HOMOLOGY MODELING OF THE *SQUALUS ACANTHIAS* AHR1 TO LOCATE STRUCTURAL DETERMINANTS OF FUNCTION IN AN ATYPICAL RECEPTOR

by Timothy Bock Jr.

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Master of Arts in

The Department of Biology

The School of Arts and Sciences

Rhode Island College

2019

Abstract

The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor that mediates a toxic response to many environmental contaminants. Cartilaginous fishes, due to gene duplication events, express 4 distinct AHR genes. *Squalus acanthias* AHR1 (SaAHR1) does not bind to typical AHR agonists, including dioxin-like chemicals (DLCs), whereas SaAHR2 binds to DLCs and subsequently upregulates genes containing dioxin responsive elements. Three-dimensional homology models were created based upon both human HIF1α and human HIF2α structural templates. These models yielded different secondary structural characteristics, most notably in the length of the β-strand on the C-terminal end of the ligand binding domain, which is centrally located in the antiparallel β-sheet that makes up a large portion of the ligand binding cavity. Residue A375 is located in a possible hinge area on this β-strand, and the mutation A375C would change this alanine residue in the nonbinding SaAHR1 to its analog in the DLC-binding SaAHR2, which may increase SaAHR1's binding affinity to DLCs. The SaAHR1 also has a novel diproline motif comprising residues 379 and 380, located on the same β-strand as residue 375. This motif is not known to be found in any other AHR isoforms. Reversion of the diproline motif to proline-leucine, its analogous residues in SaAHR2, did not restore PCB 126 induced activation of a luciferase reporter gene. Epifluorescence microscopy using a heterologous expression system and an eGFP-SaAHR1 DNA construct revealed a nuclear localization unique to this isoform. Further characterization of this non-binding receptor could help to determine species-specific sensitivities to AHR agonists, especially in animals that are unavailable for direct experimentation (i.e.: endangered species).

Table of Contents

1. Introduction

Environmental chemical pollutants are ubiquitous. They are present in the air we breathe, the water we drink, and the food we eat. Some chemicals are transient and break down by photolysis or natural bioremediation, whereas others are incredibly stable and may not break down for decades. Unfortunately, a great deal of stable chemicals are hydrophobic and lipophilic, partitioning into the fat of an organism[1]. Furthermore, hydrophobic chemicals partition into plastics[2], where they can pose a long-term exposure risk to the environment. These chemicals can accumulate in organisms and magnify exponentially up the food chain. This means that apex predators, including humans, will have high chemical body burdens relative to lower trophic levels[3].

Another characteristic of hydrophobic or lipophilic chemicals is that they easily pass through plasma membranes[1]. As such, these chemicals make their way into animal cells easily, where depending on their mechanism of action, they can affect the cell on a molecular level. This molecular change can cause damage in a multitude of ways such as altering gene transcription or blocking normal functions[4]. This dysbiosis can affect multiple downstream processes through molecular cascade effects, which ultimately magnify the overall effect. One such interface between chemicals and cellular effects is the receptor protein.

1.1 Receptor proteins

Receptor proteins bind to specific chemicals and undergo changes in structural conformation, providing chemical signals for downstream processes[4]. A chemical that

binds to a receptor protein is known as a ligand. A ligand can either be an agonist, which is a chemical that binds and mediates a downstream response, or an antagonist, which is a chemical that binds and causes the receptor to be nonfunctional in its normal downstream response. A receptor's affinity to bind a ligand is dependent on its biochemical properties, the functional domains it possesses, and the identity of key amino acid residues[5].

Receptor proteins can be localized in different cellular compartments, and as such will have different functions such as initiating and transmitting a signal or acting as a transcription factor[5]. Membrane-bound receptors recognize chemical signals outside of the cell and upon binding undergo a conformational change, normally the first step in a chemical cascade. Cytoplasmic receptors, such as the aryl hydrocarbon receptor (AHR), are localized in the cytoplasm and have various functions upon activation, including gene regulation. Nuclear receptors, such as the estrogen receptor (ER), are usually involved in gene regulation due to their proximity to DNA[6]. Some nuclear receptors, such as the thyroid hormone receptor are constitutively bound to their hormone receptor elements (HREs) on DNA even when inactive and bind to their respective hormones directly adjacent to the gene they regulate[6].

1.2 Exogenous chemicals

Since the industrial revolution, humans have been polluting the environment with large volumes of chemical waste[7]. Chemical pollutants can originate from point sources, such as discharge pipe effluent from a manufacturing plant, or non-point sources, such as

pesticide or fertilizer runoff from farmland[1]. A recent uptick in extreme weather events, attributed to global climate change, may increase the amount of chemical runoff that leads into waterways and marine environments[8]. Exogenous, otherwise known as xenobiotic, chemicals can interact with wildlife, induce toxicity in organisms, and cause deleterious effects on the ecological communities exposed[1].

1.2.1 Persistent organic pollutants

Many anthropogenic chemicals, including halogenated aromatic hydrocarbons (HAHs), are persistent organic pollutants (POPs), exhibiting high stabilities and low degradation rates[1]. These compounds persist in the environment, bioaccumulate in individuals, and biomagnify up food-web trophic levels. Bioaccumulation occurs on the organismal level, when an individual's chemical uptake occurs faster than the chemical can be metabolized or excreted. Another key factor in bioaccumulation is the hydrophobic and lipophilic nature of many persistent chemicals, which causes partitioning into an individual's fatty tissues, where any metabolism may be reduced. Biomagnification occurs at higher trophic levels when an individual consumes numerous animals of lower trophic levels, each of which have accumulated a chemical burden throughout their lifetime. Apex predators are at the highest risk of increased lipophilic chemical load due to both biomagnification and bioaccumulation[1].

1.2.2 Dioxins

Dioxins are chemical byproducts of volcanos and other combustion events and have been present in the environment for billions of years[1]. Although dioxins are

produced from these natural sources, much more has been released into the environment from industrial processes[9]. A few examples of industrial dioxin sources include effluent waste from the manufacturing of some pesticides and herbicides, and processes such as smelting and bleaching of paper pulp[9]. A well-known dispersal of dioxins occurred during the Vietnam war as an unintended ingredient in the military defoliant known as Agent Orange[10].

2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), one of the most potent and well-studied dioxins, is of special interest to AHR chemistry because this co-planar molecule binds to most mammalian AHRs with the highest affinity of any known AHR agonist[11]. Because of this high affinity to AHRs, TCDD is often used as a model coplanar chemical to study the binding properties of AHR:agonist relationships.

1.2.3 Polychlorinated biphenyls

Polychlorinated biphenyls (PCBs) are a group of environmental pollutants that were used heavily in industry for their high stability[1], which is ironically the same characteristic that makes these chemicals persist in the environment. These legacy chemicals, although banned by most countries, are still present in the environment in high concentrations, especially in sites where they were produced or dumped. One such contaminated site is New Bedford Harbor, Massachusetts, in which two electrical component plants used and improperly disposed of PCB-mixtures called Aroclors for 30 years before production was outlawed in the United States in 1978[12]. Remediation of this area is still ongoing, with approximately 485,000 cubic yards of sediment dredged and

disposed of as of September 2018[13]. Some congeners of PCBs are co-planar chemicals and are considered dioxin-like compounds (DLCs) due to their structural similarities to TCDD[14]. These co-planar PCBs are of interest to researchers due to a high affinity to mammalian AHRs[11].

1.3 The aryl hydrocarbon receptor

The aryl hydrocarbon receptor (AHR) is a cytosolic ligand-activated transcription factor in the basic-helix-loop-helix Per-ARNT-Sim (bHLH-PAS) family of environmental sensor proteins[15]. AHR binds with high affinity to a number of structurally related xenobiotic toxins, such as halogenated aromatic hydrocarbons (HAH) and polycyclic aromatic hydrocarbons (PAH), and mediates their toxic effects[16]. Potentially harmful AHR agonists have been found in many products consumers use daily, such as newspaper ink and rubber[17]. The AHR also binds to many nontoxic ligands, including endogenous chemicals such as tryptophan derivatives, and mediates normal physiological responses such as regulating immune function[18].

The superfamily of basic-helix-loop-helix Per-ARNT-Sim (bHLH-PAS) environmental sensor transcription factors control expression of homeostatic genes. This protein family also includes but is not limited to hypoxia inducible factor (HIF), single minded protein (SIM), and circadian locomotor output cycles kaput (CLOCK)[19]. In each of these proteins, the bHLH domain interacts directly with enhancer sequences on target genes to control gene activation, whereas PAS domains usually control activation or ligand binding and a subsequent dimerization step necessary to create a transcription

factor complex[19]. These domains work in conjunction to provide bHLH-PAS transcription factors with signal-induced control over a subset of the cell's transcriptional output. bHLH-PAS proteins exhibit sequence homology and as such are considered evolutionarily related[19].

1.3.1 AHR chaperones

Prior to ligand binding, the AHR is complexed with the chaperone proteins AHR-associated protein 9 (ARA9), heat shock protein 90 (HSP90), and the cochaperone prostaglandin E synthase 3 (PTGES3, also known as P23). A homodimer of HSP90, which is named due to its molecular weight of 90 kilodaltons (kDa), helps maintain an "open" conformation in the AHR ligand-binding cavity and blocks the DNA-binding domain (DBD), preventing premature DNA binding[20]. ARA9, also known as AHR-interacting protein (AIP) and xenobiotic-associated protein 2 (XAP2), has a molecular weight of 37 kDa and is thought to protect the AHR complex from proteasomal degradation and control AHR's cellular localization[21]. The cochaperone P23 is necessary for the ligand-dependent translocation of AHR, the dissociation of the HSP90 dimer, and protects the AHR from degradation[22].

1.3.2 AHR activation

Once ligand binding occurs, the AHR translocates into the nucleus, where it dimerizes with its partner protein, the aryl hydrocarbon receptor nuclear translocator (ARNT) (Figure 1). The newly formed AHR:ARNT heterodimer acts as transcription factor, upregulating genes that contain a dioxin responsive element (DRE), also known as

a xenobiotic responsive element (XRE), in their promoter regions[23]. The collection of genes under the transcriptional control of AHR, known as the AHR gene battery, includes hundreds of genes and controls many physiological functions, including xenobiotic metabolism, immune system function, growth, and development[24].

1.3.3 Physiological roles of the AHR

The AHR, besides being a xenobiotic sensor protein, has multiple roles in normal physiological development. The AHR modulates expression of the estrogen receptor 1 (ESR1), estrogen related receptor alpha (ESRRA) and estrogen related receptor gamma (ESRRG)[24]. Cytochrome P450 17A1 (CYP17A1) mRNA levels are affected by the AHR, catalyzing steps in the biosynthesis of steroid precursors[24]. Neuronal pentraxin I (Nptx1), also upregulated by AHR activation, is part of the neuronal stress pathway involved in apoptosis[24].

Untreated AHR knockout mice were shown to have compromised hepatic function and patent ductus venosus of the liver[25], suggesting the need for a functional AHR in normal liver development. Activation of the AHR has been shown to affect human hematopoietic progenitor cell differentiation, with agonism leading to expansion of erythroid differentiation and antagonism leading to differentiation into megakaryocytes[26].

Many studies suggest that the AHR plays a vital role in the regulation of the immune system. Commensal gut microbes are known to metabolize dietary tryptophan, creating kynurenines that activate the host's AHR signaling pathway, ultimately

regulating immune response and gut homeostasis[18]. It has been suggested that the human AHR has adapted to ligands produced by native gut microbiota, providing an immune response that regulates host-microbe homeostasis, and exhibits a 2:1 binding stoichiometry to indole, which has not been seen with other AHR ligand-receptor pairings[27]. The AHR may also play a future role in the treatment in autoimmune diseases, as AHR activation by TCDD has ameliorated experimentally induced colitis in mice by upregulating regulatory T cells and down regulating T-helper 17 cells[28]. Interestingly, activation of AHR with an endogenous ligand, 6-Formylindolo[3,2-b]carbazole (FICZ), caused the upregulation of T-helper 17 cells, increasing imflammation[28].

1.3.4 Evolution of the AHR

The AHR is present in all Metazoan groups and has evolved independently in these diverse lineages. Mammalian AHRs are the most well studied and differences in species have been utilized to help characterize determinants of binding affinities to AHR agonists[29]. Non-mammalian vertebrate species exhibit a greater diversity of AHRs, with many species of fish containing more than one AHR gene – often expressed in different tissues and with varying affinities to DLCs[30].

Sharks and other cartilaginous fishes have four distinct AHR genes and can synthesize multiple aryl hydrocarbon receptor (AHR) isoforms[30]. Spiny dogfish shark (*Squalus acanthias*) AHR1 does not bind to typical dioxin-like ligands, however AHR2 and AHR3 bind to dioxin-like chemicals and subsequently upregulate genes containing the dioxin responsive element (DRE) promoter region[31]. There is currently no binding

information available for the fourth spiny dogfish AHR isoform, AHR1X, which has high sequence homology to the mammalian AHR, suggesting it may be the ancestral ortholog to the human AHR[31].

1.3.5 Toxicological roles of the AHR

AHR ligands, especially synthetic and anthropogenic ligands, have been implicated in many situations where environmental and human health is at risk. One well-known risk for human health is the AHR-dependent metabolism of benzo[a]pyrene (BaP), a carcinogenic byproduct of combustion that is found in smoke and charred foods. BaP is metabolized by cytochrome P450 1A1 (CYP1A1), part of the AHR gene battery. Metabolism of BaP forms an oxidized product, which can bind to and break DNA, damaging genetic information[32]. Other highly toxic AHR ligands, such as TCDD and PCBs, can cause developmental deformities, general dysbiosis, cancer, and even death in high enough doses[33], but the mechanisms of DLC-mediated toxicity have not been fully elucidated. The AHRs of many species of vertebrates exhibit a highly variable sensitivity to DLCs, with some AHRs exhibiting no affinity for DLCs[34].

For species with functional AHRs, differences in sensitivities are predictive of toxicity. Many studies have been completed in order to characterize the determinants of sensitivity[35,36,37], which can include ligand-binding affinity as well as the gene targets of the activated AHR:ARNT transcription factor complex. In these studies, site-directed mutagenesis is often used in conjunction with homology modeling to locate amino acids of interest and mutate them to test their biochemical effects.

1.4 AHR Structure and modeling

Protein function is determined by three-dimensional structure, which can be derived via direct experimental methods, such as x-ray crystallography, nuclear magnetic resonance (NMR), or cryo-electron microscopy. In some cases, protein structures cannot be experimentally derived, due to a variety of possible hinderances. In these cases, scientists can use computational methods to predict a protein's structure *in silico* and create a three-dimensional model. These models allow researchers to accurately estimate the structure of a protein domain to guide experimental questions.

Homologous protein sequences can be computationally modelled based on a known template structure. Usually, protein structures used as scaffolding to model other proteins share a relatively high percentage of amino acid identity because the proteins are evolutionarily related.

Previous studies have shown that AHR homology models can be used to target amino-acid substitutions that may enhance binding in AHRs with low affinity to dioxin-like compounds. The *Danio rerio* AHR1a does not bind TCDD, but binding was restored with Y296H and T386A (which correspond to mouse 285 and 375) mutations, although these mutants were not able to restore AHR gene battery activation[38]. A similar study investigated the *Xenopus laevis* AHR1β, which binds TCDD with low affinity. In the *Xenopus* study, the mutations N325S, A354S, and A370S (which correspond to mouse 321, 359, and 375) all increased the AHR1β's affinity to TCDD[39]. Both of these studies used the HIF2α structural template as the basis for their homology models and

demonstrate that homology modeling, in conjunction with site-directed mutagenesis, can be used to characterize binding determinants residing in the AHR's ligand-binding domain.

1.5 *Squalus acanthias* as a human model

Elasmobranchs have long been used as vertebrate models because they are among the most primitive species exhibiting characteristics associated with human anatomy and physiology. These characteristics include the use of signaling molecules distributed via a closed circulatory system, molecular responses for salt/water homeostasis, and xenobiotic transport systems[40]. *S. acanthias*, also known as the spiny dogfish shark, is often used as a model elasmobranch due to its abundance and small size. It usually resides at depths of up to 3000 feet in cold water and may live up to 100 years. The cold temperatures of its habitat reduce the metabolic rate of this shark species and it also demonstrates resistance to normal hypoxic responses[40].

1.6 Hypotheses

The aim of this study is to characterize key amino acid residues in the *Squalus acanthias* AHR1 (SaAHR1) ligand-binding domain LBD that may restore binding of SaAHR1 to dioxin-like chemicals. By characterizing binding determinants of the *S. acanthias* AHRs, we can better understand the differential binding affinities of AHR paralogs and orthologs and how this translates to sensitivity to toxic AHR agonists. In this study, I will test hypotheses about the functional significance of specific amino acid

residues in the LBD of SaAHR1 using homology modeling and experiments with heterologous expression system-reporter gene assays.

