SCUOLA DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Department of Earth and Environmental Sciences

PhD program in Chemical, Geological and Environmental Sciences XXXI Cycle

Curriculum in Environmental Sciences

Computational approaches to study binding of xenobiotic molecules to receptors

Surname GIANI TAGLIABUE Name Sara

Registration number 720079

Tutor and Supervisor prof. Laura BONATI

Coordinator prof. Maria Luce FREZZOTTI

ACADEMIC YEAR 2017/2018

Computational approaches to study binding of xenobiotic molecules to receptors

CONTENTS

ACRONYMS

3MC	3-Methylcholanthrene
AF	Activation function
AhR	Aryl hydrocarbon Receptor
AhRR	Aryl hydrocarbon Receptor Repressor
ARNT	AhR Nuclear Translocator
AOP	Adverse Outcome Pathway
bHLH-PAS	basic helix-loop-helix/Per-ARNT-SIM
BaP	Benzo[a]Pyrene
BPA	Bisphenol A
BMAL1	Brain Muscle ARNT Like Protein 1
BNF	β-Naphthoflavone
CAR	Constitutive Androstane Receptor
CASTp	Computed Atlas of Surface Topography of proteins
CYP	Cytochrome P450
CLOCK	Circadian Locomoter Output Cycles protein Kaput
CV	Collective Variable
DBA	Dibenz(a,h)Anthracene
DBD	DNA Binding Domain
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DOPE	Discrete Optimized Protein Energy
EDC	Endocrine-Disrupting Chemicals
FDA	Food and Drug Administration
FICZ	6-Formylindolo(3,2-b)carbazole
HAH	Halogenated Aromatic Hydrocarbons
HIF	Hypoxia-Inducible Factor
HSP90	Heat Shock Protein 90
IARC	International Agency for Research on Cancer
IR	Indirubin
IUPAC	International Union of Pure and Applied Chemistry
Laq	Laquinimod
LBD	Ligand Binding Domain
LEFL	Leflunomide
KLF6	Krüppel-like Factor 6
MD	Molecular Dynamics

MM	Molecular Mechanics
MM-GBSA	Molecular Mechanics/generalized Born surface area
MM-PBSA	Molecular Mechanics/Poisson–Boltzmann surface area
NQ	Naphthoquinone
NF-κB	Nuclear factor-κB
NMR	Nuclear Magnetic Resonance
NPAS	Neuronal PAS domain protein
NR	Nuclear Receptor
PAH	Polycyclic Aromatic Hydrocarbon
PAS	PER-ARNT-SIM
PCB	Polychlorobiphenyl
PCB126	3,3′,4,4′,5-Pentachlorobiphenyl
PCDD	Polychlorinated dibenzodioxins
PCDF	Polychlorinated dibenzofurans
PDB	Protein Data Bank
PXR	Pregnane X Receptor
PXRE	PXR Response Element
QM	Quantum Mechanical
QSAR	Quantitative Structure-Activity Relationship
REACH	Registration, Evaluation, Authorization and Restriction of Chemicals
RMSD	Root Mean Square Deviation
RXR	Retinoid X Receptor
SAhRM	selective AhR modulator
SMD	Scaled MD
SIM	Single-minded protein
SRC-1	Steroid receptor co-activator
SMRT	Silencing mediator of retinoid and thyroid hormones
TCDD	2,3,7,8-Tetrachlorodibenzo-p-dioxin
TCDF	2,3,7,8-Tetrachlorodibenzofuran
Xap2	X-associated protein 2
XRE	Xenobiotic-responsive element
RESP	Restricted Electrostatic Potential
GAFF	Generalized Amber Force Field
PME	Particle Mesh Ewald
dRMSD	distance Root Mean Square Deviation

LIST OF FIGURES

1

INTRODUCTION

1.1 XENOBIOTIC RECEPTORS

Most animals are constantly exposed to a multitude of chemicals derived from the air they breath, the water they drink and the food they eat. Many of these chemicals provide useful information about the world surrounding them or are healthful and nutritive, while others are toxic and harmful. If, for any reason, some of the last ones start to circulate inside the animal body, they must be eliminated. Animals, including humans, have both constitutive expressed and inducible enzymes to facilitate the biotransformation and elimination of toxic compounds [Parkinson and Ogilvie, 2001]. Inducible enzymes are expressed only in the presence of an activating molecule, otherwise the constitutive ones are always expressed.

The biotransformation, or detoxification, enzymes are usually divided into three groups: phase I, bioactivation; phase II, conjugation; phase III, transport [Omiecinski et al., 2011]. Phase I reactions involve hydrolysis, reduction, and oxidation. These reactions expose or introduce a functional group, and usually result in a small increase in hydrophilicity; on the contrary, phase II reactions result in a large increase in xenobiotic hydrophilicity, promoting the transport and the excretion (phase III) of foreign chemicals [Parkinson and Ogilvie, 2001]. The main protein family responsible for the phase I reactions is the cytochrome P450 (CYP) monooxygenases enzyme family. The regulation of the expression level of CYP is mediated by transcriptor factors activated by ligands; because some of them can interact and be activated by xenobiotics, they are designed as xenosensors or xenobiotic receptors [Nakata et al., 2006].

1.1.1 *Functions and mechanisms*

We can say that the first step of the detoxification mechanism is the xenobiotic detection by the receptors, that include: Aryl hydrocarbon Receptor (AhR), Pregnane X Receptor (PXR or NR1I2) and Constitutive Androstane Receptor (CAR or NR1I4).

AhR belongs to the bHLH-PAS (basic helix-loop-helix/Per-ARNT-SIM) protein family and is the only one of its family that is activated by ligand binding [Perdew et al., 2018]. Family members can be grouped in two classes: class I neither homodimerize nor heterodimerize with other class I factors, it includes AhR and its repressor AhRR, hypoxia inducible factors α (HIFα), single minded proteins (SIM) circadian

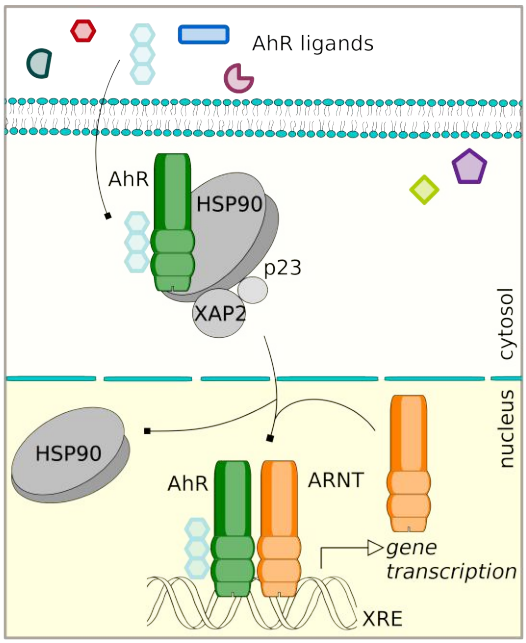

Figure 1.1 AhR activation mechanism. AhR in complex with chaperon proteins is activated by ligand binding that causes the nuclear translocation; into the nucleus AhR heterodimerizes with ARNT and the dimer binds DNA in the XRE region leading to gene transcription (modified from Bonati et al. [2017]).

locomotor output cycles kaput (CLOCK), and neuronal PAS proteins (NPAS); class II can both homodimerize and heterodimerize regardless the class of the protein partner, it includes AhR Nuclear Translocator (ARNT and ARNT2) and brain muscle ARNT like protein 1 (BMAL1) [Kewley et al., 2004].

The mechanism of action of AhR (Figure 1.1) is the prototypical of the transcription factors: it is located into the cytosol in complex with two heat shock proteins 90 (HSP90), one X-associated protein 2 (Xap2) and one p23 co-chaperon. The binding of the ligand leads to a conformational change and the cytosolic complex migrates into the nucleus. In the nucleus AhR heterodimerizes with its homologous ARNT and this dimer finally binds to DNA in a specific recognition site: the xenobiotic responsive element (XRE). The binding to DNA starts the transcription of some target genes, including the *cyp1a1* and the repressor AhRR [Ma, 2011].

The mechanism above described is usually referred to as "canonical", "classical" or "adaptive" because it is the first that was understood and it arose from the adaptation of the organisms to xenobiotics. In addition to the canonical pathway, the AhR appears to act through several non-canonical pathways [Denison et al., 2011], including, for example: the dimerization with other protein partners, e. g. the Krüppel-like Factor 6 (KLF6); the interaction of AhR:ARNT dimer directly with unliganded estrogen receptor to promote formation of a transcriptionally

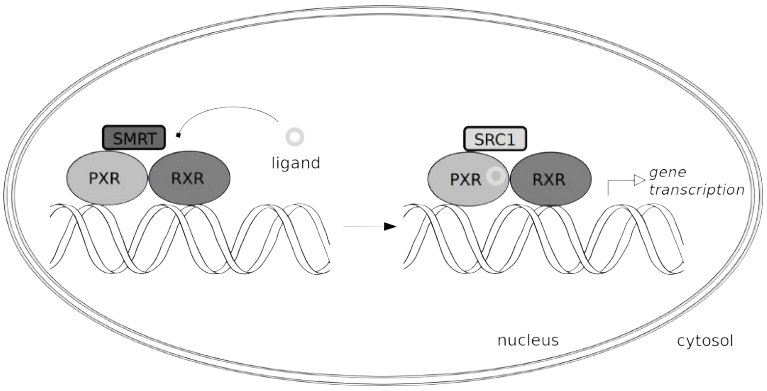

Figure 1.2 PXR, as a dimer with RXR, silences the target genes. The chemical activation is mediated by a ligand that interacts with PXR, the conformational change triggers the release of co-repressors and the recruitment of co-activators (inspired by Hernandez et al. [2009]).

active complex binding to estrogen response elements [Wright et al., 2017]; and the cross-talk between AhR transcription and Nuclear Factor-κB (NF-κB) which controls immune and inflammatory responses [Vogel et al., 2014]. AhR seems to be a very important cellular hub and it is almost impossible to describe the full spectrum of biological processes it mediates.

PXR and CAR are nuclear receptors (NRs), in particular they belong to the NR1I subfamily [Reschly and Krasowski, 2006] and their structure and mechanism of action is quite similar to other famous member of the super-family (e. g. estrogen receptor, androgen receptor) [Omiecinski et al., 2011].

Unlike steroid hormone receptors, which are highly selective and work under narrow concentration of their cognate hormone, PXR and CAR are orphan receptors, i. e. their physiological ligands are not known or do not exist [Gallastegui et al., 2015]. PXR is activated by a large variety of steroids, xenobiotics and therapeutic drugs, thus playing a central role in steroid hormone homeostasis and drug metabolism [di Masi et al., 2009]. CAR is constitutively active although it can bind different molecules, as a consequence the ligands of CAR are usually antagonists or "superagonists" [Mackowiak and Wang, 2016]. Because CAR is constitutively activated we will focus on the PXR mechanism.

It has been suggested that the human PXR acts as a gene silencer, i. e. it is constitutively bound to DNA in the PXR response element (PXRE) as heterodimer with the Retinoid X Receptor (RXR) and in this form it silences transcription of target genes (Figure 1.2). Ligand binding causes a conformational change leading to the release of co-repressors (e. g. silencing mediator of retinoid and thyroid hormones, SMRT) and the recruitment of co-activators (e. g. the steroid receptor co-activator 1, SRC-1) [di Masi et al., 2009].

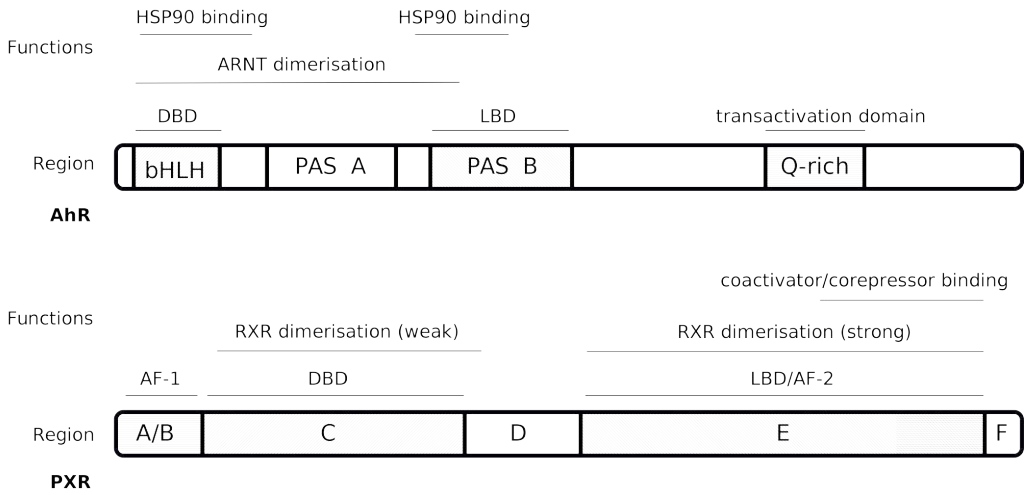

Figure 1.3 Functional domains of AhR and PXR. Regions nomenclature is derived by the protein family and the functions have been accorded to the literature [Ma, 2011, Laudet and Gronemeyer, 2002].

1.1.2 *Analogies and differences between AhR and PXR*

Both PXR and AhR are transcription factors activated by xenobiotics and start transcription of different CYPs to detoxify the organism. Though for decades the scientific community searched the endogenous ligand for AhR [Nguyen and Bradfield, 2008], it remains an orphan receptor with many endogenous candidate molecules. Also PXR is an orphan receptor [Blumberg et al., 1998], known for binding many endogenous and exogenous molecules [Chai et al., 2016]. In order to bind a large variety of the chemicals present in the environment, AhR and PXR are very promiscuous and can bind both small and hydrophobic ligands, and large and polar ones [Denison et al., 2011, Hernandez et al., 2009]. It is likely that both receptors have a large, hydrophobic and buried cavity able to accommodate different molecules. Not only PXR and AhR are activated by different ligands, but also they are species- and tissue- specific, i. e. ligands that activate transcription into a certain species or tissue might not be effective into another one. It was also found that they interact with common partners, for example both AhR and PXR interact directly with SRC-1 [Kumar and Perdew, 1999, di Masi et al., 2009]; otherwise they interfere with the regulation of NF-κB [Vogel et al., 2014, Banerjee et al., 2015], but also of estrogen receptor [Wright et al., 2017, Biswas et al., 2009].

PXR and AhR have different lengths: PXR contains $\sim$400 aminoacids, while AhR is twice longer. AhR is a bHLH-PAS protein: the bHLH is a common motif that consists of 4-6 DNA binding basic amino acids that are attached to a HLH dimerization domain [Bersten et al., 2013]; PAS domains are important signaling modules that monitor changes in light, redox potential, oxygen, small ligands, and overall energy level

of a cell and are located in the cytosol [Taylor and Zhulin, 1999]. In bHLH-PAS proteins the PAS domain is repeated twice: PASA and PASB (Figure 1.3) and, together with bHLH, PAS domains are located at the N-terminus region. In AhR the PASB is the ligand binding domain (LBD), while HLH is responsible for dimerization; HLH together with PASB enable the association with HSP90, and together with PASA allow dimerization with ARNT although PASB domain is important for initiation of AhR:ARNT dimerization and seems to have a modulatory role. At the C-terminus is located the Q-rich region that is the transactivation domain. AhRR lacks both PASB and the Q-rich, so it can compete with AhR for dimerization with ARNT but do not start any transcription activity [Kewley et al., 2004].

PXR is a NR, a family of transcription factors that share a modular structure of three conserved functional domains:

1. ligand binding/dimerization domain;
2. DNA-binding/weak dimerization domain;
3. transactivation domains (activation function 1 and 2 [AF-1, AF-2]).

The LBD is located at the C-terminus of the receptor. The DNA-binding domain (DBD) is at the N-terminus and is responsible for recognition of a receptor-specific response element in the promoter region of the target genes. Finally, the transactivation domains consist of one ligand-independent AF-1 domain at the N-terminal and one C-terminal ligand-dependent transcription AF-2 domain. These domains serve as protein-protein interfaces that guide the recruitment of transcriptional coregulators to the target gene [Laudet and Gronemeyer, 2002].

Both protein families (bHLH-PAS and NR) have a inner highly conserved structure despite low sequence identity.

Whereas the single domains of the two proteins have similar functions: LBD, DBD and transactivation domain; they are differently arranged in the protein (Figure 1.3). While the functional domains in PXR are along the whole length of the protein, in AhR they are concentrated on the N-terminal region.

Because AhR and PXR belong to different protein families, the xenobiotic detoxification mechanism might be a case of molecular convergent evolution; these two proteins evolved to detect and metabolize xenobiotics independently and on different time. Convergent evolution may explain both the molecular similarities and differences observed for these two systems.

1.1.3 *Ligands and modulators*

The best known and most studied ligands for the AhR are the environmental contaminants encompassing halogenated organic compounds. Of these, halogenated aromatic hydrocarbons (HAH), com-

posed by polychlorinated dibenzo-p-dioxins, dibenzofurans and biphenyls PCDD, PCDF, PCB), together with polycyclic aromatic hydrocarbons (PAHs) are particularly important due to their toxicity and widespread distribution in the environment [Perdew et al., 2018].

Dioxins consist of two benzene rings connected by two oxygen atoms and contain four to eight chlorines (Figure 1.4), for a total of up to 75 congeners [Schectera et al., 2006]. This family also include one of the most potent toxicants the world has ever seen: 2,3,7,8-tetrachlorodibenzo-para-dioxin or TCDD [Mandal, 2005]. Dioxins are unwanted contaminants almost exclusively produced by industrial processes, including incineration, chlorine bleaching of paper and pulp, and the manufacture of some insecticides, herbicides, and fungicides. Dioxins did not exist prior to industrialization except in very small amounts due to natural combustion and geological processes [Schectera et al., 2006]. The prototype chemical for dioxin is TCDD, which induces a broad spectrum of toxic responses: disruption of normal hormone signaling pathways, reproductive and developmental defects, immunotoxicity, liver damage, wasting syndrome, and cancer [Schectera et al., 2006, Perdew et al., 2018, Mandal, 2005]. Although a great deal is known about AhR-driven transcriptional regulation, much less is known about the mechanisms by which TCDD causes toxicity and disease [Mandal, 2005].

While the exact mechanism leading to toxicity for TCDD and HAHs remains to be clarified, in the case of PAH the genotoxicity has been better described. The conversion of PAHs into carcinogenic intermediates depends on enzyme-catalyzed biotransformation; CYPs play a pivotal role in PAHs metabolism converting them into epoxide species [Henkler et al., 2012]. The hydrolysis of the primary epoxide into a diol, and further epoxidation produces the carcinogenic metabolites of PAHs [Luch and Baird, 2010]. The prototypical carcinogenic PAH that serves as lead compound now since decades is benzo[a]pyrene (BaP) (Figure 1.4), which has been judged carcinogenic for humans (group 1) in 2012 by the International Agency for Research on Cancer (IARC), it remains the only PAH in the group 1 [IARC, 2012].

Because of their aromaticity and lack of polar substituents, parent PAHs behave as lipophilic and chemically inert compounds. Different congeners are present in the environment due to their natural occurrence in crude oil depositories; although a more relevant source, especially in cities and industrial districts, is their generation through all kinds of incomplete combustion processes [Henkler et al., 2012].

There are also non-toxic/naturally occurring ligands that are assumed from the diet and are commonly found in vegetables, fruits, berries, teas and herbal food extracts such as ginseng, ginkgo biloba and liquorice [Safe et al., 2011]. An example may be flavonoids, a large family of compounds that are widely present in vegetables and generally consumed in significant amounts; the prototypical flavonoid

that is an agonist for AhR is β-naphthoflavone (BNF) (Figure 1.4). They have anti-inflammatory activity, and many are antioxidants [Safe et al., 2011].

Some decades after the first identification of the *ahr* gene, it was found that AhR has also many physiological functions [Gasiewicz and Henry, 2011]. This hypothesis is supported by the marked degree of conservation among species and the constitutive expression during development and in adult tissues of the AhR, as well as, even more importantly, by the developmental aberrations observed in AhR knockout animal models [Guyot et al., 2013]. As a consequence it seems apparent that there must be one or several major endogenous activators of the AhR. Endogenous molecules can derive from the metabolism of the heme or the arachidonic acid [Denison and Nagy, 2003].

Halfway between exogenous and endogenous ligands there are the tryptophan derived compounds [Hubbard et al., 2015], which includes indole (Figure 1.4), indigoids, UV photoproducts of tryptophan [Nguyen and Bradfield, 2008], but also kynurenine (Figure 1.4) and its metabolites, and 6-formylindolo[3,2-b]carbazole (FICZ) (Figure 1.4) the endogenous ligand with highest affinity and, for this reason, suspected to be the physiological ligand [Wincent et al., 2009]. These compounds may derive from the organism metabolism, from dietary intake, or from the microbiota, in fact it has been suggested that the indole-like AhR ligands are the chemical mediator between the host and the microbiota [Hubbard et al., 2015].

