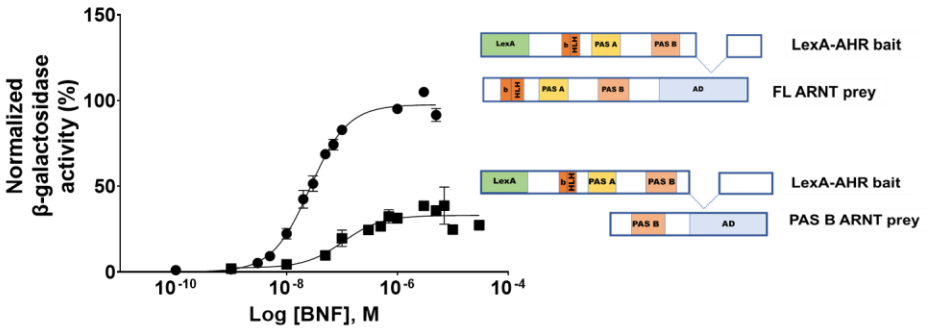


Figure 4: Controls for yeast two hybrid and mutagenesis experiments. (Left) The LexA-AHR bait alone was exposed to various concentrations of βNF to ensure ligand inducibility. Minimal auto-activation was detected. (Right). The LexA-Δ166 fused protein generates a full dose-response relationship with a top, EC_{50}, and Hillslope of: 3.7 x 10^6 RLU, 429 nM, and 2.1, respectively; with 95% CI for each parameter.



Pair	EC50, [nM]	Hillslope	Normalized Top (%)
LexA-AHR bait + FL ARNT prey	26.3	1.27	100
LexA-AHR bait + PAS B ARNT prey	105	1.25	33

Figure 5: The ARNT PAS B interacts with AHR PAS B. (Top) The LexA-AHR bait was transformed with the full length ARNT prey and exposed to the ligand, recapitulating a reliable full-dose response curve**.** The LexA-AHR bait was also transformed with the ARNT PAS B and transactivation domain, and a full dose-response curve can be observed with a right shift and lower top values when compared to the FL ARNT prey. The Hillslope for both curves is similar. Data is normalized for easy comparison

LOF mutants

Mutant	Percent Wild-type (%)	Hypothesized PAS B location
Control	100	N/A
P291L	7	Aβ-Bβ loop
C294T	3	Bβ-Cα loop
D295N	2	Bβ-Cα loop
G303D	3	Cα-Dα loop
T305I	2	Cα-Dα loop
G315E	8	Dα-Eα loop
H331Y	15	Fα
S340N	2	Gβ strand
L348F		end of 2nd Gβ
R362C	20	Hβ strand
G368R	8	Hβ-Iβ loop
P370L	1	Hβ-Iβ loop
A375V	7	Iβ strand
T381M	12	disorder after Iβ strand

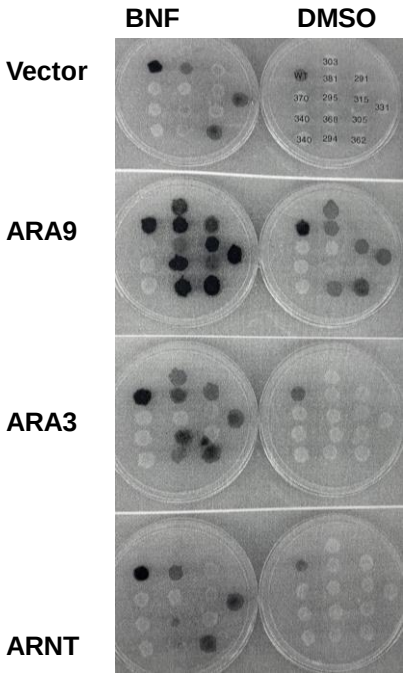


FIGURE 6: The AhR PAS B LOF mutants. (*Left*) Loss of function mutants, their response and hypothesized location based on the Bonati AhR:ARNT dimer structure are shown in this table. (*Right*) Picture of each plate for experiment described under Material and Methods.

