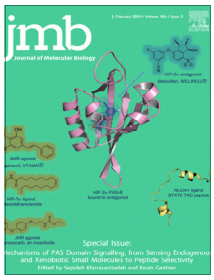


Structural Insights into the Activation of Human Aryl Hydrocarbon Receptor by the Environmental Contaminant Benzo[*a*]pyrene and Structurally Related Compounds

Hok-Sau Kwong [1,†], Matteo Paloni [1,†], Loïc Grandvuillemin [1], Savannah Sirounian [1], Aurélie Ancelin [1], Josephine Lai-Kee-Him [1], Marina Grimaldi [2], Coralie Carivenc [1], Claudia Lancey [3], Timothy J. Ragan [3], Emma L. Hesketh [3], Patrick Balaguer [2], Alessandro Barducci [1], Jakub Gruszczyk [1,*] and William Bourguet [1,*]

*1 - **CBS (Centre de Biologie Structurale),** Univ Montpellier, CNRS, Inserm, Montpellier, France*
*2 - **IRCM (Institut de Recherche en Cancérologie de Montpellier),** Univ Montpellier, Inserm, ICM, Montpellier, France*
*3 - **Leicester Institute of Structural & Chemical Biology and Department of Molecular & Cell Biology,** University of Leicester, Lancaster Rd, Leicester LE1 7HB, UK*

***Correspondence to Jakub Gruszczyk and William Bourguet:** jakub.gruszczyk@cbs.cnrs.fr (J. Gruszczyk), william.bourguet@cbs.cnrs.fr (W. Bourguet)*
https://doi.org/10.1016/j.jmb.2023.168411
Edited by Sepideh Khorasanizadeh

Abstract

The aryl hydrocarbon receptor (AHR) is a ligand-dependent transcription factor belonging to the bHLH/PAS protein family and responding to hundreds of natural and chemical substances. It is primarily involved in the defense against chemical insults and bacterial infections or in the adaptive immune response, but also in the development of pathological conditions ranging from inflammatory to neoplastic disorders. Despite its prominent roles in many (patho)physiological processes, the lack of high-resolution structural data has precluded for thirty years an in-depth understanding of the structural mechanisms underlying ligand-binding specificity, promiscuity and activation of AHR. We recently reported a cryogenic electron microscopy (cryo-EM) structure of human AHR bound to the natural ligand indirubin, the chaperone Hsp90 and the co-chaperone XAP2 that provided the first experimental visualization of its ligand-binding PAS-B domain. Here, we report a 2.75 Å resolution structure of the AHR complex bound to the environmental pollutant benzo[*a*]pyrene (B[*a*]P). The structure substantiates the existence of a bipartite PAS-B ligand-binding pocket with a geometrically constrained primary binding site controlling ligand binding specificity and affinity, and a secondary binding site contributing to the binding promiscuity of AHR. We also report a docking study of B[*a*]P congeners that validates the B[*a*]P-bound PAS-B structure as a suitable model for accurate computational ligand binding assessment. Finally, comparison of our agonist-bound complex with the recently reported structures of mouse and fruit fly AHR PAS-B in different activation states suggests a ligand-induced loop conformational change potentially involved in the regulation of AHR function.

Introduction

Humans are chronically exposed to a multitude of natural and chemical substances in various ways, including through the diet, medical therapies or direct physical contact with the immediate environment. By entering our body these compounds may affect either positively or

negatively different physiological functions, and some of them can trigger long lasting harmful effects. Living organisms have developed protein sensors allowing them to adapt to their chemical environment such as the xenobiotic receptors CAR (constitutive androstane receptor), PXR (pregnane X receptor) or AHR (aryl hydrocarbon receptor).[1] Also referred to as an "exposome receptor",[2] AHR is a crucial chemosensory protein involved in a wide range of physiological and pathological processes in response to a myriad of natural and chemical substances.[3] It belongs to the basic helix-loop-helix (bHLH)/PER-ARNT-SIM (PAS) protein family which is present ubiquitously in all kingdoms of life and predominantly involved in sensing various stimuli including endogenous compounds, foreign chemicals, light, oxygen or osmotic pressure, for example.[4–7] The AHR protein consists of around 850 amino acids and can be divided into several different domains fulfilling very specific functions. The N-terminal part of the protein consists of three structured domains. The bHLH domain is involved in DNA recognition and dimerization. The PAS region exists as a tandem of two domains, denoted PAS-A and PAS-B, respectively, each encompassing around fifty amino acids. PAS-A domain is involved in heterodimerization with other protein partners and PAS-B domain is crucial for protein activity as it is directly responsible for ligand binding. The C-terminal part of the protein is predicted to be mostly disordered and includes a transactivation domain that recruits transcriptional coactivators to the target gene promoter leading to induction of gene transcription.

AHR is historically known as the receptor for numerous xenobiotics such as halogenated aromatic hydrocarbons (HAHs, *e.g.* 2,3,7,8-tetra chlorodibenzo-p-dioxin also known as dioxin or TCDD), polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs) and other environmental contaminants collectively referred to as persistent organic pollutants (POPs). Hence, initial studies have revealed the prominent role of AHR in xenobiotic metabolism and clearance through the regulation of the expression of xenobiotic-metabolizing phase I (*e.g.* cytochrome P450 CYP1A), and phase II (*e.g.* glutathione S-transferase GST) enzymes following its interaction with chemicals. However, in addition to promoting their elimination, AHR is also known to act as a mediator of the toxicity of certain POPs through the generation of toxic metabolites, for example.[8] A sustained activation of the receptor by metabolically resistant AHR ligands (*e.g.* TCDD), in contrast with more rapidly metabolized compounds (*e.g.* PAHs), has also been suggested as a possible mechanism of AHR-mediated POPs toxicity.[9] Moreover, it was later discovered that AHR can also play a critical role in various physiological pathways, including cell growth, differentiation, immune response and barrier function, through its binding to microbiome metabolites, natural products and endogenous molecules (*e.g.* indole derivatives, alkaloids or polyphenols), and the consecutive induction of diverse sets of genes.[2,10,11] For instance, AHR activation by microbial metabolites in the gastrointestinal tract plays an important role in signaling between the enteric microflora and the immune system. Because of this, the receptor is engaged in promoting commensalism within the human gut and dysregulation of this interaction can lead to irritable bowel disease.[12–15]

In its unliganded form, AHR resides in the cytoplasm forming a complex with several other partners including the chaperone heat shock protein 90 (Hsp90) and the co-chaperones X-associated protein 2 (XAP2, also known as AIP or ARA9) and p23. Ligand binding triggers a translocation of the complex into the nucleus where AHR is released and interacts with the AHR nuclear translocator (ARNT). The AHR-ARNT heterodimer binds to the so-called "dioxin-response element" (DRE, also known as xenobiotic-response element XRE or aryl hydrocarbon-response element AHRE) DNA sequences and regulates the expression of target genes.[13,16,17] Previous structural work has revealed how AHR and ARNT bHLH/PAS-A domains heterodimerize and bind to DREs,[18–20] but provided no insights into the AHR cytosolic complex organization or the structural basis of ligand-binding and receptor activation. Indeed, full-length AHR has resisted structural studies for thirty years, and this may be due, in part, to the unstable and aggregation-prone nature of the protein, and more specifically of its PAS-B domain. Using the Sf9 insect cell expression system, protein co-expression and cryogenic electron microscopy (cryo-EM), we recently reported the structure of the AHR-Hsp90-XAP2 cytosolic complex bound to the microbial metabolite indirubin.[21] The structure showed the concerted action of Hsp90 and XAP2 in maintaining the receptor in a stable and functional form, and revealed the molecular determinants of indirubin recognition and binding. Following this initial study involving a natural AHR ligand, we now report a 2.75 Å resolution structure of the AHR-Hsp90-XAP2 cytosolic complex bound to benzo[*a*]pyrene (B[*a*]P), a prototypical member of the PAH family of environmental contaminants. B[*a*]P is a by-product formed in various combustion processes, including burning of fossil fuels, wood, and other organic materials. It is present in high levels in cigarette smoke, diesel exhaust, smoked and grilled food products, as well as certain types of industrial waste. Binding of B[*a*]P to AHR results in several toxic effects, notably through its biotransformation by AHR-induced cytochrome P450s into the carcinogenic metabolite B[*a*]P-7,8-diol-9,10-epoxide that binds covalently to DNA, thus leading to erroneous replication and mutagenesis.[8] The structure reported here reveals the binding mode

of B[*a*]P at the PAS-B that allowed a detailed structure–function analysis of the ligand-binding pocket of AHR. The B[*a*]P-bound PAS-B structure was subsequently used as a template to accurately model and evaluate the interaction of several PAH family members with the receptor. Last, comparison of our structural data with the recently published structures of AHR from various species and in different ligation states allowed us to provide the structural basis of ligand binding species specificities and to suggest potential ligand-induced regulatory structural changes.

Results and Discussion

Structure determination and architecture of the B[a]P-bound AHR-Hsp90-XAP2 complex

Human AHR lacking the disordered transactivation domain (residues 1–437), full length Hsp90, XAP2 and p23 were co-expressed using the Sf9 insect cell expression system. The human AHR-Hsp90-XAP2 cytosolic complex was prepared as previously described.[21] Briefly, following cell lysis and centrifugation, the protein extract was applied onto two consecutive HisTrap and StrepTrap affinity columns (Figure 1A, B). After tag removal, the complex was further purified using an ion exchange column and fractions containing the ternary AHR-Hsp90-XAP2 or the quaternary AHR-Hsp90-XAP2-p23 complexes were pooled separately. Our past experience on the structural characterization of the AHR cytosolic complex using cryo-EM indicated that the presence of p23 generated a high heterogeneity of the sample due to the presence of different numbers and locations of the molecule, thereby lowering the resolution of the final map.[21] Moreover, it has been recently suggested that for a similar complex between Hsp90 and the glucocorticoid receptor, the binding of p23 and immunophilin (*e.g.* XAP2) are mutually exclusive.[22] Hence, the ternary AHR-Hsp90-XAP2 complex was concentrated and injected in a size exclusion column for a final purification step and subsequent structural studies (see Material and Methods for details). B[*a*]P was added to the purified complex prior to preparation of the cryo-EM grids. B[*a*]P is a highly hydrophobic molecule consisting of five fused benzene rings (one benzene ring fused to the four-member-ring pyrene), resulting in a planar aromatic system (Figure 2A). Although less potent than indirubin or dioxin (Figure 2A), B[*a*]P is a high affinity and full agonist activating AHR with an EC_{50} of 26 nM (Figure 2B and Table S1). Using differential scanning fluorimetry (nano-DSF) that monitors the thermal stabilization of the receptor upon ligand binding, we also confirmed that B[*a*]P binds to the purified AHR-Hsp90-XAP2 complex (Figure 2C).