In recent years AhR has been considered as a new intriguing pharmaceutical target, and the scientific community started to look for selective AhR modulators (SAhRMs) [Denison et al., 2011]. The rationale for this is based on the notion that some of the biological impacts of even TCDD are such that they would be beneficial in the treatment of certain diseases if they could be separated from the toxicity. The term SAhRM has not been unambiguously defined, but it is generally used to describe compounds that exhibit tissue- or species-specific AhR activation or inactivation, and are able to only induce some of the typical responses of AhR activation, but not others (for example toxicity).

Modulation of AhR activity has been related to inflammation and regulation of other immune pathways [Denison et al., 2011, Guyot et al., 2013], moreover antagonists could be used in cancer treatment [Murray et al., 2014]. But all the physiological functions and its potential pharmaceutical use are difficult to imagine, they start from Crohn's disease [Benson and Shepherd, 2011] ending to retina protection [Gutierrez et al., 2016]. Moreover many drugs approved by the Food and Drug Administration (FDA) are AhR ligands and could be used as SAhRMs for different purposes, e. g. carbidopa, a drug for treating Parkinson's disease, inhibits pancreatic cancer cell and tumor growth [Safe, 2017]. As a consequence we should start consider the tox-

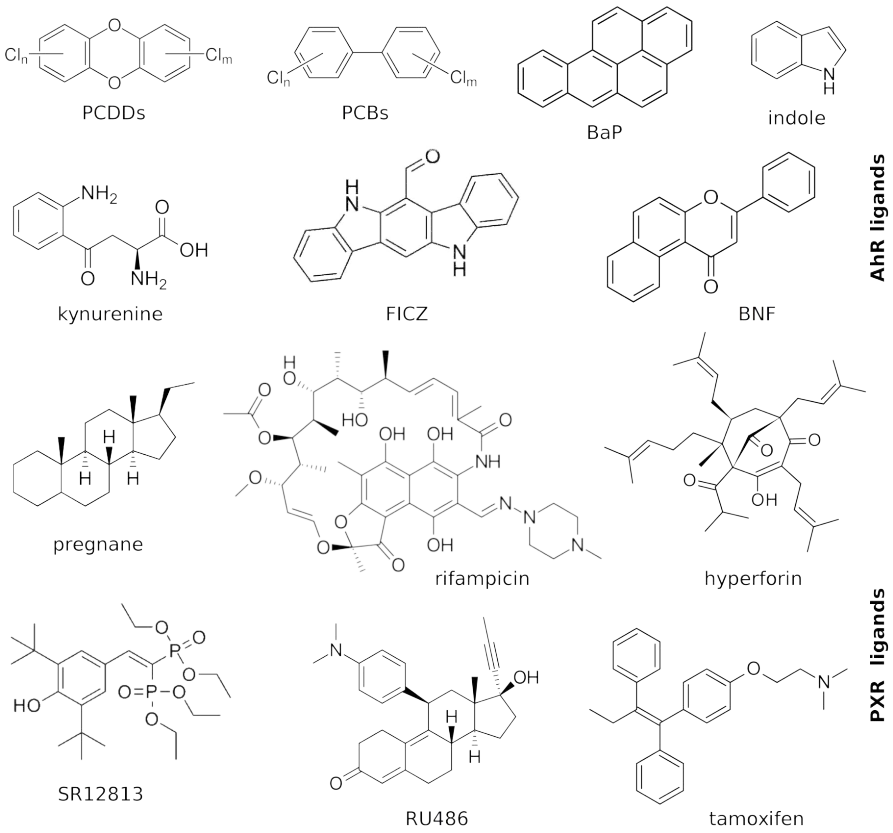

Figure 1.4 Some ligands that activate AhR and PXR. While some are pollutants: PCDDs, PCBs, BaP; some others come from the diet of from metabolism: indole, kynurenine, FICZ, BNF, pregnane; and others are used as drugs: rifampicin, hyperforin, SR12813, RU486 and tamoxifen.

icological and pharmacological consequences of these FDA approved drugs.

The name of PXR derives from its activation by the 21-carbon steroids (also known as pregnanes) (Figure 1.4). Subsequently to the first PXR discovery, researchers cloned the human PXR and called it steroid and xenobiotic receptor (SXR) because they found could also bind various xenobiotics [Blumberg et al., 1998].

Agonists of PXR have been extensively studied because the activation of PXR leads to the metabolization of many important drugs. Only to mention a few: the antibiotics rifampicin (Figure 1.4), clotrimazole, and the antiretroviral ritonavir; the antineoplastic drugs cyclophosphamide, cyproterone acetate, taxol and tamoxifen (Figure 1.4); the anti-inflammatory agent dexamethasone; the anti-type 2 diabetes drug troglitazone; the antihypertensive drugs nifedipine and spironolactone; and the sedatives glutethimide and phenobarbital; cholesterol-lowering drug SR12813 (Figure 1.4); and also RU486 (Figure 1.4) [Hernandez et al., 2009]. Moreover also molecules contained in commonly used herbal medicines activates PXR, e. g. hyperforin (Figure 1.4)

from St. John's wort [Banerjee and Chen, 2013]. In fact, in the clinic, people are cautioned about using PXR agonists because they may cause adverse drug-drug or diet-drug interactions during drug therapy; decreasing the bioavailability of the agonist, or accelerating the formation of reactive metabolites and causing undesired interactions or even toxicity [Banerjee et al., 2015]. In cancer growth and carcinogenesis, it has been suggested that PXR induces cell growth and is pro-carcinogenic, thereby acting as a possible oncogene [Biswas et al., 2009]. Different mechanisms were proposed and all support the PXR role as a protector against tissue damage, a role that may be pathophysiologic in neoplastic cells. In addition PXR has been shown to induce cancer drug resistance [Chai et al., 2016]. However the therapeutic approach of targeting PXR is disease specific. For inflammatory diseases, PXR activation has suppressive effects on NF-κB, implying the therapeutic potential of PXR agonists in treating inflammatory bowel disease. PXR antagonists are of potential therapeutic utility owing to the role of PXR in promoting tumor cell growth, chemoresistance and malignancy [Banerjee et al., 2015].

1.2 MOLECULAR MODELING OF LIGAND BINDING TO XENOBIOTIC RECEPTORS

Toxicology in the broadest sense is the study of the adverse effects of drugs or chemicals on living systems. The questions posed by this discipline include what compounds are toxic, how and why toxicity is manifested, and how might toxicity be predicted, treated, or prevented [Ekins et al., 2007]. Some of the toxicity effects are receptor-mediated, i. e. the toxic chemical binds to a receptor that initiates a cellular response. Ligand-induced modifications of the physicochemical properties or conformational changes of the receptor can trigger transcription processes or signal-transduction cascades [Perdew et al., 2010].

Computational toxicology aims to use rules, models and algorithms based on prior data for specific endpoints, to enable the prediction of whether a molecule is or is not toxic. Computational molecular modeling methods are at the core of mechanistic toxicology, allowing to understand the mechanisms through which a given chemical induces an Adverse Outcome Pathway (AOP) [Goldsmith et al., 2012, Ankley et al., 2010].

In the specific receptor-mediated toxicity the interest is focused on the pair-wise interactions between the small molecule and the target macromolecule. To evaluate these interactions certain methods are too computationally expensive/intensive and scale poorly with system size, i. e. Quantum Mechanics, for which the smallest units are electrons. If we consider the atoms as the smallest unit of relevance and represent them as spheres connected by springs (representing the

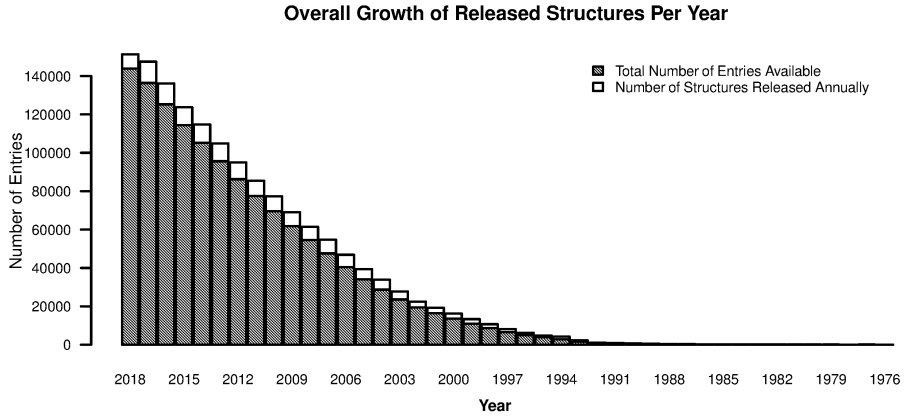

Figure 1.5 The exponential growth of the PDB. From its establishment in 1971, when contained 7 protein structures, until the update of August 29th 2018, when it contained 143,840 biological macromolecular structures.

bonds), it is sufficient a classical physics formalism, precisely a Molecular Mechanics (MM) framework to describe these systems [Goldsmith et al., 2012]. In MM methods the classical physics formalism is used to predict the energy of a molecule as a function of its conformation [Leach, 2001]. In this framework the potential energy can be expressed as the sum of bonded and non-bonded terms; the first is represented by three terms: bond stretching, bond angle bending and torsional potentials; the second as the sum of electrostatic and van der Waals interaction terms.

Receptor-mediated toxicity can be studied using computational (or *in silico*) methods based on MM, but they require a wide experimental knowledge of the biological system. This knowledge can derive from *in vitro* experiments, but often it is necessary to know the three-dimensional structure of the receptor to start the study.

The number of proteins with a known three-dimensional structure is increasing rapidly (Figure 1.5), and structures produced by structural genomics initiatives are available on public databases, e. g. the Protein Data Bank (PDB) [Berman et al., 2000]. With the exponential growth of structures also methodologies which use structural information had increased. Some structural bioinformatics tools, e. g. Homology Modeling, exploit the information contained in both sequences and structures of proteins to obtain a model representing a homologous protein which structure is not already crystallized, or that cannot be solved by experimental techniques. Molecular modeling encompasses all methods used to model the behavior of molecules, from their structures and physico-chemical properties, to dynamics and interactions. A relevant point to consider is the computer resources needed for modeling large molecular systems at the atomistic level with MM methods (a simple rule of thumb for the length of the "typical" protein might be $\sim$400 aminoacids, i. e. > 7000 atoms).

Computational methods have become a crucial component of many drug discovery programs, from hit identification to lead optimization and beyond [Kitchen et al., 2004], and are also emerging as valuable assets for *in silico* toxicology [Raunio, 2011]. Both the advancement of *in vitro* and *in silico* toxicology and the 3R principles inspired the last European legislation about chemicals, REACH (Registration, Evaluation, Authorization and Restriction of Chemicals). The 3R were proposed in 1959 by William Russell and Rex Burch; the Authors proposed the principles of Replacement, Reduction and Refinement as the key strategies to provide a systematic framework to achieve the goal of "humane" experimental techniques [Russell and Burch, 1959]. Even if the Authors originally referred to *in vitro* techniques, *in silico* toxicology can be easily inserted into this framework and molecular modeling tools can provide mechanistic information helping the explanation of the underling systems. It is interesting to recall the case of Benzo[a]pyrene; this chemical has been classified into group 1 by IARC based on "mechanistic and other relevant data"; this suggests that, in the near future, we will see more and more results of *in silico* methods being included in risk assessment documents as supporting data [Raunio, 2011].

1.2.1 *Theoretical aspects in predicting ligand binding*

Ligand binding is a complex biological event that involves both thermodynamics and kinetic properties.

Of particular interest is the binding free energy (ΔG_{bind}), which measures the favorability of a given reaction, in this case the association between ligand and receptor. The binding free energy (ΔG_{bind}) is defined as the difference between the free energy of the complex and the sum of free energies of the ligand and the receptor:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{ligand}} + G_{\text{receptor}})$$

ΔG_{bind} is a function of the temperature, pressure, ionic strength, pH, solvent, and concentration of the chemical species. Given that the most common measurement for ΔG_{bind} is through the equilibrium constant for the complex (or association constant K_a):

$$\Delta G_{\text{bind}} = -RT \ln K_a = RT \ln K_d$$

where R is the gas costant and T is the temperature in Kelvin. By this equation we linked the ΔG_{bind} with K_a that is a quantitative evaluation of ligand affinity. It is also possible to use K_d, the dissociation constant, which is simply the reciprocal of the equilibrium constant [Krumrine et al., 2005].

These thermodynamic parameters are connected to chemical kinetics:

$$K_a = k_{on}/k_{off}$$

where k_{on} is the forward rate of reaction between receptor and ligand and k_{off} is the dissociation rate.

ΔG_{bind} can also be predicted by computational methods to obtain an estimation of the ligand binding affinity.

We can also express the ΔG_{bind} in terms of changes in enthalpy (ΔH) and entropy (ΔS) upon formation of the complex.

$$\Delta G_{bind} = \Delta H - T\Delta S$$

Changes in enthalpy are due to the modification of van der Waals and Coulomb interactions as the atoms of the complex replace atoms from the solvent on complex formation. In fact both the ligand and the receptor are fully solvated before the binding event. Changes in entropy are linked to differences in translational and rotational degrees of freedom for the ligand, the receptor or the solvent molecules, moreover is due to the loss of conformational and vibrational entropy on binding for the receptor and ligand. The release of ordered water molecules around both the ligand and the receptor's active site and the resulting increase in entropy of these water molecules favors binding and is what underlies the hydrophobic effect. The binding event is a delicate balance between exchange of hydrogen bonds, establishment of van der Waals and Coulomb interactions, entropy losses of the receptor and the ligand, and gain in entropy of the solvent [Krumrine et al., 2005].

To determine the value of the ΔG_{bind} we can use molecular mechanic force field to calculate the enthalpic term and the entropic term using Boltzmann's law:

$$S = -k_B \sum_j P_j \ln P_j$$

where k_B is the Boltzmann constant and P_j is the probabily of a molecule to be in a particular microstate rather than another and is defined as:

$$P_j = \frac{e^{E_j/k_B T}}{\sum_j e^{-E_j/k_B T}}$$

Macromolecules have a high number of degrees of freedom, for this reason the entropic term is usually neglected and considered as constant. In these particular case we can write:

$$\Delta G_{bind} \approx \Delta H$$

To estimate the enthalpic term we can use Molecular Mechanics. The function useful to describe the energy is the sum of a term of intramolecular interactions between atoms (all the bonds, the angles,

the dihedrals componing the molecule), and a term of non-bonded interactions (van der Waals and Coulomb). Bonds and angles functions derive from the force law for harmonic motion (stretching and bending), while the dihedrals are described by a function which underlines the periodicity of the force. The van der Waals interactions are modeled by the Leonard-Jones function and the Coulomb interactions are described by the Coulomb's law.

$$\begin{aligned} E_{\mathrm{MM}} = & \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \\ & + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \\ & + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \\ & + \sum_{i<j} \Big[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}} \Big] \end{aligned}$$

where K_r is the harmonic force constant, r_{eq} is equilibrium bond length; K_θ is the angle bending force constant and θ_{eq} is the reference bond angle; V_n are the torsional rotation force constants, n is the periodicity, ϕ is the current torsional angle and γ is the phase angle; A and B are suitable constants, R_{ij} is the distance between the i atom and the j atom, ϵ is the permitivitty of free space and q is the charge quantity.

During the binding process, polar and nonpolar groups of the receptor and the ligand form van der Waals interactions on complex formation, and charged groups strongly interact by Coulomb interactions. The ligand binding event in buried cavities usually is driven by van der Waals interactions and the ligand is anchored to the receptor by hydrogen bonds among polar groups, because it is rare to find charged species into buried cavities.

In addition to the direct interaction between ligand and receptor, it is also relevant the role of the solvent. In case of biomolecules the solvent is almost always water and covers two possible effects: a short-range and a long-range one. The first derives from local solute-solvent interactions. The nonpolar solvation free energies, accounting for transferring a nonpolar solute from the gas phase to the solvent, are usually assumed to be proportional to the solvent accessible surface area (so that is more relevant for receptors which have a greater surface than ligands). The second effect is due to long-range Coulomb interactions and causes an attenuation of interactions between solvated charges. This "screening", usually described as a macroscopic dielectric constant, reduces electrostatic interactions in water by 80-fold from their vacuum values [Krumrine et al., 2005]. For this reason, if it is not possible to explicitly include water molecules during calculation,

solvent can be implicitly taken into account by a dielectric constant equal to 80.

1.2.2 *Molecular Docking*

One key methodology to deal with binding of small molecules to protein binding sites is molecular docking, that was pioneered during the early '80s and remains a highly active area of research [Kitchen et al., 2004]. The docking process involves the prediction of ligand conformation and orientation (posing) within a targeted binding site. Molecular docking can be thought as a combination of spatial sampling methods with scoring functions, that search over many possible interactions between a ligand and a receptor in order to identify a set of ligand poses that represent local minimum-energy positions of the ligand. If the sampling of ligand poses is adequate, and the energy scoring function is sufficiently accurate, then the global minimum-energy position of the ligand in the receptor can be selected from the set of local energy minima [Moustakas, 2007]. Using molecular docking we can estimate the active conformation of toxic compounds and conduce an in-depth study and characterization of its interactions with the receptor.

The standard docking protocol can be divided into four main components, shown in Figure 1.6:

THE TARGET STRUCTURE Many macromolecular structures have been determined both using X-ray crystallography and nuclear magnetic resonance (NMR) and are available in the PDB. When no experimental structure is known, or easily obtainable, homology modeling can provide a valid alternative. Usually it is not sufficient to have the target structures, they have to be carefully prepared to ensure structural integrity, assign the correct residue protonation and tautomeric states (especially concerning His), and inspect Asn and Gln flips.

THE COMPOUND LIBRARY The ligands subject of the study may derive from public available libraries (ZINC, DrugBank, ChEMBL), or may be built in house with a subset of interesting ligands for the investigated system. Also the ligands have to be prepared, taking into account the most representative tautomers, protomers and stereoisomers at the pH of interest.

THE DOCKING STRATEGY It has to be defined the method which will lead to the prediction of the binding geometry (pose) of the ligand inside the prepared target. It is necessary to define the level of flexibility inserted during calculation: both protein and ligands as rigid (as in the original lock and key model); rigid protein with flexible ligands; or both systems flexible. It is possible

to model the small molecules as fully flexible and only a few relevant atoms (or even sidechains) of the protein as able to move. Moreover also water molecules can be very important for the binding event; in case there is no experimental information about water molecules, some algorithms can implicitly include them during calculation with an increased computational demand. Docking sampling and scoring functions are selected depending on the aim of the study, the computational resources available and the dimension of both the target and the compound library. Both sampling and scoring had been optimized in the years in order to properly estimate the binding free energy (ΔG_{bind}), or to maximize the separation between potential ligands and inactive compounds.

COMPOUND RANKING OR RESCORING If the library is massive, molecules are ranked by docking score and chemically clustered, examined for key interaction patterns, and visually inspected. In case we are interested in the estimation of ΔG_{bind}, it is possible to rescore the top selected poses using a more sophisticated method to approximate the ΔG_{bind}, e. g. MM-PBSA or MM-GBSA (Molecular Mechanics energies combined with the Poisson-Boltzmann or Generalized Born and Surface Area continuum solvation models to treat solvation/desolvation processes) [Greenidge et al., 2014].

Molecular docking offers a number of benefits compared to traditional high-throughput wet screening approaches, such as low set-up cost, computational speed, and flexibility in changing the simulation conditions. Virtual procedures are also not affected by ligand purity, protein stability, or by the different samples and assays conditions [Spyrakis and Cavasotto, 2015].

However there are a number of limitations and open issues concerning docking. Regardind the phase of sampling, we know that we always have to find the best compromise between speed and exploration of the search space. The inclusion of flexibility and of all solvent molecules that solvate the system, leads to the combinatorial explosion of calculation. For this reason, flexibility of the receptor and solvation effects are usually neglected explicitly. Moreover during scoring we ignore the entropic term, and consider solvent as a dielectric constant.

Furthermore, it has been demonstrated that the input conformation of the ligand affects the final pose [Feher and Williams, 2009], while the final pose should ideally be unrelated to the initial conformation, especially for what concerns the reproducibility of the results. It is also known that larger ligands perform worse than smaller, more rigid molecules, and it is commonly accepted that docking hydrophobic molecules and predicting their activities using docking is quite a challenging task [Plewczynski et al., 2010]. Cross-docking experiments, *i. e.* docking a ligand into a protein structure determined in the presence

of a different ligand, also suggest that the binding site of the protein is biased toward the native ligand [Lexa and Carlson, 2012]. In fact, usually docking programs are able to find a "good" pose among the top group, but because the reproduction of the geometry is not exact (ligand-protein contacts are poorly reproduced), the scoring function is unable to predict the true binding affinities and to select the best pose [Warren et al., 2006]. But the debate about which would be the limiting step of docking, scoring [Kitchen et al., 2004] or sampling [Greenidge et al., 2014], is still open. It is certain that sampling and scoring are deeply linked and it will not be easy to solve this issue.

Molecular simulations, such as molecular docking or molecular dynamics, are everything but simple, and require a proper balance of the most appropriate tools, the adequate expertise and, most of all, an appropriate level of knowledge of the system under investigation [Spyrakis and Cavasotto, 2015].