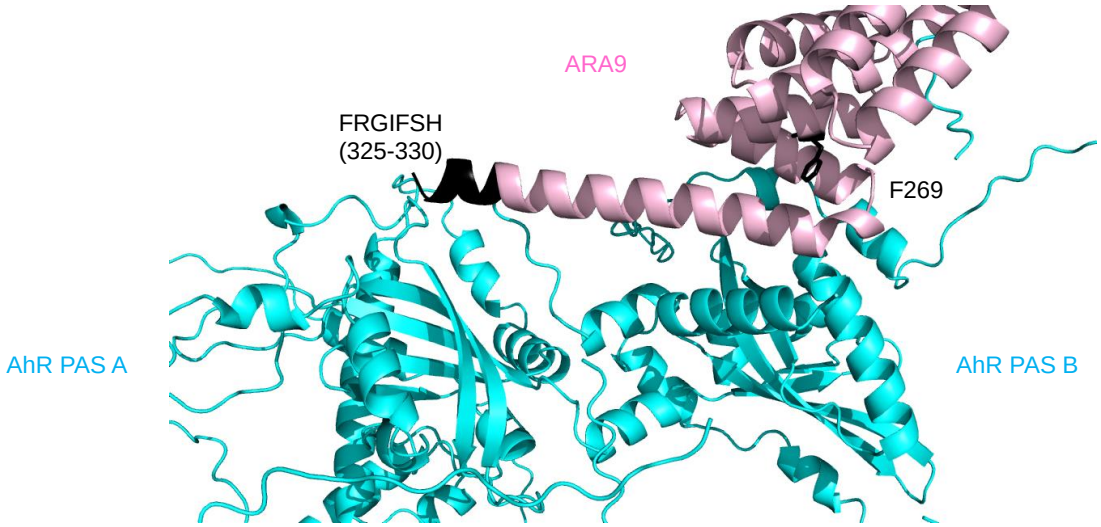


Figure 7: Predicted interactions between ARA9 and AhR. Unliganded AhR and ARA9 from AlphaFold were docked using SwarmDock. The tetratricopeptide 3 of ARA9 is predicted to bridge interconnectivity with the AhR PAS A and B repeats; however other ARA9 interactions are also forecasted with other regions of the AhR PAS B.

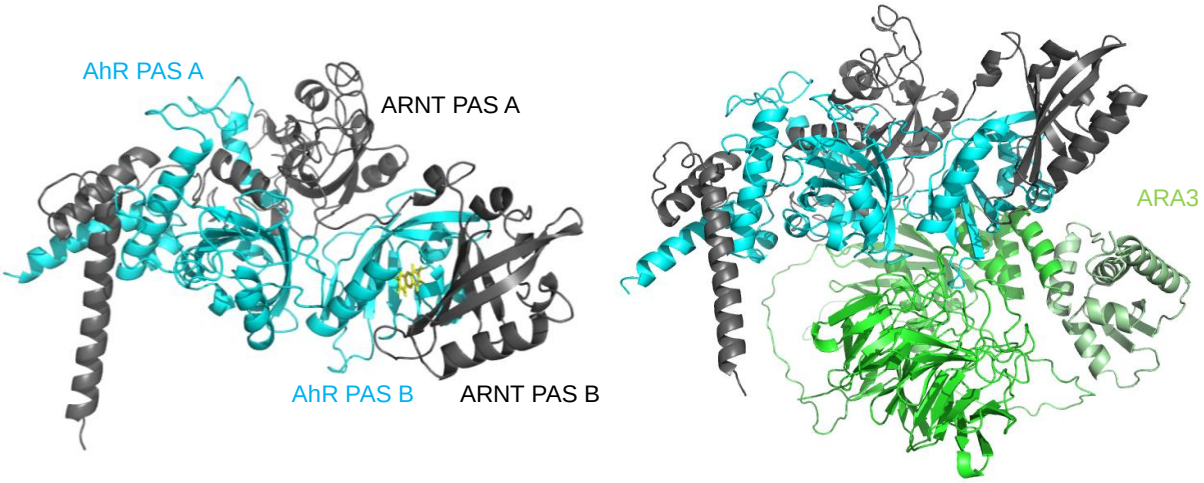


Figure 8: Predicted AhR:ARNT dimer docked to predicted ARA3. (*Left*) This AhR:ARNT predicted dimer was modeled after the resolved HIF2α:ARNT by the Bonati group. Dimer shown with TCDD ligand (yellow) in AhR PAS B. (*Right*) The predicted AhR:ARNT dimer was docked to the AlphaFold predicted ARA3 protein using SwarmDock. In this model, the ARA3 is forecasted to interconnect with the helical faces of the AhR PAS A and B repeats. No connections are apparent between ARA3 and ARNT, supporting past experiments carried out by our lab.