Cryo-EM data collected at The Midlands Regional Cryo-EM Facility in Leicester, UK, allowed us to obtain a 2.75 Å reconstruction of the AHR cytosolic complex bound to B[*a*]P (Figure 3A–D, Figure S1 and Table S2). The model was built starting from our previously published structure of the indirubin-bound AHR-Hsp90-XAP2 complex (PDB code 7ZUB). Reminiscent of this structure[21] and of the recently reported unliganded (apo) mouse AHR-Hsp90-XAP2-p23 structure,[23] the complex displays a fully closed, nucleotide-bound Hsp90 dimer (molecules denoted Hsp90A and Hsp90B) complexed with one molecule of AHR and one molecule of XAP2 that occupy the same side of the Hsp90 dimer (Figure 3E, F). The globular ligand-binding PAS-B domain of AHR is stabilized through extensive interaction surfaces with Hsp90 and XAP2, but, on the other side of the chaperone homodimer, the high dynamics of the N-terminal region prevented us from building a reliable model for the DNA-binding bHLH and dimerization PAS-A domains. In addition to the PAS-B domain, the interfaces between AHR and the other complex components involve a N-terminal fifteen-residue elongated linker connecting PAS-A and PAS-B passing through the closed Hsp90 lumen, as well as a C-terminal mostly disordered forty-residue-long loop connecting PAS-B to the transactivation domain and interacting with PAS-B and XAP2. The overall architecture of the B[*a*]P-bound cytosolic complex does not show significant deviations from the previously reported structure of the indirubin-bound AHR-Hsp90-XAP2 complex, in agreement with the similar agonistic character of the two AHR ligands.

The AHR ligand binding PAS-B domain bound to B[a]P

The B[*a*]P binding site within the AHR PAS-B ligand-binding pocket could be easily identified thanks to the characteristic shape of the cryo-EM map of the ligand. Looking at the map (Figure 4A), it appeared clearly that the bulkiest part of B[*a*]P (the four-member-ring pyrene moiety) should be positioned in the left part of the density, while the additional benzene ring would fit well in the thinner right part of the density. However, we observed that flipping the molecule by 180° along its main axis generated a second pose that fitted equally well within the density. Because the resolution of 2.75 Å at which the structure was solved did not allow discriminate the two orientations that differ only slightly, we performed multiple microsecond-long molecular dynamics (MD) simulations of the complex between AHR PAS-B (284–399) and B[*a*]P to determine which of the possible orientations or combination of them would best explain the experimental cryo-EM map (see Materials and Methods for details). The stability of the ligand in each orientation has been evaluated by monitoring the root mean square deviation (RMSD) of the B[*a*]P molecule with respect to the backbone of the protein. In simulation, the two possible orientations of B[*a*]P show comparable stability, with the most

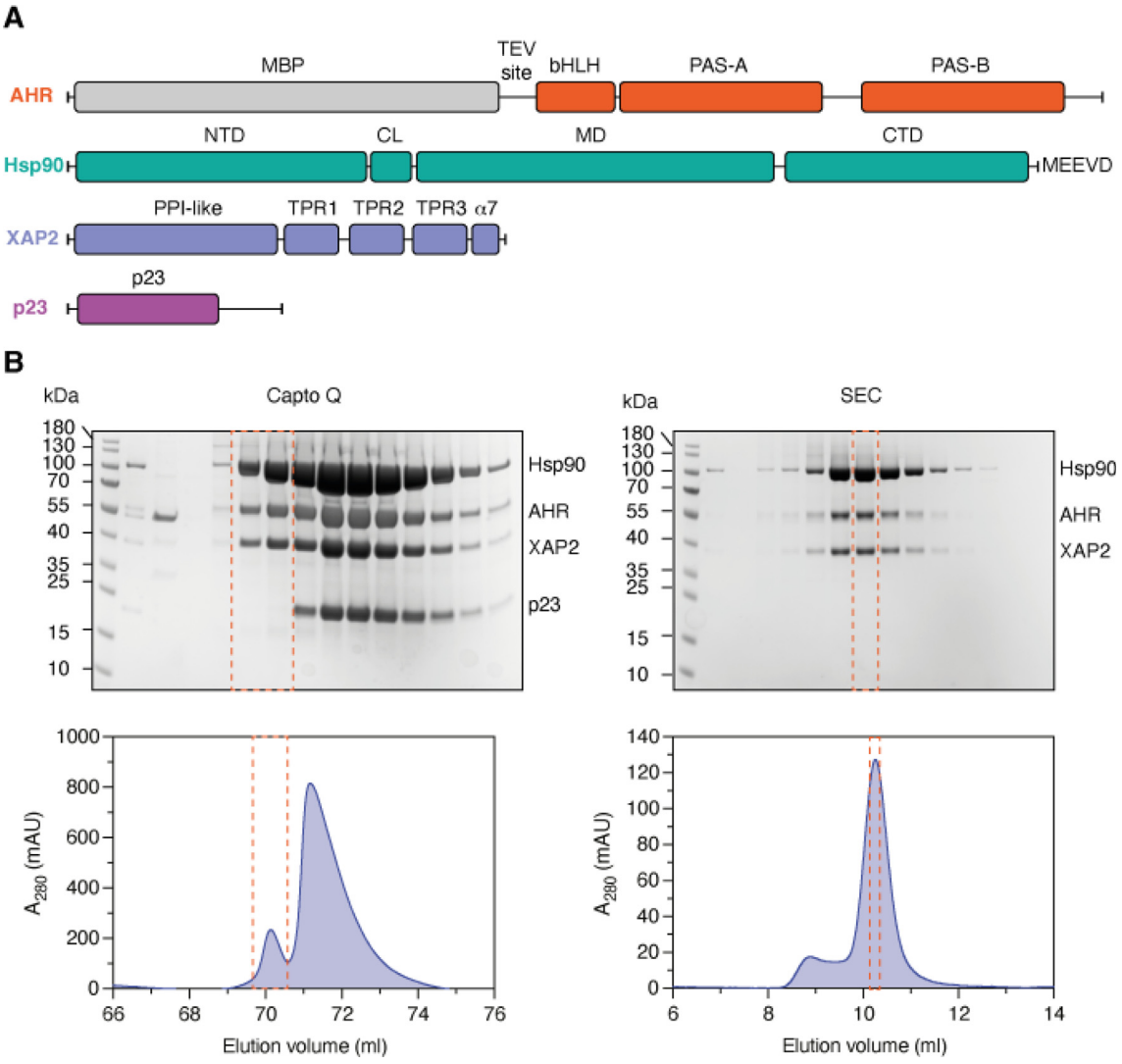


Figure 1. Preparation of the AHR-Hsp90-XAP2 complex. (A) Schematic representation of the proteins used in this study. AHR construct includes N-terminal Strep-tag, followed by MBP protein and TEV protease cleavage site; bHLH (basic helix-loop-helix motif), PAS-A and PAS-B (Per-Arnt-Sim domain A and B). Hsp90 construct includes N-terminal His-tag; NTD (N-terminal domain), CL (charged flexible linker), MD (middle domain), CTD (C-terminal domain). XAP2 and p23 were left untagged; PPI-like (peptidyl-prolyl isomerase-like domain). TPR1, 2 and 3 (tetratricopeptide repeat motif 1, 2 and 3), α7 (C-terminal helix alpha). **(B)** SDS-PAGE gels showing results obtained during the two last steps of protein purification. Fractions used for the subsequent purification step are indicated with a red dashed line. Molecular weight protein standards are indicated on the left-hand side of each gel. Gel images are representative of independent biological replicates ($n = 3$). Elution profiles from the ion exchange (Capto Q) and size-exclusion chromatography (SEC) are shown below.

populated configuration which is very close to the experimental one (Figure S2A). Nonetheless, the ligand is able to move in the pocket, exploring alternative configurations before returning to the experimental one in most of the replicates (Figure S2B). Together with the fact that the density can be more completely explained by a mixture of the two poses, we concluded from the MD analysis that B[*a*]P can bind to AHR PAS-B in two alternative, although similar, orientations (Figure 4A). Their position within the ligand binding pocket is almost identical at the working resolution of 2.75 Å, and their network of interactions with surrounding residues is almost indistinguishable, both qualitatively and quantitatively.

The elongated ligand-binding pocket extending between helices Cα and Fα is only partially occupied by the ligand that is located in close proximity to helix Fα, while the region adjacent to helix Cα remains unoccupied (Figure 4B). The B[*a*]P binding site comprises only fifteen amino acid side chains (4.2 Å distance cut-off), in line with the rather small size of the compound ($C_{20}H_{12}$,

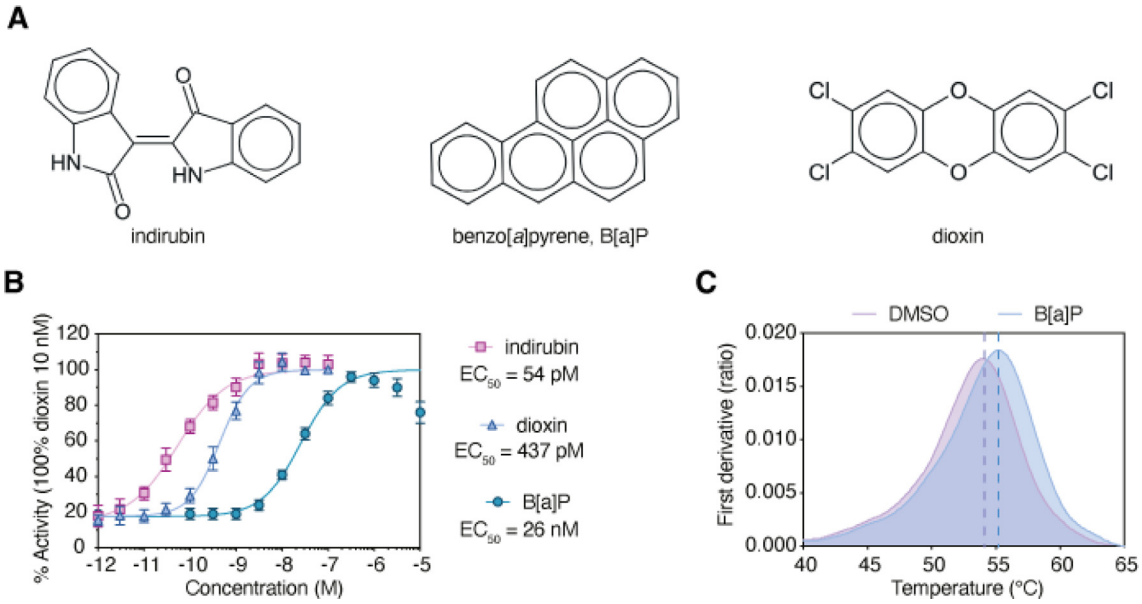


Figure 2. Activity and AHR binding properties of B[*a*]P. (A) Chemical structures of compounds used in the study. **(B)** Cell-based activity assays for three AHR ligands. The results are represented as mean value ± SD from independent biological replicates (n = 4). **(C)** Nano-DSF analysis of the interaction between the AHR-Hsp90-XAP2 complexes and B[*a*]P. The results are representative of independent biological replicates (n = 3).