Part of this study was a collaborative effort with the Bonati Laboratory. This involved making DNA constructs that code for point mutations to the SaAHR1 amino acid sequence recommended by the Bonati Laboratory's structural homology modeling (based on the HIF2α structure). These SaAHR1 mutant constructs include 3 single mutants (R310C, L342M, and A359S) and 2 double mutants (R310C\A359S and L342M\A359S)[A].

During my analysis of the amino acid sequences, I discovered that the SaAHR1 has a proline in position 380. The proline in this position is unique to SaAHR1 and does not appear in any other known isoform. Furthermore, this proline forms a diproline motif toward the C-terminal end of the LBD, which could greatly affect the secondary structure of this area considering proline's rigidity regarding side chain rotation. An analysis of diproline segments in proteins has suggested that most diproline segments exist in only a few conformations and are affected by the amino acids flanking each side[41]. My hypothesis is that the diproline segment in SaAHR1 doesn't allow for the rotational flexibility normally seen in AHR LBDs and therefore may restrict the overall structure and function of the protein. I will be creating duplicates of the aforementioned mutant constructs (from the Bonati laboratory's recommendations), but also each with the P380L mutation (P380L, R310C\P380L, L342M\P380L, A359S\P380L, R310C\A359S\P380L,

[A] Corrada D and Bonati L. Department of Earth and Environmental Sciences, University of Milano-Bicocca. Milan, Italy.

and L342M\A359S\P380L), in order to test whether or not this novel diproline motif affects SaAHR1's functional characteristics.

After my analysis of the SaAHR1 primary structure, I determined that more information could be gathered by analyzing the tertiary structure and created my own 3D structural models of the SaAHR1. The models I've created are based on the human HIF1α structure, which is different than most AHR homology modeling studies that have been done in the past, including the Bonati collaboration. These models have helped to locate amino acid residues of interest for future studies. I have created a mutant SaAHR1 construct to change one of these residues of interest (A375C). This specific residue has been implicated as a binding determinant in multiple studies of other AHR isoforms[38,39,42]. This alanine residue is part of a "faced-pair" of amino acid residues (359, 375) that point inward toward the AHR ligand-binding cavity.

If any of these specific amino acid residue changes in the LBD restore SaAHR1's binding affinity to DLCs and subsequent activation, the presence or absence of such residues may be used to predict species at risk to the toxic effects mediated by activation of the AHR pathway.

2. Methods

2.1 Homology Modeling

AHRs of multiple species were chosen for alignment based upon varying affinities for dioxin-like chemicals (Table 1). Alignments of the amino acid residues in the AHR ligand binding domains (LBDs) of multiple species were created using the ClustalW[43] alignment tool in MacVector version 15.5.4 (Figure 2).

Three-dimensional structural models (.pdb format) of the AHR LBDs were created for a subset of the aligned isoforms (SaAHR1, 2, 3 and MmAHR) using MODELLER version 9.21[44, 45] via the University of California San Francisco (UCSF) Chimera version 1.11.2[46]. The basic local alignment search tool (BLAST)[47] was used to query the Protein Data Bank (PDB)[48] to find known protein structures that share the highest percentage amino acid identity. The search result with the highest percent homology was used as a structural template to model the AHR LBDs, in this case human hypoxia inducible factor 1α (HIF1α) (PDB ID: 4H6J_A[49]) and human HIF2α (PDB ID: 1P97[50]) (Table 2). 10 output models were built for each AHR isoform with "build models with hydrogens" and "thorough optimization" options enabled. One representative model for each of the modeled ligand binding domains was chosen based upon their MODELLER-derived GA341[51], where a value greater than 0.70 is correlated to a 95% probability that a model has the correct fold[52], and zDOPE[53] (normalized discrete optimized protein energy), where negative values are preferable scores[52] (Appendix 1). A

qualitative analysis comparing both the amino acid alignments and the structural models was undertaken to identify residues of interest for site-directed mutagenesis.

2.2 Mutagenic Primer Design

Specific amino acid changes (R310C, L342M, A359S, R310C\L342M, and L342M\A359S) were recommended by a computational analysis performed by the Bonati Laboratory. Other residues of interest (A375C and P380L) were identified via the computational analysis I performed. Mutagenic primers were designed using the QuikChange Primer Design online tool available from Agilent (Table 3). Polyacrylamide gel electrophoresis (PAGE) purified primers were ordered from Integrated DNA Technologies (IDT).

2.3 Chemicals

3,3',4,4',5-Pentachlorobiphenyl (PCB 126) in dimethyl sulfoxide (DMSO), 2 mg/ml, was obtained from AccuStandard, Inc (CAS# 57465-28-8, 99.5% purity).

2.4 Cell Culture

Mammalian cell-cultures were maintained at 37°C with 5% supplemental CO_2. COS7 cells (green monkey kidney epithelial cells, ATCC CRL-1651) were maintained in Dulbecco's Modified Eagle's Medium (DMEM, Sigma D7777), supplemented with 3.7 g/L NaHCO3, with 10% fetal bovine serum (FBS) added. C35 cells (mouse liver hepatoma cells, derived from Hepa-1c1c7, ATCC CRL-2715) were maintained in alpha

minimum essential medium (αMEM, ATCC 30-2605), supplemented with L-glutamine (2mM final concentration), with 10% heat-inactivated FBS added.

2.5 Expression Constructs

Full length *Squalus acanthias* AHR1 and AHR2 were cloned into DEST53 expression vectors using the Gateway system prior to this project by the Merson Laboratory (Table 4). This DNA construct is expressed under the control of a cytomegalovirus (CMV) promoter and encodes an N-terminal eGFP-AHR fusion protein. An inducible firefly luciferase, pGudLuc 6.1, was a gift from M. Denison from University of California, Davis. A constitutive *Renilla* luciferase, pRL-TK, was included in the Promega Dual Luciferase Reporter Assay Kit. Full length *Fundulus heteroclitus* FhAHR2 and FhARNT2 were a gift from M. Hahn from Woods Hole Oceanographic Institution. A constitutive firefly luciferase, pGudLuc 4.13, was a gift from Promega technical support and was used to verify adequate transfection.

2.6 SaAHR1 Site-Directed Mutagenesis

Site-directed mutagenesis was performed on the pDEST53SaAHR1 DNA construct previously created by the Merson Laboratory. Site-directed Mutagenesis was performed using a QuikChange Lightning Multi Site-Directed Mutagenesis Kit using an Eppendorf thermal cycler. The pDEST53SaAHR1 wildtype construct (50 ng), the mutagenic primer (100 ng), a dNTP mix (1 µl), the QuikChange Lightning Multi enzyme blend (1 µl), and the 10x QuikChange Lightning Multi reaction buffer (2.5 µl) was placed into a reaction tube in MilliQ water to a final volume of 25 µl. This mixture was cycled

(Table 5) and a proprietary DpnI restriction enzyme (10 units) was added afterward to digest the parental DNA strands. Digested DNA was transformed into ultracompetent XL-10 gold *Escherichia coli*, which were included in the kit, plated onto Luria broth (LB)-ampicillin (100 μg/ml) agar plates, and incubated overnight at 37 °C. Two colonies were chosen from each plate and grown in 5 ml LB-ampicillin (50 μg/ml) cultures overnight at 37°C with shaking at 250 rpm. Plasmid DNA was isolated from 5 ml cultures via QIAGEN Plasmid Mini Spin Kit (12125) using manufacturer's protocols. Multiple mutant eGFP-SaAHR1 DNA constructs were created (Table 4) for future analysis. Each mutant construct was analyzed via DNA gel electrophoresis (0.8% agarose, 150V, 1 hour), and nanodrop spectroscopy. Specific mutations were confirmed via Sanger sequencing and subsequent analysis with MacVector version 15.5.4.

2.7 Immunoblotting

To confirm that the expression constructs were active in the cell line, pDEST53SaAHR1 wildtype and mutant constructs (6 μg DNA/transfection) were transfected into COS7 cells in confluent T25 flasks using X-tremeGENE HP (6 μl/transfection) transfection reagent. After 72 hours, cells were lysed using 700 μl Cellytic M (Sigma C2978) with cell-scraping necessary, and 10 μl protease inhibitor cocktail (Sigma P8340) was added. Lysates (10 μl/sample) were prepared with 2X protein loading dye (9.5 μl/sample), β-mercaptoethanol (0.5 μl/sample), and boiled for 5 minutes. Lysate preparations (20 μl/well) were resolved via SDS-PAGE using BIO-RAD mini-PROTEAN Tetra apparatus with precast BIO-RAD Mini-PROTEAN TGX stain-free gradient gels (4-15%). The resolved proteins were transferred onto a GE Amersham

Protran nitrocellulose membrane (0.45 μm pore size) using a BIO-RAD Trans-Blot semi-dry transfer apparatus (15V, 0.9 hours). Western blotting, using rabbit anti-GFP-N-terminal primary antibody (Sigma G1544, 1:500 concentration) and goat anti-rabbit horseradish peroxidase conjugated secondary antibody (Sigma A275, 1:10,000 concentration), was performed to confirm eGFP-SaAHR1 fusion protein identity. The Western blot was visualized using BIORAD Clarity (170-5061) enhanced chemiluminescence reagents and a BIORAD ChemiDoc MP.

2.8 Dual luciferase reporter gene assays

C35 cells were seeded in 48-well plates at 30,000 cells/well 24 hours prior to transfection (Figure 3). At the time of transfection, medium was refreshed using αMEM containing 10% heat-inactivated-FBS. pDEST53SaAHR1 wildtype or mutant constructs were co-transfected along with pGudLuc 6.1, pRL-TK, and FhARNT2 (Table 6, assay 6) using the lipofectamine 2000 (1 μl/well) transfection reagent in serum-free αMEM. We used the pDEST53SaAHR2 construct as an inducible positive control and pcDNA empty vector as a negative control. 24 hours post-transfection, 1.5 μl of either a DMSO only control or PCB 126 in DMSO (100 nM final concentration) was added to each well, using three replicate wells for each treatment group. A passive lysis buffer (100 μl) was added to each well and the lysates used for luciferase reporter gene AHR activation assays. A Turner Biosystems 20/20 luminometer and Promega Dual-Luciferase Reporter Assay System Kit was used to quantify luminescence in order to assess AHR activation (Figure 4). The pGudLuc 4.13 plasmid was also used as a constitutively expressed positive control to confirm adequate transfection.

2.9 Epifluorescence Microscopy and AHR Localization Assays

Two 12-well glass bottom plates were seeded with COS7 cells (7.5 x 10^6 cells/well, 24 hours prior to transfection). Medium was refreshed just prior to transfection with DMEM containing 10% FBS. Each well was transfected with pDEST53SaAHR1 wildtype, mutants, or pDEST53SaAHR2 (1 μg DNA/well) using X-tremeGENE HP transfection reagent (3 μl/well) in serum-free DMEM. One 12-well glass bottom plate was used as a vehicle-only (2.5 μl DMSO/well) control, and one 12-well glass bottom plate was dosed with PCB 126 (2.5 μl of 20 μM PCB 126 in DMSO, for a final concentration of 50 nM). An Olympus IX81 microscope was used to visualize the eGFP-SaAHR fusion proteins via epifluorescence on the FITC channel (excitation: 469-509 nm, emission: 488-528 nm). Images were taken with a Retiga camera using Q Imaging software 32 hours after dosing using the microscope's 20X objective.

3. Results

3.1 Homology models suggest residues of interest for SaAHR1 ligand binding

To determine amino acid residues that differ among various isoforms of AHR with differential binding affinities, a ClustalW alignment of ligand binding domains (residues 278-385) was generated. This alignment revealed distinct amino acid residues of interest for further analysis (Figure 2). The residue P380 is of great interest in the SaAHR1, because it is not known to be present in any other AHR isoforms. This residue creates a novel diproline motif toward the C-terminal side of the LBD, which could restrict the overall structure, and therefore affect the function, of the SaAHR1.

BLAST results of LBDs selected for 3D structural modelling (SaAHR1, SaAHR2, SaAHR3, and MmAHR) show that the proteins with known structure and the highest percent homology are human hypoxia inducible factor 1α (HIF1α, PDB ID: 4H6J_A[49]) and human hypoxia inducible factor 2α (HIF2α. PDB ID: 1P97[50]) (Table 2). The two sets of structural models created from these scaffolds allowed for further scrutinization of residues of interest, such as their 3D location, and the possible relation of each residue's sidechain to the ligand binding cavity (Figures 5-8, 14). In the analysis that follows, I have adopted the secondary structure nomenclature used in other AHR LBD studies[39,54] (Figure 5).

My analysis of these two modeling scaffolds reveals overall agreement in much of the modelled ligand binding domain. There are some basic secondary structure differences in the models. The HIF1α-based models (Figure 6) split both the Gβ and Hβ

strands into two separate strands each, which I will refer to as $G\beta_1$, $G\beta_2$, $H\beta_1$, and $H\beta_2$ (Figure 5). The HIF2α-based models maintain a closer structural resemblance to the HIF2α template (Figure 6). Another notable difference between HIF1α and HIF2α based models is the length of the β-strand on the C-terminal end of the LBD, Iβ. All of the models generated based upon HIF1α terminate the Iβ-strand around residue 375, bending the Iβ-strand away from the interior of the ligand binding cavity, disagreeing with the original HIF1α template The models based on HIF2α more closely resemble the Iβ-strand in the HIF2α template. The exception to the larger Iβ-strand in the HIF2α-based models is the SaAHR1 WT. This model bends the Iβ-strand away from the ligand binding cavity around residue 375, similar to the HIF1α models (Appendix 2).

The mutations recommended by the Bonati Laboratory were visualized in the HIF1α-based models, comparing the residue side chains between SaAHR1 and SaAHR2 (Figure 8 A-C). For each of the three mutations (R310C, L342M, and A359S) the SaAHR1 residue was mutated to its analogous residue in the SaAHR2. The mutation R310C (Figure 8A) is pointed away from the ligand-binding cavity in the flexible belt portion of the AHR LBD. The mutations L342M and A359S (Figures 8B, 8C) are located on the internal side of the β-sheet that creates the ligand-binding cavity.

According to my analysis of the 3D models, the SaAHR1 residue A375 (Figure 8D) is of interest due to its side chain chemistry, location, and importance in previous studies of other AHR isoforms. It is centrally located, pointing inward toward the ligand binding cavity. This residue is a cysteine in the SaAHR2 isoform, which readily binds to DLCs, and the cysteine's sidechain chemistry may restore binding affinity to the

SaAHR1 A375C mutant construct (which was created during this study but not yet tested).

3.2 SaAHR1 mutant constructs contain verified mutations and express protein in COS7 cells.

In order to verify the presence and size for each DNA mutagenesis construct plasmid preparation, each construct was resolved via gel electrophoresis (Figure 9). This DNA gel shows the presence of each mutant plasmid isolated from an individual plated colony, and the size of each mutant plasmid is consistent with the wildtype pDEST53SaAHR1 (8935 bp). To verify successful introduction of specific mutations, each mutant construct (except A375C) was analyzed via Sanger sequencing and did contain their respective mutations (Figure 10). Western blotting of a subset of mutants suggests that the mutant protein is expressed and is consistent in size with the wildtype pDEST53SaAHR1 eGFP fusion protein (calculated at 131.5 kDa) (Figure 11).

3.3 Reversion of the SaAHR1 diproline motif with the P380L mutation does not restore inducible activation with PCB 126.

To determine whether the diproline motif impaired DLC binding and SaAHR1 activation, a dual luciferase reporter gene assay was conducted. There was no difference in activation of the reporter in SaAHR1 WT (with diproline present) and SaAHR1 P380L (with diproline reverted) (Figure 12), suggesting that the diproline alone was not implicated in the nonbinding character and lack of transactivation activity. This experiment was performed with FhARNT2 co-transfected to determine if ARNT was a

limiting factor in previous experiments (Appendix 3), with similar results reported both with and without ARNT.

Luciferase assays were also used to assess activation of the pDEST53SaAHR1 double mutants R310C/A359S and L342M/A359S (Appendix 4). These results were not included here due to low raw luminescence readings, possibly due to low transfection efficiency of the luciferase-encoding plasmids (Appendix 5).

3.4 Epifluorescence microscopy shows a unique localization of SaAHR1 WT and mutants.

To determine the localization of the SaAHR1 protein in the heterologous expression system, transfected cells were visualized via N-terminal eGFP-SaAHR fusion protein and imaged (Figure 13). The images show some portion of the GFP-tagged protein, representing SaAHR1 WT, SaAHR1 mutants, and SaAHR2, is translocated to the nucleus in the COS7 cells, even in vehicle-only control wells. This is unexpected due to the mechanism of AHR activation, normally translocating to the nucleus only after agonist-binding. The P380L mutant seemed to exhibit the lowest amount of nuclear localization of any of the tested constructs.

4. Discussion

The models produced with the MODELLER function of the UCSF Chimera program produced interesting results. One notable difference between using human HIF1α as a template and using human HIF2α as a template is the size of the modelled ligand binding cavity. In the HIF1α based models, the overall size of the ligand binding cavity seems greatly reduced compared to the HIF2α models. This is in part due to the length of the β-strand on the C-terminal end of the LBD, Iβ. The HIF1α-based models all terminate the Iβ-strand around residue 375, presumably due to the lack of homologous residues past that point (Appendix 8). These models all bend the Iβ-strand away from the interior of the ligand binding cavity, creating a much smaller antiparallel β-sheet and thus a smaller surface area for binding to occur.