Dealing with promiscuity of receptor proteins requires additional attention in the use of docking protocols, also because it is more difficult to achieve a deep knowledge of these systems. Promiscuous proteins, i. e. with alternative activities, can become important for the survival of an organism, if the conditions in the environment change. This can give a selective advantage to members of the population where the secondary activity is higher, thus resulting in the incorporation of such activity under evolutionary selective pressure [Nobeli et al., 2009]. Perhaps the most important mechanism by which promiscuity can be achieved is structural flexibility. Promiscuous proteins are usually highly flexible, can display multiple binding sites with the possibility of accommodating multiple ligands in a variety of ways. It comes apparent that to deal with promiscuous proteins is necessary to include as flexibility as possible in the computational simulation to better describe the binding event.

1.2.3 *Beyond classical molecular docking*

In order to consider conformational dynamics during the binding process, different models have been proposed. The first lock and key model proposed by Fischer (1890) began to "rust" after a century [Jorgensen, 1991]. Alternative models able to take into account flexibility were proposed. The induced fit model found its theoretical statement on the hypothesis that the bound protein conformation is induced by ligand intervention [Koshland, 1958]. On the contrary the conformational selection model considers the ligand as the selector of the most complementary protein conformation among a pre-existing ensemble of metastable states; the selection shifts the dynamic population equilibrium toward the selected conformation [Boehr et al., 2009]. While "following 50 years of debate and experimentations, the vast majority of the data support the conformational selection

scheme of signal transduction" [Changeux and Edelstein, 2011], it may be interesting to draw a parallel between this debate and the Gould-Dawkins one about evolution of organism [Shanahan, 2017]. In both debates it is reasonable to think that each case is a special case, and the two proposed mechanisms may coexist in the same process, with a mechanism being more relevant in one case and less in another.

The simplest way to include flexibility during binding calculation is soft docking, that is based on the induced-fit model. In soft docking the van der Waals radii of the protein are scaled to a certain percent, so that more poses can be sampled preventing steric clashes; this first step is usually followed by a local energy minimization to better accommodate the ligand and propose a more realistic complex [Jiang and Kim, 1991]. The flexibility of the protein was explicitly introduced in docking calculations at first using a library of sidechain rotamers instead of a single conformation, and also considering their random movements during calculation [Lexa and Carlson, 2012]. The generation of new conformations of the binding site was implemented in the induced-fit docking protocol as an iteration of sampling ligand orientation and sampling protein conformations until the final score overcome a certain threshold [Sherman et al., 2006]. More sophisticated methods can include not only sidechain- but also backbone-flexibility, coupling a first raw placement of the ligand with a Monte Carlo refinement, in this case the ligand library that can be analyzed is very limited in number. These methods include the RosettaLigand and PELE [Davis and Baker, 2009, Borrelli et al., 2005] approaches. Finally it is possible to include flexibility *a posteriori*, performing a short Molecular Dynamics (MD) simulation of a pose previously determined using docking. A MD simulation numerically solves Newton's equations of motion, thus allowing to follow the evolution in time of structural fluctuations [De Vivo et al., 2016].

There are also approaches based on the conformational selection theory; the simplest is ensemble docking, which uses a pre-existing ensemble of conformations to be submitted to docking [Totrov and Abagyan, 2008, Amaro et al., 2018]. The ensemble can derive from experiments (different PDB structures of the same protein), or from conformations extracted from a MD trajectory, or from multiple homology models. Results obtained from ensemble docking show that without a "correct" protein conformation included in the ensemble, ligands cannot be docked well. Thus, it seems essential to better represent protein flexibility in the quest for better docked ligands [Greenidge et al., 2014, Motta and Bonati, 2017].

To explicitly study the molecular binding mechanism it would be necessary to simulate the whole process using MD. Such approach relies on the hypothesis that simulating "long enough" the unbound system with the ligand, they will diffuse into water until they form the energy favored protein-ligand complex.

Even if the complex is thermodinamically favored, its formation is associated to the overcome of a free-energy barrier that determines the rate of transition, and thus the timescale of the process. The highest is the barrier, the lowest is the probability to observe the transition. In the case of a protein-ligand complex formation, the process can require from milliseconds to even seconds to happen. Nowadays it is possible to simulate until microseconds on high-performance computers, but simulations of milliseconds are still rare. Despite these limitations, long plain MD simulations actually were employed to study the ligand binding events [Dror et al., 2012].

During last decades many methods have been designed to speed up the conformational sampling, thus reducing the computational demand. The simplest way is to raise the temperature of the simulation; both simulated annealing and replica exchange methods exploit this idea [Adcock and McCammon, 2006]. It is also possible to speed sampling by introducing artificial biases into the model upon which the simulation is based. Many of these methods need to define a reaction coordinate *a priori* to correctly simulate the binding event; for this reason it is important to reach an appropriate level of knowledge of the system under investigation [Sinko et al., 2013]. The introduced biases guide the sampling toward high energy states (rarely sampled). For example, in steered MD a pulling force is used to facilitate the conformational change along a selected reaction coordinate; in umbrella sampling different simulations are run in parallel, each biased to a specific range of the reaction coordinate; in accelerated MD, a bias potential is added to raise low-energy states, reducing the energy required for the transition; in metadynamics, an history-dependent bias is added to the system to fill up the valleys of the energy landscape [Adcock and McCammon, 2006, Sinko et al., 2013].

All these techniques allow to compute the underlying free energy landscape of the system analyzed, providing important information about thermodynamics and kinetics of the process of interest, but also mechanistic insights about the process under investigation.

1.3 AIMS AND OUTLINE OF THE THESIS

Aim of the studies performed in this thesis was to understand the mechanism of action of some known agonists of both the AhR and PXR proteins using molecular modeling methods, in the framework of computational/mechanistic toxicology. The inclusion of flexibility and plasticity of the protein during ligand binding was tackled in different ways depending on the system investigated. Specific computational protocols were developed to overcome the known limitations of classical docking.

In Chapter 2, I present the results on computational modeling of ligand binding to the AhR. The studies performed on the receptors

of different species, from the invertebrate *C. elegans*, to amphibians, rodents and human, helped in elucidating the differences in ligand binding to different species, which may lead to different biological responses. I demonstrate how molecular modeling approaches contributed to explain the species-specificity of AhR. Given that AhR is not only species-specific, but also ligand-specific, I present the results of an in-depth study of binding to AhR of a set of diverse ligands, aimed to find the molecular determinants that differentiate their binding modes inside the cavity of AhR. Here I describe a novel computational protocol I developed to perform docking to homology models taking into account receptor flexibility.

In Chapter 3, I present computational studies on binding to PXR. Firstly I compare the results obtained with a classical docking protocol with those derived using an ensemble docking approach, and I define the specific limitations of docking in this study case. Secondly I describe the use of enhanced MD methods to determine the path of entrance into the binding cavity of the most studied agonist of PXR and to try to define its preferred binding mode.

Finally, in Chapter 4, I draw some general conclusions about the use of molecular modeling tools in computational toxicology with specific comments about the two study cases here analyzed.

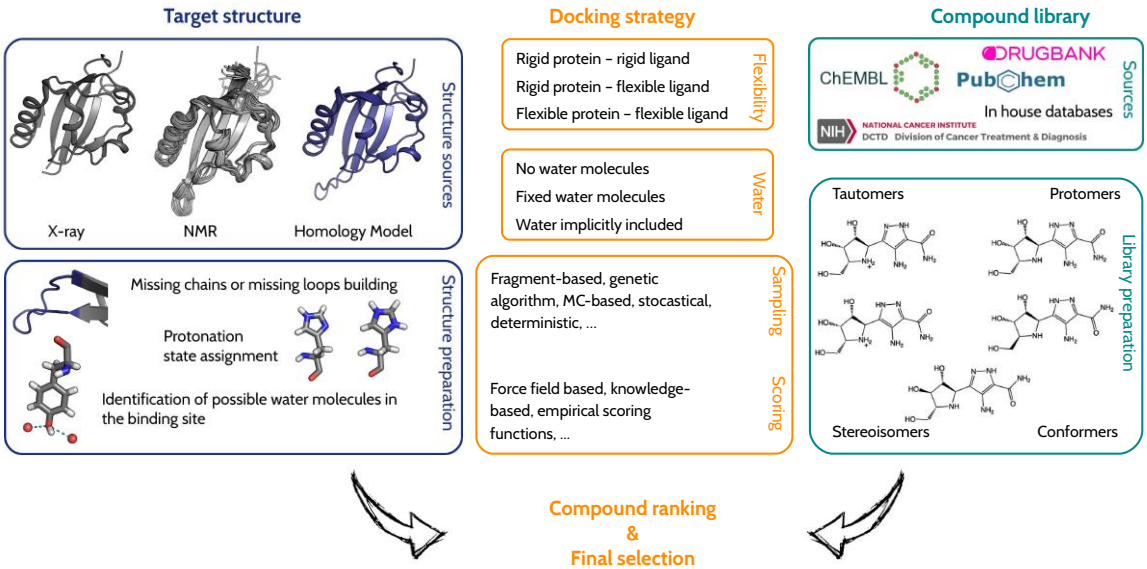

Figure 1.6 General docking protocol. Target structures and ligands have to be prepared before treating them with the selected docking strategy (Image inspired by Spyrakis and Cavasotto [2015]).

2

BINDING OF DIVERSE LIGANDS TO DIFFERENT AHRS

2.1 INTRODUCTION

In order to understand how different ligands in different species act in different ways, an in-depth study of the mechanism of action of AhR is necessary. Molecular modeling can be crucial to rationalize experimental studies, or to model at a molecular level the AhR activation and transformation events induced by ligands.

Unfortunately, to date, no experimental structure of the AhR-PASB domain, where the LBD is located, is available. However, during the last years structural information about the bHLH-PAS proteins has rapidly increased and recently the full-length structure of the three N-terminal domains, bHLH, PASA and PASB, has become available for some members of the family. First the CLOCK:BMAL1 [Huang et al., 2012], then the HIFα:ARNT [Wu et al., 2015], and finally the NPAS:ARNT [Wu et al., 2016] dimer structures have been resolved; most of these structures are bound to DNA. Recently also the AhR:ARNT dimer has been partly solved, and the first X-ray depositions include the bHLH and PASA domains of both AhR and ARNT [Schulte et al., 2017, Seok et al., 2017]. It is also interesting to cite the AhRR:ARNT dimer structure, that includes the ARNT PASB [Sakurai et al., 2017] and shows an asymmetric intertwined domain organization.

Because our interest is focused on ligand binding, the lack of any experimental structures of the AhR PASB domain (either isolated or in a dimeric form) has hampered in depth studies for many years; however, information derived from other bHLH-PAS proteins, when available, has been exploited for modeling the LBD structure and interactions.

The first model of the mouse AhR (mAhR) was built in 2002 on the basis of a distant PAS protein template structure [Procopio et al., 2002]. The deposition of the evolutionary closer templates HIF2α and ARNT lead to a new and more reliable model, because both these structures belong to bHLH-PAS family and have higher sequence identity to the AhR. This model was validated with mutagenesis studies [Pandini et al., 2007], and made possible to identify the "TCDD-binding fingerprint", i. e. the group of residues associated to optimal TCDD binding [Pandini et al., 2009].

These early approaches elucidated some structural feature of the AhR-LBD, which shares the typical PAS fold shown in Figure 2.1, and identified the internal cavity as the binding site for TCDD.

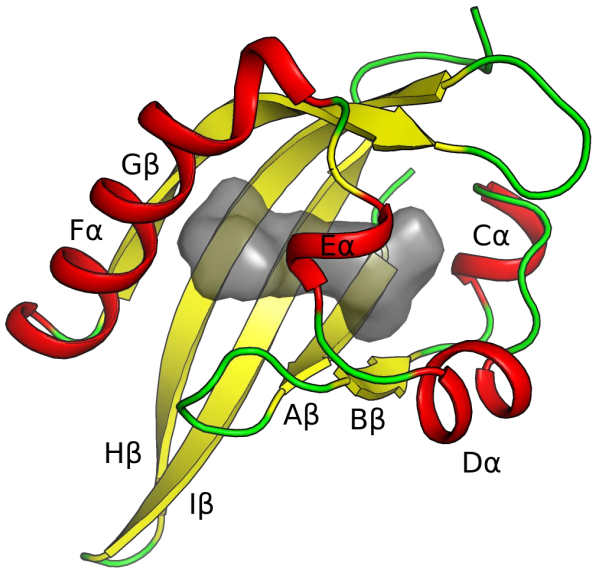

Figure 2.1 The modeled mAhR-PASB domain. It shows the typical PAS fold, composed by five strands, three small helices and a long helical connector. The binding site is located between the β-sheet and the helices and is a buried cavity, not exposed to solvent. The secondary-structure nomenclature for PAS domains here reported was proposed in Möglich et al. [2009].

Molecular docking to the homology models of mAhR was used in a virtual screening campaign to identify ligands in a database of 498 natural compounds [Bisson et al., 2009]. The top-scored molecules resulted flavonoids, a group of molecule already known as AhR ligands. The same computational protocol was used also to rationalize binding to AhR of: agonists, e. g. the anti-inflammatory drug leflunomide [O'Donnell et al., 2010, Goodale et al., 2012] or ligands showing anticancer properties [O'Donnell et al., 2014]; antagonists, like GNF351 [Smith et al., 2011]; and SAhRMs [Murray and Perdew, 2011].

Afterwards, to more effectively capture the features of the binding site that are important for ligand recognition, our group set up new computational protocols, which made use of proteins co-crystallized with ligands as templates for homology modeling. This choice allowed to take into account the conformational rearrangements induced by a ligand ("induced-fit" effects, see section 1.2.3). The use of HIF2α bound to ligands [Scheuermann et al., 2009, Key et al., 2009] as templates to model the mAhR LBD structure improved the description of the binding cavity [Motto et al., 2011]. Moreover, the employment of the ensemble docking technique (section 1.2.3), based on docking to four different AhR modeled conformations, allowed us to take into account a certain degree of protein flexibility [Motto et al., 2011].

Flexibility was explicitly introduced performing MD simulations on the docking poses in a study aimed at identifying differences between AhR agonists and antagonists [Perkins et al., 2014] and, in

another application, to refine the poses obtained from docking of some agonists [Nuti et al., 2014]. Molecular modeling approaches were also used to shed light on the role of AhR in the induction of endotoxin tolerance [Bessede et al., 2014] and in the signaling between microbiota and host [Moura-Alves et al., 2014].

Promiscuous proteins are known to be flexible; for this reason, one of the main goals in treating systems like AhR is to include the protein flexibility into the computational protocol in order to better describe the process of protein-ligand binding.

The modeling support is essential also to rationalize the diversity in ligand binding among AhRs in the metazoan taxa. The reconstruction of the evolutionary relationships among the living organisms can give insights into the function and the role of AhR. AhR has existed for more than 600 million years, it is present both in vertebrate and invertebrate animals, but it is absent in plants. In living invertebrates, AhR plays a role in the development of sensory structures, including sensory neural systems; it could be supposed that these are the most ancient roles of metazoan AhR. The adaptive functions seem to be a vertebrate innovation, perhaps driven by a need to detoxify halogenated marine natural products, while invertebrates do not bind the typical environmental pollutants (PCDDs, PCDFs, PCBs, PAHs) [Hahn et al., 2017]. In fact AhR can also serve as a target for natural selection in presence of pollutants, e. g. atlantic killifish population in contaminated sites possesses AhR variants adapted to respond to complex toxicant mixtures [Reid et al., 2016]. Given that AhR can rapidly evolve under selective pressure, it can be suggested that the presence of AhR duplicates may relax evolutionary constraints on AhR function [Hahn, 2011]. For example four alleles have been described for the mouse AhR (mAhR). One of them, the AhRd, has a lower binding affinity for the TCDD; this difference is attributed to a single amino acid variation inside the ligand binding cavity [Poland et al., 1994].

Recently, using phylogenetic analyses and information from shared synteny, different genes of AhR in almost all metazoan (animal) taxa were identified and classified. AhR is referred to vertebrate orthologs of the AhR originally identified in mammals; AhR1 and AhR2 are paralogs derived from a tandem gene duplication; AhR3 is a novel AhR found originally in elasmobranchs (a subclass of cartilaginous fishes encompassing true sharks, skates, and rays); AhRR is distinct from AhR, AhR1, AhR2, and AhR3 and acts as a repressor [Hahn et al., 2017], its emergence was ascribed to one of the whole genome duplication occurring early in vertebrate evolution (∼510-600 MYA) [Hahn and Karchner, 2006].

The study of AhR in different species can help in reconstructing the evolutive history of this pleiotropic protein, thus giving insights on the physiological roles and ligands of AhR.

Computational modeling was used to rationalize species-specific differences in ligand binding affinity. For fishes, it was mainly studied an environmental model: *Danio rerio* (zebrafish). Comparison between the homology models of different isoforms of zebrafish AhRs (zfAhR1a and zfAhR2) allowed the identification of a reduced volume of the binding pocket in zfAhR1a, which does not bind TCDD, due to variation of key amino acids lining the cavity. Once the residues were changed as in zfAhR2 and mAhR, this mutant zfAhR1a bound TCDD [Fraccalvieri et al., 2013]. Moreover a single point mutation, corresponding to the single difference of the mouse AhRd allele, in sturgeon confers resistance to PCDDs, PCDFs and PCBs [Doering et al., 2015]. Amphibian AhRs have been investigated with the support of homology modeling, in both Xenopus laevis, [Odio et al., 2013] and *Ambystoma mexicanum* [Shoots et al., 2015]. For both these cases, three internal residues are responsible for the relative insensitivity of amphibian AhRs to TCDD. Also AhR in birds has been investigated with the help of molecular docking to homology models: three avian species were compared and ranked for PCDD, PCDF and PCB binding affinity, finding that chicken AhR is more sensitive compared to albatross and cormorant [Hirano et al., 2015, Kim et al., 2016]. To predict sensitivity of wild birds to PCDDs, PCDFs and PCBs, the structures of the AhR LBDs of pheasant and quail were modeled and investigated using mutagenesis [Farmahin et al., 2013].

2.2 METHODS

2.2.1 *Homology Modeling (Par. 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5)*

Due to the lack of structural information about the AhR-PASB it was necessary to exploit structures of homologous proteins belonging to the same superfamily bHLH-PAS to obtain a model of its three-dimensional structure.

As templates for modeling the AhR LBDs in different species (Par. 2.3.2, 2.3.3, 2.3.4), we used three different depositions of HIF2α (3f10:A, 3h7w:A, 3h82:A) in complex with ligands, as proposed in [Motto et al., 2011]. Given that *C. elegans* (ceAhR) is a distant homologous, we proposed different templates to describe as good as possible its LBD (Par. 2.3.1). The template structures were searched in the PDB using PSI-BLAST until convergence and only proteins belonging to bHLH-PAS class I proteins were selected. The final templates were: NPAS3 (5sy7 chain B), CLOCK (4f3l:B), HIF2α (3h82:A, 4zp4:B, 3f1n:A) and HIF1α (4h6j:A).

The model of mAhR to study binding to diverse ligands (Par. 2.3.5), is obtained by a new protocol with homology modeling exploiting also new deposition of HIF2α (PDB id: 3f1n:A, 3f10:A, 3f1p:A [Scheuermann et al., 2009]; 3h7w:A, 3h82:A [Key et al., 2009]; 4ghi:A [Scheuer-

mann et al., 2013]; 4gs9:A [Rogers et al., 2013]; 4xt2:C [Scheuermann et al., 2015]; 4zp4:B, 4zqd:B [Wu et al., 2015]).

The structures of the selected chains of the templates were downloaded from the PDB [Berman et al., 2000] and prepared using the Protein Preparation Wizard [Schrödinger LLC, 2016b] included in the Schrödinger suite. This tool adds the hydrogen atoms, optimizes the hydrogen bond network, minimizes the energy with the OPLS3 force field [Madhavi Sastry et al., 2013], removes atomic clashes and checks the most probable tautomer of histidines at a certain pH using Epik [Shelley et al., 2007]. All the atoms inside the cavities were kept (ligands for holo structures and water molecules able to form at least one hydrogen bond with the protein for apo structures).

Then, it was necessary to build an alignment between the template and the target protein.

We obtained the sequences of the target AhRs from our experimental partners or from UniProt in case of mouse AhRb-1 (mAhR, id: P30561) and human AhR (hAhR, id: P35869). To align all the sequences we used MUSCLE; the algorithm includes fast distance estimation using k-mer counting, progressive alignment using a profile function called the log-expectation score, and refinement using tree-dependent restricted partitioning [Edgar, 2004]. In case of need the alignment obtained by MUSCLE was manually curated.

The aligned sequences obtained by MUSCLE were given as input to MODELLER [Sali and Blundell, 1993], obtaining as output 100 models of the AhR (Par. 2.3.1, 2.3.2, 2.3.3, 2.3.4), and 500 for the mAhR (Par. 2.3.5). MODELLER is a comparative protein modeling method designed to find the most probable structure using spatial restraints derived from the alignment and expressed as probability density functions. Ligands and water molecules included in the template structures were maintained during the modeling procedure using the BLK function implemented in MODELLER. In Par. 2.3.5 we also used the loop-model routine to refine loops; this routine combines optimization steps and molecular dynamics with simulated annealing [Fiser et al., 2000].