252.31 Da). Next to helix Fα, the pyrene moiety is primarily involved in three stacked or T-shaped π-π interactions with F295 on one face, F351 on the other face, and Y322 at the edge (Figure 4B, C). The fifth benzene ring is sandwiched between L353 and H291, and forms a fourth π-π interaction with the latter. The remaining pocket residues (G321, F324, I325, C333, S336, H337, I349, S365, V381 and Q383) are engaged in van der Waals contacts with all parts of B[*a*]P. Interestingly, the unique spatial arrangement of six amino acids on both faces of B[*a*]P, and tentatively defined as the "floor" (H291, F295) and the "roof" (I325, C333, F351, L353) of the binding site, provides a clear explanation for the marked preference of AHR for compounds displaying a planar structure (Figure 4C). We exploited MD simulations of B[*a*]P-bound PAS-B to estimate the relative energetic contribution of the various contacting residues to the interaction with the environmental ligand. The energetic analysis confirms that both possible orientations are not significantly different, in fact they have similar interaction energies with the residues of the ligand binding pocket (Figure 4D). B[*a*]P is an apolar moiety that strongly interacts with aromatic and hydrophobic residues through van der Waals contact, such as those corresponding to the stacking arrangement with F295 (between −25 and −30 kJ/mol), for example. This is consistent with what has been estimated for the interaction between the same residue and indirubin.[21]

We next superimposed the B[*a*]P- and indirubin-bound AHR-Hsp90-XAP2 complex structures to compare the binding modes of the two ligands and their impact on the AHR PAS-B domain structure. Indirubin is a dietary-derived endogenous ligand of AHR produced from tryptophan by intestinal bacteria[24] and displaying an asymmetric double indole structure (Figure. 2A). Like B[*a*]P and most AHR ligands, indirubin is a small planar molecule, and the two compounds are of approximately the same size. As a consequence, the binding sites of the two ligands largely overlap, and residues interacting with B[*a*]P are in contact with indirubin as well (Figure 4E). Comparison of the per-residue interaction energies of indirubin (Figure 6H of 21 and B[*a*]P (Figure 4D) with AHR PAS-B shows that the two ligands share pattern and magnitude of van der Waals contacts. At variance with B[*a*]P, indirubin also forms two additional electrostatic interactions with S365 and Q383 of about 30 kJ/mol each,[21] ascribable to the formation of stable hydrogen bonds with these two amino acids that can rationalize the roughly two orders of magnitude higher affinity of indirubin compared to B[*a*]P (Figure 2B). These findings are also in qualitative agreement with the observation that a single ligand–protein hydrogen bond can increase binding affinity by about one order of magnitude.[25] To validate these computational data, S365 or Q383 were mutated into alanine, and the capacity of the corresponding AHR mutants to bind indirubin was monitored by nano-DSF. The data displayed in Figure 4F show that each mutation individually decreases the interaction of AHR with indirubin by nearly 50%, suggesting an even more drastic effect of the double mutation. As a whole, these structural, *in silico*

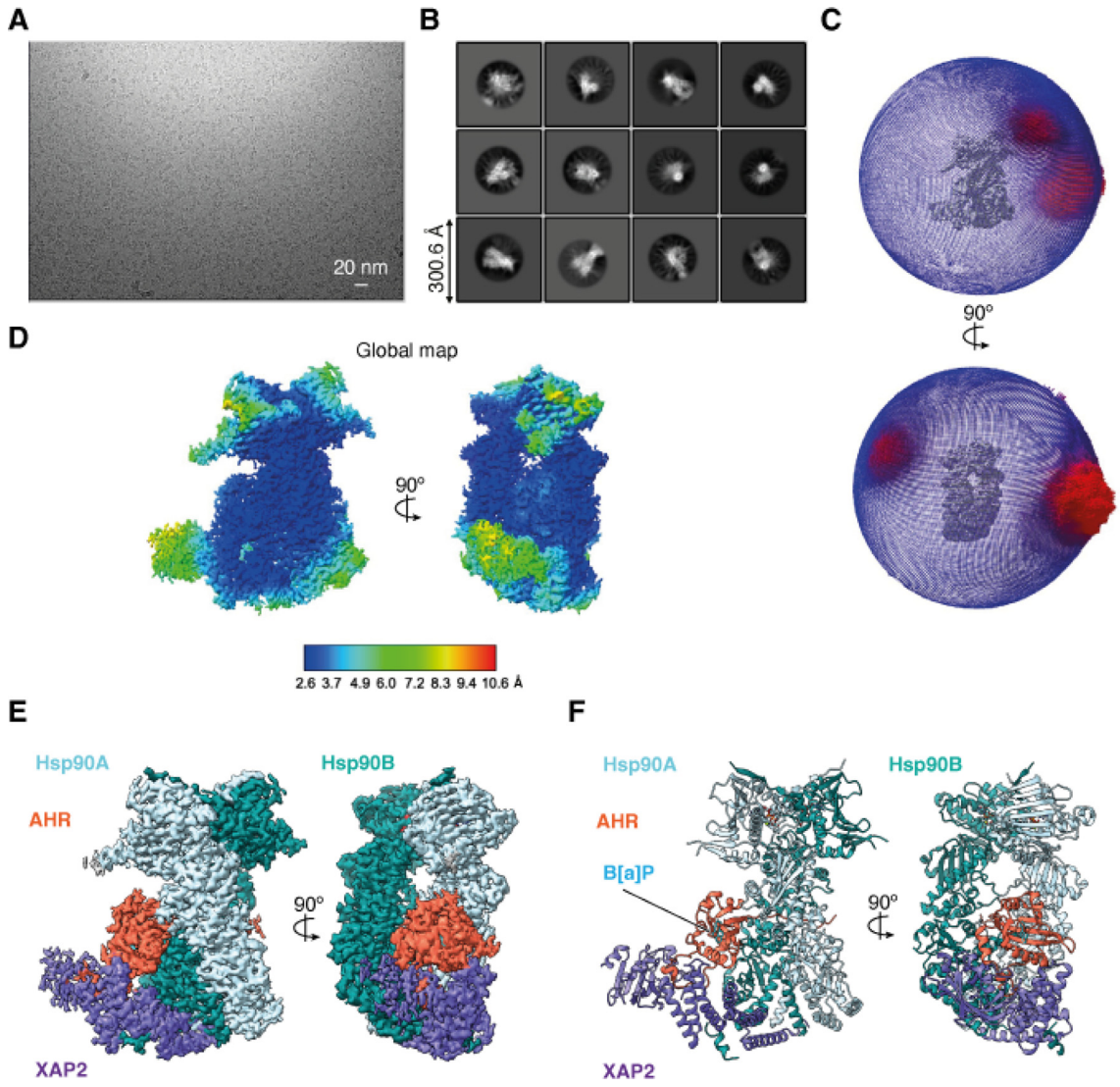


Figure 3. Cryo-EM data and overall architecture of the B[*a*]P-bound complex. (A) Representative micrograph of the sample. The scale bar represents 20 nm. **(B)** Representative 2D class averages. 2D class averages exhibit different projections corresponding to each orientation. **(C)** Euler angle distribution for the final reconstruction. The model is shown in two orthogonal views. **(D)** Local resolution estimation diagram of the final map (global map). **(E)** A composite map of the AHR-Hsp90-XAP2 complex in two orthogonal views. Hsp90A (light blue), Hsp90B (dark blue), XAP2 (purple), AHR (orange). The same color scheme is used throughout the manuscript unless stated otherwise. (**F)** The atomic model of the complex in cartoon representation (the orientation of the molecule same as in **E**.

and biochemical data highlight the importance of hydrogen bonding in the interaction between indirubin and AHR, and are consistent with the weaker binding affinity of compounds not engaged in such type of contacts (*e.g.* B[*a*]P). Note in this respect that picoberin, a compound resembling indirubin and predicted to interact with AHR through a very similar mechanism, including hydrogen bonding with S365 and Q383, was recently shown to bind to the receptor with an affinity in the picomolar range.[26] Comparison of the indirubin- and B[*a*]P-bound PAS-B structures also revealed only minor side chain reorientations between the binding sites (Figure 4E), and a well conserved conformation of the PAS-B domain with a RMSD value of 0.47 Å over 136 equivalent Cα atoms (Figure S3).

As a whole, our structural studies allowed define a bipartite ligand-binding pocket consisting of (i) a primary binding site that provides the necessary molecular environment to control ligand-binding specificity and affinity, and discriminating AHR binders based on their aromaticity and planar shape, and (ii) a secondary binding site in the region adjacent to helix Cα that is vacant in the presence of B[*a*]P or indirubin and allowing AHR to accommodate larger compounds. Such organization of the ligand-binding pocket could explain the high binding promiscuity of AHR which