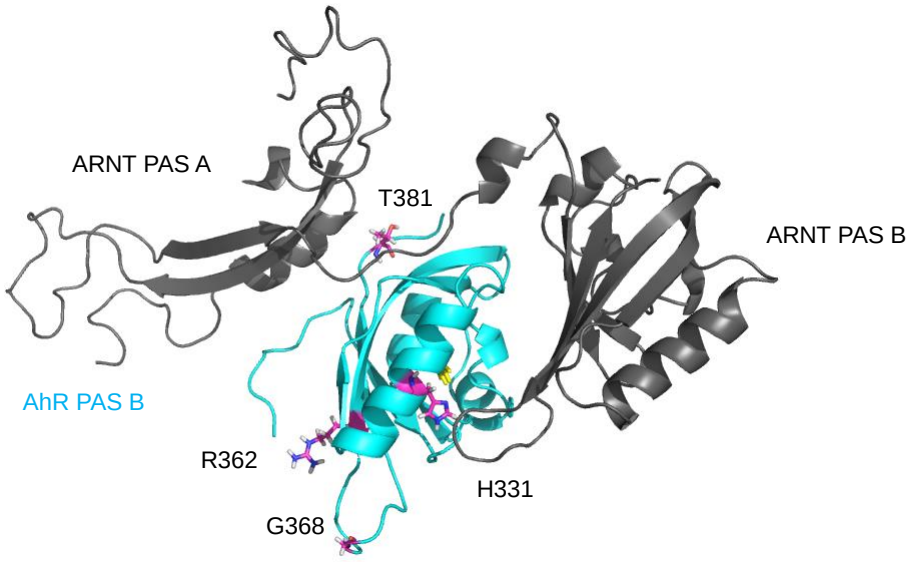


Figure 9: Mutations on the AhR PAS B repeat affecting ARNT binding. Using the Bonati AhR:ARNT heterodimer and the PyMOL mutagenesis wizard, we visualized the location of each mutation with respect to the dimer. Image shows the original amino acid in pink. Mutations appear to affect both ARNT PAS A and B repeat interactions.

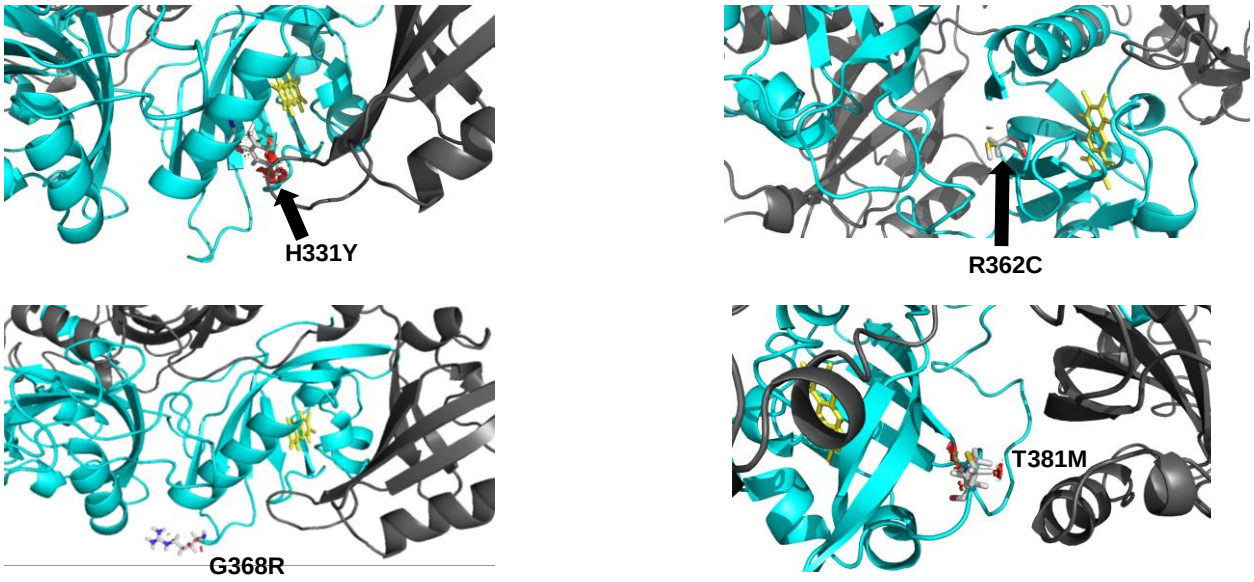


Figure 10: Close up of mutations in AhR PAS B that affect ARNT binding. In this depiction, we mutated each location with the corresponding substitution. In top left (H331Y), steric hindrance appears to have an effect. In R362C and G368R, there are no obvious problems with the substitutions, thus no reason to conclude an increase or decrease based on this model of heterodimerization. Lastly, with T381M, there appears to be steric hindrance within the AhR PAS B and may also influence ARNT binding.