The best models were identified using the DOPE score. Then we validated the models using both PROCHECK [Laskowski et al., 1993] and ProSA-web [Wiederstein and Sippl, 2007]. PROCHECK inspects the stereochemistry of a protein structure, thus giving an assessment of its overall quality; in ProSA, the energy plots and the quality scores of the modeled protein are displayed in the context of all known protein structures, highlighting potential problems. Models were then prepared using the Preparation Wizard tool.

To find internal cavities and calculate their volumes we used CASTp [Dundas et al., 2006], a web server which characterizes geometric and topological properties of protein structures using an analytical approach [Tian et al., 2018]. To determine the secondary structures of

a protein we used DSSPcont [Carter et al., 2003]. DSSPcont introduces a continuous assignment of secondary structure by predicting the hydrogen bond network; changing the hydrogen bond threshold you get different states. A DSSPcont assignment for a particular residue is a percentage likelihood of eight secondary structure states, derived from a weighted average of the ten DSSP assignments.

The obtained models of mAhR in Par. 2.3.5 were subjected to energy minimization with MacroModel [Schrödinger LLC, 2015b] (after removal of ligand and water molecules). In this last step, the sidechains of residues lining the cavity, identified by CASTp, were free to move, while their backbone and the rest of the protein were constrained by a force constant of 200 $kJ\,mol^{-1}\,Å^{-2}$. We set up a protocol that did not cause the collapse of the internal cavity and that tried to remove the initial bias due to the presence of the template ligand (or water molecules). The convergence is observed always after a maximum of 50 iterations.

2.2.2 *Molecular Docking (Par. 2.3.3, 2.3.4 and 2.3.5)*

The structures of the ligands were downloaded from PubChem and then prepared with the LigPrep utility [Schrödinger LLC, 2015a]. Their protonation states were determined with the Epik tool for pKa prediction included in Maestro [Greenwood et al., 2010, Shelley et al., 2007, Schrödinger LLC, 2016a], that is based on PROPKA as heuristic pKa calculator. Ligand structures were then optimized using MacroModel [Schrödinger LLC, 2015b] with the OPLS3 force field in implicit water.

In the case of three ligands with particularly complex conformational and electronic characteristics (Par. 2.3.4) we used Jaguar [Bochevarov et al., 2013] to perform conformational analysis using ab initio Quantum Mechanical (QM) calculations at the RHF/6-31G* level. The torsional angle was increased of 15° each step and solvation was implicitly included with the Poisson-Boltzmann (PB) model.

Performing docking to protein homology models may lead to accurate prediction of the binding poses depending on the reliability of the model. Homology models based on holo template structures are more able to take into account the induced-fit effects in the description of the binding region [McGovern and Shoichet, 2003]; they are particularly performing if the sequence identity calculated on the internal residues is high [Bordogna et al., 2010].

Docking calculations were performed using Glide XP with OPLS3 force field. Glide uses a series of hierarchical filters to search for possible locations of the ligand in the binding-site region including ligand flexibility. The properties of the protein are represented using grids built with different probes; the description of the site by these grids is progressively more accurate moving forward the hierarchy [Friesner et al., 2004]. Glide performs exhaustive sampling through

initial greedy positioning of the ligand and subsequent optimization; Glide XP adds an anchor-and-grow sampling of the ligand and finally include the explicit water description [Friesner et al., 2006]. Glide uses empirical scoring functions designed to maximize separation between high affinity ligands and low affinity ones. The scoring functions include empirically-based functions that account for different interactions (e. g. lipophilic- lipophilic, hydrogen-bond, metal-ligand terms) and also incorporate force field-based functions describing Coulomb and van der Waals contributions to the interaction energies [Halgren et al., 2004].

The grid was centered on the center of mass of three conserved internal amino acids: glutamine, histidine and threonine.

For mAhR, hAhR and rAhR (Par. 2.3.3 and 2.3.4) we clustered the output models from MODELLER on the basis of the backbone structural similarity using the Self Organizing Map (SOM) approach described in Fraccalvieri et al. [2011] and found four clusters; the model with the best DOPE for each cluster was selected as representative of a specific conformation of the domain. The four representatives were then used in the ensemble docking protocol as proposed in Motto et al. [2011].

2.2.3 *Post-docking rescoring (Par. 2.3.3 and 2.3.4)*

To rescore the docking poses we calculated the ΔG_{bind} using the MM-GBSA approach implemented in Prime-MMGBSA [Jacobson et al., 2002]. In this method, the ΔG_{bind} is obtained as the sum of energy associated with complex formation in the gas-phase, calculated with a MM force field, and the difference in solvation free energies between the complex and the unbound monomers.

The solvation terms are calculated as a sum of an electrostatic contribution (Generalized Born, GB) and a non-electrostatic contribution (Surface Area, SA) [Homeyer and Gohlke, 2012]. We used the Single Frame Protocol in Prime, i. e. both ligand and protein conformations were obtained from the optimized structure of the complex instead of performing distinct optimizations of the three different states (ligand, protein, and complex). Therefore the strains of both the ligand and the protein were neglected. We observed that the results obtained with this approximation are more consistent in an ensemble docking protocol, because considering the protein strain can lead to overestimate the contribution of the different protein conformations within the ensemble.

2.2.4 *Molecular Dynamics Simulations (Par. 2.3.5)*

Each selected docking pose was prepared for simulation using the tleap module of the AMBER14 package [Case et al., 2005, Salomon-Ferrer et al., 2013] and the ff14SB [Maier et al., 2015] force field with

TIP3P [Jorgensen et al., 1983] water placed up to 12 Å from the solute and neutralizing the system with 7 Cl^- ions. Parametrization of the ligands was performed using the Antechamber module of AMBER14, [Wang et al., 2006] using the Generalized Amber Force Field [Wang et al., 2004] (GAFF) to assign the atom-types and the AM1-BCC method [Jakalian et al., 2000, 2002] to assign charges. A prior multistage equilibration approach was used to remove unfavorable contacts and provide a reliable starting point for the simulations. The systems were subjected to 1000 steps of steepest descent energy minimization, followed by 1000 steps of conjugate gradient with restraint applied to backbone and ligand atoms (100 kcal mol^{-1} $Å^{-1}$). Subsequently, a 750 ps MD simulation was used to gradually heat the system from 0 K to 100 K in the NVT ensemble with backbone restraint lowered to 10 kcal mol^{-1} $Å^{-1}$ and from 100 K to 300 K in NPT ensemble with backbone restraint lowered to 2 kcal mol^{-1} $Å^{-1}$. Finally, the systems were equilibrated with a 1.0 ns NPT simulation mantaining the backbone restraint of 2 kcal mol^{-1} $Å^{-1}$. All the restraints were removed for the production runs. In all the stages, the temperature was controlled by the Langevin temperature equilibration scheme [Loncharich et al., 1992] with a collision frequency of 2.0 ps^{-1} and pressure targeted to 1 bar using a Berendsen barostat [Berendsen et al., 1984]. A time step of 2.0 fs was used, together with the SHAKE algorithm [Ryckaert et al., 1977] to constrain the bonds connecting the hydrogen atoms. The Particle Mesh Ewald method [Darden et al., 1993] was used to treat the long-range electrostatic interactions with the cutoff distances set to 9 Å. Production runs were carried out for 10 ns in the case of some ligands, and 20 ns in the case of other ligands, to allow a complete equilibration of the different poses.

MD trajectories were visually inspected using VMD [Humphrey et al., 1996]. Images were generated with Pymol [Schrödinger LLC, 2010].

2.2.5 *Binding Free Energy Calculations (Par. 2.3.5)*

The binding free energy (ΔG_{bind}) for complex formation was evaluated in implicit solvent by means of the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method [Kollman et al., 2000, Srinivasan et al., 1998] implemented in the AMBER software package. The ΔG_{bind} is determined omitting the entropic term and include an implicit solvent model. The polar solvation term was approximated with the Generalized Born (GB) model [Hawkins et al., 1996] using OBC rescaling of the effective Born radii [Onufriev et al., 2004]. The non-polar solvation term was calculated as the product of the surface tension parameter and the solvent accessible surface area (SA) evaluated using the Linear Combination of Pairwise Overlap (LCPO) algorithm [Weiser et al., 1999]. The single-trajectory approach was selected, i e.

the conformational ensemble was extracted from the single trajectory of the complex, instead of the three-trajectory approach (that uses the separate trajectories of the complex, receptor and ligand). The per-residue energy decomposition analysis was used to extract the contributions of single residues to the ΔG_{bind}, and those of residues within the binding site allowed identification of the hotspot residues for the different ligands. For this purpose, an ensemble of 500 conformations regularly sampled in the last 8 ns of the different simulations was used in calculations.

2.3 RESULTS

2.3.1 *Modeling of the Caenorhabditis elegans AhR LBD*

Although invertebrate proto-AhRs do not appear to bind typical vertebrate ligands and there is some evidence in supporting a constitutive and ligand-independent activity, it is also possible that some endogenous molecules can effectively bind to the proto-AhR of *C. elegans* (ceAhR) [Hahn et al., 2017]. Because invertebrate AhRs do not bind PCDDs, PCDFs, PCBs or PAHs, they do not properly function as "aryl hydrocarbon receptors"; anyway all the evidence currently available is consistent with the idea that invertebrate and vertebrate AhRs are orthologs, descending from the same gene in the most recent common bilaterian ancestor.

It was recently found that AhR promotes aging phenotypes in an evolutionarily conserved manner [Eckers et al., 2016]. Here it was shown that ceAhR loss or its mutation extends healthy life span in *C. elegans*. Moreover the treatment of *C. elegans* with curcumin also extend its lifespan and it seems that could act similarly to what the ahr-mutant does; in fact curcumin can not extend lifespan in the *C. elegans* AhR mutant (Dr. Ventura, personal communication).

For these reasons it is interesting to study the ceAhR-LBD and to determine while its activation could be mediated by direct ligand binding. We identify PASB, where the LBD is located, from two isoforms 1a and 1b (`http://www.worbase.com`). The two putative AhRs are identical in the PASB, but differ in total lenght; the ceAhR1b isoform is truncated in N-terminus, but it maintains the PASB. This form presumably could not bind DNA because it lacks a portion of the bHLH domain. The PASB sequence was then compared to previously studied vertebrate AhRs [Fraccalvieri et al., 2013, Shoots et al., 2015, Odio et al., 2013]; it becomes clear how invertebrate and vertebrate differ (Figure 2.2).

In the alignment were included some representatives of the main vertebrate families that have been earlier studied, and two invertebrates: *C. elegans* and *D. melanogaster* (Figure 2.2). The main peculiarity of invertebrates is a gap in the alignment in correspondence of the

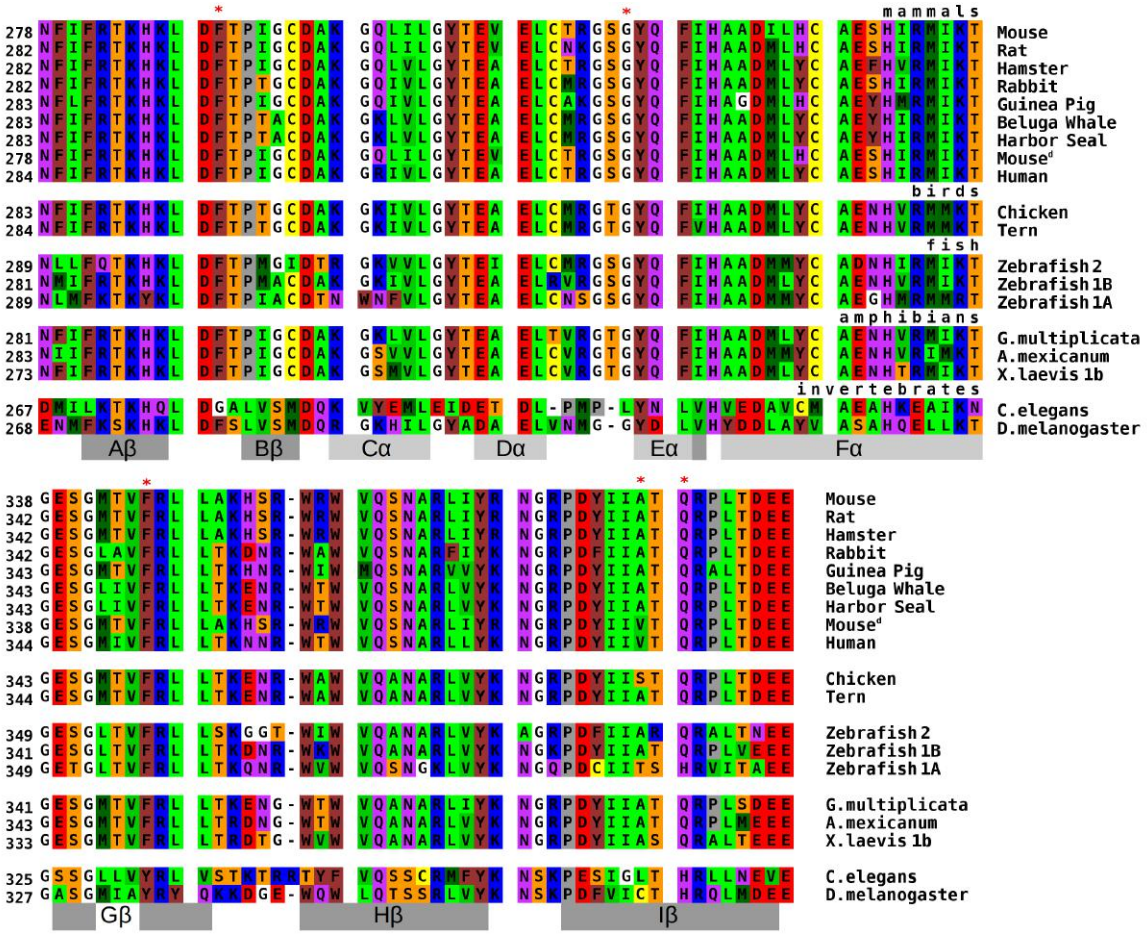

Figure 2.2 AhRs from different species are aligned and divided into mammals, birds, fish, amphibians and invertebrates. The red asterisks sign the candidate residues which may determine the non binding of the invertebrate AhRs.

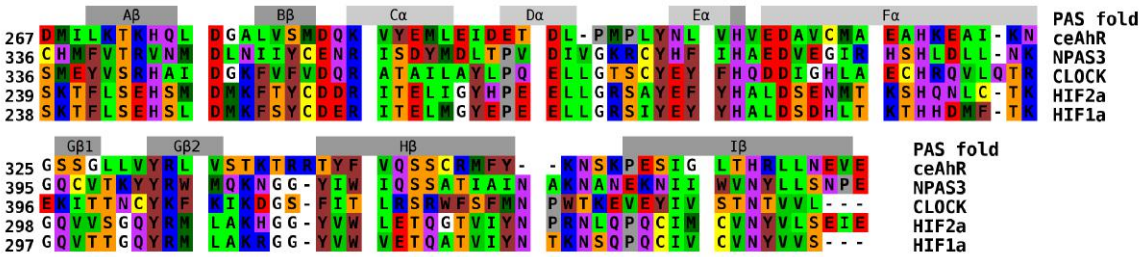

Figure 2.3 Target and templates alignment used for homology modeling. Aminoacids are colored depending on their physico-chemical properties. The first line represents the typical PAS fold with its nomenclature: dark gray are strands and light gray helices.

most disordered region of the PAS fold: the helical bundle embracing Dα, Eα and the short loops connecting these elements. It is interesting to note also that between ceAhR and *D. melanogaster* AhR (dmAhR) the sequence identity is not much higher than the one calculated between the ceAhR and mAhR. While ceAhR and dmAhR share 41% of identical residues (and 63% of similarity using the BLOSUM80 matrix), ceAhR and mAhR share 36% of identity (and 56% similarity). Among the vertebrates, the AhR-LBD is more conserved; between the mouse (*Mus musculus*) and the frog (*Xenopus laevis*) AhRs we observe an identity of 80% (and a similarity of 91%). This is compatible with the hypothesis that vertebrate AhRs evolved as a xenobiotic sensor, while in the invertebrate AhR assumes a large variety of functions.

A specific multi-template approach was used for modeling ceAhR, because it is very different from the previously modeled AhRs, for which we used only HIF2α (with about 30% of sequence identity) as template. The templates, chosen on the basis of their sequence identity with the ceAhR (about 20%), belong to the bHLH/PAS class I family: NPAS3, CLOCK, HIF2α and HIF1α. The alignment between the target ceAhR and the templates is shown in Figure 2.3. While vertebrate AhR are characterized by an insertion in the Dα/Eα loop, when aligned to the template HIF2α, ceAhR has a deletion in the same site.

The output model of ceAhR presents the typical PAS fold (Figure 2.4), with a shorter Dα compared to other AhRs because of the deletion already mentioned. The internal cavity is truncated in half by the hydrogen bond between H365 and H274, moreover the presence of the side-chains of Y332, L363 and L302 further reduces the internal volume of the cavity. We hypothesized that these residues are the most relevant to explain the difference of ceAhR compared to other AhRs. It is interesting also to point out that L363 corresponds to A375 in mAhR^{b-1} and V375 in mAhRd; this particular residue has a great impact on the ligand binding ability of AhR, in fact mAhR^{d} is known to not bind "classical" ligands like PCDDs, PCDFs, PCBs and PAHs. Two other differences noticeable differences in the ceAhR sequence are L302 and G278: L302 corresponds to a glycine, and the G278 to a

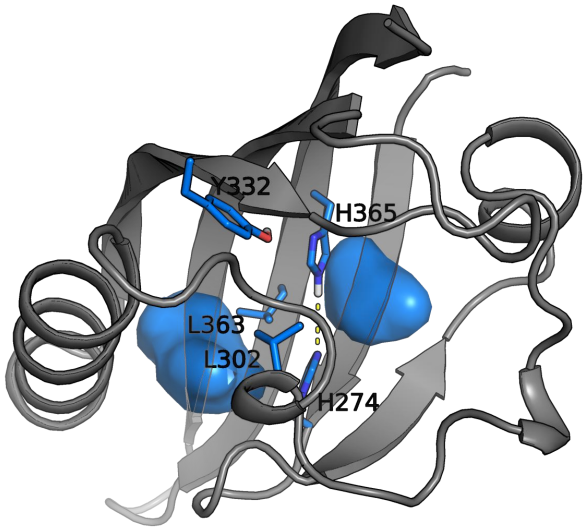

Figure 2.4 Three dimensional model of ceAhR LBD. The typical PAS fold is maintained, but the internal cavity is truncated in half by a hydrogen bond. The internal cavity is shown as blue surface and in stick are displayed the residues that most likely contribute to the uniqueness of this receptor.

phenylalanine in all other AhRs (Figure 2.2). These six residues are highlighted by a red asterisk in Figure 2.2, and are the main candidate to explain the different behavior of this receptor.

To date only one other AhR PASB domain has been described with an internal cavity truncated in half by internal sidechains: zebrafish AhR1a (*Danio rerio*).The zebrafish AhR1 (zfAhR1) was first identified in [Andreasen et al., 2002], and subsequently renamed AhR1a due to the identification of an other AhR1, called AhR1b [Karchner et al., 2005]. According to recent discoveries it is likely that AhR1a is the ortholog of mammalian AhRs [Hahn et al., 2017]. As already touched on pag 24, zfAhR1a was previously modeled [Fraccalvieri et al., 2013], and three residues were identified as a "barrier" truncating the cavity. The mutation of these residues to the corresponding ones in zfAhR2 restored the TCDD binding ability. In the case of ceAhR, the differences are presumably too many, so it is unlikely that this receptor could bind the classical AhR ligands. Anyway there are small and flexible ligands known to activate AhR, like leflunomide, which binds also zfAhR1a. We can therefore speculate that ceAhR could bind small and flexible ligands and activate transcription of target genes.

collaboration with Dr. Natascia Ventura, Institute of Clinical Chemistry and Laboratory Diagnostic Medical Faculty, Heinrich Heine University and the IUF-Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany

2.3.2 *Modeling of the Gymnopis multiplicata AhR LBD*

Amphibians are largely insensitive to the toxic effects of classical AhR ligands like PCDDs, PCDFs, PCBs and PAHs. This phenotype is driven by sequence and structural properties of the AhR-LBD. Fish, which evolved prior to amphibians, typically possess high-affinity AhRs and are sensitive to dioxin toxicity. Reptiles and birds, which arose after amphibians, also have high-affinity AhRs and high sensitivity to dioxin toxicity. Thus, the loss of dioxin-binding capacity common to amphibians AhRs seems to be lineage specific.