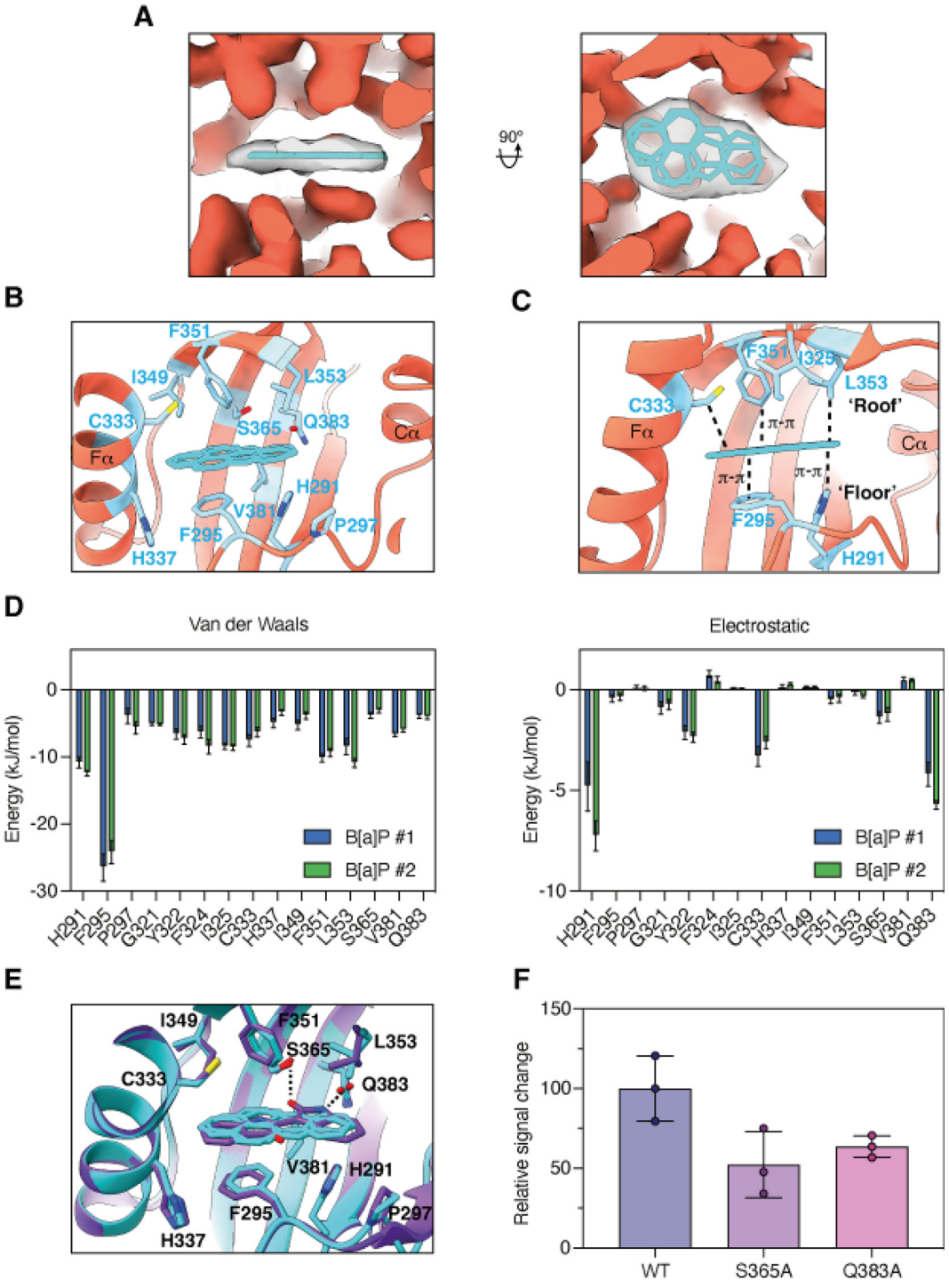


Figure 4. Interaction between B[*a*]P and the PAS-B domain of AHR. (A) Two orthogonal views of the cryo-EM map for the region corresponding to B[*a*]P. The experimental cryo-EM map and MD simulations suggest that B[*a*]P can adopt two alternative orientations (shown as cyan sticks) **(B)** Close-up view of the B[*a*]P binding site. Amino acid residues in contact with the ligand are shown as cyan sticks. **(C)** Amino acid residues forming the roof (upper layer) and the floor (lower layer) of the ligand binding site. **(D)** Van der Waals and electrostatic interaction energies between B[*a*]P and the AHR residues. Data are presented as mean ± standard deviation of MD simulation replicas ($n = 10$). **(E)** Superimposition of the B[*a*]P- (cyan) and indirubin-bound (purple) structures. Hydrogen bonds are indicated as dashed lines. (**F**) Nano-DSF analysis of the interaction between the AHR-Hsp90-XAP2 S365A and Q383A mutant proteins and indirubin. The results are representative of independent biological replicates ($n = 3$).

can be bound and activated by a vast array of compounds, ranging from small indole derivatives with a molecular weight below 200 Da to large molecules of around 600 Da, like for example bilirubin.[3]

Structure-activity analysis of B[a]P and structurally related compounds

B[*a*]P is the main representative of the PAHs family of compounds which are conjugated aromatic chemicals composed of multiple benzene rings. They are POPs of critical concern due to their widespread environmental occurrence and toxicity.[27,28] PAHs are carcinogenic and genotoxic agents strongly suspected to be involved in the development of breast, colorectal, lung or prostate cancers.[29] Our cryo-EM data revealed the binding mode of B[*a*]P at AHR PAS-B. We thus took advantage of this knowledge to assess the interaction between some B[*a*]P congeners and AHR through computational approaches. Using the HAhLH reporter cell line, we first investigated the activity of two B[*a*]P-related compounds belonging to the PAH family, the benzo[k]fluoranthene (B[k]F) and benzo[*a*]anthracene (B[*a*]A). With an EC_{50} value of around 1.5 nM, B[k]F is the most potent AHR agonist followed by B[*a*]P (EC_{50} 26 nM), and finally B[*a*]A (EC_{50} 0.8 μM) (Figure 5A and Table S3). We next carried out MD simulations of AHR PAS-B in complex with either B[k]F or B[*a*]A, superimposing them to the structures of B[*a*]P, in order to evaluate their binding modes and their relative binding energies. Starting from the experimental pose of B[*a*]P allowed us to model poses of alternate ligands that are stable in the microsecond timescale (RMSD of around 0.1 nm for B[k]F and RMSD of around 0.2 nm for B[*a*]A, Figure 5B). Overall, the interaction energies of the three ligands with AHR PAS-B is in qualitative agreement with the measured biological responses, in fact B[k]F has the lowest (most favorable) interaction energy, while B[*a*]A has the highest (least favorable) one (Figure 5C). Energetic analysis at the residue level indicates that the interaction pattern of all three compounds is similar (Figure 5D) although B[*a*]A forms weaker interactions with several pocket residues, namely H291, P297, F324, L353, and Q383, in qualitative agreement with a lower affinity. This trend is further corroborated by the geometric analysis of the interactions with aromatic sidechain with the largest energy contributions (*i.e.* H291 and F295) suggesting that stacked (F295) or T-shaped (H291) arrangements are observed for all the ligands but less pronounced in the case of B[*a*]A (Figure S4). Even though these results do not consider other factors that could influence the binding free energy of the ligands for the protein (*e.g.* solvation effects), they show that the B[*a*]P-bound PAS-B structure can be exploited as a model for the assessment of ligand binding relative stability.

Comparison of human, mouse and fruit fly AHR PAS-B structures

Structural analyses of the AHR PAS-B domains from two different species and in different ligation states have been recently published, providing us with the opportunity to carry out a comparative study with the hope to gain insights into the mechanisms of AHR activation and ligand-binding specificity. On one hand, structures of the fruit fly AHR PAS-B domain (*Dm*PAS-B) in its unliganded (apo) and α-naphthoflavone (αNF)-bound forms have been solved by X-ray crystallography,[30] while on the other hand, the structure of the unliganded mouse AHR PAS-B domain (*Mm*PAS-B) was solved by cryo-EM in the context of the AHR-Hsp90-XAP2-p23 cytosolic complex.[23] Superimposition of the PAS-B structures of the various organisms revealed that they all exhibit the canonical fold of PAS domains comprising a five-stranded antiparallel β-sheet (Aβ, Bβ, Gβ, Hβ and Iβ) flanked by four consecutive α-helices of variable lengths (Cα, Dα, Eα and Fα), plus an additional helix Jα present at the carboxy-terminus of the domain (Figure 6A). αNF-bound *Dm*PAS-B, apo *Dm*PAS-B and apo *Mm*PAS-B superimposed on their human counterpart (*Hs*PAS-B) with RMSD values of 0.91 Å, 1.37 Å and 1.99 Å, respectively. Interestingly, in contrast with mammalian AHRs, the transcriptional activity of invertebrate AHRs (*e.g.* fruit fly or *C. elegans*) is ligand-independent,[31,32] and it was previously shown that *Dm*PAS-B is unable to bind any known human AHR ligands such as indirubin or B[*a*]P, for example.[30] αNF was recently identified as a "neutral" *Dm*AHR ligand that is unable to modulate the transcriptional activity of the receptor either positively or negatively.[30] Comparison of the B[*a*]P-bound *Hs*PAS-B and αNF-bound *Dm*PAS-B structures revealed that both ligands bind to the primary binding site defined above, occupying similar space and lying in the same plane (Figure 6B). However, among the fifteen residues contributing to the B[*a*]P binding site in *Hs*PAS-B, only five are strictly conserved in *Dm*-PAS-B (Figure 6C, D). Note that two of the conserved residues belong to the "floor" of the binding site (H291, F295; human numbering), while none of the conserved residues belong to the "roof". The most notable differences are the substitutions F351/Y334 (roof), C333/V316 (roof), L353/Y336 (roof), S365/T348 and Q383/H366 (*Hs*/*Dm* numbering) that all generate steric clashes with B[*a*]P (Figure 6C). Other differences are the substitutions P297/L281, F324/L307, I325/V308 (roof), S336/A319 and V381/C364 (Figure 6D). Moreover, the volume of the ligand-binding pocket, comprising both the primary and secondary binding sites and determined by CASTp,[33] appeared to be much smaller in *Dm*PAS-B (440 $Å^3$) than in its human counterpart (682 $Å^3$), in agreement with the much higher ligand-binding capacity of the latter. The situation is very similar in *C. elegans* where the

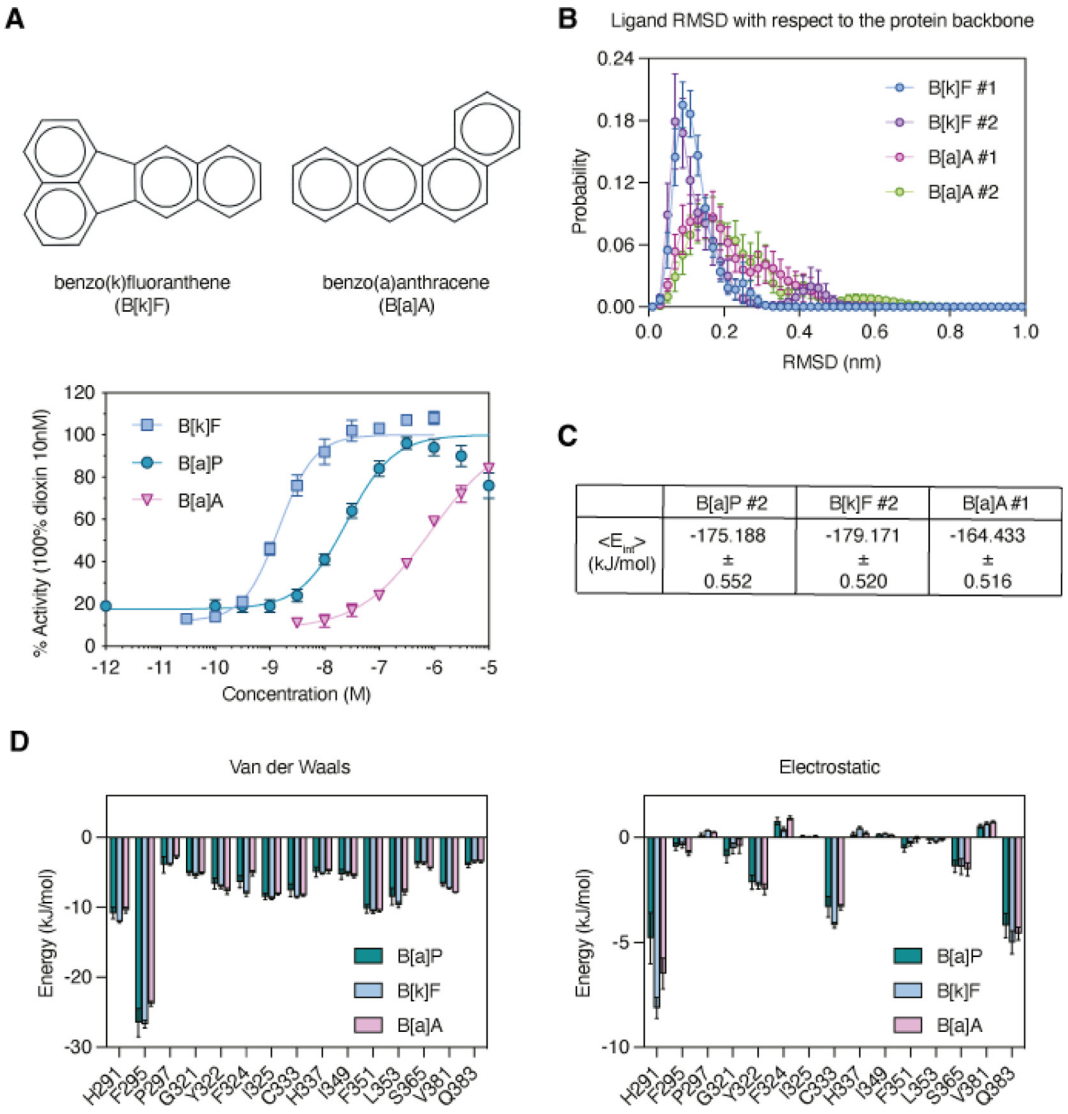
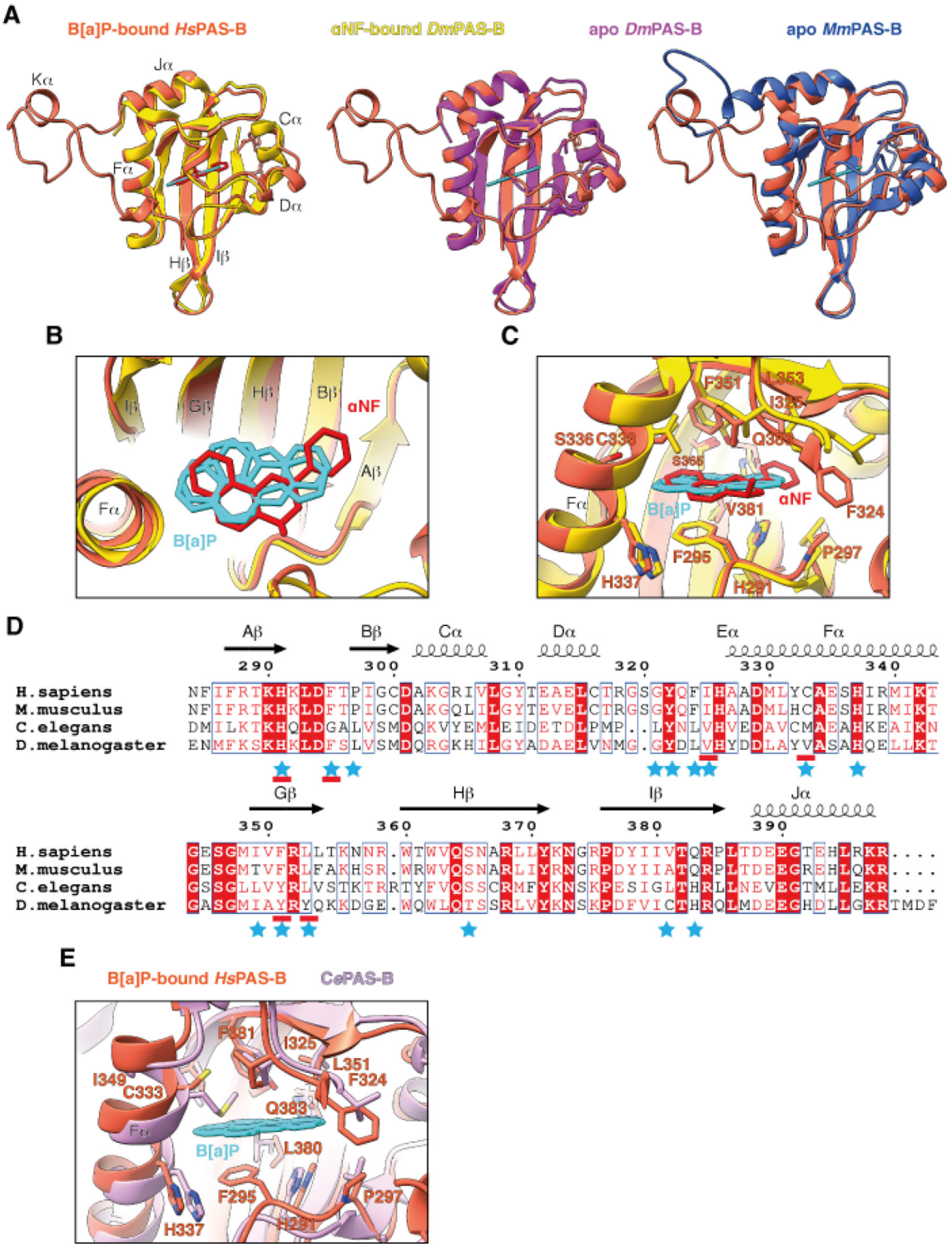