The HIF2α template allows for the SaAHR2 and SaAHR3 models to more closely conform to the Iβ-strand in the HIF2α structure due to a homologous leucine 380 residue. Interestingly, this is the same position that creates the diproline motif in the SaAHR1 isoform. Presumably, this lack of the homologous leucine 380 residue causes the SaAHR1 HIF2α-based model to more closely resemble the HIF1α-based models, bending the Iβ-strand away from the ligand binding cavity (Appendix 2). Usually, β-strands in antiparallel β-sheets will have a primary amino acid structure that alternates amino acid side chain character between hydrophobic and hydrophilic[5]. The homology modelling programs could more closely rely on homology alone, forcing models to conform to the known structure rather than using the actual side chain biochemical properties to determine secondary and tertiary structure formation. If the real structure of some AHRs

more closely resemble that of HIF1α, the HIF2α homology models could be greatly overestimating the size of the β-sheet that makes up a large portion of the ligand binding cavity.

Both HIF1α-based and HIF2α-based structural models have a similar "flexible belt area", which was previously theorized to be critical for ligand binding[54]. The amino acid residues that Xing, *et al.* hypothesized to be critical to ligand binding in MmAHR (F318 and I319) are conserved in all SaAHR isoforms. The same study also found that G315 is critical for MmAHR belt flexibility, and this residue is also conserved in all SaAHR isoforms. If the flexible belt area is well conserved in most AHR isoforms, and is critical for ligand binding, then the area opposite the flexible belt (the β-sheet) could cause the differences in binding affinities that we see in some AHR isoforms. This concept has been tested in many studies, and multiple residues in the interior-facing portion of the β-sheet have been mutated and resulted in a higher binding affinity to DLCs[38,39,42].

Some of the studies that have increased binding affinity with certain mutations to residues in the β-sheet share these homologous residues in the SaAHR1 LBD. In birds, the AHR1 has been suggested to have a faced residue pairing of great importance (359 and 375)[42]. When one of these residues is alanine and the opposite is serine, the AHR will have a higher binding affinity than if both were alanine. In *Xenopus laevis*, AHR1β was mutated to increase its affinity for TCDD using mutations analogous to A359S and A375S[39].

The residues on the interior portion of the ligand binding cavity could be interacting to form a charged surface. The HIF1α models of the SaAHRs bend the Iβ-strand away from the interior of the ligand-binding cavity around residue 375. In SaAHR2, which binds to DLCs, the cysteine 375 residue in that location pulls away from the interior cavity further than the alanine 375 in the SaAHR1 model. If these models are accurate, the Iβ-strand could be key to understanding the variable ligand binding properties of the SaAHRs. If this Iβ-strand is part of a hinge that imparts some amount of flexibility into the SaAHRs, it could "close" upon ligand binding, thus enlarging the antiparallel β-sheet and providing a more sterically favored area for dimerization with ARNT.

Residue 375 is also adjacent to other residues of interest, 342 and 359, and together these residues could create a localized area that is more favorable for binding DLCs. This may be a good example of how multiple residues interact to form a charged surface in a ligand binding cavity (Figure 14). I've created a pDEST53SaAHR1 construct with the mutation A375C for future study to test the effect of this residue's chemistry on the binding ability of the SaAHR1's LBD.

The luciferase reporter gene assays testing the SaAHR1 P380L mutant suggests no difference in activation when compared to SaAHR1 WT (Figure 12). This experiment was previously performed without cotransfection of a *Fundulus heteroclitus* ARNT2 construct. The addition of FhARNT2 in assay 6 (Figure 12) was to determine if ARNT was a limiting factor in the C35 cell line used for these experiments. The results were similar with and without ARNT cotransfection (Appendix 3), so it is reasonable to infer

that ARNT cotransfection is not required for AHR-based luciferase assays while using the C35 cell line. It is important to note that this assay only correlates to AHR activation and does not assess binding affinity. This means that the mutation in question may or may not increase binding affinity to PCB 126, but the results are not determinative of this question.

Luciferase reporter gene assays assessing activation of the SaAHR1 double mutants R310C/A359S and L342M/A359S were also conducted (Appendix 4). These results were not included in this paper due to very low raw luminescence readings for both the induced firefly luciferase as well as the constitutive *Renilla* luciferase (Appendix 5). This could be due to low transfection efficiency of the luciferase encoding plasmids in these experiments. Because some of the luciferase assay experiments seemed unreliable, we switched from expression in the C35 cell line to expression in the COS7 cell line. Epifluorescence microscopy experiments and western blotting were utilized to confirm adequate expression of the pDEST53SaAHR constructs in the COS7 cell line.

Interestingly and quite unexpectedly, the pDEST53 eGFP-SaAHR1 fusion proteins (wildtype and mutants) are localized in the nucleus when visualized via epifluorescence microscopy (Figure 13). This localization occurred in both PCB 126 dosed wells and vehicle-only control wells for most of the pDEST53 eGFP-SaAHR1 fusion proteins, with the exception of the SaAHR1 P380L mutant (Figure 13). The P380L mutation may have changed the shape of the protein enough to retain it in the cytoplasm, or this result could simply be due to low transfection efficiency. Most characterized AHR isoforms exhibit cytoplasmic localization and only translocate to the nucleus upon

binding with an agonist. The aberrant localization seen in the epifluorescence microscopy images may be due to the proximity of the eGFP to the AHR's N-terminal nuclear localization sequence. In order to test this, the C-terminal SaAHR1-eGFP fusion protein (which was already created in the Merson Lab, pDEST47SaAHR1) should be tested alongside the pDEST53SaAHR1 construct in a nuclear localization assay. If the C-terminal SaAHR1-eGFP fusion protein revealed similar results to the N-terminal eGFP-SaAHR1 fusion protein, this abnormal localization could be due to the SaAHR1 evolving a different function than most vertebrate AHRs. Another experiment that may change this localization would involve co-transfecting AHR chaperones, such as ARA9, in order to preserve the three-dimensional fold of SaAHR1, which may keep the NLS hidden from cellular machinery prior to ligand-binding. Another possibility regarding the localization of the eGFP-AHR fusion proteins could involve unknown endogenous ligands being present in the medium or in the COS7 cell line.

There are many ways to further probe the structure/function characteristics of SaAHR1. One such way would be to create a chimeric protein substituting domains of the DLC-binding SaAHR proteins, perhaps first concentrating on a chimeric SaAHR1 with an SaAHR2 LBD and a chimeric SaAHR1 with a SaAHR2 transactivation domain (TAD). This would allow us to determine whether SaAHR1 can be a functioning transcription factor even if the LBD is changed to the point of restoring agonist binding. Other labs have deleted the entire LBD of mouse AHR, which prevents binding to the chaperone HSP90, and ultimately causes "a ligand-independent constitutively active AHR"[55].

If it is determined that the SaAHR1 chimera containing the SaAHR2 LBD does not activate DREs, it would be interesting to test the library of mutant constructs created during the course of this study in ligand-binding assays. This would allow the binding affinities to be measured independent of any activation-induced gene transcription events. Ultimately, we may find that the SaAHR1 has evolved a different function than other AHRs, but further experimentation will be needed to make this determination.

5. Acknowledgments

Research reported in this thesis was supported in part by the Institutional Development Award (IDeA) Network for Biomedical Research Excellence from the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103430 through an investigator subaward to Dr. Rebeka Merson, and Rhode Island NSF Established Program to Stimulate Competitive Research (EPSCoR) Genomics and Sequencing Center, supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-1655221.

Special thanks to Dr. Michael Denison (University of California, Davis) for the pGudLuc 6.1 plasmid, Dr. Mark Hahn (Woods Hole Oceanographic Institution) for the FhAHR2 and FhARNT2 plasmids, and Dr. Dario Corrada and Dr. Laura Bonati (University of Milano-Bicocca) for the collaboration that provided mutation recommendations.

Thank you to Dr. Rebeka Merson, for teaching me how to conduct scientific research in a laboratory and for providing me many opportunities that I would have never otherwise had. Thank you to my thesis committee members: Dr. Sibel Karchner, Dr. William Holmes, and Dr. Mary Baker. Thank you to Dr. Lloyd Matsumoto for inspiring me to become a biologist during his Biology 111 introductory class. And last, but certainly not least, thank you to my wife Anna Bock, for always believing in me and being the best partner imaginable.

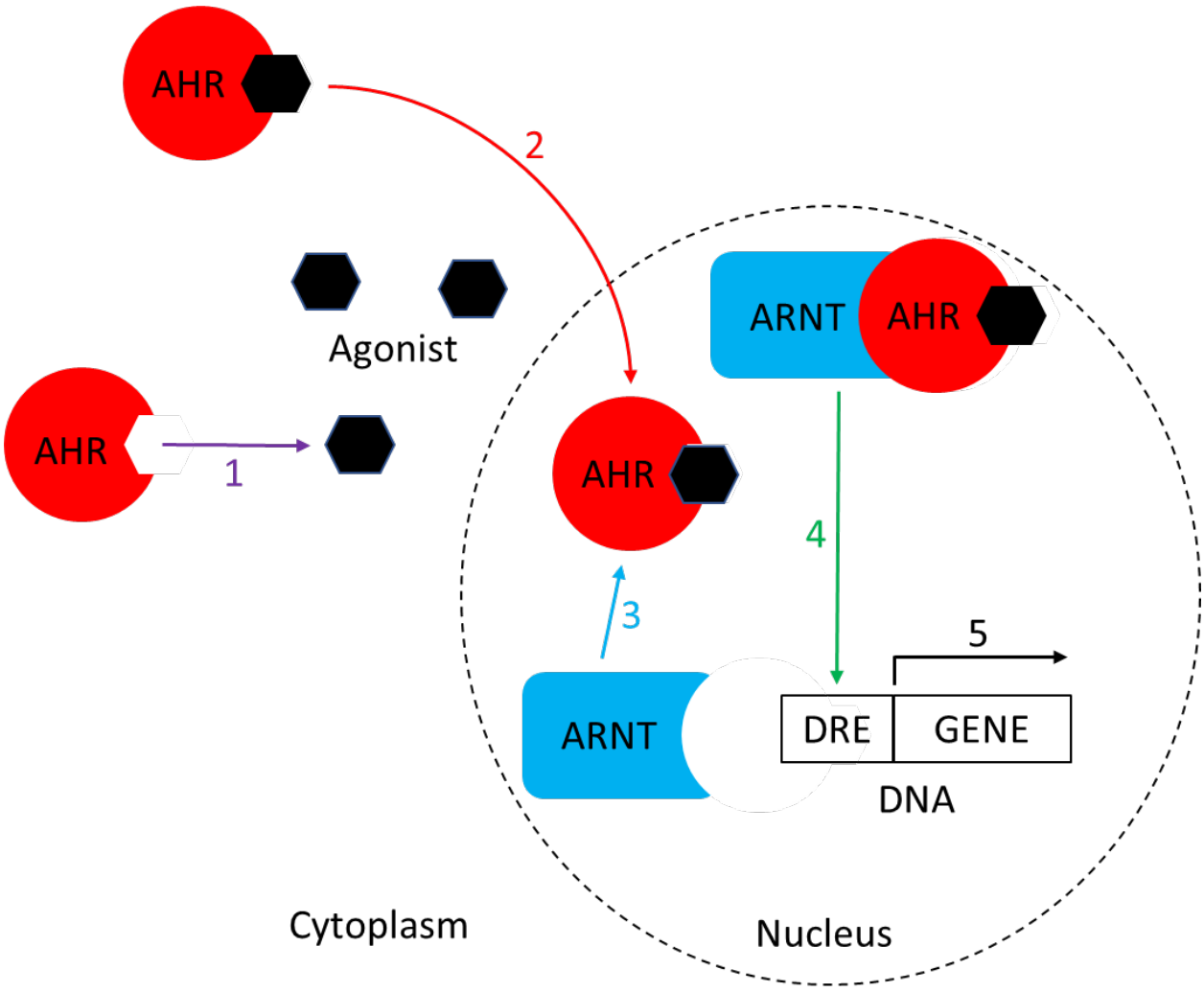


Figure 1: **AHR is a ligand-activated transcription factor.** (1) AHR binds agonist, (2) AHR translocates into nucleus, (3) ARNT dimerizes with AHR, (4) AHR:ARNT transcription factor complex binds to DREs on DNA, (5) genes under control of DREs are transcribed.

Table 1: AHR sequences chosen for analysis via amino acid alignment

Abbrevation	Species	Common Name	Isoform	Affinity to DLCs	Citation
Mm	*Mus musculus*	mouse	AHR	high	Denison *et al*. 1986
Hs	*Homo sapiens*	human	AHR	medium	Ema *et al*. 1994
Gg	*Gallus gallus*	chicken	AHR1	high	Karchner *et al*. 2006
Sh	*Sterna hirundo*	common tern	AHR1	low	Karchner *et al*. 2006
Xl	*Xenopus laevis*	African clawed frog	AHR1α	low	Odio *et al*. 2013
			AHR1β	low	Odio *et al*. 2013
Dr	*Danio rerio*	zebrafish	AHR1a	none	Fraccalvieri *et al*. 2012
			AHR1b	medium	Fraccalvieri *et al*. 2012
			AHR2	high	Fraccalvieri *et al*. 2012
Sa	*Squalus acanthias*	spiny dogfish	AHR1	none	Merson (unpublished)
			AHR2	high	Merson (unpublished)
			AHR3	high	Merson (unpublished)

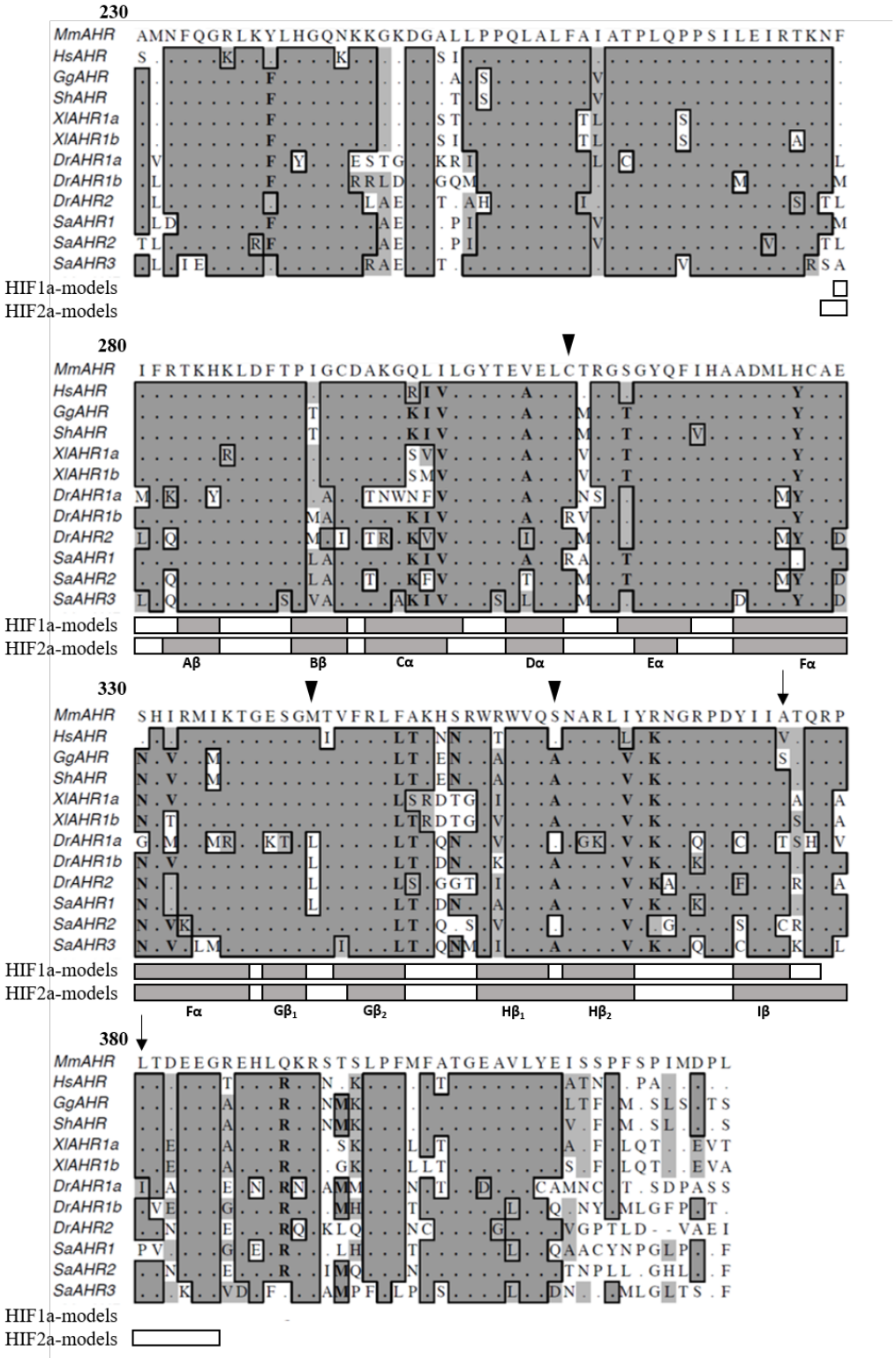


Figure 2: Amino acid alignment of AHR LBDs. Numbering conforms to full length mouse AHR amino acid sequence. Triangles denote mutations recommended by the Bonati Laboratory (R310C, L342M, and A359S) and arrows denote mutations determined by my analysis of homology models (A375C and P380L). Species abbreviations available in Table 1. Modelled area and secondary structures of models noted.