Amphibians are divided into three orders: Apoda (caecillians), Caudata (salamanders), and Anura (frogs and toads). The Apoda order is the first to diverge from the common amphibian lineage, after amphibian divergence from the rest of vertebrates but before the split between frogs and salamanders. Therefore, characterizing caecilian AhR structure, function, and ligand binding is crucial to determine when the low affinity binding phenotype evolved, and if it is a fingerprint of all amphibians, or it is proper of salamanders and frogs.

A representative of the Apoda order is *G. multiplicata*, a species of caecilian in the family Dermophiidae found in central America. Firstly it was necessary to identify the *G. multiplicata* AhR (gmAhR). This was made in the prof. Powell laboratoris isolating a 3811 base pair cDNA sequence from gmAhR, which includes an 843 residue open reading frame encoding a 92.7 kDa AhR. Using degenerate primers it was possible to determine a contiguous sequence of the gmAhR (Prof. Powell, Kenyon College, Gambier, OH (USA), private communication).

Once obtained the gmAhR sequence it was possible to build a homology model of the AhR-LBD; we built also the models of *A. mexicanum* (amAhR, salamander; order Urodela), *X. laevis 1β* (xlAhR, frog; order Anura), *G. gallus* (ggAhR, the common chicken), and *M. musculus* (mAhR) with the same protocol (described in section 2.2) for an appropriate comparison (Figure 2.5).

Only three residues with sidechains protruding into the modeled binding cavity of gmAhR differ from at least one of the sequences showing high affinity for the TCDD (mAhR and ggAhR): N333, A362, A378 (Fig. 11). These residues are identical to the corresponding residues of xlAhRs and amAhR that were already identified as responsible for low TCDD affinity [Odio et al., 2013, Shoots et al., 2015].

The EC_{50} value for TCDD deduced from the transactivation assays with gmAhR is consistent with low binding affinity (see Par. A.1.1 for experimental details). gmAhR had an EC_{50} value of 22.7 nM $\pm$ 0.2 nM, which is similar to the xlAhR1β and amAhR EC_{50} values of 24.2 nM $\pm$ 0.2 nM and 34.1 nM $\pm$ 0.1 nM, respectively. As a comparison, the EC_{50} of mAhR is 0.13 nM $\pm$ 0.2 nM, a value sharply lower meaning much higher affinity for TCDD. If it is true that usually the range of

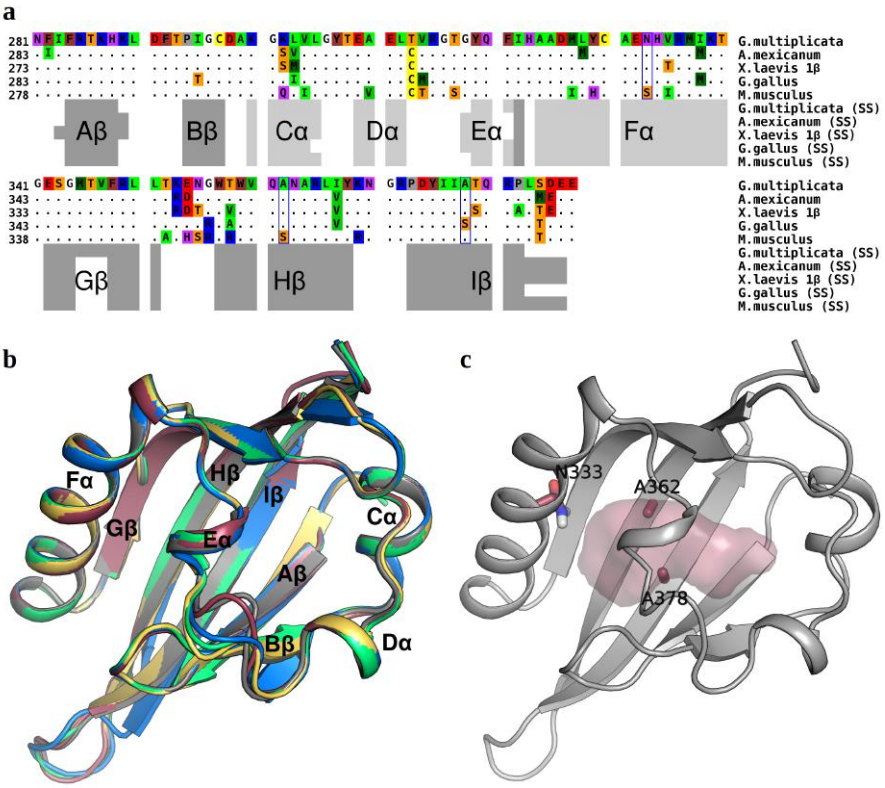

Figure 2.5 gmAhR LBD model in comparison with the models of other amphibian AhRs and of the high-affinity receptors of mouse (mAhR) and chicken (ggAhR). **a.** Aligned sequences of the gmAhR, amAhR, xlAhR, ggAhR and mAhR ligand binding domains. Only residues that differ from the *G. multiplicata* AhR sequence are shown, while dots indicate conservation. Blue boxes highlight the variable residues which sidechains protrude into the modeled binding cavity. The secondary-structures (SS) of each species are reported below the sequence: dark gray are β-sheets, and light gray helices. **b.** Comparison of cartoon rendering for modeled gmAhR (dark magenta), amAhR (green), xlAhR (yellow), ggAhR (blue) and mAhR (gray). **c.** Cartoon representation of gmAhR showing as shaded surface the internal ligand-binding cavity. The three residues that differ from both mAhR and ggAhR are shown as sticks and labeled.

nM affinity is considered good (for example many pharmaceutical drugs fall in this range), for AhR we know that frogs are insensitive to TCDD toxicity [Collier et al., 2008] and show an EC_{50} of approximately 20 nM ([Odio et al., 2013] and the present study). Moreover a comparison among the four different alleles of mAhR demonstrated that the equilibrium dissociation constants (K_D) ranged from 6 to 10 pM for AhR^b, whereas for the AhR^d allele the K_D was 37 pM, roughly 4-5 times larger, enough to greatly reduce the sensitivity to TCDD [Poland et al., 1994].

Taken together, these data predict that *G. multiplicata* is insensitive to dioxin toxicity. Moreover we can conclude that low-affinity AhR arose before the divergence of amphibians from the common lineage; this indicates that dioxin insensitivity is shared among all amphibians.

Collaboration with prof. Wade H. Powell, Kenyon College 202 N. College Rd.; Gambier, OH 43022 USA
Paper in preparation

2.3.3 *Comparison of ligand binding to the human and murine AhR LBDs*

A parallel analysis of the human AhR (hAhR) and the mouse AhR (mAhR) was conduced on two different type of ligands: the endogenous indirubin and the exogenous naphthoquinones.

Indirubin (IR) belongs to the class of indigoids that have been suggested to be endogenous AhR agonists. Indirubin is the most potent of this chemical class and is particularly intriguing to study because it shows different potency on mAhR and on hAhR. IR has been shown to have a 10-fold greater potency as an inducer of AhR-dependent gene expression than TCDD in human hepatoma cells, but it is ∼10-fold less potent than TCDD in mouse hepatoma cells. Additionally, IR was shown to be a more efficacious activator than TCDD of the hAhR *in vitro* and in cell culture [Faber et al., 2018].

Naphthoquinones are secondary metabolites of naphthalene, a ubiquitous air pollutant to which humans and animals are widely exposed via numerous anthropogenic sources. Naphthalene is a probable human carcinogen and it activates the CYPs leading to epoxides. Secondary metabolites are a mixture of different naphthoquinones (NQs), which result in a varied biochemical, pharmacological and toxicological effects [IARC, 2002]. 1,2-naphthoquinone (1,2-NQ) and 1,4-naphthoquinone (1,4-NQ) not only produce toxicity via redox cycling and oxidative stress, but they can modulate cell signaling pathways and gene expression [Abiko et al., 2016, 2015].

Indirubin binding

We hypothesized that the enhanced potency of IR as an agonist of the hAhR, compared to the mAhR, results from species-specific differences

in the interaction of IR with residues within the human and mouse AhR LBDs.

To examine this hypothesis, our experimental collaborators (Prof. Denison laboratory) first compared the ability of IR and TCDD to stimulate *in vitro* transformation and DNA binding of the mAhR and hAhR by gel retardation analysis. Actually IR can stimulate transformation and DNA binding of both *in vitro* synthesized mouse and human AhRs and is a more potent and efficacious activator than TCDD of the hAhR, similar to the gene expression results.

To confirm that the increased responsiveness of the hAhR is due to its LBD and not to other regions of the AhR, they examined the ability of TCDD and IR to stimulate DNA binding of a chimeric mAhR in which the mAhR-LBD has been replaced with the corresponding region of the hAhR-LBD.The relative potency of IR in the chimeric AhR was not dramatically different from that of TCDD, as observed with the hAhR, but it resulted still more potent (Figure A.1). These results suggest that the hAhR-LBD plays a role in ligand-selective activation by IR. However, since the chimeric AhR do not completely recapitulate the potency difference in IR activation of the hAhR, it is conceivable that additional regions of the hAhR play a role and/or that protein-protein interactions of the hAhR with ARNT or other partners contribute to enhancing IR-dependent hAhR transformation efficiency and DNA binding.

Given the highly conserved nature of the AhR across species among vertebrates and the role of the hAhR LBD in ligand-selective activation by IR, it is likely that amino acid differences within the hAhR are responsible for the enhanced activation by IR. Sequence analysis of the LBDs of the mouse and human AhRs identified 13 nonconserved residues (refer to Figure 2.2 for the alignment and to Figure 2.6 for the three dimensional model). Among these, only the mAhR A375 (V381 in the hAhR) has the sidechain directly pointing into the AhR binding cavity.

Sequence identity and similarity between hAhR and mAhR are high (87% and 93%, respectively). As can be seen in Figure 2.6, the physicochemical nature of the non-conserved residues is almost always conserved. For this reason to determine which of the different residues play a role in the enhanced response of the hAhR to IR, we mutated the residues in the mAhR-LBD to the corresponding hAhR-LBD residues (Figure A.2). The relative ability of increasing concentrations of IR and TCDD to stimulate transformation and DNA binding of the mutant mAhRs was determined by gel retardation analysis.

These experiments identified two mAhR to hAhR mutations (H326Y and A349T) that enhanced the efficacy of IR-dependent DNA binding of the mAhR by IR, and one mutation (A375V) that increased the relative potency, but not efficacy, of IR, such that IR was now equipotent to that

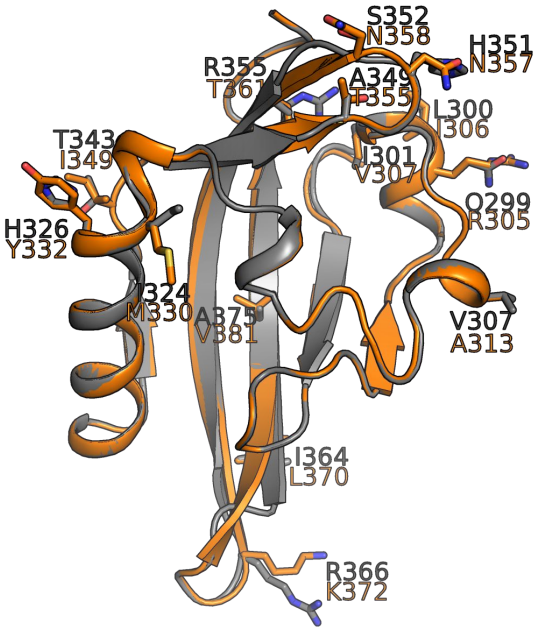

Figure 2.6 Comparison between mAhR (gray) and hAhR (orange) LBD models. The 13 residues that differ are shown as sticks and labeled. Only A375/V381 lines the binding cavity, while all the other differences are located outside.

of TCDD (Figure A.2). All the other mutations did not significantly alter IR activation from that of wild-type mAhR.

Mutation of histidine 326 to tyrosine (H326Y) and alanine 349 to threonine (A349T) enhanced the ability of IR to stimulate transformation of the mAhR into its high-affinity DNA binding form (efficacy) and increased the relative potency of IR for the mAhR to greater than that of TCDD. The A349T mutation also resulted in the greatest increase in the relative potency of IR as compared to TCDD and this appeared to result primarily from a significant decrease in the ability of TCDD to stimulate AhR DNA binding.

The enhancement of response of A349T to IR may be due to ligand binding ability or to other subsequent biological processes. To determine if A349T alters the relative binding ability of TCDD or IR prof. Denison research team used a competitive [^{3}H]TCDD ligand binding analysis. [^{3}H]TCDD binding to the A349T mutant mAhR was not significantly different from that of the wild-type mAhR, indicating that this hAhR specific mutation had no significant effect on overall [^{3}H]TCDD specific binding, confirming our molecular modeling prediction that A349 cannot be directly associated with ligand binding. While A349 is not involved in AhR ligand binding, the increased DNA binding response suggests that it plays a role in ligand-stimulated AhR transformation/DNA binding.

Despite the fact that IR maintains the characteristic planarity of TCDD and other dioxin-like AhR ligands, it shows different molecular

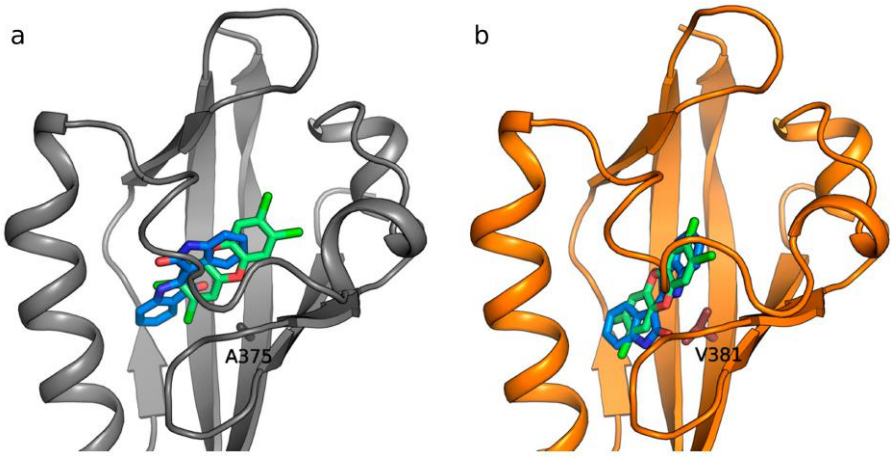

Figure 2.7 Docking poses of TCDD (green) and IR (blue) in **a.** mAhR and in **b.** hAhR. The only nonconserved residue is shown and labeled.

dimensions and shape, higher polarity, and has functional groups that may form hydrogen bonds. To better understand how these differences might modify the ligand binding pose within the LBD we predicted both TCDD and IR binding geometries in mAhR and hAhR.

Docking analysis revealed binding geometries of IR in the mAhR LBD distinctly different from that of TCDD (Figure 2.7a).

While TCDD showed stabilizing interactions with residues both in the middle and in the inner part of the cavity (Cα/Dα), the IR binding site was predicted to be closer to the entrance of the cavity and this resulted in different stabilizing interactions of IR with internal residues. Interestingly, docking to the hAhR revealed that the V381 residue, known to adversely affect the binding of TCDD, perturbs both the TCDD and IR binding poses causing a similar placement of the two ligands within the cavity (Figure 2.7b).

V381 residue does not hinder the binding of IR, but induces a rotation of the molecular plane. Taken together, these findings suggest that alanine or valine similarly affect the binding affinity of IR to the AhR in agreement with experimental evidence and suggest that IR may have a higher relative binding affinity to the hAhR compared to TCDD. This prediction is consistent with the [^{3}H]TCDD competitive binding results that suggest higher binding affinity of IR to the hAhR LBD.

The differential binding poses of IR and TCDD within the cavity also likely contribute, along with other sequence and structural differences in the AhRs, to the subsequent transformation events responsible for the enhanced potency and efficacy of IR.

Collaboration with prof. Michael S. Denison Department of Environmental Toxicology, University of California Davis, CA (USA)

This work on IR was published in Faber et al. [2018].

Naphthoquinones

To test the hypothesis that NQs interact with residues within the AhR ligand binding cavity differently compared to TCDD, we predicted their binding geometries using molecular modeling methods. Ligand docking to the mAhR LBD generated by homology modeling showed that both 1,2-NQ and 1,4-NQ fit within the binding cavity but with different arrangements and interactions compared to TCDD (Figure 2.8).

TCDD occupies the central part of the cavity and its chlorine atoms reach the most internal hydrophobic region (L302 and L309 residues). In contrast, both NQs are predicted to bind at the entrance of the cavity with the highest contribution to their stabilization provided by F289, M334 and M342 residues (Figure 2.8b). While it was demonstrated that F289 also contributes to TCDD binding to the mAhR, the two methionine residues lie far from the TCDD-binding region. NQ placements are similar in both species and the same residues (F295, M340, M348) contribute to the complex stabilization. However, higher NQ binding affinities for this receptor than for the mAhR have been determined, which could be due to the additional stabilization given by both hydrophobic interactions with V381 (corresponding to A375 in mAhR) and hydrogen-bonding of a carbonyl oxygen atom with S365 (Figure 2.8d).

To examine the role of each of the amino acids (F289, M334, and M342) that were predicted to be involved in NQ-selective AhR activation by molecular modeling, our experimental collaborators (Prof. Denison research team) determined the effect of their mutation to alanine on ligand-AhR-dependent gene transcription in COS-1 cells. COS-1 cells were transiently co-transfected with pGudLuc6.1, an AhR-responsive reporter gene plasmid and wild-type mAhR or mutant mAhR (F289A, M334A, or M342A) and induction by TCDD, 1,2-NQ, or 1,4-NQ determined (see A.1.1 and 2.9 for experimental details).

Interestingly, while F289A negatively affected ligand-inducible, AhR-dependent gene expression by TCDD and both NQs, only NQ-dependent gene expression was reduced by the M334A and M342A mutations (Figure 2.9). F289A was previously shown to negatively affect ligand-dependent AhR transformation and DNA binding by diverse ligands. The NQ-selective reduction in AhR-dependent gene expression by the M334A and M342A mutations is consistent with differential binding of TCDD and NQ within the AhR ligand binding cavity. Additionally, the greater reduction in activation by 1,4-NQ compared to 1,2-NQ (Figure 2.9) also suggests differences in their specific binding interactions within the binding cavity, consistent with the modeling results.

Naphthalene is a known environmental contaminant and probable human carcinogen that is activated by various CYPs in a tissue-specific manner to generate metabolites involved in redox cycling and oxidative stress. As secondary naphthalene metabolites, 1,2-NQ and 1,4-NQ have been shown to influence oxidative status in rodent models and

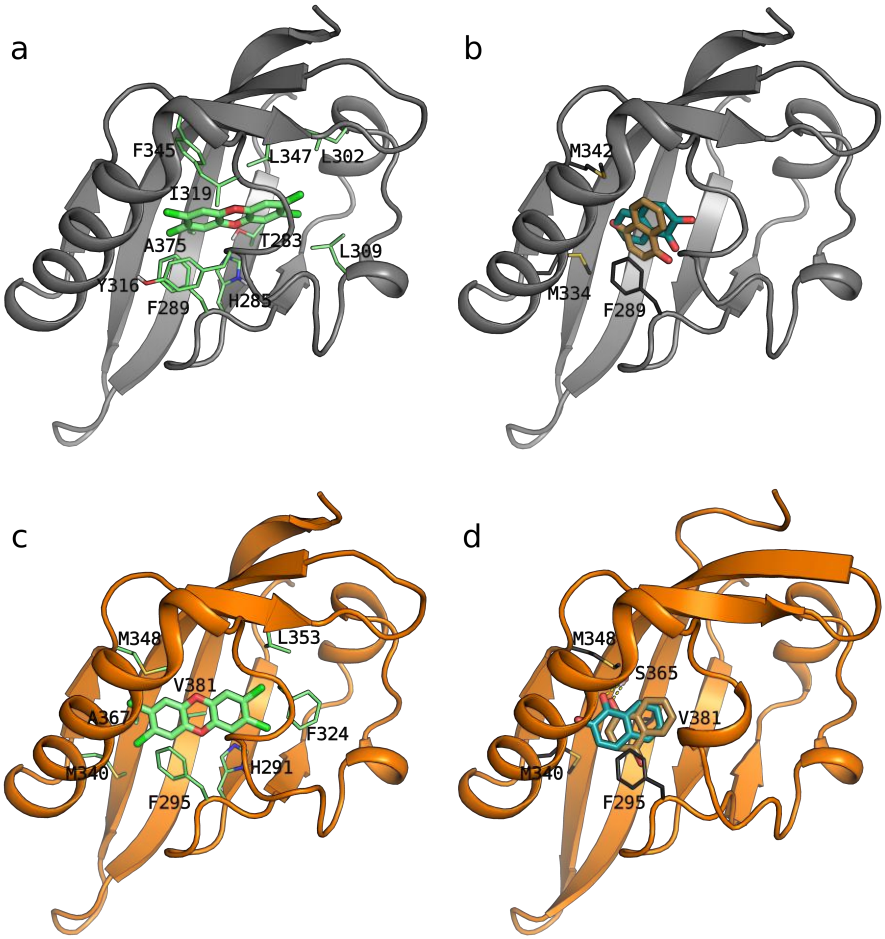

Figure 2.8 Predicted binding poses of TCDD, 1,2-NQ and 1,4-NQ in mAhR and hAhR LBDs. The models are shown as cartoons (mAhR in gray, hAhR in orange); ligands and residues that mainly contribute to ΔG_{bind} are shown as sticks. **a.** TCDD pose in the mAhR LBD, taken as reference: the key role of the residues shown has been previously validated by mutagenesis. **b.** NQs bind differently in the mAhR LBD compared to TCDD: the main interactions are with M334, M342, and F289. **c.** TCDD pose in the hAhR: due to the steric hindrance of V381 both the position and the laying plane of the molecule within the cavity are different compared to the pose in mAhR. **d.** The NQs poses in the hAhR are in the same binding site as in the mAhR LBD, but additional stabilization is provided by ligand interactions with V381 and S365.