Figure 5. Modeling of the interaction between B[*a*]P congeners and AHR PAS-B. (A) Chemical structures of benzo[k]fluoranthene (B[k]F) and benzo(a)anthracene (B[*a*]A), and cell-based activity assays for B[*a*]P, B[k]F and B[*a*]A. The results are represented as mean value ± SD from independent biological replicates ($n = 4$). **(B)** Histogram of the RMSD of the positions of B[k]F and B[*a*]A with respect to the backbone of AHR PAS-B observed in MD simulations. Data are presented as mean ± error of the mean of MD simulation replicas ($n = 10$). **(C)** Total interaction energies between ligands and AHR PAS-B. Data are presented as mean ± error of the mean of MD simulation replicas ($n = 10$). **(D)** Van der Waals and electrostatic interaction energies between ligands and PAS-B individual residues. Data are presented as mean ± error of the mean of MD simulation replicas ($n = 10$).

PAS-B domain displays many sequence differences, particularly in the "floor" and "roof" regions where four out of the six human residues are not conserved in the worm (Figure 6D, E). As a whole, these data provide a rational basis to explain the different ligand-binding properties of AHR from various species and highlight the critical role of the "floor" and "roof" residues in this process.

We next analyzed various AHR PAS-B structures in the light of their respective functional states. Indeed, because *Dm*AHR is a constitutively active receptor and considering that αNF binding has no effect on its transcriptional activity, we reasoned that the *Dm*PAS-B (both apo and ligand-bound) together with the B[*a*]P-bound *Hs*PAS-B structures would represent the active form of the domain, whereas the structure of the unliganded *Mm*PAS-B could be viewed as an inactive conformation. As indicated by the RMSD values obtained from superimpositions of the four structures, all AHR PAS-B domains share a well conserved fold. However, a careful visual inspection revealed that the loop connecting helices Dα and Eα (Dα-Eα loop) follows a significantly different path in apo *Mm*PAS-B (inactive form) when compared with the B[*a*]P-bound *Hs*PAS-B structure (active form),


A
B[a]P-bound HsPAS-B
αNF-bound DmPAS-B
apo DmPAS-B
apo MmPAS-B
Kα
Jα
Cα
Fα
Dα
Hβ
Iβ
B
Iβ
Gβ
Hβ
Bβ
αNF
Aβ
Fα
B[a]P
C
F351
I353
I325
S336
C333
Q383
S385
Fα
B[a]P
V381
αNF
F324
F295
P297
H337
H291
D
Aβ
Bβ
Cα
Dα
Eα
Fα
290
300
310
320
330
340
H.sapiens NFIFRTKHKLDFTPIGCDAKGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKT
M.musculus NFIFRTKHKLDFTPIGCDAKGQLILGYTEVELCTRGSGYQFIHAADMLHCAESHIRMIKT
C.elegans DMILKTKHQLDGALVSMDQKVYEMLEIDETDLPMP..LYNLVHVEDAVCMAEAHKEAIKN
D.melanogaster ENMFKSKHKLDFSLVSMDQRGKHILGYADAELVNMG..GYDLVHYDDLAYVASAHQELLKT
Gβ
Hβ
Iβ
Jα
350
360
370
380
390
H.sapiens GESGMIVFRLLTKNNR.WTWVQSNARLIYKNGRPDYIIVTQRPLTDEEGTEHLRKR....
M.musculus GESGMTVFRLFAKHSR.WRWVQSNARLIYRNGRPDYIIATQRPLTDEEGREHLQKR....
C.elegans GSSGLLVYRLVSTKTRRTYFVQSSCRMFYKNSKPESIGITHRLLNEVEGTMLLEKR....
D.melanogaster GASGMIAYRYQKKDGE.WQWLQTSSRLVYKNSKPDFVICTHRQLMDEEGHDLLGKRTMDF
E
B[a]P-bound HsPAS-B
CePAS-B
F331
I325
L351
F324
I349
C333
Q383
Fα
B[a]P
L380
F295
P297
H337
H291

most particularly between R318 and F324 (human numbering) (Figure 7A). We noticed that three residues, namely G321, Y322 and F324, that are involved in close contacts with B[*a*]P, are displaced by several Å in the apo *Mm*PAS-B (G315, Y316, F318), suggesting that ligand-binding stabilizes a particular conformation of the loop. In line with this interpretation, a loop arrangement similar to that found in the agonist-bound *Hs*PAS-B structure was also observed in the apo and ligand-bound *Dm*PAS-B structures (active forms) with G304 (G321 in human) and Y305 (Y322 in human) interacting with the ligand αNF (Figure 7B). Our structures of the B[*a*]P- and indirubin-bound *Hs*PAS-B also revealed that the Dα-Eα loop serves as a docking surface for the C-terminal extension connecting PAS-B to the transactivation domain. This forty-residue-long sequence forms a long loop folding back onto the PAS-B domain (Figure 7C) and participates in the interface with XAP2 (Figure 3F). In contrast, the intramolecular interaction between the PAS-B Dα-Eα loop and the C-terminal linker is not visible in the unliganded *Mm*PAS-B structure, suggesting that it could be dependent on the presence of a ligand stabilizing a productive conformation of the Dα-Eα loop.

In summary, this comparative structural analysis suggests that binding of an agonist to AHR PAS-B could trigger a conformational change of the Dα-Eα loop with possible consequences on protein interactions and receptor activation. Note in addition that data obtained from a previous molecular modeling study comparing the structural dynamics of the Dα-Eα loop in HIF-2α, another member of the bHLH/PAS family, and in AHR suggested that this long and flexible "belt" could contribute to the plasticity of the ligand-binding pocket and to the ability of AHR to accommodate a wide diversity of ligands.[34] Finally, using MD simulations we previously found that the AHR PAS-B region encompassing helices Cα and Dα exhibits a high degree of flexibility in the unliganded form and undergoes stabilization upon ligand binding, thus suggesting a potential entry site for the ligand.[21] The flexibility of the Dα-Eα loop that directly follows helix Dα suggests that it could be involved in the regulation of ligand entry as well. Together, these observations suggest a ligand-related regulatory role for the Dα-Eα loop.