Table 2: 3D homology modeling and template information

Template	PDB Accession Number	Method	Metrics	Organism	Modelling		
					AHR isoform	Residues Modelled	Percent Identity
HIF1α	4H6J_A	X-ray diffraction	1.52 Å resolution	*Homo sapiens*	MmAHR	279-377	33.7% (281-374)
					SaAHR1	279-377	31.2% (281-375)
					SaAHR2	279-377	30.2% (281-375)
					SaAHR3	279-377	32.6% (281-374)
HIF2α	1P97	Solution NMR	20 conformers	*Homo sapiens*	MmAHR	278-385	31.4% (281-384)
					SaAHR1	278-385	26.0% (281-375)
					SaAHR1 P380L	278-385	25.7% (281-384)
					SaAHR2	278-385	27.6% (281-384)
					SaAHR3	278-385	29.6% (278-384)

Table 3: Site-directed mutagenesis primers

AA Mutation	DNA Mutation(s)	Primer Sequence
L342M	c1027a	5'- TCG GAA CAG AGT CAT GCC GCT CTC TCC TG -3'
A359S	g1078t	5'- CAG TCT TGC GTT TGA TTG GAC CCA GGC CC -3'
A375C	g1126t, c1127g, a1128c	5'- GTA CAA AAA TGG AAA ACC GGA TTA TAT CAT ATG CAC GCA GAG GCC TCC AG -3'
P380L	c1142t	5'- CTT CCT CGT CGA CTA GAG GCC TCT GCG TT -3'
		5'- AAC GCA GAG GCC TCT AGT CGA CGA GGA AG -3'

Table 4: Expression constructs

Name	Description	Source
pDEST53SaAHR1 WT	*Squalus acanthias* AHR1 full length, wildtype, N-terminal GFP fusion protein	Merson Laboratory, previous work[B]
pDEST53SaAHR1 R310C	*Squalus acanthias* AHR1 full length, R310C mutant, N-terminal GFP fusion protein	Merson Laboratory, previous work[B]
pDEST53SaAHR1 L342M	*Squalus acanthias* AHR1 full length, L342M mutant, N-terminal GFP fusion protein	Merson Laboratory, previous work[B]
pDEST53SaAHR1 A359S	*Squalus acanthias* AHR1 full length, A359S mutant, N-terminal eGFP fusion protein	Merson Laboratory, previous work[B]
pDEST53SaAHR1 A375C	*Squalus acanthias* AHR1 full length, A375C mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 P380L	*Squalus acanthias* AHR1 full length, P380L mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 R310C/A359S	*Squalus acanthias* AHR1 full length, R310C/A359S double mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 R310C/P380L	*Squalus acanthias* AHR1 full length, R310C/P380L double mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 L342M/A359S	*Squalus acanthias* AHR1 full length, L342M/A359S double mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 L342M/P380L	*Squalus acanthias* AHR1 full length, L342M/P380L double mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 A359S/P380L	*Squalus acanthias* AHR1 full length, A359S/P380L double mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 R310C/A359S/P380L	*Squalus acanthias* AHR1 full length, R310C/A359S/P380L triple mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR1 L342M/A359S/P380L	*Squalus acanthias* AHR1 full length, L342M/A359S/P380L triple mutant, N-terminal eGFP fusion protein	This project
pDEST53SaAHR2	*Squalus acanthias* AHR2 full length, wildtype, N-terminal eGFP fusion protein	Merson Laboratory, previous work[B]
pcDNA EV	pcDNA empty vector, same backbone as pDEST53 sans eGFP	Merson Laboratory, previous work
pGudLuc 6.1	Inducible firefly luciferase, under control of multiple dioxin responsive element promoters	Gift from M. Denison, UC Davis
pGudLuc 4.13	Constitutive firefly luciferase	Gift from Promega technical support
pRL-TK	Constitutive renilla luciferase	Promega Dual Luciferase Reporter Assay Kit
FhAHR2	*Fundulus heteroclitus* AHR2 full length	Gift from M. Hahn, WHOI
FhARNT2	*Fundulus heteroclitus* ARNT2 full length	Gift from M. Hahn, WHOI

[B] Hersey S. 2008. Cellular Localization of *Squalus acanthias* aryl hydrocarbon receptor 2. Undergraduate honors project, Rhode Island College.

Table 5: Thermal cycling parameters

Segment	Cycles	Temp. (°C)	Time (minutes)
Denaturation	1	95	2:00
Denaturation	30	95	0:20
Annealing		55	0:30
Extension		65	4:30
Final Extension	1	65	5:00

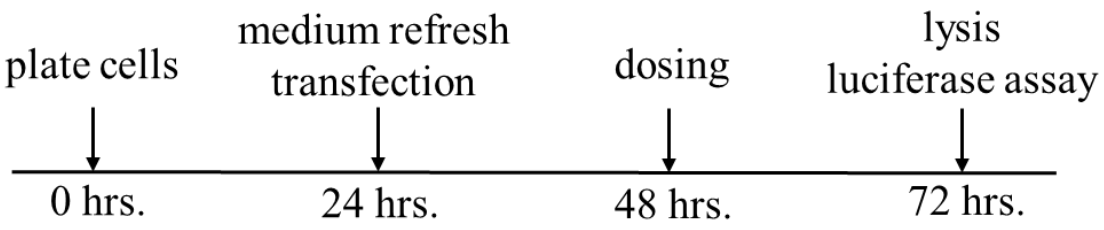


Figure 3: Experimental design for luciferase reporter gene assays. C35 cells were seeded into 48-well plates at 30,000 cells/well. At 24 hrs post-plating, medium was refreshed and DNA constructs co-transfected with Lipofectamine 2000. At 48 hrs post-plating, cells were dosed with either PCB-126 in DMSO (100 nM final concentration) or vehicle-only control. At 72 hrs post-plating, cells were lysed and lysates used for dual luciferase reporter gene assays.

Table 6: Luciferase assays performed during the course of this study.

Assay	Date	Cell type	Transfected plasmids	DNA (ng/well)	Transfection reagent	Objective of experiment
1	2/19/2015	C35	pDEST53Sa2	50	Turbofect	Assessment of AHR1 double mutants
			pDEST53Sa1 WT	50		(R310C/A359S, L342M/A359S).
			pDEST53Sa1 R310C/A359S	50		
			pDEST53Sa1 L342M/A359S	50		
			pGudLuc 6.1	20		Results shown in Appendix 4, 5
			pRL-TK	3		
			pcDNA empty vector	50		
2	3/16/2015	C35	pDEST53Sa2	15	Lipofectamine 2000	Assessment of AHR1 double mutants
			pDEST53Sa1 WT	50		(R310C/A359S, L342M/A359S).
			pDEST53Sa1 R310C/A359S	50		
			pDEST53Sa1 L342M/A359S	50		
			pGudLuc 6.1	20		Results shown in Appendix 4, 5
			pRL-TK	3		
			pcDNA empty vector	50		
3	4/10/2015	C35	pDEST53Sa2	15	Lipofectamine 2000	Optimization of experimental timeline: Dosing 6
			pDEST53Sa2	30		or 24 hours post-transfection.
			pGudLuc 6.1	20		
			pRL-TK	3		Results shown in Appendix 6, 9
4	4/23/2015	C35	pDEST53Sa2	15	Lipofectamine 2000	Optimization of experimental timeline: Dosing 6
			pDEST53Sa2	30		or 24 hours post-transfection.
			pGudLuc 6.1	20		
			pRL-TK	3		Results shown in Appendix 6, 9
5	7/7/2015	C35	pDEST53 Sa2	15	Lipofectamine 2000	Assessment of AHR1 with diproline segment
			pDEST53 Sa1 WT	50		reverted to P-L.
			pDEST53 Sa1 P380L	50		
			pGudLuc 4.13	50		
			pGudLuc 6.1	20		Results shown in Appendix 3, 7
			pRL-TK	3		
			pcDNA EV	50		
6	7/16/2015	C35	pDEST53Sa2	15	Lipofectamine 2000	Assessment of AHR1 with diproline segment
			pDEST53Sa1 WT	50		reverted to P-L, with FhARNT2 added to all
			pDEST53Sa1 P380L	50		wells.
			FhAHR2	20		
			FhARNT2	50		
			pGudLuc 4.13	50		Results shown in Figure 12, Appendix 7
			pGudLuc 6.1	20		
			pRL-TK	3		
			pcDNA EV	50		

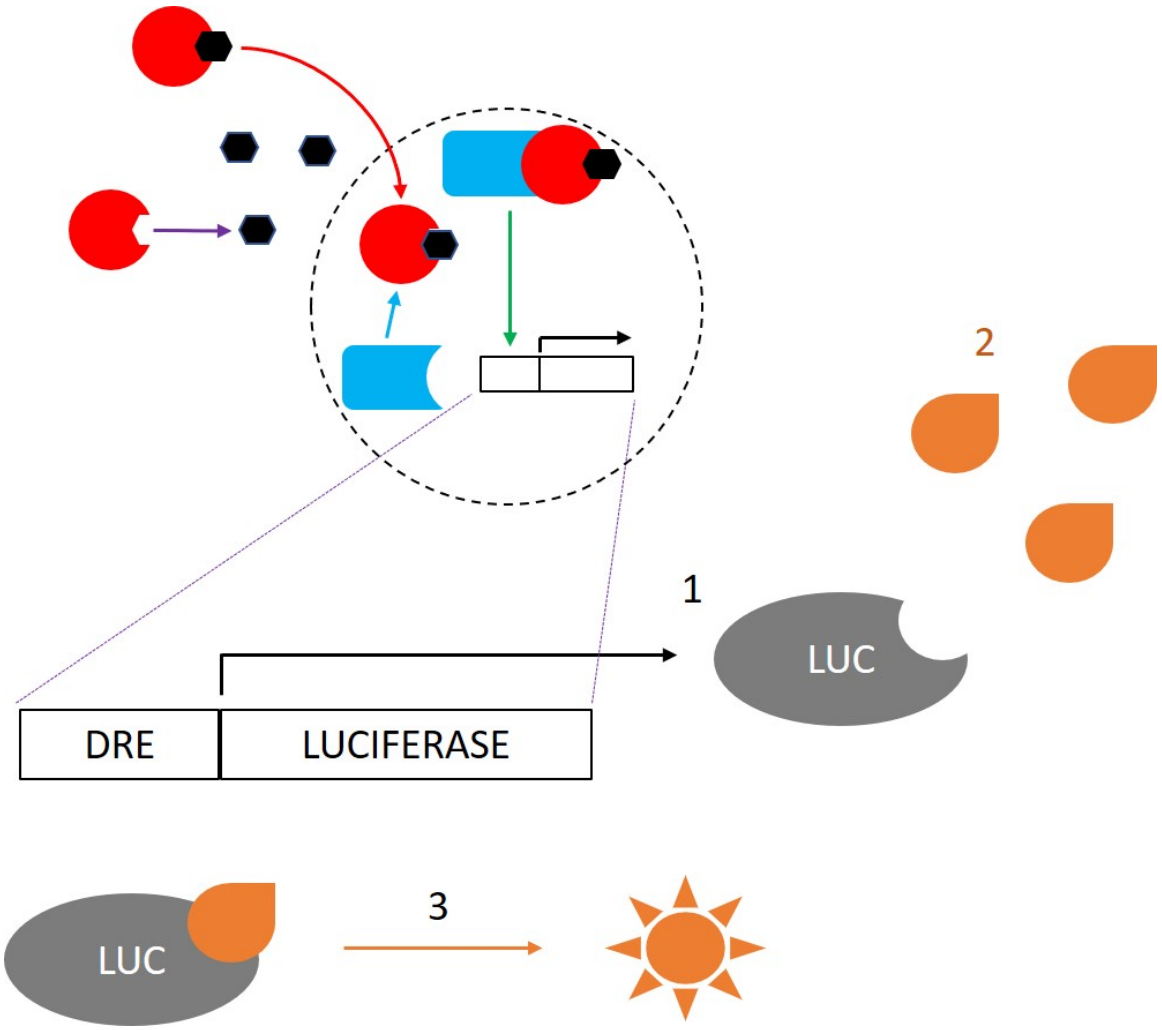


Figure 4: Luciferase reporter gene assay of AHR activation. (1) Activated AHR:ARNT transcription factor complex aids in expression of firefly luciferase, (2) luciferin substrate added to lysates, (3) luciferase catalyzes reaction on substrate, which produces luminescence.

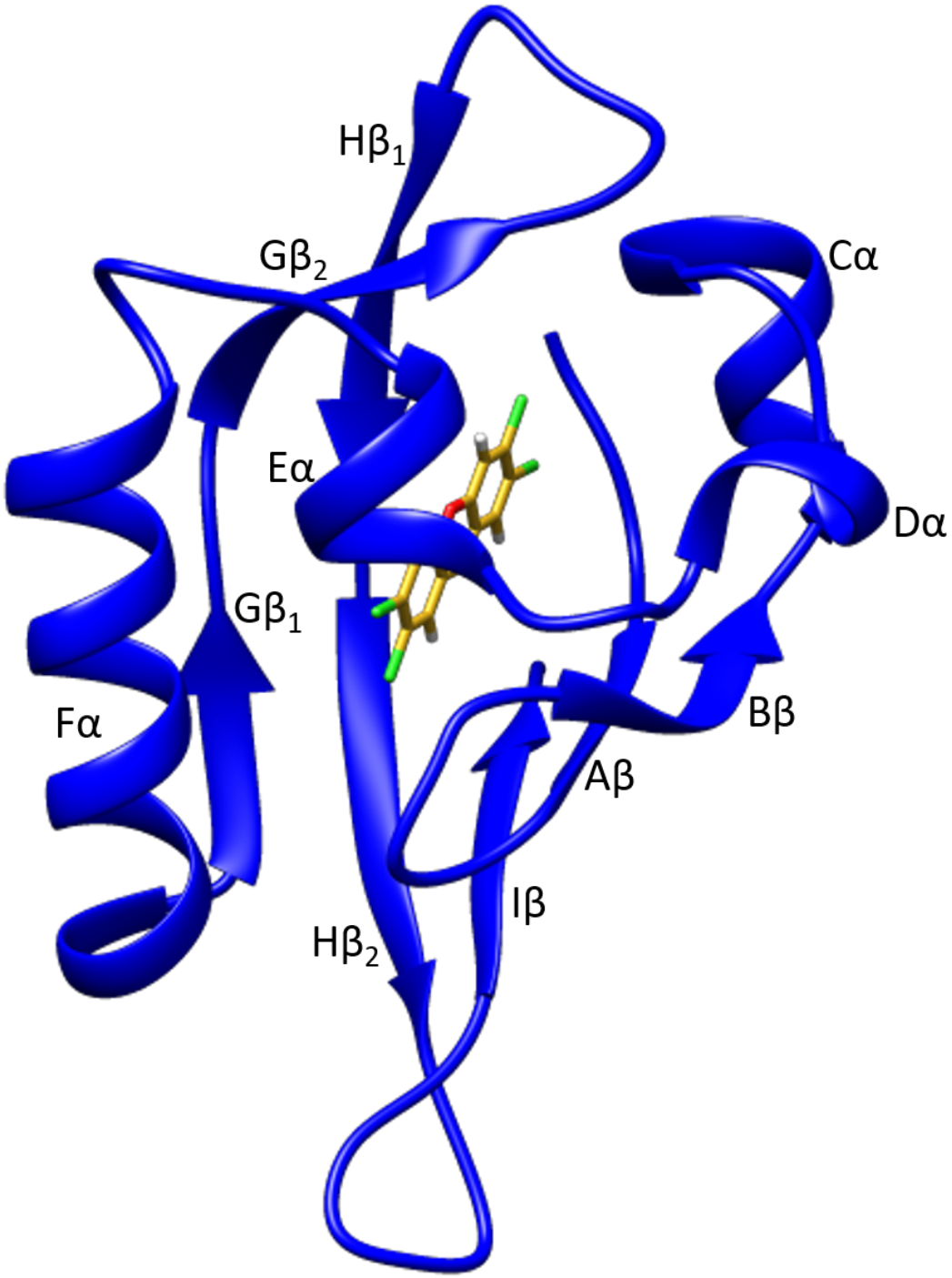


Figure 5: A HIF1α-based 3D model of the SaAHR2 LBD complexed with TCDD. Nomenclature of secondary structure is overlaid. Residues 279-377.