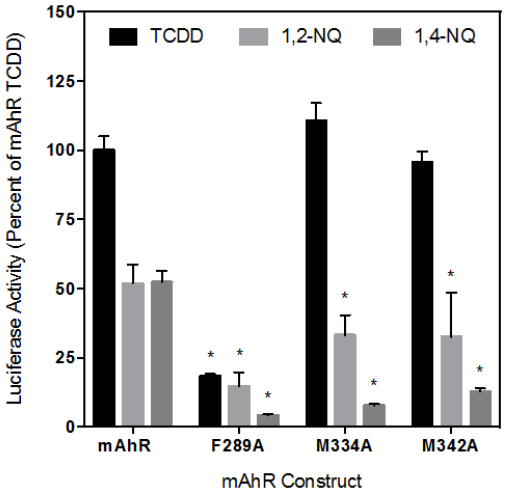

Figure 2.9 Key mAhR ligand binding domain residues are involved in stabilizing 1,2-NQ and 1,4-NQ within the pocket. COS-1 cells transiently transfected with wild-typfraccalvierie or mutant mAhR and XRE-containing reporter pGudLuc6.1 were treated with solvent control DMSO (0.1%, v/v), TCDD (10 nM), 1,2-NQ (5 μM), or 1,4-NQ (5 μM) for 18 to 20 h. Cells were lysed, and lysates were analyzed for firefly luciferase activity. Asterisks indicate the values that are significantly lower than the mAhR, as indicated by Two-Way ANOVA with $P < 0.05$. Luciferase activity (relative light units; RLU) was measured and corrected for background (DMSO) and normalized to mAhR TCDD levels. Values represent the mean $\pm SD$ of triplicates from three independent experiments.

cells in culture, and have been linked to toxicity. Their potential to activate the AhR signaling pathway *in vivo* could lead to increased production of reactive metabolites as a result of increasing expression of AhR-responsive CYPs (*e. g.* CYP1A1, CYP2A5) and other enzymes known to be important in the metabolism and activation of naphthalene. This study shed some lights on the possible mechanism of toxicity of these compounds combining the molecular modeling and a mutagenesis validation.

Collaboration with prof. Michael S. Denison Department of Environmental Toxicology, University of California Davis, CA (USA)
Submitted paper

2.3.4 *Comparison of binding of SAhRMs to the human and rat AhR LBDs*

Two novel SAhRMs in Sprague Dawley rats have been recently characterized: IMA-08401 (N-acetyl-N-phenyl-4-acetoxy-5-chloro-1,2-dihydro-1-methyl-2-oxo-quinoline-3-carboxamide; later referred to as C2 for simplicity; Figure 2.10) and IMA-07101 (N-acetyl-N-(4-trifluoromethylphenyl)-4-acetoxy-1,2-dihydro-5-methoxy-1-methyl-2-oxo-quinoline-3-carboxamide; later C4) [Mahiout et al., 2017]. C2 and C4 represent diacetyl prodrugs of N-hydrogen metabolites of the immunomodulatory drug compounds laquinimod and tasquinimod, of which laquinimod has been studied in phase II/III clinical trials for treatment of multiple sclerosis and Chron's disease, and is currently in phase II studies for treatment of Huntington's disease. *In vivo*, it was demonstrated that even at the highest doses practically achievable, C2 and C4 appeared considerably less toxic than TCDD (at a some 1000-fold lower single dose) after acute and subacute (repeated 5-day) dosing, while acting as effective activators of the AhR, as evaluated by *cyp1a1* gene induction [Mahiout et al., 2017].

Some major characteristic toxicities of dioxins that C2 and C4 lacked, at least at the dose levels tested, were hypercholesterolemia, reduced plasma thyroxine levels, acute lethality, wasting syndrome, grave liver and testis lesions, hypoglycaemia, and elevated plasma free fatty acid levels. The main adverse effects seen with high doses of C2 and C4 *in vivo* were thymic atrophy, alterations in serum triglyceride and 3-hydroxybutyrate levels, and changes in liver and kidney retinol and

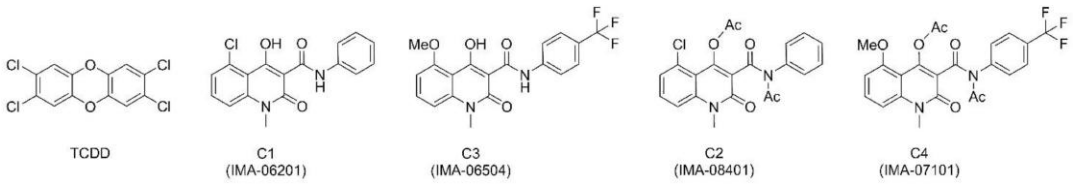

Figure 2.10 Chemical structures of TCDD and the novel SAhRMs C1 and C3.

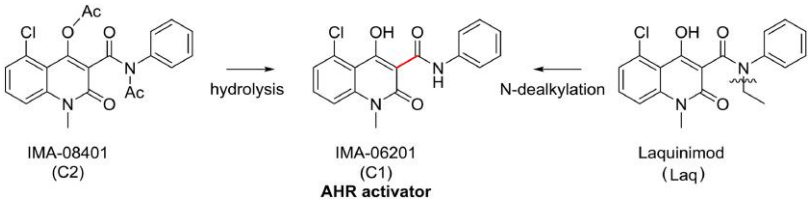

Figure 2.11 The SAhRM C1 is formed *in vivo* by hydrolysis of the diacetate prodrug C2, but also in small amounts from the drug compound laquinimod by N-dealkylation. C3 is similarly formed *in vivo* from the prodrug C4 and the drug compound tasquinimod.

retinyl palmitate concentrations. Furthermore, there were modulations in the expression of selected genes of the AhR battery. Intriguingly, all of the dioxin-like effects observed *in vivo* fell into the "type I" category previously demonstrated to be largely indifferent to structural variation at the transactivation domain of the AhR in TCDD-treated rats [Pohjanvirta et al., 2011].

To gain further understanding of the effects and potency of these compounds, it is presented here a computational study of binding to the rat AhR (rAhR) of the respective active deacetylated metabolites of C2 and C4, named C1 and C3 (Figure 2.10). For comparative purposes, the computational study of binding of the laquinimod (Laq) is also presented, despite the possible activation of AhR by Laq is still under study. To give more insights and to evaluate the impact of these compounds on human AhR, molecular modeling was carried out also on hAhR and the results are here compared to those on rAhR.

The chemical relationships between C2, Laq and C1 are depicted in Figure 2.11.

C1 and C3 are characterized by an extended electron conjugation. For this reason, we carried out conformational analysi susing *ab initio* Quantum Mechanical (QM) calculations at the RHF/6-31G* level. The minimum energy conformations were obtained by systematic search of the torsional angle highlighted in red for C1 in Figure 2.11 with complete geometry optimization at every step (scan at every 15°). Geometry optimization was carried out in water solution using a continuum solvent model.

For both C1 and C3 compounds, the absolute minimum corresponds to a value of 0° for this torsional angle and a planar geometry for most of the molecular structure, due to electron conjugation. Moreover there are two intra-molecular hydrogen-bonds stabilizing this conformation. Indeed, energy starts to increase while overcoming 30°, which is the common cut-off for H-bonding. In correspondence of 180° the steric clash between the two hydrogen atoms causes a maximum in the energy profile.

In the case of laquinimod, the global minimum corresponds to a value of 30° for the same torsional angle; this allows formation of

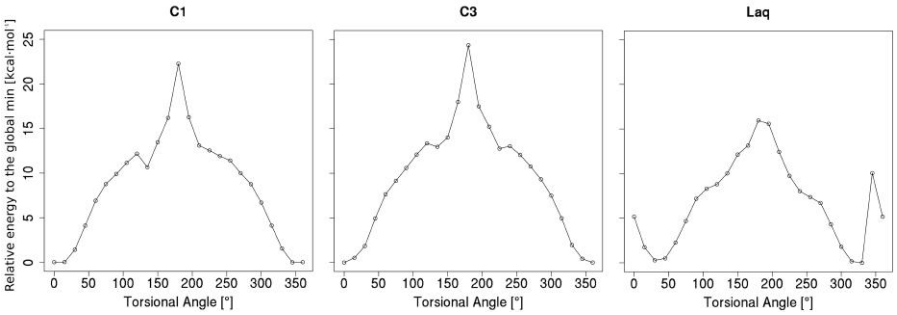

Figure 2.12 Systematic conformational search of the torsional angle C4-C3-C11-O for C1, C3 and Laq. The relative energy to the global minimum is shown.

one intra-molecular H-bond (the amidic carbonyl with the hydroxyl group) but prevents steric clashes of the ethyl group. We found a very good agreement with a previous QM analysis [Jansson et al., 2006]. On the basis of these results, to assure a correct treatment of these molecular structures, in the subsequent docking calculations geometric constraints derived from the QM analysis were imposed to the torsional angles analyzed and the QM net atomic charges were used in describing the ligand electron distribution.

The models of human and rat AhRs were obtained as described in section 2.2; for each species we selected four representative modeled conformations as input for the ensemble-docking protocol. It is interesting to point out that rAhR and mAhR LBDs are very similar, indeed they share the 97% of identity and 99% of similarity in the alignment (Figure 2.2). The ligand binding cavity of hAhR is smaller than the rAhR one, mainly because of the different internal residue (valine instead of alanine, that is present in both rAhR and mAhR, in Figure 2.6).

Docking calculation for these ligands was slightly modified to consider additional flexibility of the protein during binding simulations. To this aim we applied a scaling factor of 0.8 to the van der Waals radii of the protein atoms (soft docking). The final pose for each ligand was selected on the basis of the best ΔG_{bind} calculated using Prime-MMGBSA.

Docking to the rAhR LBD predicted that all the three ligands occupy the central region of the binding cavity (Figure 2.13). C1 and laquinimod lie on a plane very similar to that of the TCDD (Figure 2.7) and orient the main axis of the molecule in the same way, so that the phenyl groups are in the same position within the cavity. The methoxy and the three-fluoromethyl functional groups of C3 bring the molecule nearer to the helical connector, thus translating the rest of the molecule forward compared to the other ligands. It is also interesting to note that we found two poses of laquinimod in the rAhR (Figure 2.13c-d) with approximately the same binding free-energy but with opposite orientations in the cavity: one is very similar to that of

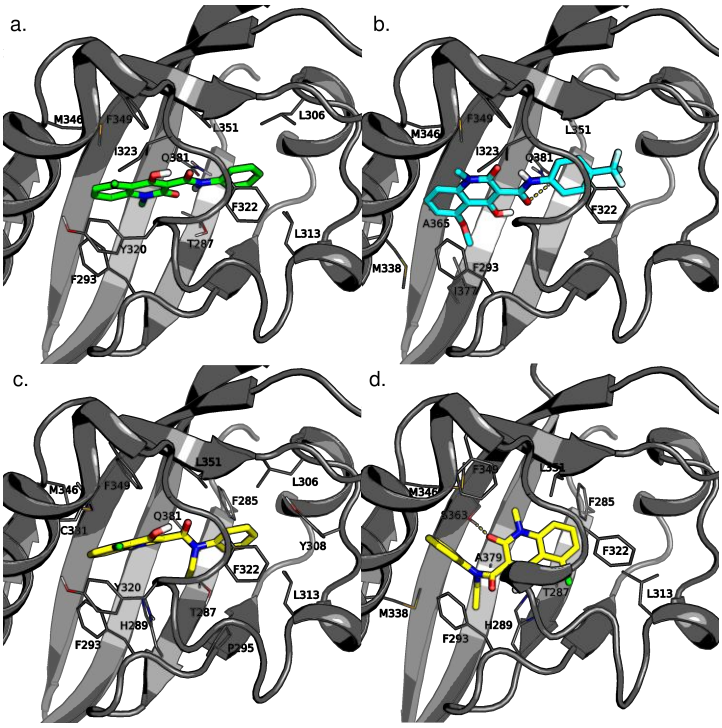

Figure 2.13 Docking poses of C1, C3 and Laq in rat AhR. **a.** C1 (green), **b.** C3 (cyan), **c.** and **d.** laquinimod (yellow). Residues that mainly contribute to the pose stabilization are shown and labeled.

C1 (Figure 2.13c), the other is nearer to the helical connector and is turned upside down (Figure 2.13d). The molecule adopts two different conformations (Z and E, described for the unbound molecule in Jansson et al. [2006]) in the two poses. It is conceivable that the binding cavity of rAhR is big enough to host the ligand in both the orientations and to allow conformational flexibility. The bound ligands maintain the intramolecular H-bonds of the unbound form, so their stabilization within the binding cavity is mainly due to hydrophobic interactions (van der Waals, and aromatic stacking) with residue sidechains.

Most of the residues that were found to be involved in the C1, C3, and Laq stabilization belong to the group of highly conserved residues found in the binding cavities of several mammalian AhRs with high affinity for the TCDD (including rAhR) that were proved to be necessary for optimal TCDD binding by mutagenesis and functional analysis studies ("TCDD binding-fingerprint") [Pandini et al., 2009]. Therefore these computational results support the hypothesis that these novel SAhRM metabolites effectively bind to the AhR and act as agonists.

Docking studies of the three ligands in the hAhR (Figure 2.14) predicted that C1 lies on the same plane in the AhR of both species (Figure 2.13a and Figure 2.14a), but the molecule is rotated of 180° around the main axis, so that in hAhR the chlorine atom points toward the β-sheets instead of toward the helical boundle (as observed in rAhR); however the polar groups of the ligand are maintained in the

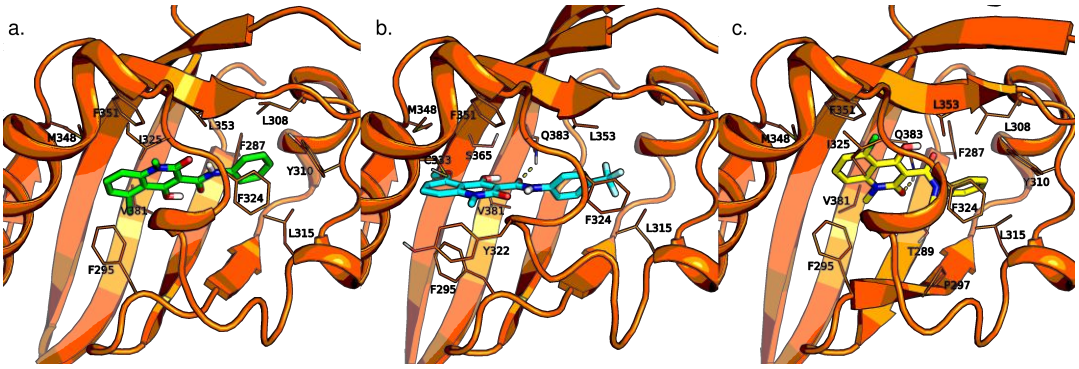

Figure 2.14 Docking poses of C1, C3 and Laq in human AhR. **a.** C1 (green), **b.** C3 (cyan), **c.** and **d.** laquinimod (yellow). Residues that mainly contribute to the pose stabilization are shown and labeled.

same position within the cavity, we can say the C1 binds in the same way in both the species. The shift of C3 toward the helical connector due to the three-fluoromethyl group, observed in the rAhR pose, is maintained in hAhR (Figure 2.13b and 2.14b). We also observed that the phenyl ring is rotated, thus slightly perturbing the planarity of the molecule.

In contrast, Laq adopts a more planar conformation in hAhR compared to the pose in rAhR, with the ethyl group pointing toward the helices and the phenyl ring toward the back of the cavity, similarly to what observed in one of the poses in rAhR (Figure 2.13c). While C1 and laquinimod interact with the residues at the bottom of the cavity (L308, Y310 and L315), C3 interacts more with the residues of the helical connector. Both C3 and Laq are predicted to form a H-bond with Q383 (yellow dashes in Figure 2.14b and Figure 2.14c). On the whole, the interactions with the residues internal to the binding cavity result very similar for the three ligands.

From the comparison of the binding poses, we concluded that C1, C3 and Laq could similarly bind the rat and human AhRs, because we observed the same type of interaction in both the AhRs. These similarities could be a starting point or a clue that these SAhRMs have similar effects on rats and humans.

Altogether, the results of computational and experimental studies on the rAhR [Mahiout et al., 2018] demonstrated that while C1 and C3 show negligible *in vitro* toxicity, they are potent and effective AhR agonists, and bind to the AhR in a manner similar to that of TCDD. Therefore these SAhRMs appear interesting candidates for therapeutic uses, and could also be valuable tools in further elucidating the multi-faceted physiological roles of the AhR, and the underlying molecular mechanisms.

Collaboration with prof. Raimo Pohjanvirta, Dept. of Food Hygiene & Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Helsinki

Part of this work was published in Mahiout et al. [2018]

2.3.5 *Modeling the binding of diverse ligands within the mAhR LBD*

Several docking studies coupled with experiments were performed during the last years (Par. 2.1 and Bonati et al. [2017]), but it is still lacking a systematic analysis of the binding modes associated with different classes of chemicals. The rationalization of common interaction patterns between ligands and binding cavity of AhR LBD would contribute to understand the differential modulation of AhR function.

We performed a comparative analysis of different ligand-binding modes obtained by applying molecular docking to a group of chemicals representative of the major classes of AhR ligands, with the aim of classifying ligands into groups according to their binding characteristics. For this purpose, we proposed a novel computational protocol for ligand docking to AhR homology models, specifically developed to take into account the receptor flexibility involved in binding process as well as to obtain a more comprehensive view of the key ligand-receptor interactions within the LBD. The residues that are predicted to play a critical role in binding of a group of ligands were then validated by site-directed mutagenesis and functional analysis followed by ligand binding analysis of the mutant AhRs.

A set of 10 representative ligands belonging to different chemical classes were selected for detailed analysis (Figure 2.15 and Table A.1). TCDD, TCDF and PCB126 belong to the class of HAHs, prototypical and high affinity AhR ligands. While TCDD and TCDF show elongated and planar structures, the shorter PCB126 molecule is characterized by rotation of the chlorinated rings around the central bond, that leads to a higher three-dimensional hindrance. BaP, 3MC and DBA are PAHs characterized by bulky planar structures with extended electron conjugation. BNF is a synthetic flavonoid, FICZ is a photoproduct of tryptophan, IR is a naturally-occurring indole-containing compound, and LEFL is an immuno-modulator. In contrast to the very hydrophobic HAHs and PAHs, these latter compounds contain several functional groups (carbonylic, aminic, heteroaromatic) that confer a certain degree of polarity to the molecules. All these compounds are known to be AhR agonists [Soshilov and Denison, 2014].

To develop homology models of the AhR LBD useful to study binding of ligands with a wide range of diverse structural and physicochemical characteristics, they must take into account the flexibility and plasticity of the domain, and thus cannot be described with a single structural model. To this end, we built homology models of the LBD of mAhR using ten different HIF-2α depositions (PDB id: 3f1n:A, 3f10:A, 3f1p:A [Scheuermann et al., 2009]; 3h7w:A, 3h82:A [Key et al., 2009]; 4ghi:A [Scheuermann et al., 2013]; 4gs9:A [Rogers et al., 2013]; 4xt2:C [Scheuermann et al., 2015]; 4zp4:B, 4zqd:B [Wu et al., 2015]) and

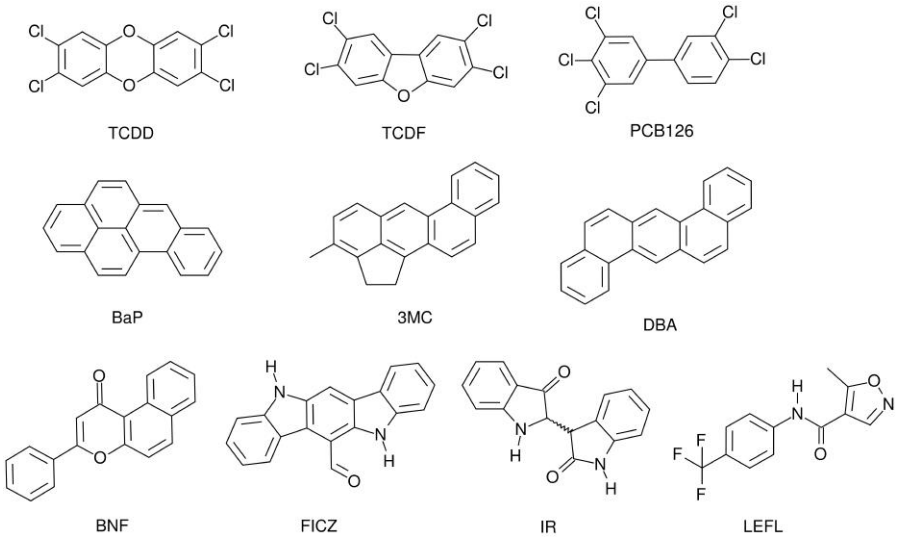

Figure 2.15 2D representation of the set of ligands analyzed in this study. On top: HAHs; in the middle: PAHs; on bottom: polar ligands.

these were employed in a multi-template ensemble docking approach to study ligand binding. During homology modeling, ligands and water molecules present in each HIF-2α PASB structure (Table A.3) were retained inside the binding cavity in order to take into account the induced fit effects of the different compounds [Fan et al., 2012]. Only water molecules falling directly inside the ligand cavity and displayed hydrogen bonds with the protein.