Concluding remarks

AHR plays many important roles in the environmental adaptation, normal physiology and development of diverse pathological conditions. To achieve these functions AHR binds and is activated by a multitude of environmental and natural substances. First identified in 1976 as a dioxin-binding protein,[35] it was later cloned in 1992[36] and found to be a member of the bHLH/PAS family of transcription factors. Although the 3D structures of many PAS domains have been reported, either as isolated domains or in the context of full-length proteins,[37] thus providing a wealth of information regarding their mechanisms of dimerization, ligand- and DNA-binding or protein–protein interactions, many attempts at solving the structure of the entire AHR or of its isolated PAS-B domain have failed for thirty years. The reasons why AHR PAS-B is more prone to aggregation than other PAS domains are still not fully understood, but co-expression of proteins of the core AHR cytosolic complex in the Sf9 insect cell expression system allowed to circumvent this problem. Using this approach, we were able to solve high-resolution cryo-EM structures of human AHR in complex with one representative natural ligand, indirubin,[21] and one prototypical exogenous ligand B[*a*]P (this work). The obtained structures show that the two types of ligands bind to the same subsite of a larger PAS-B binding pocket that is perfectly designed to accommodate small planar mostly aromatic compounds in a highly specific fashion. In agreement with the prevalently aromatic nature of AHR ligands, the binding mechanisms involve a number of π-stacking and C–H/π interactions, in addition to van der Waals and a few hydrogen bonds. The structures also reveal that the size of the ligand-binding pocket largely exceeds that of indirubin and B[*a*]P indicating how bigger compounds can bind to AHR PAS-B by exploiting the pocket extension left empty in the presence of these two small ligands

Figure 6. Comparison of the AHR ligand-binding site from various species. (A) Superimposition of human B[*a*]P-bound (PDB code 8QMO), fruit fly apo (PDB code 7VNA) and αNF-bound (PDB code 7VNH), and mouse apo (PDB code 8H77) AHR PAS-B structures showing the well conserved fold of PAS domains. **(B)** Close-up view showing that B[*a*]P (cyan) and αNF (red) bind to the same region of the PAS-B ligand-binding pocket in the human and fruit fly receptors, respectively. **(C)** Superimposition of human B[*a*]P-bound (PDB code 8QMO, orange), and fruit fly αNF-bound (PDB code 7VNH, yellow) AHR PAS-B structures showing some of the many residue differences in the ligand-binding pockets of the two receptors species. **(D)** Sequence alignment of AHR PAS-B from human (with numbering), mouse, *C. elegans* and fruit fly. Residues that are strictly conserved between all species are highlighted in red. Residues involved in the interactions with B[*a*]P are marked with a blue star. Residues defining the "floor" and the "roof" of the B[*a*]P binding site are indicated by a red bold line above or below the blue star, respectively. **(E)** Superimposition of human B[*a*]P-bound (PDB code 8QMO, orange), and *C. elegans* (AlphaFold model, violet) AHR PAS-B showing some of the many residue differences in the ligand-binding pockets of the two receptors species.

and referred to as the secondary binding site. A detailed view of the ligand binding modes at the *Hs*-PAS-B allowed us to explain interspecies differences in activation of AHR by these compounds (this work)[21] and will provide a solid framework for future docking studies, as exemplified by our work on the B[*a*]P congeners reported here. Overall, the near atomic-level information provided by our structures will likely facilitate rational design of different classes of ligands for therapeutic use and will aid the development of structure-based *in silico* methods aimed at predicting the interaction of AHR with natural, chemical or environmental compounds. Although the current study was aimed primarily at deciphering the mechanisms underlying the ligand-binding specificity and promiscuity of AHR, the important question of how interaction with an agonist triggers receptor activation was also addressed thanks to the recent report of several AHR PAS-B structures from two different organisms and in various activation states.[22,29] It is believed that ligand binding induces conformational changes in AHR PAS-B that could be transduced through Hsp90 to the N-terminal part of the receptor to expose the nuclear localization signal (NLS) to the importin complex, thereby triggering translocation of the complex to the nucleus.[13,16,17] Several lines of evidence suggest a ligand-dependent rearrangement of the Dα-Eα loop that might play a role in various processes, including ligand recognition and binding, protein–protein interaction, and possibly receptor activation. These hypotheses rely, in part, on the comparison of recently reported crystallographic or cryo-EM structures, some of which being solved at a medium (around 4 Å) resolution.[23] Further structural and functional studies will be required to substantiate the involvement of the Dα-Eα loop in the regulation of AHR function.

Materials and Methods

Protein expression and purification

Protein expression and purification were carried out for the AHR, Hsp90, and XAP2 complex,

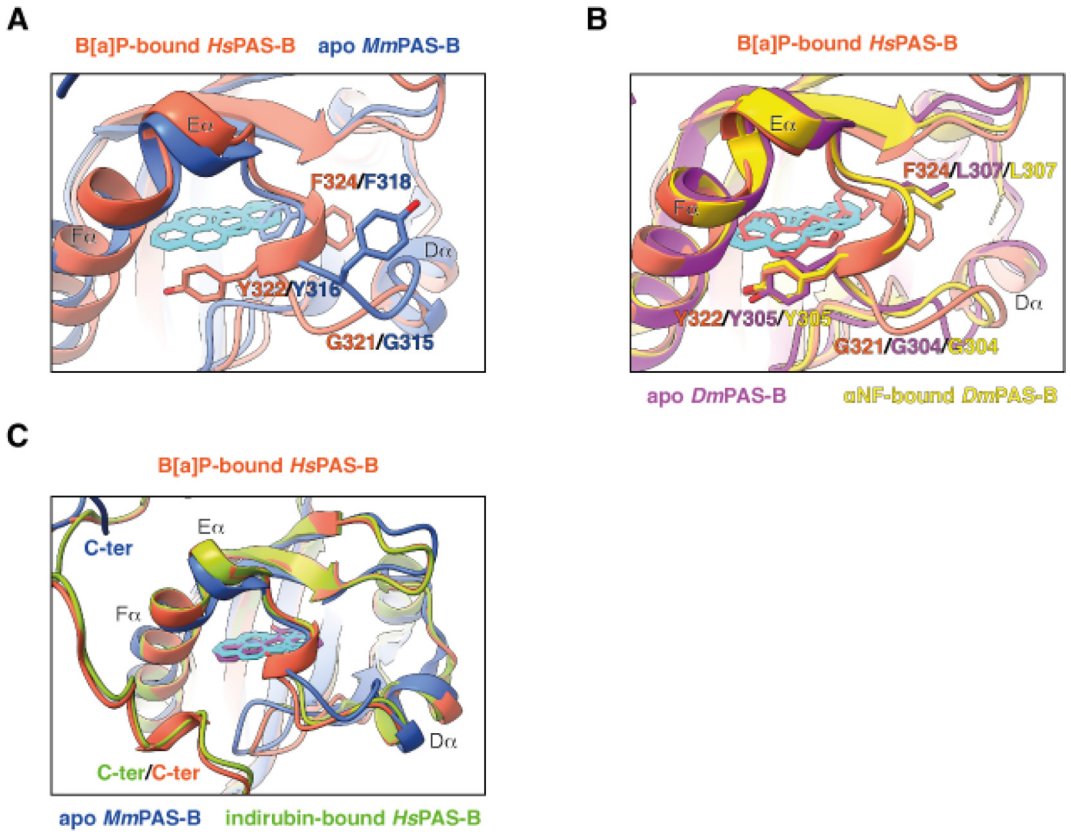


Figure 7. Comparison of Dα-Eα loops in various AHR PAS-B structures. (A) Superimposition of human B[*a*]P-bound (PDB code 8QMO) and mouse apo (PDB code 8H77) AHR PAS-B structures showing the different paths of the Dα-Eα loops in the two structures and large differences in the positioning of equivalent residues. **(B)** Superimposition of human B[*a*]P-bound (PDB code 8QMO), fruit fly apo (PDB code 7VNA)- and αNF-bound (PDB code 7VNH) AHR PAS-B structures showing a similar conformation of the Dα-Eα loops. **(C)** Superimposition of human B[*a*]P-bound (PDB code 8QMO), indirubin-bound (PDB code 7ZUB) and unliganded mouse (PDB code 8H77) AHR PAS-B structures showing the docking site of the C-terminal extension on the Dα-Eα loop in the human agonist-bound structures.

following an established protocol.[21] In brief, *Spodoptera frugiperda* Sf9 cells were co-transfected with a plasmid construct and flashBAC ULTRA viral DNA. Protein expression was induced by infecting the cell culture with the second passage stock of the virus. After 36 hours, the cells were harvested, and subjected to immediate protein purification without storage. For the purification process, the cell pellets were resuspended in a freezing buffer containing 2 mM PMSF (Sigma-Aldrich), DNase (Sigma-Aldrich), and a protease inhibitor cocktail tablet (Roche). Sonication was used for cell lysis, followed by centrifugation, filtration, and application of the supernatant to a HisTrap HP column (Cytiva). After washing the column, the protein complexes were eluted using an imidazole buffer. Subsequently, the eluted fractions were applied to a Strep-Tactin column and eluted with biotin (IBA Lifesciences). The protein was then incubated with TEV protease to ensure tag cleavage. The sample underwent filtration and was loaded onto a Capto HiRes Q column (Cytiva), separation of protein complexes was achieved using a NaCl gradient. Next, the fractions containing the desired protein complexes were pooled, concentrated, and further purified using a size exclusion column. Throughout the protein purification steps, a constant temperature of 4 °C was maintained. The purified protein fractions were either used immediately or stored at −80 °C. Protein concentration was determined using the absorption coefficient obtained from ProtParam.

Sample preparation for cryo-EM and data collection

Benzo[*a*]pyrene was purchased from Sigma-Aldrich (catalog number B1760). The compound was dissolved at 20 μM in DMSO (Sigma-Aldrich) and added to the protein solution at a 5x molar excess and 0.25% (v/v) final DMSO concentration. Before preparation of the grids, the complexes were incubated on ice for 1 h to allow sufficient time for binding. Cryo-EM data collection was performed at The Midlands Regional Cryo-EM Facility in Leicester, UK using FEI Titan Krios transmission electron microscope (Thermo Fisher Scientific) operating at 300 kV and equipped with BioQuantum K3 direct electron detector camera and GIF BioQuantum energy filter (Gatan) set to a slit width of 20 eV. EPU v.2.11 software was used for automatic data collection. Images were collected at nominal magnification of 105,000 x in counting mode with a calibrated pixel size of 0.835 Å, 2 s exposure time and an accumulative dose of 43 e-/Å^2 and 50 frames per movie. The applied defocus range varied between −0.8 and −2.3 in 0.3 μm internals. Details of the data collection parameters are presented in Supplementary Table S2.

Cryo-EM data processing

In this study, RELION3.1[38] was employed for the cryo-EM data processing. This included gain correction, drift correction, motion correction, dose weighting, and alignment performed on 9,849 micrographs using MotionCor2.[39] For each micrograph, initial contrast transfer function (CTF) values were estimated using CTFFIND4.1,[40] and low-quality images were discarded. Subsequently, reference-free particle picking was carried out in RELION using the Laplacian-of-Gaussian auto-picking mode, resulting in the selection of 6.3 million particles that were 4x binned (3.34 Å/pix). After two rounds of 2D classification with these particles, a subset of 6.2 million particles was selected. 3D classification was then applied to the chosen particles, and one class containing 1.8 million particles (30.5%) was identified for further analysis. This class was reprocessed with 2x binning (1.67 Å/pix) for another round of 3D classification, resulting in the categorization of particles into three classes totaling 807,352 particles. These particles were subsequently subjected to re-extraction without binning (0.835 Å/pix), followed by another round of 3D classification, resulting in the selection of 630,113 particles. The chosen particles were employed in a 3D refinement process, utilizing a 20 Å low-pass filtering, leading to the generation of a 3.7 Å resolution map. To ensure accurate particle alignment, an additional round of 3D classification was performed without alignment. After performing CTF refinement, Bayesian polishing with dose weighting, and a final 3D refinement round, a high-quality global map with a resolution of 2.76 Å was successfully obtained for further comprehensive analysis.

Model building and refinement

The initial model used was the AHR-Hsp90-XAP2 model, which included indirubin (PDB access code 7ZUB). The model for B[*a*]P and the constraints file were generated through Grade (https://grade.globalphasing.org). Indirubin in the ligand binding pocket of the AHR PAS-B domain was replaced with B[*a*]P. The electron density of the ligand was adequate to assign it as B[*a*]P; however, it was insufficient to distinguish between the two possible mirror orientations. The model refinement was conducted using Phenix. To calculate the model versus map Fourier Shell Correlation (FSCs), Phenix was employed. Additionally, Molprobity was utilized to validate the geometry of the final structure.