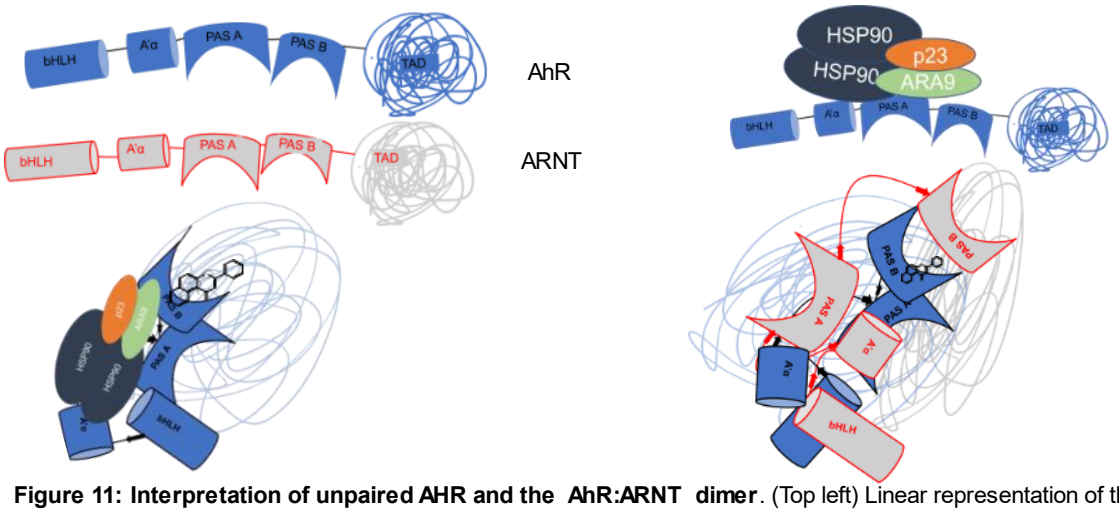


Figure 11: Interpretation of unpaired AHR and the AhR:ARNT dimer. (Top left) Linear representation of the AhR and ARNT proteins with respective bHLH domains, the accessory motif A' α helix, the PAS domain comprised of PAS A and B repeats, and the transactivation domains on each. (Top right) The AhR depicted in linear form with chaperones: heat shock protein 90 (x2), ARA9, and p23. (Bottom Left) Hypothesized AhR after transformation and activation steps upon ligand binding; β-napthoflavone shown in PAS B. Transactivation domain interacts with other domains in a net-like configuration (Bottom Right) Hypothesized AhR interacting regions and heterodimer configuration upon shedding chaperones and ARNT binding; βNF shown between AhR PAS B and ARNT PAS B. Arrows indicate direction and both net -like transactivation domains are shown here. .