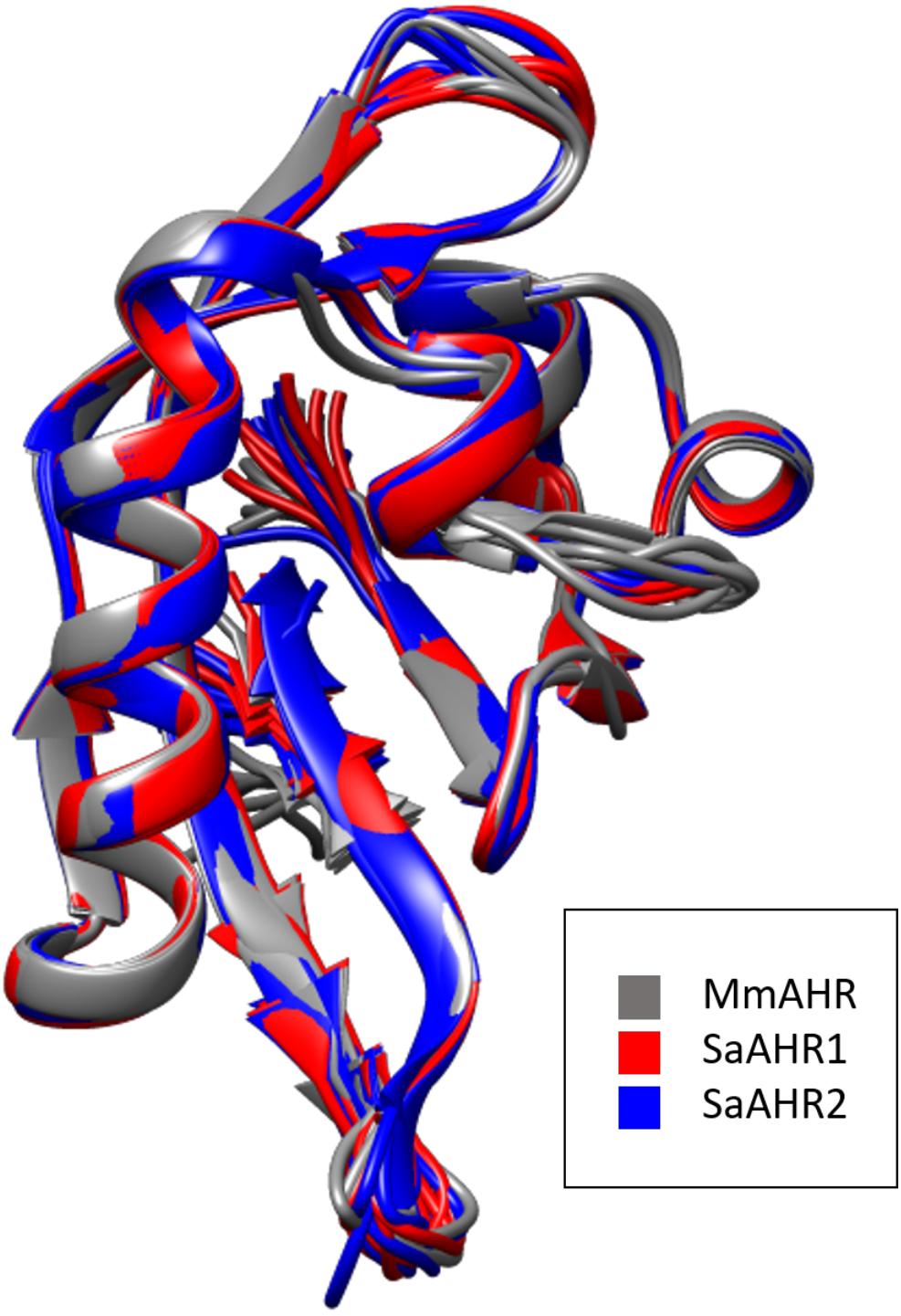


Figure 6: A cluster of HIF1α-based AHR models. 10 Models for each MmAHR WT (gray), SaAHR1 WT (red), and SaAHR2 WT (blue). Residues 279-377.

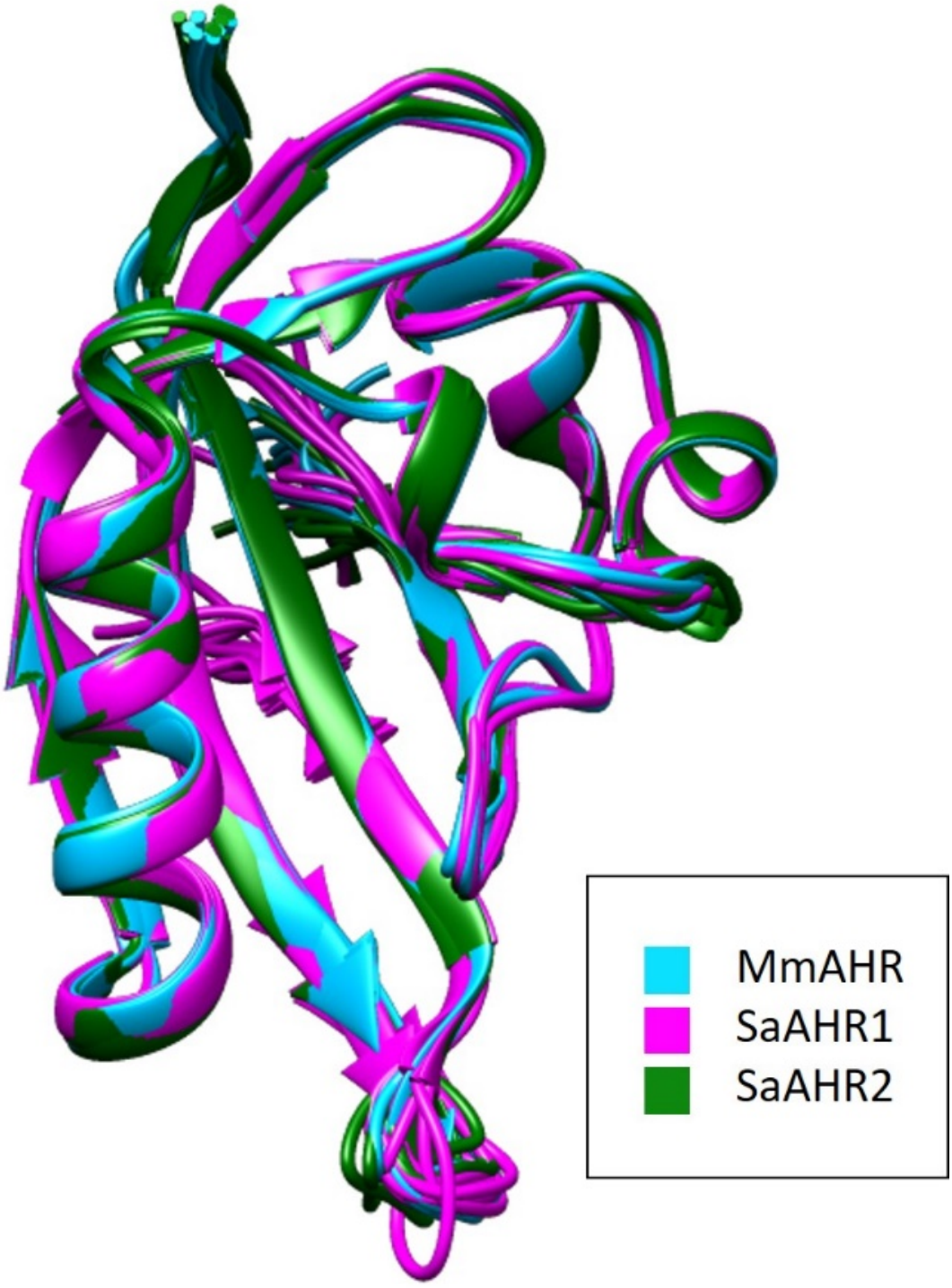


Figure 7: A cluster of HIF2α-based AHR models. 10 Models for each MmAHR WT (light blue), SaAHR1 WT (magenta), and SaAHR2 WT (dark green). Residues 278-385. (Ribbon positions 379-385 hidden in SaAHR1 for ease of viewing.

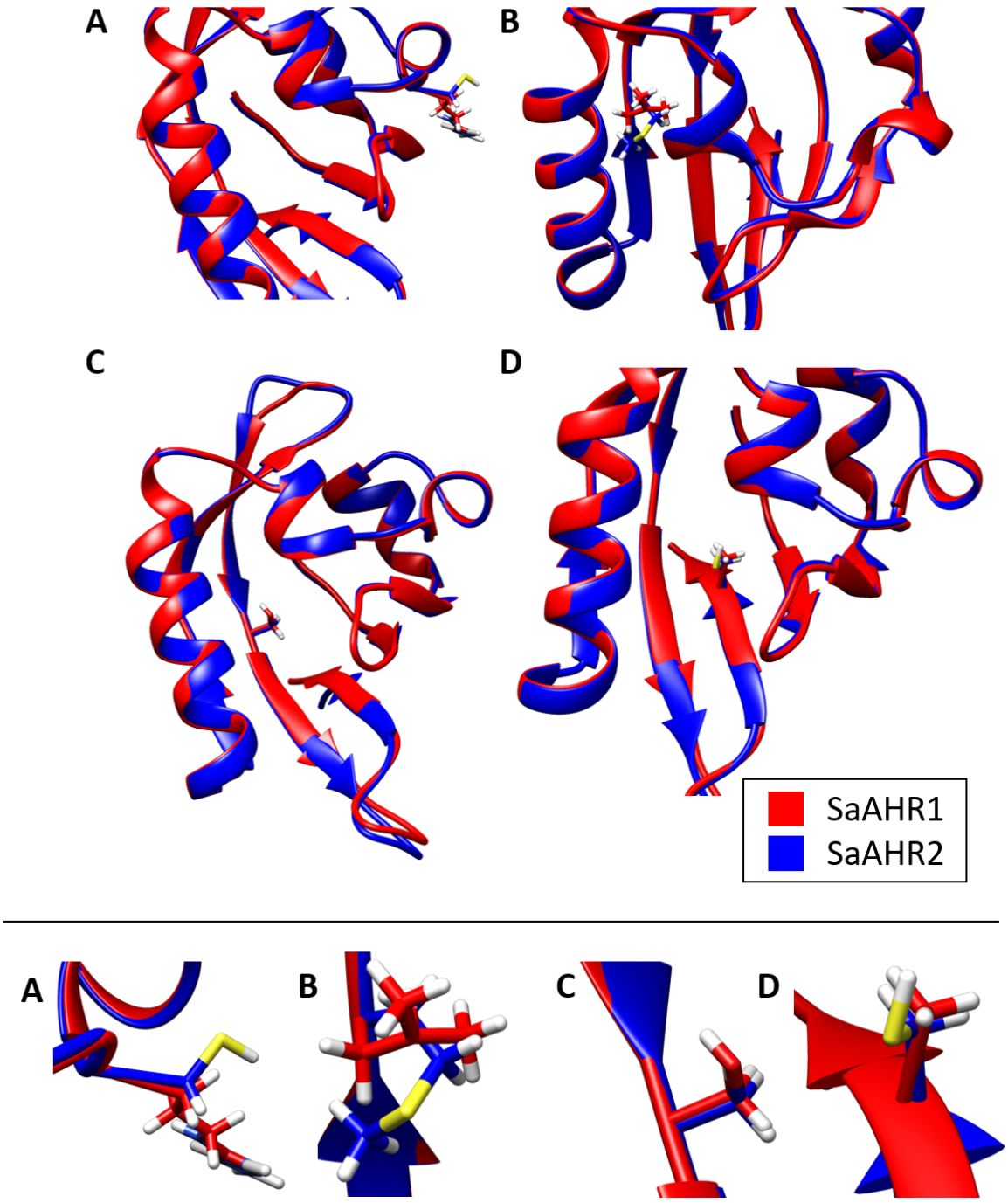


Figure 8: HIF1α-based structural models of SaAHR1 (red) and SaAHR2a (blue): Molecular structure of residues of interest shown (A) R310C, (B) L342M, (C) A359S, (D)A375C.

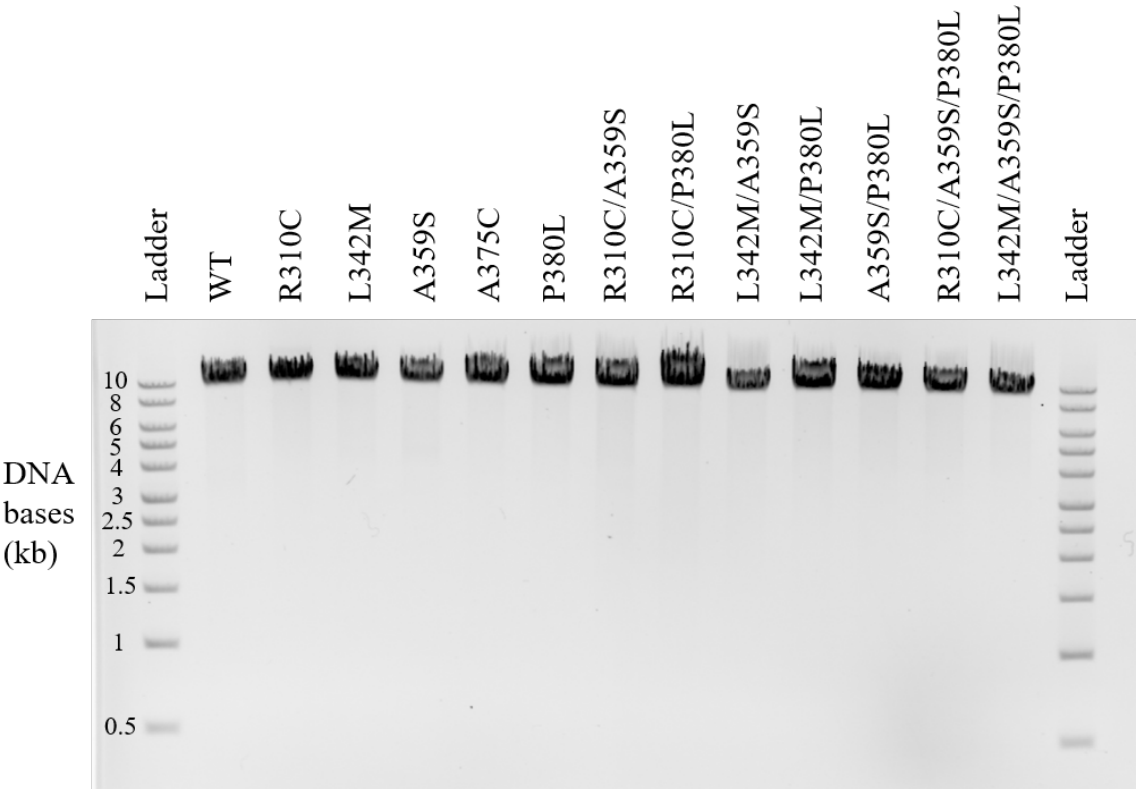


Figure 9: DNA gel electrophoresis of pDEST53SaAHR1 WT and mutant constructs. Mutant constructs are present and consistent in size with wildtype.