Thermal stability measurements (nano-DSF)

Thermal denaturation of the complexes in the presence of B[*a*]P was analysed using a Tycho NT.6 device (NanoTemper Technologies). The protein sample was diluted in the buffer containing 20 mM Bis-Tris–HCl pH 7.0, 50 mM NaCl, 10 mM

KCl, 10 mM $MgCl_2$, 20 mM Na_2MoO_4 and 2 mM BME to a final protein concentration of 2.5 μM. B[*a*]P solution at 20 mM was prepared in DMSO and added to the sample to a concentration of 12.5 μM. For the experiments with AHR S365A and Q383A mutant proteins, the samples were prepared the same way except the final concentration of indirubin was set at 2.5 μM. DMSO alone was used as a reference at a concentration of 5% (v/v). Upon addition of the ligand, samples were incubated at room temperature for 10 min. Denaturation curves were recorded in a range from 35 to 95 °C with 1 °C/min steps and plotted as a first derivative of the fluorescence ratio at 330 nm/350 nm against temperature. All measurements were performed in triplicates.

Reporter cell line assay

HAhLH reporter cells were seeded in 96-well white opaque tissue culture plates at a density of 50,000 cells per well in 150 μl of the test medium. Test compounds were prepared at 4× concentration in the same medium, and 50 μl per well were added 24 h after seeding. Cells were incubated for 8 h in the presence of the compounds at 37 °C. At the end of incubation, the medium containing test compounds was removed and replaced with test culture medium containing 0.3 mM luciferin and luminescence was measured in intact living cells for 2 s. Experiments were performed in quadruplicate for each ligand concentration. Each compound was tested in at least four independent experiments. Results were expressed as a percentage of maximal luciferase activity. Maximal luciferase activity (100%) was obtained in the presence of 10 nM dioxin. Dose response curves were modeled using GraphPad Software (version 5.0 Inc., San Diego, CA). Effective concentrations (ECs) are derived from the Hill equation. For a given compound, EC_{50} is defined as the concentration inducing 50% of its maximal effect. The EC_{50} values were calculated including the adjustment for the basal activity of the cell line.

Molecular dynamics simulations

MD simulations were performed with GROMACS 2022.[41] As we have recently done for the modeling of the complex between PAS-B and indirubin,[21] we have modeled the protein with the ff19sb force field,[42] the water molecules with the OPC force field,[43] while GAFF2[44] was used to model the ligands (see Tables S4–S6 for ligand parameters). Parameters for protein and water molecules have been assigned with the tleap utility in AmberTools20[45] in AMBER format and then converted to GROMACS format with the amb2gro_top_gro.py script included in AmberTools20. Parameters for the ligand have been estimated with the tools in AmberTools20 and then converted to GROMACS format with the parmed library included in AmberTools20. Charges for the ligands have been estimated at the AM1-BCC level. In MD simulations of the full complex, five replicates of 100 ns have been performed for each configuration of B[*a*]P following the same protocol that has been used in the MD simulations of the full complex with indirubin.[21]

MD simulations of complexes of PAS-B (284–399) with the ligands (B[*a*]P, B[*a*]A, and B[k]F) were prepared with the following protocol. The complex was solvated in a rhombic dodecahedron box of around 330 nm^3 and the net charge was neutralized with 0.1 M of NaCl. We performed an energy minimization with the steepest descent algorithm, followed by 100 ps equilibration in the NVT ensemble (T = 300 K) and 10 ns of equilibration in the NPT ensemble (P = 1 atm) restraining the positions of the heavy atoms of protein and ligand molecules. The velocity-rescale algorithm[46] was used to control temperature, while the Parrinello-Rahman algorithm[47] was used to control pressure. All bonds involving hydrogen atoms were constrained with the LINCS[48] algorithm which allowed for a time step of 2 fs for the integration of the equations of motion. A cut-off of 1 nm was used for Lennard-Jones interactions, while electrostatic interactions were computed with the particle mesh Ewald[49] algorithm with a direct space cut-off of 1 nm.

We initially performed MD simulations of B[*a*]P-bound PAS-B complexes without any restraints on the positions of the atoms, but we noticed that the ligand was leaving the experimental pose, with rearrangements of some of the side chains of the protein interacting with the ligand. These motions are not seen in the simulations of the full complex, indicating that the presence of the other molecules in the full complex stabilizes the configurations of PAS-B and of B[*a*]P observed in the experimental structure. In order to reproduce the presence of the full complex, we applied a soft position restraint (10 kJ/mol/nm2) on the heavy atoms of PAS-B that have an average RMSF < 0.1 nm in the simulations of the full complex in order to have fluctuations comparable to those observed in the full complex (Figure S5). We then performed 10 one-microsecond long simulations of each possible orientation, saving 1 configuration every 10 ps for further analysis.

The initial configurations of B[*a*]A and B[k]F in complex with PAS-B have been estimated by superimposing these ligands to B[*a*]P in both of the possible configurations with the LS-align tool[50] available online as a web-server. In order to let the ligand adjust in the structure of PAS-B, before the production run two additional steps of equilibration have been added after the NPT stage with position restraints on all heavy atoms of protein and ligand. We ran 1 ns equilibration with a position restraint on the heavy atoms of the protein and no restraint on the position of the ligand, followed by 1 ns equi-

libration with a position restraint of the heavy atoms of the main chain of the protein and no restraints on protein side chain atoms nor on ligand atoms. We then performed 10 1-microsecond long simulations for each ligand applying the soft position restraint (10 kJ/mol/nm^2) on the same heavy atoms of PAS-B as done for the complex with B[*a*]P in order to mimic the presence of the full complex. We saved configurations every 10 ps for further analysis. Comparison between the ligands has been performed selecting the pose with the lowest average protein–ligand interaction energy.

Heavy atoms of PAS-B on which we applied a soft position restraint were defined after computing the root mean-square fluctuations (RMSF) of heavy atoms using the gmx rmsf tool in GROMACS 2022. The motion of the ligands has been quantified by computing the root mean-square deviation (RMSD) with respect to the backbone atoms of PAS-B by means of the gmx rms tool available in GROMACS 2022. Van der Waals and Coulomb interaction energies between the ligands and the residues of the ligand pocket of PAS-B have been estimated with the gmx energy tool in GROMACS 2022. Relative orientations and minimum distances between the planar ligands and the rings of the aromatic amino acids of the ligand pocket were evaluated by means of gmx gangle and gmx mindist tools in GROMACS 2022, respectively. Distance-angle probability distributions have been normalized by a factor $D^2\sin(\theta)$[51] in order to obtain a flat distribution in case of neither distance nor angle preference.

Accession numbers

Atomic coordinates for the B[*a*]P-bound AHR-Hsp90-XAP2 complex have been deposited in the Protein Data Bank (PDB code 8QMO). The cryo-EM map obtained in this study has been deposited in the Electron Microscopy Data Bank (EMDB code EMD-18498).

Funding

This work received support from the French government, managed by the Agence Nationale de la Recherche (ANR) as part of the future investment programme under the reference “ANR-21-EXES-0005”, from the Occitanie Region and from the ExposUM Institute at the University of Montpellier. The work was also supported by funding from ATIP-Avenir 2020 grant No. R20059SP and the ANR (the French National Research Agency) “Investissements d’avenir” programme reference No. ANR-16-IDEX-0006. The Centre de Biologie Structurale is a member of the French Infrastructure for Integrated Structural Biology (FRISBI), a national infrastructure supported by the French National Research Agency (ANR-10-INBS-04-01). This work was supported by the European Union’s Horizon 2020 and 2021–2027 research and innovation programs under grant agreements No. 825489 (GOLIATH) and No. 101057014 (PARC).

CRediT authorship contribution statement

Hok-Sau Kwong: Writing – review & editing, Resources, Methodology, Investigation. **Matteo Paloni:** Writing – review & editing, Methodology, Investigation. **Loïc Grandvuillemin:** Resources, Methodology. **Savannah Sirounian:** Resources, Methodology. **Aurélie Ancelin:** Methodology. **Josephine Lai-Kee-Him:** Methodology. **Marina Grimaldi:** Methodology. **Coralie Carivenc:** . **Claudia Lancey:** Methodology. **Timothy J. Ragan:** Methodology. **Emma L. Hesketh:** Methodology. **Patrick Balaguer:** Supervision, Methodology. **Alessandro Barducci:** Supervision, Methodology. **Jakub Gruszczyk:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **William Bourguet:** Writing – original draft, Validation, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We acknowledge the Midlands Regional Cryo-EM Facility at the Leicester Institute of Structural and Chemical Biology (LISCB) and major funding from MRC (MC_PC_17136).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2023.168411.

Received 27 September 2023;
Accepted 18 December 2023;
Available online 20 December 2023

Keywords:
AHR;
cryo-EM;
Polycyclic aromatic hydrocarbon;
PAH;
B[*a*]P

† Equal contribution.

Abbreviations:
AHR, aryl hydrocarbon receptor; XAP2, X-associated protein 2; Hsp90, heat shock protein 90; cryo-EM, cryogenic electron microscopy; bHLH, basic helix-loop-helix; PAS, PER-ARNT-SIM; HAH, halogenated aromatic hydrocarbon; PAH, polycyclic aromatic hydrocarbon; PCB, polychlorinated biphenyl; POP, persistent organic pollutant; ARNT, AHR nuclear translocator; B[*a*]P, benzo [*a*]pyrene; MD, molecular dynamics