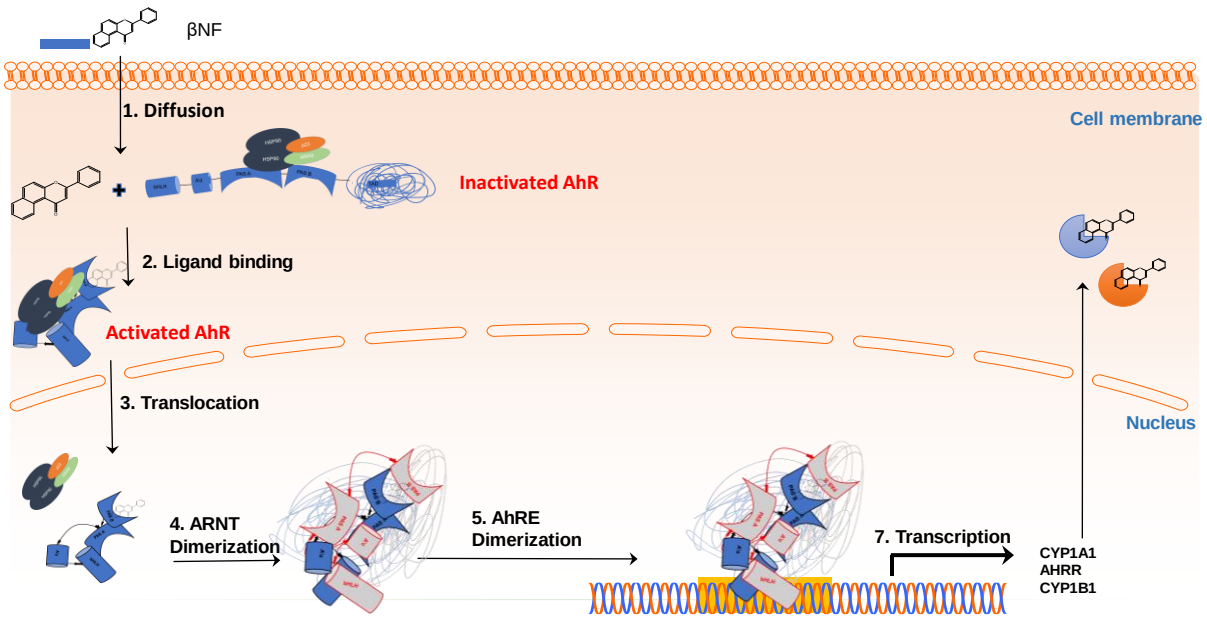


Figure 12: Depiction of the new AhR signaling pathway. In **step 1**, the hydrophobic ligand diffuses through the cell membrane and into the cytoplasm, where the inactivated and chaperone bound AhR resides held. (**2**) Upon ligand binding, AhR activates and tranforms into an ARNT accepting conformation; the complex translocates to the nucleus (**3**), where the chaperones are shed and replaced by ARNT(**4**). Together, they bind responsive elements (**5**) on the DNA which can lead to AhR being exported for degradation (**6**) or the initiation of transcription for metabolizing enzymes, like *CYP1A1* or *CYP1B1* and/or the AhR repressor (**7**).

REFERENCES

1 Bradfield, C. A., Glover, E. & Poland, A. Purification and N-terminal amino acid sequence of the Ah receptor from the C57BL/6J mouse. *Molecular Pharmacology* **39**, 13-19 (1991).

2 Vazquez-Rivera, E. *et al.* The aryl hydrocarbon receptor as a model PAS sensor. *Toxicology Reports* **9**, 1-11, doi:https://doi.org/10.1016/j.toxrep.2021.11.017 (2022).

3 McIntosh, B. E., Hogenesch, J. B. & Bradfield, C. A. Mammalian Per-Arnt-Sim Proteins in Environmental Adaptation. *Annual Review of Physiology* **72**, 625-645, doi:10.1146/annurev-physiol-021909-135922 (2010).

4 Perdew, G. H. & Bradfield, C. A. Mapping the 90 kDa heat shock protein binding region of the Ah receptor. *Biochem Mol Biol Int* **39**, 589-593, doi:10.1080/15216549600201651 (1996).

5 Dunham, E. E., Stevens, E. A., Glover, E. & Bradfield, C. A. The aryl hydrocarbon receptor signaling pathway is modified through interactions with a Kelch protein. *Mol Pharmacol* **70**, 8-15, doi:10.1124/mol.106.024380 (2006).

6 Carver, L. A., LaPres, J. J., Jain, S., Dunham, E. E. & Bradfield, C. A. Characterization of the Ah receptor-associated protein, ARA9. *Journal of Biological Chemistry* **273**, 33580-33587, doi:10.1074/jbc.273.50.33580 (1998).

7 LaPres, J. J., Glover, E., Dunham, E. E., Bunger, M. K. & Bradfield, C. A. ARA9 modifies agonist signaling through an increase in cytosolic aryl hydrocarbon receptor. *Journal of Biological Chemistry* **275**, 6153-6159, doi:10.1074/jbc.275.9.6153 (2000).

8 Mimura, J., Ema, M., Sogawa, K. & Fujii-Kuriyama, Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes & Development* **13**, 20-25, doi:10.1101/gad.13.1.20 (1999).

9 Seok, S. H. *et al.* Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 5431-5436, doi:10.1073/pnas.1617035114 (2017).

10 Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **524**, 303-308, doi:10.1038/nature14883 (2015).

11 Wu, D., Su, X., Potluri, N., Kim, Y. & Rastinejad, F. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. 1-15, doi:10.7554/eLife.18790 (2016).