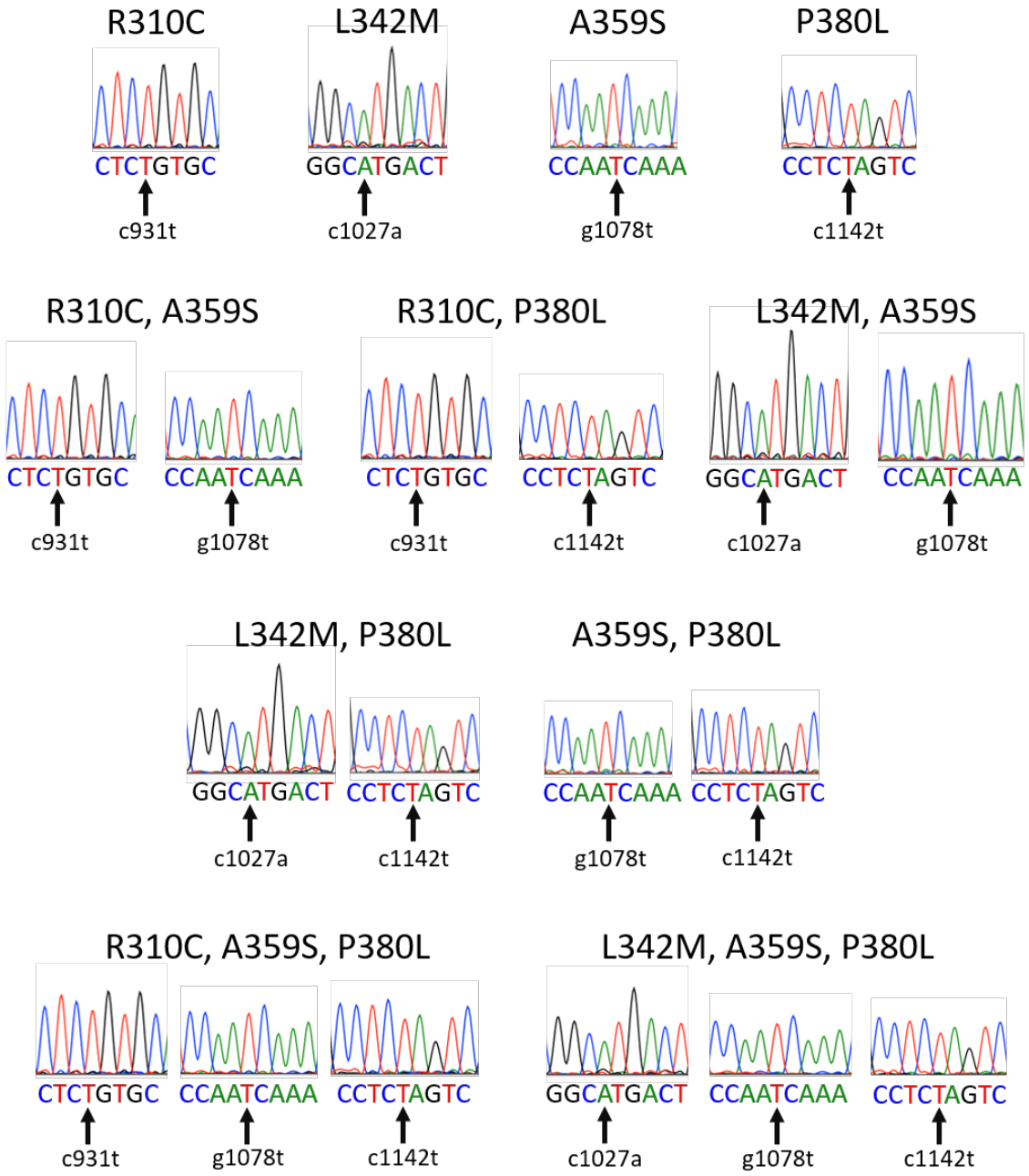


Figure 10: Chromatograms of Sanger sequencing reads show the presence of specified mutations in each mutant DNA construct, with the exception of A375C (no data available).

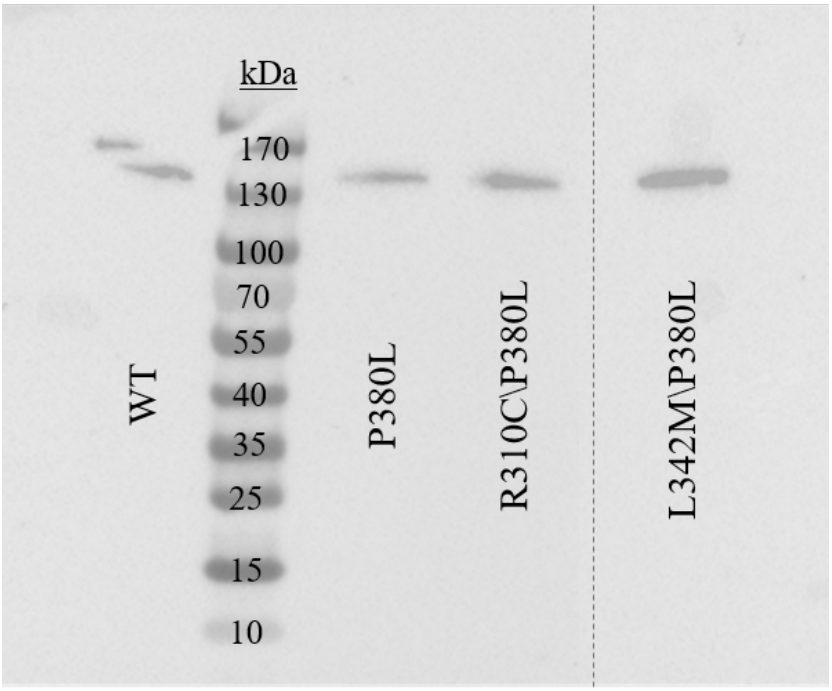


Figure 11: Western blot of pDEST53SaAHR1 WT and mutant constructs show expressed protein with a size consistent with pDEST53SaAHR1 WT calculated size (131.5 kDa). Blotting performed using rabbit anti-GFP-N-terminal primary antibody. Dotted line represents area of blot that was removed in image editing software.

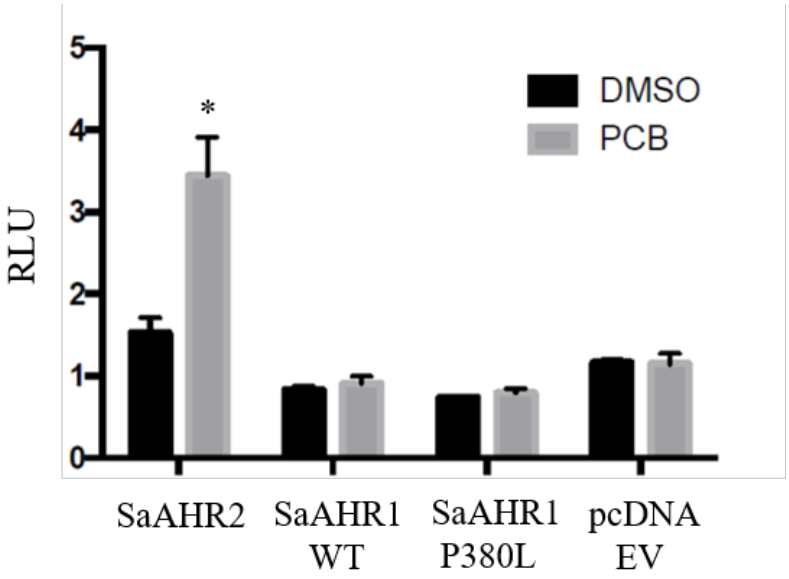


Figure 12: Luciferase assay 6, the P380L mutation did not cause PCB 126 induced activation in dual luciferase reporter gene assay. SaAHR2 was used as a positive control and a pcDNA empty vector was used as a negative control. All replicates were cotransfected with FhARNT2. Bars denote mean value of triplicate wells, error bars denote 1 standard deviation. PCB 126 final concentration of 100 nM. Asterisk denotes values significantly different ($P<0.01$) from DMSO only control, Student's *t* test.

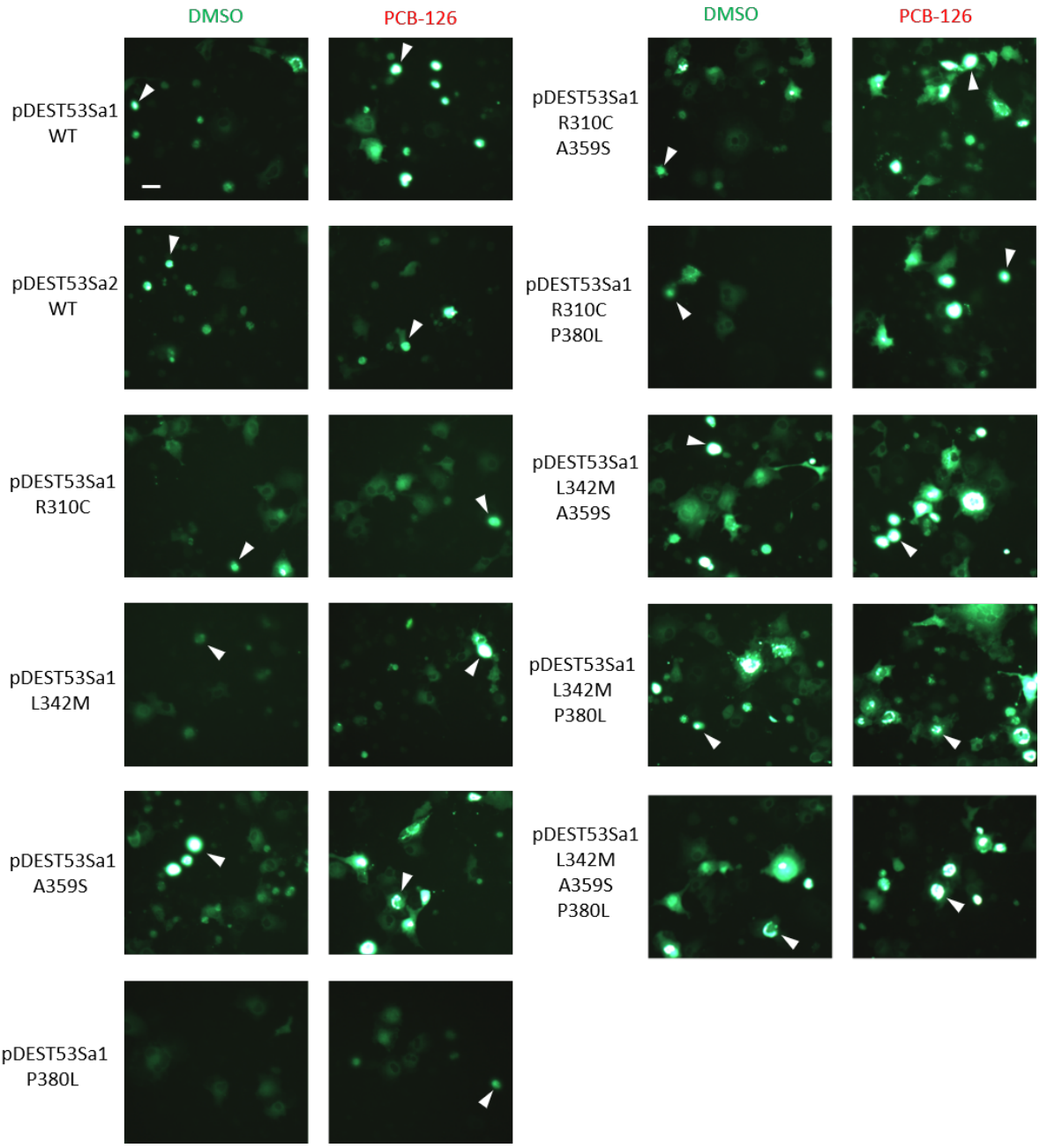


Figure 13: Epifluorescence microscopy (20X objective) of eGFP-AHR-transfected COS7 cellular localization assay. Images were taken 30 hours post dosing (50 nM final concentration) with either PCB-126 in DMSO or vehicle only. White triangles denote possible nuclear localization. Scale bar in top left image measures 50 µm.

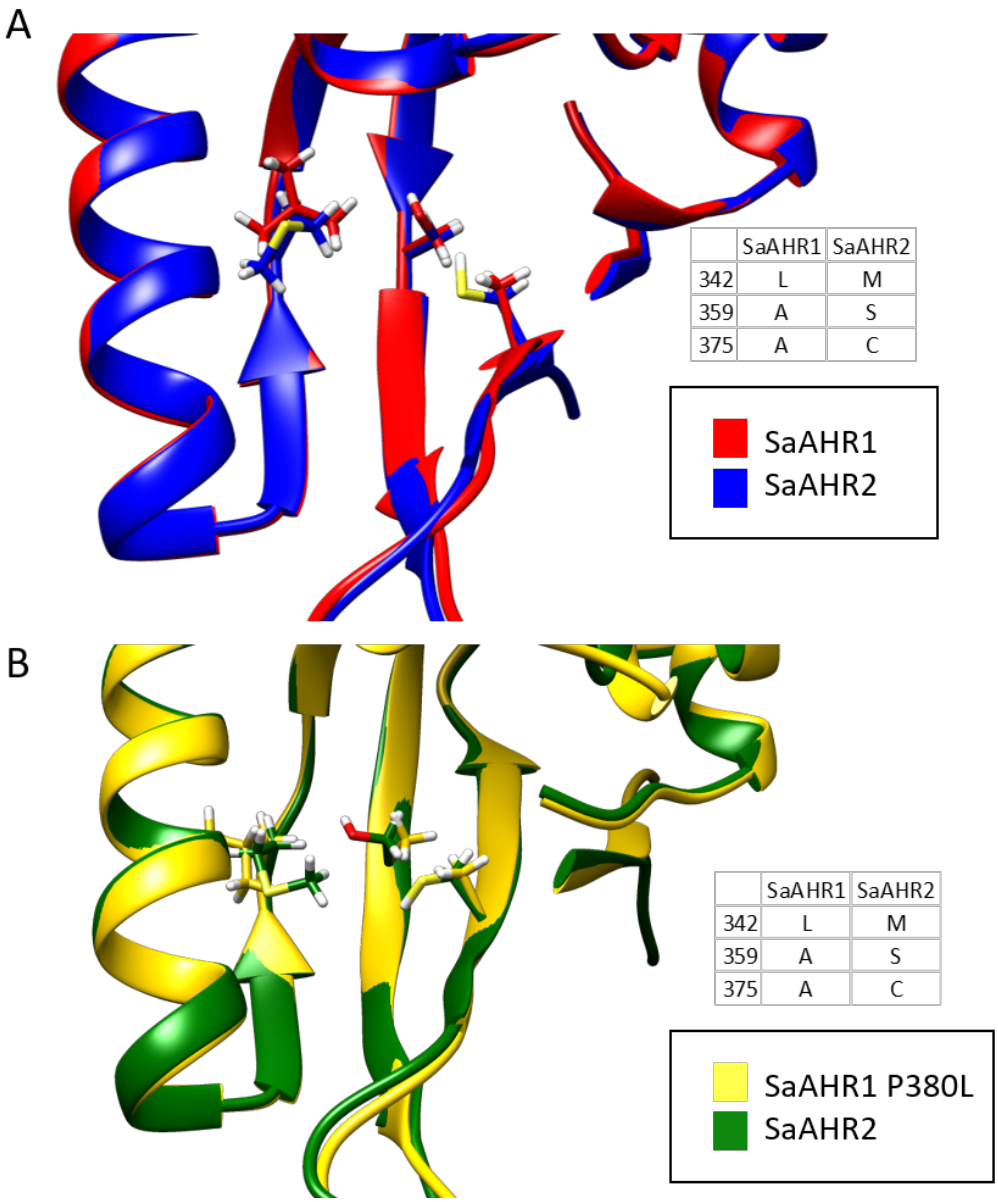


Figure 14: Residues 342, 359, and 375 are adjacent and could form a charged surface for ligand binding. (A) HIF1α-based models, (B) HIF2α-based models. Ribbon positions 284-291 hidden for ease of viewing.

References

---

[1] Walker CH, Sibly RM, Hopkin SP, Peakall DB. 2012. *Principles of Ecotoxicology,* 4th edition. CRC Press. 360 Pages.

[2] Perron MM, Burgess RM, Suuberg EM, Cantwell MG, Pennell KG. 2013. Performance of passive samplers for monitoring estuarine water column concentrations 1. Contaminants of concern. *Environmental Toxicology and Chemistry* 32(10): 2182–2189.

[3] Lu Z, Fisk AT, Kovacs KM, Lydersen C, McKinney MA, Tomy GT, Rosenburg B, McMeans BC, Muir DCG, Wong CS. 2014. Temporal and spatial variation in polychlorinated biphenyl chiral signatures of the Greenland shark (*Somniosus microcephalus*) and its arctic marine food web. *Environmental Pollution* 186: 216-225.

[4] Hall JE. 2016. *Guyton and Hall Textbook of Medical Physiology,* 13th edition. Elsevier. 1145 pages.

[5] Garrett RH, Grisham CM. 2013. *Biochemistry*, 5th edition. Cengage Learning. 1169 pages.

[6] Goodman HM. 2009. *Basic Medical Endocrinology*, 4th edition. Academic Press. 309 Pages.

[7] US EPA. Report on the environment: Greenhouse gas concentrations. Available from https://cfpub.epa.gov/roe/indicator_pdf.cfm?i=24. Accessed July 2019.

[8] Nadal M, Marquès M, Mari M, Domingo JL. 2015. Climate change and environmental concentrations of POPs: A review. *Environmental Research* 143: 177-185.

[9] World Health Organization. 2016. Dioxins and their effects on human health. Available from https://www.who.int/news-room/fact-sheets/detail/dioxins-and-their-effects-on-human-health. Accessed May 2019.

[10] US Department of Veteran Affairs. Agent Orange active ingredients and characteristics. Available from https://www.publichealth.va.gov/exposures/agentorange/basics.asp. Accessed July 2019.

[11] Safe S. 1990. Polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds: Environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). *Critical Reviews in Toxicology* 21(1): 50-88.

[12] Pesch C, Voyer R, Latimer J. 2011. Imprint of the past: ecological history of New Bedford Harbor. US EPA.

[13] US EPA. EPA cleanups: Communities around New Bedford Harbor. Available from https://www.epa.gov/new-bedford-harbor. Accessed June 2019.

[14] Agency for Toxic Substances and Disease Registry. 2014. Polychlorinated biphenyls (PCBs) toxicity. What are polychlorinated biphenyls (PCBs)? Available from https://www.atsdr.cdc.gov/csem/csem.asp?csem=30&po=4. Accessed March 2019.