The sequence of the mAhR LBD (Uniprot id: P30561, residues 276-384) was aligned to the human and mouse HIF-2α PASB sequences (Uniprot id: Q99814 and P97481, residues: 239-349), that are identical in the PASB region. Sequence identity of mAhR with the template is 31% and similarity 52% (BLOSUM62), with only two indels (one insertion and one deletion) in the alignment (Figure A.5).

For each template, we generated 500 putative models using MODELLER and selected one with the best DOPE score. Indels were further refined using the loop-modeling routine, obtaining 500 loop conformations per model and again selecting the best one by DOPE score (Table A.4). The models obtained were validated using PROCHECK and ProSa and all were identified as good-quality models (Table A.5). For the 4ghi template, two good models with comparable DOPE scores were identified and both were included in the analysis.

The structures of the eleven models obtained are very similar in their backbone geometry (Figure 2.16), confirming the highly conserved 3D structure of the templates used. Also the backbone of the refined loops tends to converge into a similar structure in all the models (Figure 2.16), and does not affect the cavity volume. What makes each model different from the others are the sidechain conformations. In fact, the binding cavity volumes, a good measure of the

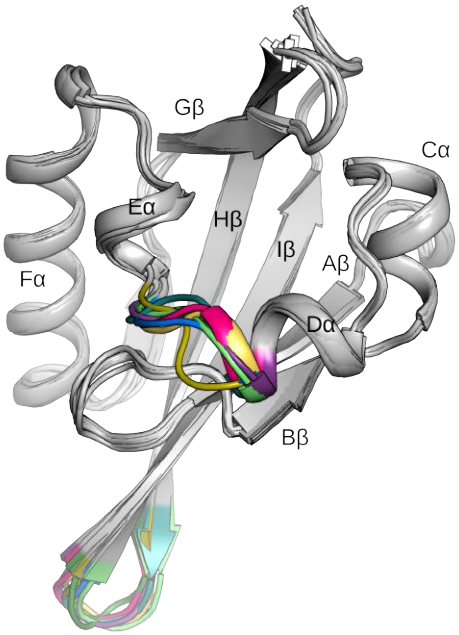

Figure 2.16 Cartoon representation of all the mAhR PASB models superimposed. The region refined by loop modeling is colored according to the model. Secondary structure elements are labeled according to the PAS domain nomenclature [Möglich et al., 2009].

conformational characteristics of the residues inside the cavity, present very different values (Table A.4) and thus the modeled cavities show a wide spectrum of sizes and shapes (Figure 2.17). Part of the variability is due to the different ligands present in the template structures and maintained during the modeling step: the smallest cavities derive from the template apo structures (4zp4 and 3f1p) and the largest one from the holo structure with the bulkiest co-crystallized ligand (4xt2).

Before performing molecular docking in the prepared mAhR models, the structures of the test ligands in Figure 2.15 were prepared. The analysis performed with Epik led to identification of the most probable conformers for all the molecules in water solution at pH 7. Three possible forms were predicted for IR: trans, cis, and a charged form (Figure A.6), and all of them were submitted to docking calculations. For each of the other ligands, a single conformer was identified and analyzed by docking.

To take into account the LBD flexibility involved in binding, we selected the ensemble-docking approach, that consists of repeating ligand docking to an ensemble of different receptor conformations [Totrov and Abagyan, 2008, Fan et al., 2009]. We previously used this approach to study TCDD binding to the AhR, but in that case only a limited conformational variability of the LBD was considered [Motto et al., 2011]. In order to study binding of ligands that have a variety of structural and chemical characteristics, we extended this approach by including a larger set of different conformations of the

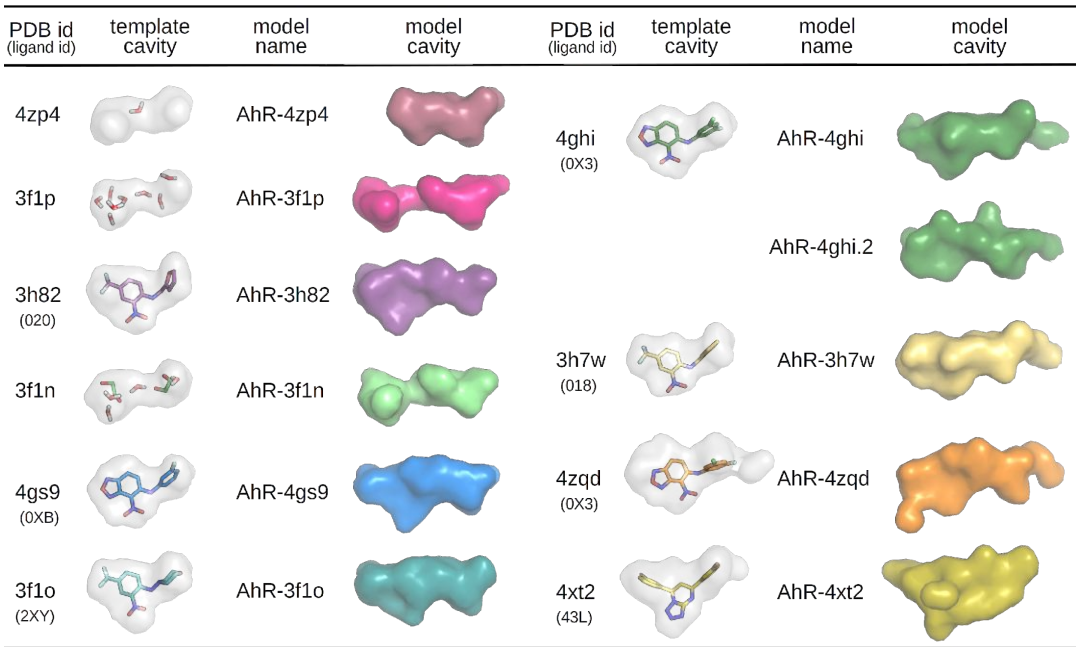

Figure 2.17 Internal surfaces of the binding cavities for the HIF-2α templates (gray and transparent with the ligand inside) and for the mAhR models (colored and solid). Models are ordered by increasing cavity volumes (Table A.4).

AhR LBD. This was obtained by using the eleven homology models derived from different experimental template structures, including both apo and holo forms with different ligands in the binding cavities. Ensemble-docking results for the set of ligands investigated here are recapitulated in Figure A.3, that presents the poses obtained for each ligand in all the models as a matrix.

All the ligands bind in the modeled cavity, consistent with the experimental data that confirms their competitive binding with TCDD. In fact, previous experimental ligand binding analysis revealed that all ten ligands could effectively compete with [^{3}H]TCDD for binding to the mAhR (see section A.1.1 at page 89 for the details), and that these chemicals are relatively potent AhR agonists, and that exists a relationship between ligand binding affinity and potency to stimulate AhR transformation/DNA binding (Figure A.4).

For all the ligands, we obtained several alternative docking poses depending on the size, shape and conformational characteristics of the receptor model used. This confirms the importance of the ensemble-docking approach to analyze alternative ligand binding possibilities allowed by receptor flexibility. As we expected, we observed that as the cavity volume increases, the less selective the model becomes. In fact, we obtained docking poses for all the ligands into the largest cavity (AhR-4xt2). In contrast, there is one model that did not produce any docking poses for any ligands (AhR-3f1p), and three models that gave poses only for the smallest ligand, LEFL; these models were derived

either from apo template structures or have a very small binding cavity (AhR-3h82). However, the size of the binding cavity by itself is not sufficient to explain the results reported in the matrix. In fact, we obtained few poses for rigid ligands, regardless of the cavity size. Therefore, also shape complementarity between ligand and cavity, that is related to the arrangement of internal sidechains, has a role in determining the ligand binding ability.

We visually inspected and manually selected two representative poses for each ligand (red rounds in Figure A.3) suitable to summarize its alternative binding geometries in the most optimal way. The selected poses are shown in Figure A.7. For IR, only two representative poses obtained for the IR-trans tautomeric form are shown here because the subsequent ΔG_{bind} calculation demonstrated that the IR-trans form is the most stable inside the cavity. It can be observed that, in some cases (*e. g.* TCDD, TCDF and DBA), the poses of the same ligand differ depending on a translation of the molecule inside the cavity and, in other cases (e. g.
PCB126, BaP, BNF and LEFL) depending on the ligand rotation of 180° around the minor molecular axis, that inverts the molecular "head" and "tail".

Given that the obtained docking scores (XP GlideScore, in Table A.6) were not able to clearly discriminate among the alternative poses of the same ligand on the basis of the approximate evaluation of their stability, the two representative poses were further analyzed by short MD simulations. We used MD as a short refinement of the poses obtained by docking and to evaluate if this approach could overcome the encountered limitations of docking scores.

This was accomplished by calculating the ΔG_{bind} values on the basis of the MD trajectory to analyze the dynamic behavior of the poses into the cavity and verify possible convergence of similar poses toward a unique geometry, as well as to find the most stable pose, in cases of distinctly different poses and similar docking scores.

While we did not observe the convergence of the MD simulations of the two representative poses into a single pose for most of the ligands, for TCDD, BaP and IR-trans the poses sampled during the two simulations tended to overlap. Moreover, different behaviors occurred during the simulations. Some ligands (TCDF, BNF, LEFL) remain quite close to the docked conformation, while others (BaP, IR, DBA) rapidly modify their initial conformation and then remain stable for the rest of the time; for example, BaP moves away from the docking pose, reaching the bottom of the cavity in both the simulations. Finally, other ligands move away from the initial binding geometry but then come back to the original pose. An example is a docking pose of TCDD at the center of the cavity (TCDD-3h7w in Figure A.7) that reaches 4 Å RMSD during the MD simulation while it is located at the Bβ, Cα, Dα

site (at the bottom of the cavity, see Figure 2.16) and then comes back to 2 Å RMSD, nearer to the initial region (green graph in Figure A.8).

A possible explanation of the changes introduced by MD simulations in the docking poses is that the modeled AhR conformations used did not completely describe the binding site flexibility (*e. g.* they prevented the AhR ligands from reaching the bottom of the cavity) as a consequence of the different binding regions and related induced fit effects of the template ligands. In fact, HIF-2α shows a PASB cavity smaller than the AhR one in both the apo and holo structures (Table S4) and its ligands bind mostly in the central part or near to the entrance of the cavity (at the Fα/Gβ site). Therefore, the final MD refinement, that allows for local deformations induced by the AhR ligands in the binding cavity, was needed to reach a more complete description of binding.

To identify the most stable pose between the two selected, we calculated the binding free energy (ΔG_{bind}) with the MM-GBSA method. In particular, for the comparison we used the average value of ΔG_{bind} in the last 8 ns of the simulations (Table A.7), which represent a stable portion of simulation for all the complexes (Figure A.8). The pose with the most favorable ΔG_{bind} was further analyzed to rationalize the binding mode.

MD simulations gave us also the opportunity to obtain a dynamic view of ligand behavior inside the cavity, thus improving the description of the binding modes and the identification of all the interacting residues.

Our previous docking studies predicted a TCDD binding mode inside the mAhR cavity that was validated by extensive mutagenesis studies [Pandini et al., 2007, 2009, Motto et al., 2011]. With the MD simulations here performed we obtained a new dynamic representation of TCDD binding, that we can compare to the previous static view [Motto et al., 2011] (Figure 2.18). During MD, TCDD moves inside the cavity by translating from one side (the inner part) to the other (the center of the cavity) and samples also the previously determined pose, lying in an average position compared to those obtained by MD.

The consistency of the TCDD results encouraged us to further characterize the poses obtained with the new "dynamic" protocol for all the diverse ligands, to obtain a comprehensive description of all the protein-ligand interactions. To this aim, we analyzed the different profiles obtained by per-residue decomposition of the ΔG_{bind} for each ligand pose (Figure 2.19a). It was possible to identify a set of residues that are present in all the ligand energetic profiles, most of them belonging to the TCDD-binding fingerprint originally published [Pandini et al., 2009]. In the present study we especially highlight the importance of T283, F289, F318, I319, F345, L347 and Q377 in binding of almost all the ligands under investigation. Moreover, we found addi-

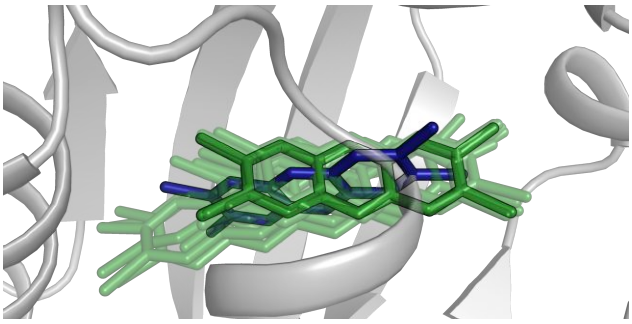

Figure 2.18 Dynamic view of a TCDD pose inside the binding cavity; ligand is shown as sticks, protein as gray cartoons. In blue, the pose previously obtained by docking [Motto et al., 2011]; in dark green, one of the docking poses obtained in this work; in green, 10 snapshots taken from the last 8 ns of the MD simulation.

tional interactions that characterize binding of each ligand and these were used to gather ligands into three different groups (Figure 2.19).

The first group includes TCDD, TCDF and BaP, that bind at the bottom of the cavity (Bβ, Cα, Dα region) and are stabilized by hydrophobic interactions. In the second group, 3MC, PCB126, and DBA bind nearer to the Fα/Gβ site and also have hydrophobic interactions. Finally, ligands of the third group (BNF, FICZ, IR, LEFL) lie in the same site as group 2, but establish hydrogen-bonds and polar interactions with several residues (Figure 2.19).

In particular, ligands in group 1 are stabilized both by hydrophobic interactions common to all the ligands of this study, in the central part of the cavity (F289, F318, I319, F345 and L347), and by others typical of this group with residues at the bottom of the cavity: L302, L309 and P291. In fact, as already pointed out, TCDD moves inside the cavity (Figure 2.18 and Figure 2.19a) interacting with residues near to the Cα and Dα (L302 and L309) and near to the Aβ and Bβ (F289 and P291). Similar movements are observed for TCDF (Figure 2.19b), that however mostly occupies the Cα, Dα site, as in the original docking pose, and BaP (Figure 2.19c).

Ligands in group 2 are very stable in the Fα/Gβ site, except for PCB126 that moves also at the center of the cavity, where it assumes planar conformations (Figure 2.19e). The two residues that can better describe these binding modes are M334 (on Fα) and M342 (on Gβ). As evident in Figure 2.19a, all the per-residue ΔG_{bind} profiles of group 2 have few high bars related to key interactions, and stabilization can be attributed to many residues in the whole cavity (except in the zone at the bottom of the cavity). In fact, 3MC and DBA are bulky ligands and tend to occupy almost the entire volume of the cavity, and PCB126 can contact many residues due to its conformational flexibility around the central bond and the hindrance of five chlorine atoms. It is also interesting the role of H331, because during the simulations this histidine sidechain flips inside the cavity giving stabilizing interactions

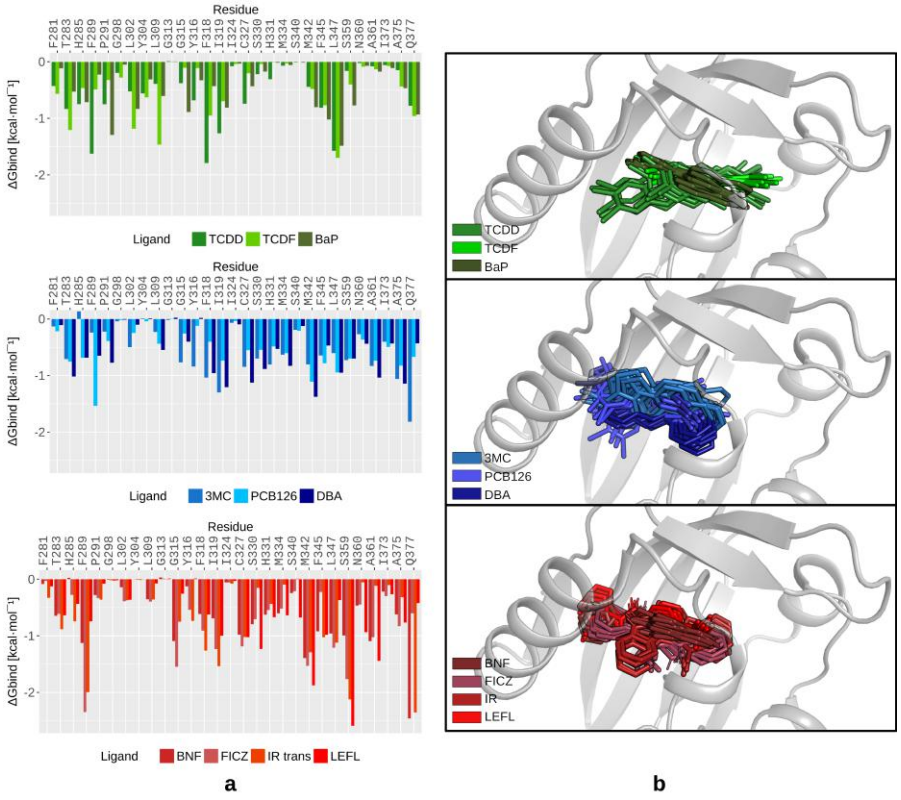

Figure 2.19 Ligands are gathered into three groups defined by the occupancy of different sites inside the AhR cavity and by characteristic residue interactions during MD simulations. **a.** Per-residue decomposition of ΔG_{bind}. Values are obtained as the averages on the last 8 ns of simulation. In the plots, only residues lining the internal cavity are shown. **b.** Ten snapshots sampled during the last 8 ns of simulation are shown for each ligand belonging to the three groups; ligands are represented as sticks.

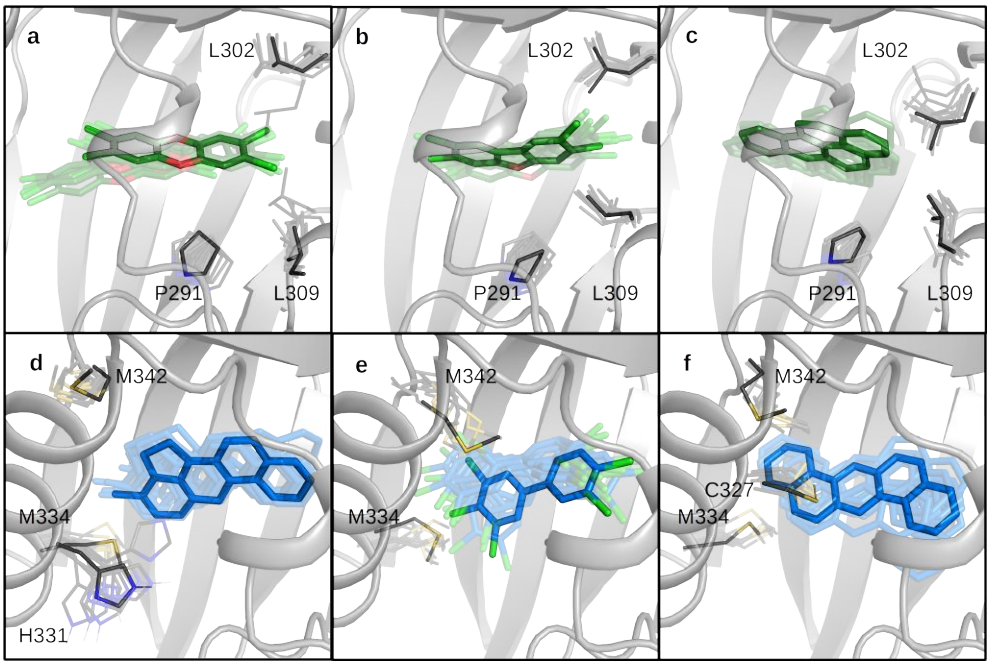

Figure 2.20 Dynamic view of the binding poses of ligands in group 1, **a.** TCDD, **b.** TCDF, **c.** BaP and group 2, **d.** 3MC, **e.** PCB126, **f.** DBA. Ligands are represented as sticks and 10 snapshots extracted during the last 8 ns of MD simulation are shown in transparency; solid sticks indicate the most sampled poses in the MD simulations. The most relevant residues identified by per-residue decomposition of ΔG_{bind} are shown as lines and labeled.