12 Fukunaga, B. N., Probst, M. R., Reisz-Porszasz, S. & Hankinson, O. Identification of functional domains of the aryl hydrocarbon receptor. *Journal of Biological Chemistry* **270**, 29270-29278, doi:10.1074/jbc.270.49.29270 (1995).

13 Reisz-Porszasz, S., Probst, M. R., Fukunaga, B. N. & Hankinson, O. Identification of functional domains of the aryl hydrocarbon receptor nuclear translocator protein (ARNT). *Molecular and Cellular Biology* **14**, 6075-6086, doi:10.1128/mcb.14.9.6075 (1994).

14 Huang, N. *et al.* Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337**, 189-194, doi:10.1126/science.1222804 (2012).

15 Kucera, N. *et al.* Unwinding the differences of the mammalian PERIOD clock proteins from crystal structure to cellular function. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 3311-3316, doi:10.1073/pnas.1113280109 (2012).

16 Hennig, S. *et al.* Structural and functional analyses of PAS domain interactions of the clock proteins Drosophila PERIOD and mouse period2. *PLoS Biology* **7**, 0836-0853, doi:10.1371/journal.pbio.1000094 (2009).

17 Corrada, D., Soshilov, A. A., Denison, M. S. & Bonati, L. Deciphering Dimerization Modes of PAS Domains: Computational and Experimental Analyses of the AhR:ARNT Complex Reveal New

Insights Into the Mechanisms of AhR Transformation. *PLoS Computational Biology* **12**, doi:10.1371/journal.pcbi.1004981 (2016).

18 Hollenberg, S. M., Sternglanz, R., Cheng, P. F. & Weintraub, H. Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Molecular and Cellular Biology* **15**, 3813-3822, doi:10.1128/mcb.15.7.3813 (1995).

19 Carver, L. A. & Bradfield, C. A. Ligand-dependent interaction of the aryl hydrocarbon receptor with a novel immunophilin homolog in vivo. *Journal of Biological Chemistry* **272**, 11452-11456, doi:10.1074/jbc.272.17.11452 (1997).

20 Hogenesch, J. B. *et al.* 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**, 8581-8593 (1997).

21 Hoffman, E. C. Cloning of a Factor Required for Activity of the Ah (Dioxin) Receptor. *Science* **252**, 954-958 (1991).

22 Masdor, N. A.

23 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

24 Gu, Y.-Z., Hogenesch, J. B. & Bradfield, C. A. The PAS Superfamily: Sensors of Environmental and Developmental Signals. *Annual Review of Pharmacology and Toxicology* **40**, 519-561, doi:10.1146/annurev.pharmtox.40.1.519 (2000).

25 Jones, S. An overview of the basic helix-loop-helix proteins. 1-6 (2004).

26 Evans, R. M. & Mangelsdorf, D. J. Review Nuclear Receptors , RXR , and the Big Bang. *Cell* **157**, 255-266, doi:10.1016/j.cell.2014.03.012 (2014).

27 Ong, K. M., Blackford, J. A., Kagan, B. L., Simons, S. S. & Chow, C. C. A theoretical framework for gene induction and experimental comparisons. *Proc Natl Acad Sci U S A* **107**, 7107-7112, doi:10.1073/pnas.0911095107 (2010).

28 Sakurai, S., Shimizu, T. & Ohto, U. cro The crystal structure of the AhRR – ARNT heterodimer reveals the structural basis of the repression of AhR-mediated. **292**, 17609-17616, doi:10.1074/jbc.M117.812974 (2017).

29 Meyer, B. K. & Perdew, G. H. Characterization of the AhR-hsp90-XAP2 core complex and the role of the immunophilin-related protein XAP2 in AhR stabilization. *Biochemistry* **38**, 8907-8917, doi:10.1021/bi982223w (1999).

30 Bell, D. R. & Poland, A. Binding of aryl hydrocarbon receptor (AhR) to AhR-interacting protein. The role of hsp90. *J Biol Chem* **275**, 36407-36414, doi:10.1074/jbc.M004236200 (2000).

31 Pollenz, R. S. & Dougherty, E. J. Redefining the role of the endogenous XAP2 and C-terminal hsp70-interacting protein on the endogenous Ah receptors expressed in mouse and rat cell lines. *J Biol Chem* **280**, 33346-33356, doi:10.1074/jbc.M506619200 (2005).