[15] Burbach KM, Poland A, Bradfield CA. 1992. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proceedings of the National Academy of Sciences of the United States of America* 89: 8185-8189.

[16] Hankinson O. 1995. The aryl hydrocarbon receptor complex. *Annual review of pharmacology and toxicology* 35: 307-340.

[17] Zhao B, Bohonowych JES, Timme-Laragy A, Jung D, Affatato AA, Rice RH, Di Giulio RT, Denison MS. 2013. Common commercial and consumer products contain activators of the aryl hydrocarbon (dioxin) receptor. *PLOS ONE* 8(2): e56860.

[18] Zelante T, Iannitti RG, Cunha C, De Luca A, Giovannini G, Pieraccini G, Zecchi R, D'Angelo C, Massi-Benedetti, Fallarino F, Carvalho A. 2013. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* 39: 372-385.

[19] Kewley RJ, Whitelaw ML, Chapman-Smith A. 2004. The mammalian basic-helix-loop-helix/PAS family of transcriptional regulators. *The International Journal of Biochemistry and Cell Biology* 36: 189-204.

[20] Pongratz I, Mason GG, Poellinger L. 1992. Dual roles of the 90-kDa heat shock protein hsp90 in modulating functional activities of the dioxin receptor. *The Journal of Biological Chemistry* 267(19): 13728-13734.

[21] Meyer BK, Perdew GH. 1999. Characterization of the AhR-hsp90-XAP2 core complex and the role of the immunophilin-related protein XAP2 in AhR stabilization. *Biochemistry* 38: 8907–8917.

[22] Nguyen PM, Wang D, Wang Y, Li Y, Uchizono JA, Chan WK. 2012. p23 co-chaperone protects the aryl hydrocarbon receptor from degradation in mouse and human cell lines. *Biochemical Pharmacology* 84(6) :838-850.

[23] Denison MS, Fisher JM, Whitlock JP. 1988. The DNA recognition site for the dioxin-Ah receptor complex. *The Journal of Biological Chemistry* 263(33): 17221-17224.

[24] Tijet N, Boutros PC, Moffat ID, Okey AB, Tuomisto J, Pohjanvirta R. 2006. Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries. *Molecular Pharmacology* 69: 140-153.

[25] Harill JA, Hukkanen RR, Lawson M, Martin B, Gilger B, Soldatow V, LeCluyse EL, Budinsky RA, Rowlands JC, Thomas RS. 2013. Knockout of the aryl hydrocarbon receptor results in distinct hepatic and renal phenotypes in rats and mice. *Toxicology and Applied Pharmacology* 272: 503–518.

[26] Smith BW, Rozelle SS, Leung A, Ubellacker J, Parks A, Nah SK, French D, Gadue P, Monti S, Chui DHK, Steinberg MH, Frelinger AL, Michelson AD, Theberge R, McComb

ME, Costello CE, Kotton DN, Mostoslavsky G, Sherr DH, Murphy GJ. 2013. The aryl hydrocarbon receptor directs hematopoietic progenitor cell expansion and differentiation. *Blood* 122: 376-385.

[27] Hubbard TD, Murray IA, Bisson WH, Lahoti TS, Gowda K, Amin SG, Patterson AD, Perdew GW. 2015. Adaptation of the human aryl hydrocarbon receptor to sense microbiota-derived indoles. *Scientific Reports* 5: 12689.

[28] Singh NP, Singh UP, Singh B, Price RL, Nagarkatti M, Nagarkatti PS. 2011. Activation of aryl hydrocarbon receptor (AHR) leads to reciprocal epigenetic regulation of FoxP3 and IL-17 expression and amelioration of experimental colitis. *PLOS ONE* 6(8): e23522.

[29] Hahn, ME. 2002. Aryl hydrocarbon receptors: diversity and evolution. *Chemico-Biological Interactions* 141: 131 – 160

[30] Hahn ME, Karchner SI, Evans BR, Franks DG, Merson RR, Lapseritis JM. 2006. Unexpected diversity of aryl hydrocarbon receptors in non-mammalian vertebrates: Insights from comparative genomics. *Journal of Experimental Zoology* 305A: 693-706.

[31] Merson RR, *et al*. (in prep)

[32] Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. 2004. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *The Journal of Biological Chemistry* 279(23): 23847–23850.

[33] Walden R, Schiller CM. 1985. Comparative toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in four (sub)strains of adult male rats. *Toxicology and Applied Pharmacology* 77(3): 490-495.

[34] Hahn, ME. 1998. The aryl hydrocarbon receptor: a comparative perspective. Comparative biochemistry and physiology. Part C, Pharmacology, toxicology and endocrinology 121(1-3): 23-53.

[35] Andreasen EA, Tanguay RL, Peterson RE, Heideman W. 2002. Identification of a critical amino acid in the aryl hydrocarbon receptor. *The Journal of Biological Chemistry* 277(15): 13210-13218.

[36] Pandini A, Denison M, Song Y, Soshilov A, Bonati L. 2006. Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. *Biochemistry* 46: 696-708.

[37] Denison MS, Soshilov AA, He G, DeGroot DE, Zhao B. 2011. Exactly the same but different: Promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicological Sciences* 124(1): 1-22.

[38] Fraccalvieri D, Soshilov A, Karchner SI, Franks DG, Pandini A, Bonati L, Hahn ME, Denison MS. 2013. Comparative analysis of homology models of the Ah receptor ligand binding domain: verification of structure-function predictions by site-directed mutagenesis of a nonfunctional receptor. *Biochemistry* 52: 714-725.

[39] Odio C, Holzman SA, Denison MS, Fraccalvieri D, Bonati L, Franks DG, Hahn ME, Powell WH. 2013. Specific ligand binding domain residues confer low dioxin responsiveness to AHR1b of *Xenopus laevis*. *Biochemistry* 52: 1746-1754.

[40] Mattingly C, Parton A, Dowell L, Rafferty J, Barnes D. 2004. Cell and molecular biology of marine elasmobranchs: *Squalus acanthias* and *Raja erinacea*. *Zebrafish* 1(2): 111-120.

[41] Saha I, Shamala N. 2011. Investigating diproline segments in proteins: occurrences, conformation, and classification. *Biopolymers* 97(1): 54-64.

[42] Karchner SI, Franks DG, Kennedy SW, Hahn ME. 2006. The molecular basis for differential dioxin sensitivity in birds: Role of the aryl hydrocarbon receptor. *PNAS* 103(16): 6252-6257.

[43] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 21: 2947-2948.

[44] Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology* 234: 779-815.

[45] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25(13): 1605-1612.

[46] Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

[47] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J*ournal of Molecular Biology* 215: 403-410.

[48] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Research* 28: 235-242.

[49] Cardoso R, Love R, Nilsson CL, Bergqvist S, Nowlin D, Yan J, Liu KK, Zhu J, Chen P, Deng YL, Dyson HJ, Greig MJ, Brooun A. 2012. Identification of Cys255 in HIF-1α as a novel site for development of covalent inhibitors of HIF-1α/ARNT PasB domain protein-protein interaction. *Protein Science* 21(12): 1885-1896.

[50] Erbel PJ, Card PB, Karakuzu O, Bruick RK, Gardner KH. 2003. Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor. *Proceedings of the National Academy of Sciences of the United States of America* 100(26): 15504-15509.

[51] Melo F, Sanchez R, Sali A. 2001. Statistical potentials for fold assessment. *Protein Science* 11: 430-448.

[52] Chimera User's Guide. 2015. Comparative modelling. Accessed from https://www.rbvi.ucsf.edu/chimera/docs/UsersGuide/tutorials/dor.html

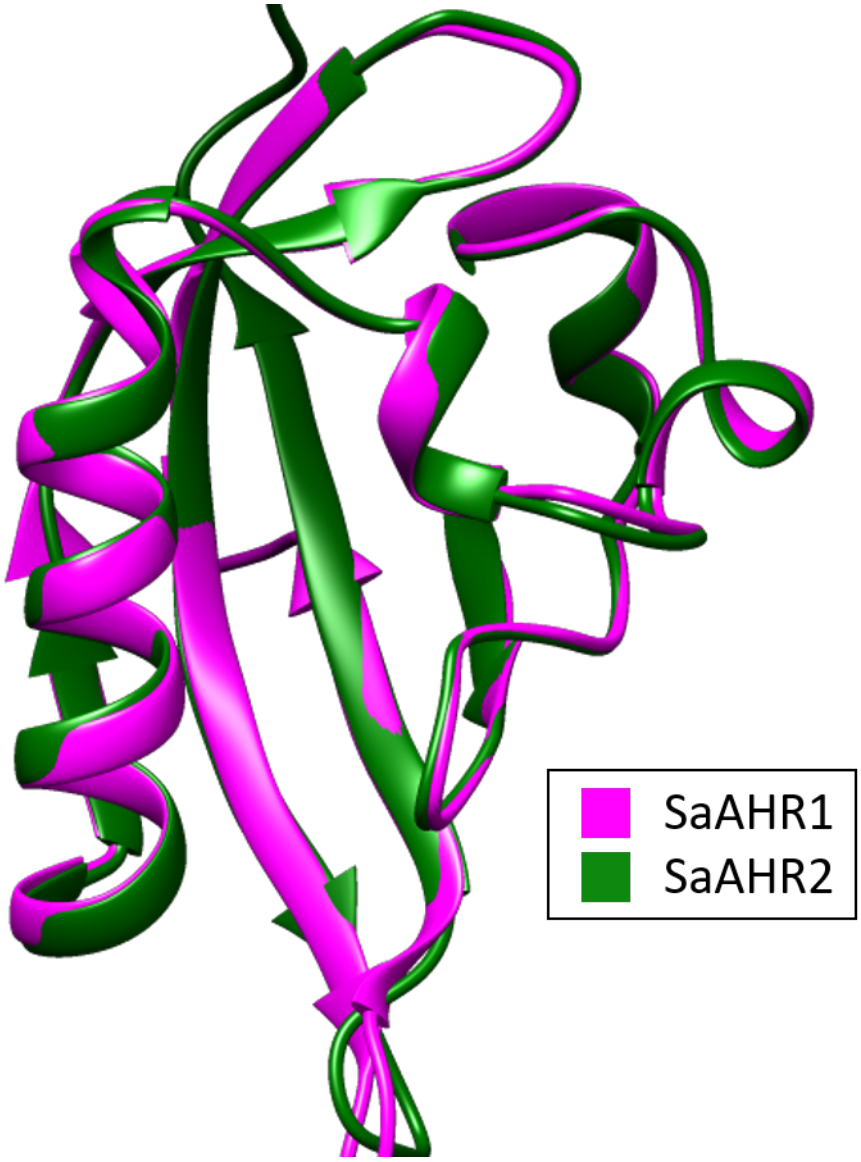
[53] Shen MY, Sali A. 2006. Statistical potential for assessment and prediction of protein structures. *Protein Science* 15(11): 2507-2524.

[54] Xing Y, Nukaya M, Satyshur KA, Jiang L, Stanevich V, Korkmaz EN, Burdette L, Kennedy GD, Cui Q, Bradfield CA. 2012. Identification of the Ah-receptor structural determinants for ligand preferences. *Toxicological Sciences* 129(1): 86-97.

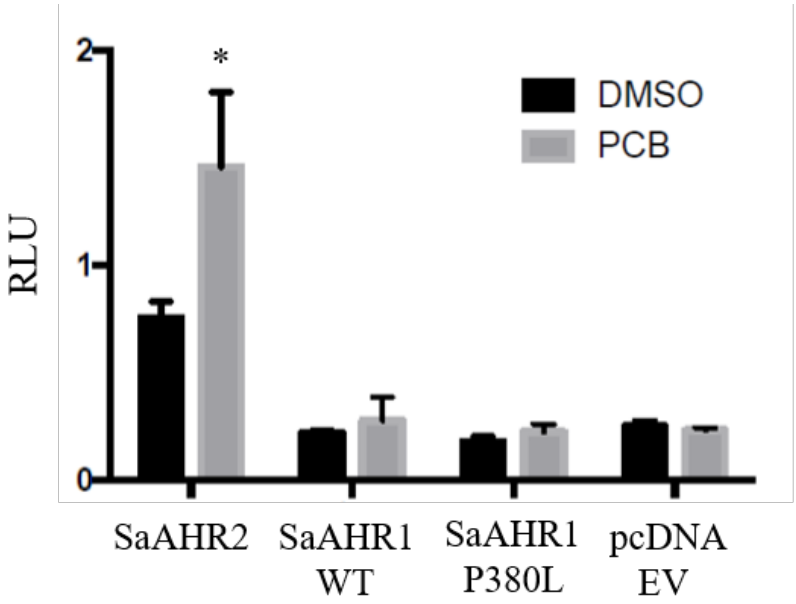
[55] McGuire J, Okamoto K, Whitelaw ML, Tanaka H, Poellinger L. 2001. Definition of a dioxin receptor mutant that is a constitutive activator of transcription: delineation of overlapping repression and ligand binding functions within the PAS domain. *The Journal of Biological Chemistry* 276: 41841-41849.

Appendix 1: Raw scores from Modeller results. Bold font denotes one representative model from each isoform.

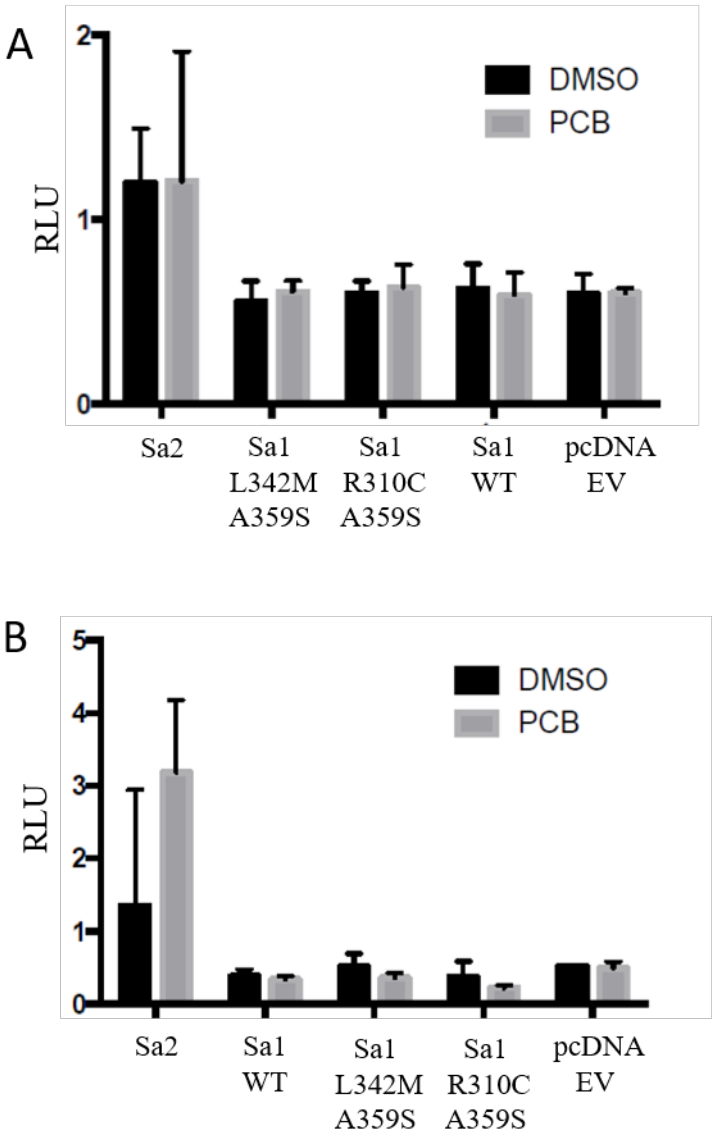
	HIF1α-based			HIF2α-based		
Isoform	Model	GA341	zDOPE	Model	GA341	zDOPE
MmAHR	1	0.81	-0.07	**1**	**0.93**	**-0.41**
	2	0.70	0.00	2	0.88	-0.64
	3	0.59	-0.04	3	0.82	-0.47
	4	0.85	0.04	4	0.87	-0.39
	5	0.87	-0.03	5	0.92	-0.39
	6	0.87	0.01	6	0.81	-0.48
	7	0.63	-0.02	7	0.83	-0.57
	8	**0.91**	**-0.10**	8	0.82	-0.41
	9	0.62	0.04	9	0.86	-0.65
	10	0.74	0.06	10	0.82	-0.32
SaAHR1	1	0.93	-0.21	**1**	**0.92**	**-0.03**
	2	0.93	-0.23	2	0.77	0.08
	3	0.76	-0.45	3	0.88	-0.17
	4	0.89	-0.10	4	0.91	-0.10
	5	0.83	-0.40	5	0.99	0.05
	6	0.90	-0.37	6	0.85	-0.12
	7	**0.96**	**-0.37**	7	0.79	-0.07
	8	0.94	-0.18	8	0.91	-0.23
	9	0.86	-0.01	9	0.73	-0.19
	10	0.94	-0.18	10	0.90	-0.17
SaAHR1 P380L	Models not created. Mutation is outside of modelled area			1	0.96	-0.34
				2	**0.99**	**-0.56**
				3	0.85	-0.39
				4	0.80	-0.50
				5	0.89	-0.45
				6	0.96	-0.53
				7	0.95	-0.46
				8	0.86	-0.36
				9	0.83	-0.33
				10	0.93	-0.48
SaAHR2	1	0.54	-0.15	**1**	**0.93**	**-0.31**
	2	0.50	-0.02	2	0.82	-0.25
	3	0.55	-0.06	3	0.58	-0.34
	4	**0.82**	**-0.02**	4	0.81	-0.29
	5	0.57	-0.12	5	0.79	-0.34
	6	0.58	0.00	6	0.91	-0.52
	7	0.54	-0.04	7	0.72	-0.48
	8	0.35	-0.13	8	0.75	-0.26
	9	0.79	-0.06	9	0.83	-0.43
	10	0.80	0.03	10	0.74	-0.45
SaAHR3	1	0.85	0.14	**1**	**0.98**	**-0.19**
	2	0.67	0.09	2	0.77	-0.42
	3	0.73	0.14	3	0.92	-0.37
	4	0.71	0.16	4	0.84	-0.41
	5	0.84	0.17	5	0.93	-0.32
	6	0.65	0.12	6	0.88	-0.44
	7	**0.84**	**-0.08**	7	0.91	-0.14
	8	0.77	0.17	8	0.95	-0.25
	9	0.74	0.07	9	0.90	-0.39
	10	0.58	0.08	10	0.86	-0.31



Appendix 2: Representative models of HIF2α-based SaAHR1 WT and SaAHR2 WT. Strand Iβ pulls away from the β-sheet earlier in the SaAHR1 WT model due to the presence of the diproline motif.