References

1. Pascussi, J.-M., Gerbal-Chaloin, S., Duret, C., Daujat-Chavanieu, M., Vilarem, M.-J., Maurel, P., (2008). The tangle of nuclear receptors that controls xenobiotic metabolism and transport: crosstalk and consequences. *Annu. Rev. Pharmacol. Toxicol.* **48**, 1–32. https://doi.org/10.1146/annurev.pharmtox.47.120505.105349.
2. Larigot, L., Benoit, L., Koual, M., Tomkiewicz, C., Barouki, R., Coumoul, X., (2022). Aryl Hydrocarbon Receptor and Its Diverse Ligands and Functions: An Exposome Receptor. *Annu. Rev. Pharmacol. Toxicol.* **62**, 383–404. https://doi.org/10.1146/annurev-pharmtox-052220-115707.
3. Denison, M.S., Nagy, S.R., (2003). Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.* **43**, 309–334. https://doi.org/10.1146/annurev.pharmtox.43.100901.135828.
4. Gu, Y.Z., Hogenesch, J.B., Bradfield, C.A., (2000). The PAS superfamily: sensors of environmental and developmental signals. *Annu. Rev. Pharmacol. Toxicol.* **40**, 519–561. https://doi.org/10.1146/annurev.pharmtox.40.1.519.
5. Vazquez-Rivera, E., Rojas, B., Parrott, J.C., Shen, A.L., Xing, Y., Carney, P.R., Bradfield, C.A., (2022). The aryl hydrocarbon receptor as a model PAS sensor. *Toxicol. Rep.* **9**, 1–11. https://doi.org/10.1016/j.toxrep.2021.11.017.
6. Nebert, D.W., (2017). Aryl hydrocarbon receptor (AHR): "pioneer member" of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of "sensors" of foreign and endogenous signals. *Prog. Lipid Res.* **67**, 38–57. https://doi.org/10.1016/j.plipres.2017.06.001.
7. McIntosh, B.E., Hogenesch, J.B., Bradfield, C.A., (2010). Mammalian Per-Arnt-Sim proteins in environmental adaptation. *Annu. Rev. Physiol.* **72**, 625–645. https://doi.org/10.1146/annurev-physiol-021909-135922.
8. Bukowska, B., Mokra, K., Michałowicz, J., (2022). Benzo[*a*]pyrene-environmental occurrence, human exposure, and mechanisms of toxicity. *Int. J. Mol. Sci.* **23**, 6348. https://doi.org/10.3390/ijms23116348.
9. Okey, A.B., (2007). An aryl hydrocarbon receptor odyssey to the shores of toxicology: the Deichmann Lecture, International Congress of Toxicology-XI. *Toxicol. Sci. Off. J. Soc. Toxicol.* **98**, 5–38. https://doi.org/10.1093/toxsci/kfm096.
10. Esser, C., Lawrence, B.P., Sherr, D.H., Perdew, G.H., Puga, A., Barouki, R., Coumoul, X., Receptor, O., (2018). Old Receptor, New Tricks-The Ever-Expanding Universe of Aryl Hydrocarbon Receptor Functions. Report from the 4th AHR Meeting, 29−31 August 2018 in Paris, France. *Int. J. Mol. Sci.* **19** (2018), E3603. https://doi.org/10.3390/ijms19113603.
11. Hubbard, T.D., Murray, I.A., Bisson, W.H., Lahoti, T.S., Gowda, K., Amin, S.G., Patterson, A.D., Perdew, G.H., (2015). Adaptation of the human aryl hydrocarbon receptor to sense microbiota-derived indoles. *Sci. Rep.* **5**, 12689. https://doi.org/10.1038/srep12689.
12. Esser, C., Rannug, A., (2015). The aryl hydrocarbon receptor in barrier organ physiology, immunology, and toxicology. *Pharmacol. Rev.* **67**, 259–279. https://doi.org/10.1124/pr.114.009001.
13. Rothhammer, V., Quintana, F.J., (2019). The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. *Nature Rev. Immunol.* **19**, 184–197. https://doi.org/10.1038/s41577-019-0125-8.
14. Sanmarco, L.M., Chao, C.-C., Wang, Y.-C., Kenison, J.E., Li, Z., Rone, J.M., Rejano-Gordillo, C.M., Polonio, C.M., et al., (2022). Identification of environmental factors that promote intestinal inflammation. *Nature* **611**, 801–809. https://doi.org/10.1038/s41586-022-05308-6.
15. Dong, F., Perdew, G.H., (2020). The aryl hydrocarbon receptor as a mediator of host-microbiota interplay. *Gut Microbes* **12**, 1859812. https://doi.org/10.1080/19490976.2020.1859812.
16. Beischlag, T.V., Luis Morales, J., Hollingshead, B.D., Perdew, G.H., (2008). The aryl hydrocarbon receptor complex and the control of gene expression. *Crit. Rev. Eukaryot. Gene Expr.* **18**, 207–250.
17. Denison, M.S., Soshilov, A.A., He, G., DeGroot, D.E., Zhao, B., (2011). Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci. Off. J. Soc. Toxicol.* **124**, 1–22. https://doi.org/10.1093/toxsci/kfr218.
18. Schulte, K.W., Green, E., Wilz, A., Platten, M., Daumke, O., (2017). Structural basis for aryl hydrocarbon receptor-mediated gene activation. *Struct. Lond. Engl.* **1993** (25), 1025–1033.e3. https://doi.org/10.1016/j.str.2017.05.008.
19. Seok, S.-H., Lee, W., Jiang, L., Molugu, K., Zheng, A., Li, Y., Park, S., Bradfield, C.A., Xing, Y., (2017). Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5431–5436. https://doi.org/10.1073/pnas.1617035114.
20. Wu, D., Potluri, N., Kim, Y., Rastinejad, F., (2013). Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain. *Mol. Cell. Biol.* **33**, 4346–4356. https://doi.org/10.1128/MCB.00698-13.
21. Gruszczyk, J., Grandvuillemin, L., Lai-Kee-Him, J., Paloni, M., Savva, C.G., Germain, P., Grimaldi, M., Boulahtouf, A., et al., (2022). Cryo-EM structure of the agonist-bound Hsp90-XAP2-AHR cytosolic complex. *Nature Commun.* **13**, 7010. https://doi.org/10.1038/s41467-022-34773-w.
22. Noddings, C.M., Johnson, J.L., Agard, D.A., (2023). Cryo-EM reveals how Hsp90 and FKBP immunophilins co-regulate the glucocorticoid receptor. *Nature Struct. Mol. Biol.*. https://doi.org/10.1038/s41594-023-01128-y.
23. Wen, Z., Zhang, Y., Zhang, B., Hang, Y., Xu, L., Chen, Y., Xie, Q., Zhao, Q., et al., (2023). Cryo-EM structure of the cytosolic AhR complex. *Struct. Lond. Engl.* **1993** (31), 295–308.e4. https://doi.org/10.1016/j.str.2022.12.013.

24. Hubbard, T.D., Murray, I.A., Perdew, G.H., (2015). Indole and tryptophan metabolism: endogenous and dietary routes to ah receptor activation. *Drug Metab. Dispos. Biol. Fate Chem.* **43**, 1522–1535. https://doi.org/10.1124/dmd.115.064246.
25. Williams, M.A., Ladbury, J.E., (2003). Hydrogen Bonds in Protein-Ligand Complexes. *In: Protein-Ligand Interact.* John Wiley & Sons Ltd, pp. 137–161. https://doi.org/10.1002/3527601813.ch6.
26. Flegel, J., Shaaban, S., Jia, Z.J., Schulte, B., Lian, Y., Krzyzanowski, A., Metz, M., Schneidewind, T., et al., (2022). The highly potent AhR agonist picoberin modulates Hh-dependent osteoblast differentiation. *J. Med. Chem.* **65**, 16268–16289. https://doi.org/10.1021/acs.jmedchem.2c00956.
27. Vondráček, J., Machala, M., (2021). The role of metabolism in toxicity of polycyclic aromatic hydrocarbons and their non-genotoxic modes of action. *Curr. Drug Metab.* **22**, 584–595. https://doi.org/10.2174/1389200221999201125205725.
28. Kim, K.-H., Jahan, S.A., Kabir, E., Brown, R.J.C., (2013). A review of airborne polycyclic aromatic hydrocarbons (PAHs) and their human health effects. *Environ. Int.* **60**, 71–80. https://doi.org/10.1016/j.envint.2013.07.019.
29. Jarvis, I.W.H., Dreij, K., Mattsson, Å., Jernström, B., Stenius, U., (2014). Interactions between polycyclic aromatic hydrocarbons in complex mixtures and implications for cancer risk assessment. *Toxicology* **321**, 27–39. https://doi.org/10.1016/j.tox.2014.03.012.
30. Dai, S., Qu, L., Li, J., Zhang, Y., Jiang, L., Wei, H., Guo, M., Chen, X., Chen, Y., (2022). Structural insight into the ligand binding mechanism of aryl hydrocarbon receptor. *Nature Commun.* **13**, 6234. https://doi.org/10.1038/s41467-022-33858-w.
31. Kudo, K., Takeuchi, T., Murakami, Y., Ebina, M., Kikuchi, H., (2009). Characterization of the region of the aryl hydrocarbon receptor required for ligand dependency of transactivation using chimeric receptor between Drosophila and Mus musculus. *Biochim. Biophys. Acta* **1789**, 477–486. https://doi.org/10.1016/j.bbagrm.2009.06.003.
32. Larigot, L., Bui, L.-C., de Bouvier, M., Pierre, O., Pinon, G., Fiocca, J., Ozeir, M., Tourette, C., et al., (2022). Identification of modulators of the *C. elegans* aryl hydrocarbon receptor and characterization of transcriptomic and metabolic AhR-1 profiles. *Antioxid. Basel Switz.* **11**, 1030. https://doi.org/10.3390/antiox11051030.
33. Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **46**, W363–W367. https://doi.org/10.1093/nar/gky473.
34. Xing, Y., Nukaya, M., Satyshur, K.A., Jiang, L., Stanevich, V., Korkmaz, E.N., Burdette, L., Kennedy, G.D., et al., (2012). Identification of the Ah-receptor structural determinants for ligand preferences. *Toxicol. Sci. Off. J. Soc. Toxicol.* **129**, 86–97. https://doi.org/10.1093/toxsci/kfs194.
35. Poland, A., Glover, E., Kende, A.S., (1976). Stereospecific, high affinity binding of 2,3,7,8-tetrachlorodibenzo-p-dioxin by hepatic cytosol. Evidence that the binding species is receptor for induction of aryl hydrocarbon hydroxylase. *J. Biol. Chem.* **251**, 4936–4946.
36. Burbach, K.M., Poland, A., Bradfield, C.A., (1992). Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 8185–8189. https://doi.org/10.1073/pnas.89.17.8185.
37. Wu, D., Rastinejad, F., (2017). Structural characterization of mammalian bHLH-PAS transcription factors. *Curr. Opin. Struct. Biol.* **43**, 1–9. https://doi.org/10.1016/j.sbi.2016.09.011.
38. Scheres, S.H.W., (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530. https://doi.org/10.1016/j.jsb.2012.09.006.
39. Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., Agard, D.A., (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331–332. https://doi.org/10.1038/nmeth.4193.
40. Rohou, A., Grigorieff, N., (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221. https://doi.org/10.1016/j.jsb.2015.08.008.
41. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindahl, E., (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software X* **1–2**, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
42. Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Migues, A.N., Bickel, J., Wang, Y., et al., (2020). ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552. https://doi.org/10.1021/acs.jctc.9b00591.
43. Izadi, S., Anandakrishnan, R., Onufriev, A.V., (2014). Building water models: A different approach. *J. Phys. Chem. Lett.* **5**, 3863–3871. https://doi.org/10.1021/jz501780a.
44. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., Case, D.A., (2004). Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174. https://doi.org/10.1002/jcc.20035.
45. Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., et al., (2005). The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688. https://doi.org/10.1002/jcc.20290.
46. Bussi, G., Donadio, D., Parrinello, M., (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, https://doi.org/10.1063/1.2408420 014101.
47. Parrinello, M., Rahman, A., (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190. https://doi.org/10.1063/1.328693.
48. Hess, B., (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122. https://doi.org/10.1021/ct700200b.
49. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593. https://doi.org/10.1063/1.470117.

50. Hu, J., Liu, Z., Yu, D.-J., Zhang, Y., (2018). LS-align: an atom-level, flexible ligand structural alignment algorithm for high-throughput virtual screening. *Bioinf. Oxf. Engl.* **34**, 2209–2218. https://doi.org/10.1093/bioinformatics/bty081.
51. Marsili, S., Chelli, R., Schettino, V., Procacci, P., (2008). Thermodynamics of stacking interactions in proteins. *Phys. Chem. Chem. Phys. PCCP* **10**, 2673–2685. https://doi.org/10.1039/b718519g.

Glossary

Exposome: The exposome is defined as the composite of every exposure to which an individual is subjected from conception to death. The chemical exposome is a fraction of the exposome that can be summarized as the sum of all xenobiotics that humans are exposed to throughout a lifetime