Appendix 3: Luciferase assay 5, the P380L mutation did not cause PCB 126 induced activation in dual luciferase reporter gene assay. SaAHR2 was used as a positive control and a pcDNA empty vector was used as a negative control. This experiment did not include FhARNT2 cotransfection. Bars denote mean value of triplicate wells, error bars denote 1 standard deviation. PCB 126 final concentration of 100 nM. Asterisk denotes value significantly different ($P<0.02$) from DMSO only control, Student's *t* test.



Appendix 4: Luciferase assays 1(A) and 2(B), the L342M/A359S and R310C/A359S double mutants did not cause PCB 126 induced activation in dual luciferase reporter gene assay. These results were not included in the body of this work due to low and possibly unreliable raw luminescence values. SaAHR2 was used as a positive control and a pcDNA empty vector was used as a negative control. This experiment did not include FhARNT2 cotransfection. Bars denote mean value of triplicate wells, error bars denote 1 standard deviation. PCB 126 final concentration of 100 nM.

Appendix 5: Luciferase assay luminescense raw values for assays 1 and 2

Assay 1		firefly	*Renilla*	RLU
Sa2	DMSO	457	370	1.2344
		365	247	1.476
		214	239	0.8954
	PCB 126	810	401	2.0175
		253	297	0.8524
		151	201	0.7504
Sa1 L342M + A359S	DMSO	124	218	0.5688
		120	275	0.4387
		135	204	0.661
	PCB 126	128	202	0.6333
		128	196	0.6531
		117	215	0.5426
Sa1 R310C + A359S	DMSO	118	218	0.5447
		132	197	0.6726
		135	229	0.5877
	PCB 126	188	243	0.772
		121	202	0.5971
		121	227	0.5347
Sa1 WT	DMSO	206	404	0.5121
		244	315	0.776
		147	251	0.586
	PCB 126	150	251	0.5971
		279	393	0.7102
		208	444	0.4697
pcDNA EV	DMSO	143	302	0.4742
		150	222	0.6746
		143	222	0.6454
	PCB 126	149	245	0.6077
		145	252	0.577
		135	216	0.6244

Assay 2		firefly	*Renilla*	RLU
Sa2	DMSO	158	279	0.5667
		208	683	0.3051
		7201	2261	3.1846
	PCB 126	12162	5918	2.055
		20632	5815	3.5478
		19630	4980	3.9418
Sa1 L342M + A359S	DMSO	210	350	0.6006
		187	295	0.6337
		458	1511	0.3036
	PCB 126	486	1636	0.2975
		604	1703	0.3548
		746	1756	0.425
Sa1 R310C + A359S	DMSO	235	1167	0.2018
		146	238	0.6144
		567	1798	0.3157
	PCB 126	430	2102	0.2048
		366	1784	0.2054
		606	2305	0.2633
Sa1 WT	DMSO	1196	4306	0.2778
		170	372	0.4574
		1553	3809	0.4078
	PCB 126	1771	5188	0.3414
		1494	5116	0.2921
		1504	3913	0.3845
pcDNA EV	DMSO	1245	2289	0.544
		846	1650	0.513
		1986	3934	0.5049
	PCB 126	5149	9306	0.5533
		5003	11559	0.4328
		Third replicate value lost		

Appendix 6: Luciferase assay luminescense raw values for assays 3 and 4

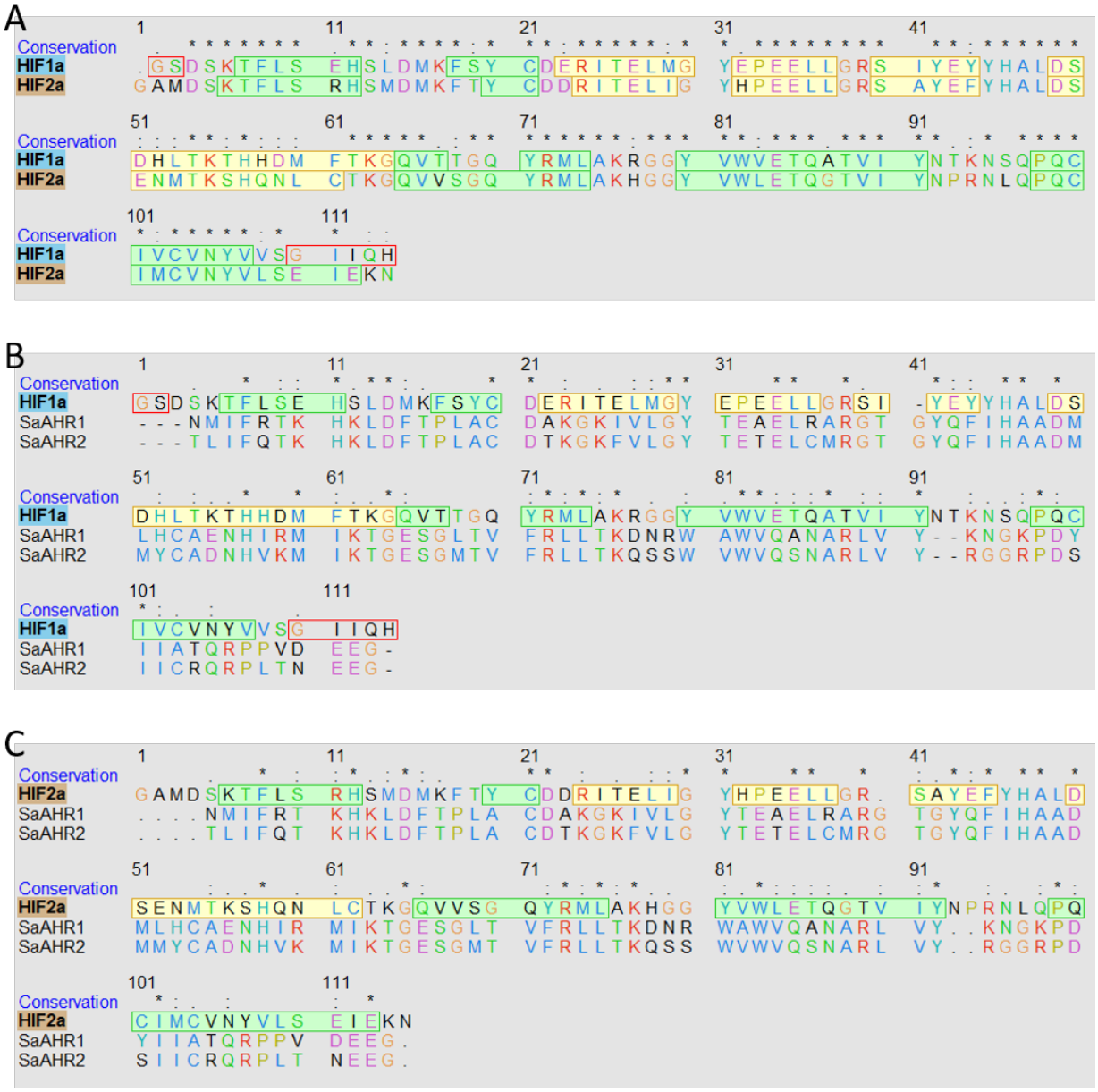
Assay 3		firefly	*Renilla*	RLU
Sa2 6 hours 15 ng DNA	DMSO	281	1044	0.2692
		4042	846	4.7752
		278	507	0.549
	PCB 126	131	278	0.4729
		508	765	0.6641
		279	742	0.3772
Sa2 6 hours 30 ng DNA	DMSO	321	539	0.5968
		420	697	0.6038
		141	474	0.298
	PCB 126	477	484	0.9864
		553	606	0.9115
		1400	1054	1.3274
Sa2 24 hours 15 ng DNA	DMSO	16335	7315	2.2331
		12957	9244	1.4017
		9675	8499	1.1384
	PCB 126	12247	7046	1.7382
		16965	8442	2.0096
		14245	5127	2.7783
Sa2 24 hours 30 ng DNA	DMSO	23989	8943	2.6824
		25186	7757	3.2467
		953	1288	0.7401
	PCB 126	6440	3084	2.0881
		15128	8673	1.7442
		28943	9165	3.1578

Assay 4		firefly	*Renilla*	RLU
Sa2 6 hours 15 ng DNA	DMSO	7400	6300	1.1746
		6205	5907	1.0505
		5224	5720	0.9133
	PCB 126	44086	4340	10.1564
		22721	3678	6.1763
		1222	1820	0.6713
Sa2 6 hours 30 ng DNA	DMSO	35049	12989	2.6983
		30433	12198	2.4948
		36829	12742	2.8903
	PCB 126	181076	15031	12.0465
		181953	15954	11.4045
		200447	13849	14.4729
Sa2 24 hours 15 ng DNA	DMSO	3561	2275	1.5653
		58316	4739	12.3055
		14684	4581	3.2052
	PCB 126	15597	2476	6.2977
		65985	5706	11.5629
		11068	1251	8.8413
Sa2 24 hours 30 ng DNA	DMSO	59953	11916	5.031
		46801	9836	4.758
		84294	7779	10.8358
	PCB 126	323273	13075	24.7239
		240893	12530	19.2245
		266409	12291	21.6751

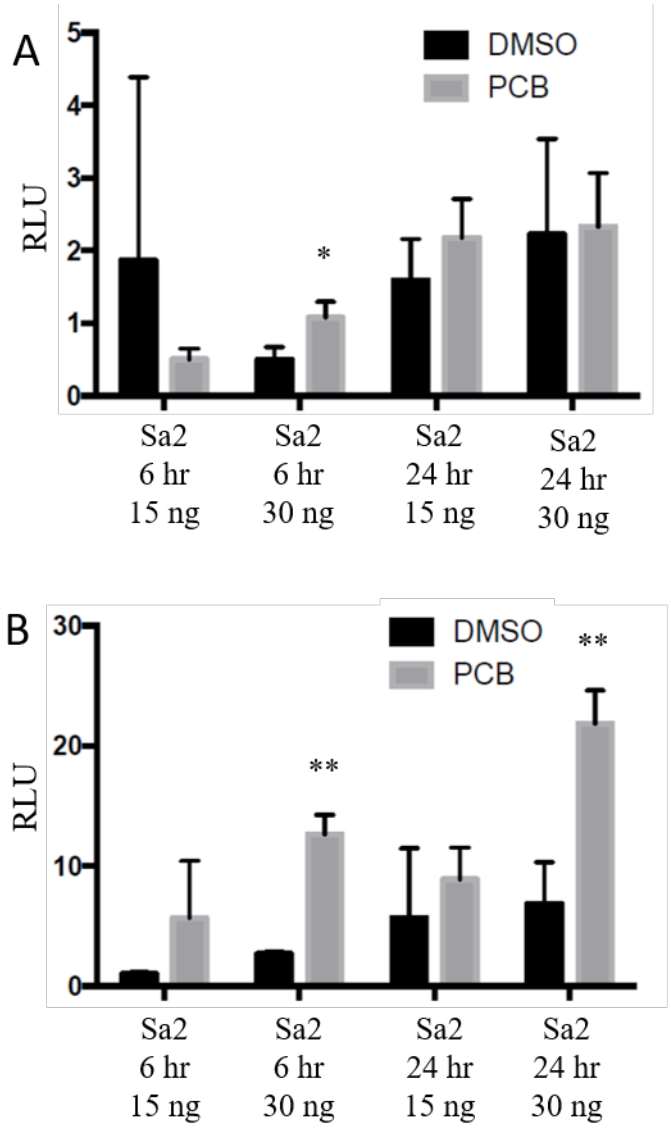
Appendix 7: Luciferase assay luminescense raw values for assays 5 and 6

Assay 5		firefly	*Renilla*	RLU
Sa2	DMSO	31977	38250	0.836
		26390	38649	0.6828
		21382	28717	0.7446
	PCB 126	14526	13600	1.0681
		50330	28828	1.7459
		53635	34505	1.5544
Sa1 P380L	DMSO	3828	22756	0.1683
		4947	24028	0.2059
		3625	22173	0.1635
	PCB 126	2926	12654	0.2312
		5476	21379	0.2562
		4902	26665	0.1839
Sa1 WT	DMSO	7691	33814	0.2275
		6533	31008	0.2107
		7947	35288	0.2252
	PCB 126	8633	21419	0.4031
		6001	28363	0.2116
		6308	29765	0.2119
pGL 4.13	DMSO	20560674	13136	1565
		35686976	23896	1493
		36266048	20220	1793
	PCB 126	5195250	6845	758.901
		34280424	20845	1644
		33775972	21542	1567
pcDNA EV	DMSO	17949	80147	0.224
		15827	61235	0.2585
		19092	70299	0.2716
	PCB 126	10055	46166	0.2178
		15803	66438	0.2379
		15958	67104	0.2378
COS7 untransfected control	DMSO	295	341	0.8672
		772	361	2.1364
		381	341	1.1188
	PCB 126	203	343	0.593
		207	351	0.5908
		189	335	0.5652

Assay 6		firefly	*Renilla*	RLU
Sa2	DMSO	88867	53236	1.6693
		115113	72995	1.577
		79749	60513	1.3179
	PCB 126	319767	81464	3.9253
		186850	54938	3.4011
		127319	42526	2.9939
Sa1 P380L	DMSO	24639	34114	0.7223
		23895	31904	0.749
		29158	39675	0.7349
	PCB 126	32087	41349	0.776
		28402	33430	0.8496
		24144	32078	0.7527
Sa1 WT	DMSO	36203	45395	0.7975
		37184	46222	0.8045
		46174	52258	0.8836
	PCB 126	46953	57553	0.8158
		39106	39387	0.9929
		30432	33496	0.9085
pGL 4.13	DMSO	153667024	58237	2638
		148657344	52085	2854
		165592592	64682	2560
	PCB 126	170774336	65619	2602
		162415440	59972	2708
		144758880	53262	2717
pcDNA EV	DMSO	120165	102773	1.1692
		81386	71168	1.1436
		111004	93026	1.1933
	PCB 126	140270	116439	1.2047
		74448	73613	1.0113
		83627	67479	1.2393
FhAHR2	DMSO	2017297	54276	37.167
		2530627	62795	40.2992
		3068355	88995	34.4776
	PCB 126	3333815	66278	50.2999
		4563634	72341	63.0842
		4105533	68834	59.6439



Appendix 8: Alignments of structural templates (HIF1α and HIF2α) vs. model sequence (SaAHR1 and SaAHR2). Screenshots taken from UCSF Chimera program. Green highlighted sequence denotes β-strand in structural template. Yellow highlighted sequence denotes α-helix in structural template. AHR sequences start at residue 278. Conservation symbols: asterisk denotes homologous residue, colon denotes highly similar residue, period denotes less similar residue. (A) HIF1α vs. HIF2α, (B) HIF1α vs. SaAHR1 and SaAHR2, (C) HIF2α vs. SaAHR1 and SaAHR2.



Appendix 9: Luciferase assays 3(A) and 4(B), optimization of SaAHR2 transfection amount (15 ng vs. 30 ng) and dosing timepoint (6 hr vs. 24 hr post transfection). This experiment did not include FhARNT2 cotransfection. Bars denote mean value of triplicate wells, error bars denote 1 standard deviation. PCB 126 final concentration of 100 nM. Asterisks denote values significantly different from DMSO only control, Student's *t* test (* =P<0.02, ** =P<0.005).