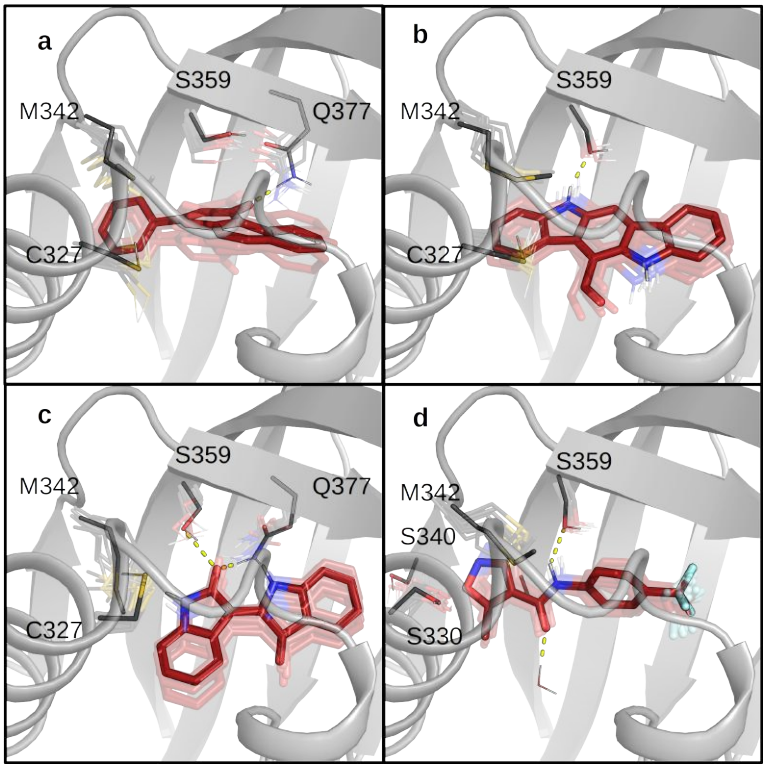

Figure 2.21 Dynamic view of the binding poses of ligands in group 3, **a.** BNF, **b.** FICZ, **c.** IR, **d.** LEFL. Ligands are represented as sticks and 10 snapshots extracted during the last 8 ns of the MD simulation are shown in transparency; solid sticks are the most sampled poses in the MD simulations. The most relevant residues identified by per-residue decomposition of ΔG_{bind} are shown as lines and labeled.

with the ligand (e. g.
with 3MC, Figure 2.20d), or outside, exposed to the solvent.

Group 3 is the most heterogeneous, and like group 2, they occupy the Fα/Gβ site where ligands are stabilized by interactions with some residues belonging to the Fα (C327 or S330) and Gβ (M342). Moreover, the peculiarity of this group is given by the polar interactions and hydrogen-bonds established by ligands with two internal sidechains: Q377 and S359. Given that the mAhR cavity is hydrophobic, polar residues (S359, Q377, T283, H285) tend to interact with each other or with water molecules by hydrogen bonding. Some ligands are able to interfere with this interaction chain (BNF and IR in Figure 2.21a and Figure 2.21c); in contrast, others act as hydrogen-bond donor toward S359 while preserving the H-bond between S359 and Q377 (FICZ and LEFL, Figure 2.21b and 2.21d). For LEFL we also observed a water-bridged H-bond between the oxygen atom of its amidic group and the hydroxyl of Y316.

To obtain an experimental validation of the different binding modes predicted by our analysis for ligands with different characteristics, for each group we selected a small set of residues to be tested in mutage-

nesis studies. Non-selective residues that give stabilizing interactions with all the ligands under study were not included in this validation. Therefore, for group 1 we selected L302, L309 and P291; for group 2, M334 and M342 (in addition H331 for 3MC, and C327 for DBA), and for group 3, C327, M342 and S359 (in addition S330 for LEFL).

Mutagenesis is often used to validate docking results, but it is also important to note that the mutation of some sidechain could change receptor conformation or leads to unexpected interactions, that are absent in the wild type. To be as consist as possible prof. Denison laboratories performed both the DNA and the ligand binding assay. We obtain a rough proof that the mutant receptor maintains its conformation by the fact that it still binds the XRE. Using a competitive assay for ligand binding we verify that no other binding sites are explored by the ligands under investigation, neither with mutant receptors.

The ligand-dependent transformation/DNA binding demonstrated that six out of the eight amino acid mutations (P291L, C327A, H331A, M334A, M342A, and S359A) resulted in ligand-dependent AhR:ARNT:XRE specific complex formation greater than 50% of wt/mAhR activated by TCDD. The ligand dependency demonstrates that the mutant AhRs are not constitutively induced and that probably the mutations do not alter the receptor conformation. Two mutations (L302A and L309A) eliminated ligand-stimulated AhR DNA binding or resulted in less than 25% TCDD-induced AhR. Therefore L302A and L309A could not be used for the subsequent competitive binding analysis.

To assess the influence of the remaining six residues in binding diverse ligands within the AhR ligand binding pocket, [^{3}H]TCDD competitive ligand binding was carried out with increasing concentrations of each ligand and their relative affinity (IC_{50}) calculated from the competitive binding curves (Table A.8 and Figure 2.22). Interestingly, ligand binding analysis revealed that P291L substitution dramatically enhanced the relative affinity of TCDF for the AhR (compare 20 nM for wild-type AhR (Table A.2) to 0.04 nM for the P291L AhR (Table A.8)), but while the mutation suggests an increase in BaP binding, the result was not statistically significant. The relative binding affinity of PCB126, IR and 3MC were reduced with the M342A substitution, but the relative affinity of LEFL and FICZ were significantly increased. In contrast, DBA and BNF were not affected by the M342A mutation. The M334A substitution significantly reduced the relative affinity of PCB126 and decreased 3MC binding, but had no significant effect on the binding of DBA. Interestingly, while the H331A mutation dramatically increased the relative affinity of 3MC for the AhR, S330A had no significant effect on LEFL AhR binding. The results with the S359A substituted AhR were similar to that of M342A in that it significantly increased the relative binding affinity by one ligand (BNF) and suggests a decrease in the binding of another (FICZ). Similarly, the C327A mutation significantly

decreased BNF, but had no significant effect on that of IR, DBA or FICZ. Overall, the results of these mutational analysis reveal significant differences in ligand-specific, amino acid-dependent binding to the AhR.

Comparison between computational and experimental results allowed us to analyze in detail each group of ligand.

Group 1 (formed by TCDD, TCDF and BaP) is stabilized within the AhR LBD mainly by several interactions with residues in the central part of the cavity, belonging to the "TCDD-binding fingerprint" previously determined [Pandini et al., 2009], that we detected for all the ligands under investigation (Figure 2.19). Additional characteristic contributions to the binding free energy of these ligands are provided by interactions with some hydrophobic residues at the bottom of the cavity, in particular L302 and L309 (Figure 2.20a, 2.20b, 2.20c). These were identified thanks to the MD refinement of the docking poses, that allowed us to describe a local deformation of the inner part of binding cavity induced by these ligands. In fact, previous docking analyses indicated the TCDD placement at the center of the cavity [Pandini et al., 2009, Bisson et al., 2009, Motto et al., 2011, Xing et al., 2012, Perkins et al., 2014, Nuti et al., 2014], whereas our simulations suggested that this ligand can move inside the cavity by translating from the center to the inner part and vice versa (Figure 2.18). Also the initial docking pose of BaP, that we found in the center of the cavity consistently with other docking studies [Xing et al., 2012], was dramatically changed by MD simulations, that allowed the ligand to reach the bottom of the cavity. Conversely, TCDF was already predicted to reach the inner region by docking calculations, and MD refinement did not alter this placement.

Mutagenesis studies supported the predicted binding geometry of the group 1 ligands. We previously observed that the L302A, L309A mutations dramatically reduced TCDD binding activity, whereas C327A, M334A and S359A, at the entrance of the cavity, only partially affect it [Motto et al., 2011]. In the present work we demonstrated that these two mutations eliminated or greatly reduced the AhR DNA binding induced by TCDD, TCDF and BaP (Figure A.9). Given the correlation observed between binding affinity and potency to stimulate AhR DNA binding (Figure A.4) we could confirm the role of the long hydrophobic leucine sidechains in the inner part of the cavity in stabilization of these three ligands. The P291L mutant was designed to further validate the predicted binding poses through direct evaluation of the relative ligand binding affinities. The IC_{50} data (Figure 2.22) revealed that indeed P291, lying at a lateral side of the hydrophobic region, is not deeply involved in the stabilization of this group of ligands. In fact, the P291L mutation had limited effect on BaP binding and the observed enhancement of the TCDF affinity could be related

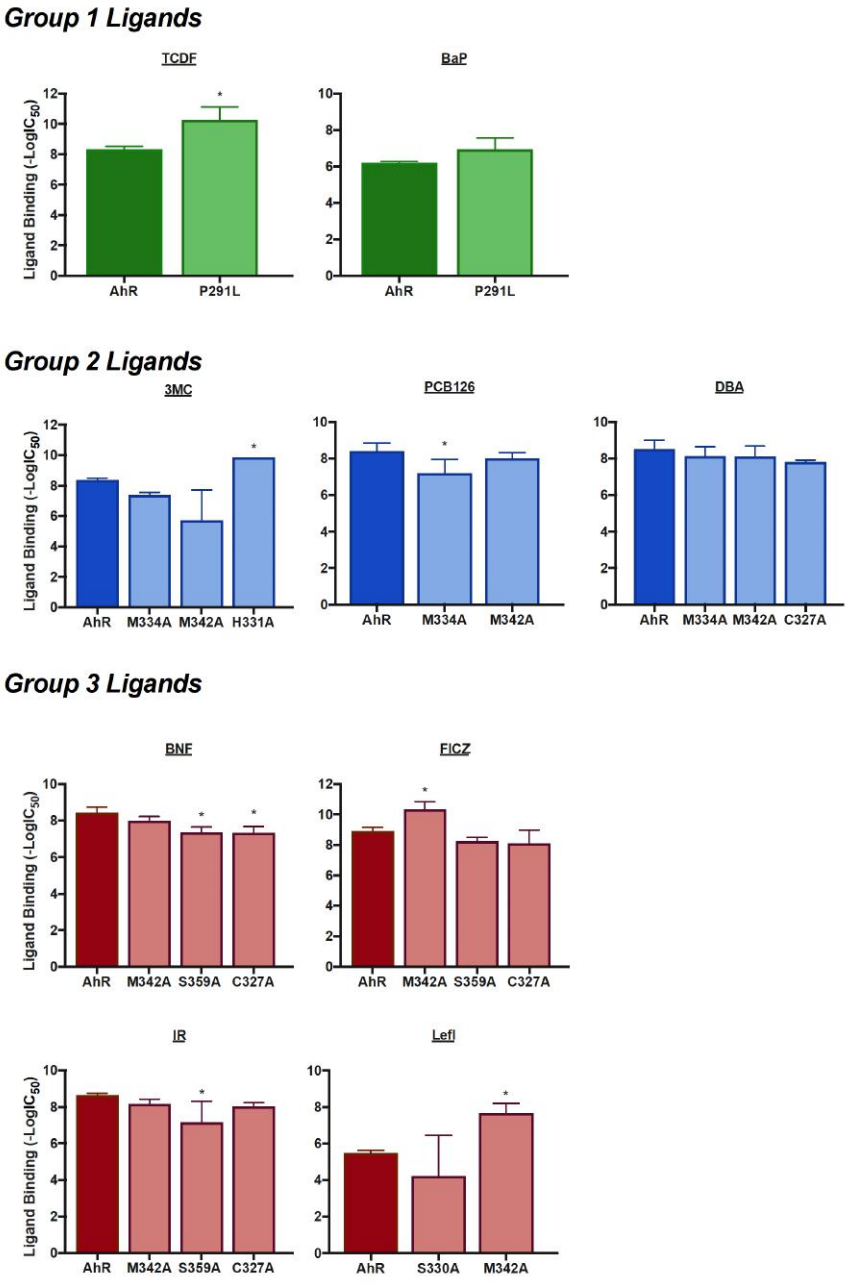

Figure 2.22 Relative binding affinity for group 1, 2 and 3 ligands relative to wild-type and mutant AhRs. The relative affinity ($-log\text{IC}_{50}$) of each test chemical for the AhR ligand was determined from concentration-dependent inhibition curves obtained using [^{3}H]TCDD ligand binding analysis. The mean IC_{50} ± standard deviation was determined using three-parameter non-linear regression of nine independent reactions. (*) Represents a significant ($p \pm 0.05$) change in ligand binding affinity relative to wild-type mAhR in One-Way ANOVA Multiple Comparison Test.

to the more effective stabilization produced by the longer leucine sidechain.

The second group of chemicals sharing common binding features includes 3MC, PCB126 and DBA. Docking analysis followed by MD refinement predicted that the two bulky PAHs tend to occupy almost the entire space within the cavity except the inner region, with a number of stabilizing hydrophobic interaction with the internal sidechains (Figure 2.19). The PCB126 molecule, with very different structural characteristics, occupies the same region thanks to a certain translational freedom into the cavity and its adaptability given by rotation around the central bond. On the whole, all the three ligands, in addition to interactions with central residues, showed characteristic interactions with hydrophobic residues at the entrance of the cavity: M334 and M342 at the Fα/Gβ site. Moreover, 3MC contacts several times the flexible sidechain of H331, on the Fα, during the MD simulation (Figure 2.20d, 2.20e, 2.20f).

Mutagenesis studies supported the predicted binding poses of 3MC and PCB126. In fact, the lower binding affinities measured for the M334A and M342A mutants in comparison with the wt AhR (Figure 2.22) confirmed the hypothesis that substitution of the long hydrophobic methionine sidechains with alanine decreases stabilization of the two ligands. Contrary to expectations, the H331A mutation increased the 3MC affinity. This is probably due to the removal of unfavorable interactions with a polar sidechain. Evaluations of IC_{50} for DBA binding to M334A and M342A (Figure 2.22) revealed that computational predictions in that case were incorrect; in fact, the DBA binding affinity for the wt AhR remained unaltered upon both mutations. It is conceivable that the alternative docking pose in which the DBA molecule reaches the inner part of the cavity (similarly to group 1 ligands), that was discarded during the selection of the best pose, might be more reliable.

The third group of ligands, including BNF, FICZ, IR and LEFL, shows completely different physico-chemical characteristics. Accordingly, while docking calculations predicted their binding at the Fα/Gβ site similarly to ligands of group 2, their stabilization derived from different contributions, (i e. electrostatic interactions and hydrogen-bonds with the polar residues located in this region: C327, S330, S359, Q377) (Figure 2.21). MD refinement indicated that these strong interactions determine poor mobility of the ligands into the cavity. MD simulations also revealed that BNF and IR are able to break an inter-residue hydrogen bond between S359 and Q377 by establishing new ligand-residue H-bonds, and that FICZ and LEFL partially interfere with the same H-bond network.

Given that the mutation of Q377 to alanine was proved to abolish TCDD specific binding [Pandini et al., 2007], this mutant could not be used to evaluate the relative affinity of BNF and IR by competitive

binding analysis with [^{3}H]TCDD. However, the observed lowering of the BNF and IR affinities for the S359A and C327A mutants compared to the wt AhR (Figure 2.22) confirmed both the predicted binding poses and the key role of the electrostatic interactions in ligand stabilization. While the S359 residue appeared to play a less important role in stabilization of FICZ binding, the observed higher affinity of this ligand for M342A compared to the wt AhR (Figure 2.22) could confirm its placement in the same binding site of the other ligands of this group (at the Fα/ Gβ site). In fact, it is conceivable that the long hydrophobic methionine sidechain in the Gβ strand could interfere with ligand stabilization. Noticeably, experiments for determination of the relative binding affinity of LEFL for S359A from concentration-dependent inhibition curves resulted not converging (Table A.8), thus indicating a very low affinity of LEFL for this mutant in agreement with our prediction. A further confirmation of the LEFL binding pose could derive from the significant increase of LEFL binding affinity upon mutation of M342 to alanine, similar to that observed for FICZ (Figure 2.22).

While other investigators predicted binding characteristics similar to those here described for FICZ and IR using different docking approaches, but they suggested that these ligands share the same binding site of TCDD [Bisson et al., 2009, Perkins et al., 2014, Nuti et al., 2014]. Here we demonstrated that, while TCDD can reach the inner hydrophobic region of the AhR cavity thanks to induced-fit effects, FICZ, IR and the other group 3 chemicals bind nearer to the entrance of the cavity and form H-bonds with residues in this region.

Another confirmation of differential binding of TCDD (group 1) and LEFL (group 3) ligands derives from a study on ligand binding to different zebrafish AhR paralogues (isoforms: 1a, 1b and 2), where it was found that the zebrafish AhR1a (zfAhR1a) was unable to bind TCDD but can bind and be activated by LEFL [Goodale et al., 2012]. Our previous comparative analysis of the AhR LBDs homology models of different species revealed that zfAhR1a has a dramatically shortened binding cavity compared to the mouse AhR [Fraccalvieri et al., 2013] and this is due to the sidechains of three residues at the center of the cavity, which reduce the available internal space only to the region at the entrance of the cavity (here called the Fα/Gβ site). A confirmation of the role of this reduction on differential binding was provided by the evidence that mutation of these residues in zfAhR1a to those present in mouse AhR restored the zfAhR1a ability to bind TCDD [Fraccalvieri et al., 2013]. These findings, in addition to the mutagenesis data here presented, confirm our prediction that the binding site of LEFL (and of the other group 3 chemicals) is located near to the entrance of the AhR binding cavity.

We can conclude that our novel approach for ligand docking to the AhR homology models, that includes ensemble-docking to different

modeled conformations of the LBD and post-docking refinement by MD simulations, was effective in capturing the differences in binding of diverse agonists. Structure-driven site-directed mutagenesis followed by evaluation of the relative ligand binding affinities for the obtained mutants provided confirmation of most of the computational predictions.

Three groups were identified with different binding sites and interactions with residues within the binding site. The main difference was observed between groups 1 and 3, that show completely different physico-chemical characteristics and accordingly yield interactions with hydrophobic residues in the inner part of the cavity (group 1, including TCDD, TCDF and BaP) or alternatively with polar residues at the Fα/Gβ site (group 3, including BNF, FICZ, IR and LEFL). Another characteristic of group 3 chemicals revealed by MD simulations was the ability of these ligands to break an inter-residue H-bond network existing at the center of the cavity. An additional determinant of group 1 binding was their ability to penetrate deeply into the cavity using the flexibility and plasticity of the inner zone of the cavity. In contrast, 3MC and PCB126 (group 2), despite the hydrophobic stabilization similar to that of group 1, cannot reach the inner part of the cavity due to their high steric hindrance and thus they occupy a binding site nearer to the entrance of the cavity.

While we were able to demonstrate differences in ligand binding of this set of chemicals, what remains to be determined is whether the different pattern of interactions with residues within the AhR binding cavity can result in differential effects on AhR conformational changes and interactions with protein partners that may propagate downstream in the AhR signaling pathway. It is conceivable that ligands that bind near to the entrance of the cavity could have a more favorable dissociation kinetics (high koff constants) that could lead to a lower stability of the complexes or reduced rates of AhR transformation. On the other hand, our modeling studies on the AhR:ARNT dimer, based on the HIF-2α:ARNT template, predicted that the Fα/Gβ site may be involved in the dimerization interface [Corrada et al., 2016, 2017]. Therefore, specific perturbation effects on the dimer structure and stability could be produced by ligands that are in contact with this interface, similarly to what suggested for the HIF-2α:ARNT dimer [Motta et al., 2018b]. These and other hypotheses will be addressed in future studies aimed at analyzing the effects of differential ligand binding on the ligand-specific AhR transformation and dimerization. Understanding the effects of differential binding in modulating the AhR functionality would allow the design of new ligands targeted to promote specific alterations of the AhR mechanism useful for medicinal chemistry.

3

BINDING OF LIGANDS TO HUMAN PXR

3.1 INTRODUCTION

The Pregnane X Receptor (PXR or NR1I2) is a nuclear receptor (NR) that has important roles in drug metabolism and drug-drug interactions. In particular it binds the PXR response element (PXRE) in the promoter region of *cyp3a4*. In humans, CYP3A4 is involved in the metabolism of more than 50% of all drugs as well as a variety of other xenobiotics and endogenous substances, including steroids [Jones et al., 2000]. The activation of the CYPs metabolizing enzymes may lead to an undesired decrease in the bioavailability of many prescribed drugs [Banerjee et al., 2015]. For this reason, elucidation of the exact molecular mechanism that underlies PXR activation has important implications for drug development processes [Willson and Kliewer, 2002]. Moreover, PXR has been strongly associated with cancer and with metabolic and inflammatory diseases, [Banerjee et al., 2015] making PXR an intriguing new target for drug design studies aimed at developing both antagonist [Biswas et al., 2009] and agonist molecules.

The high promiscuity and the divergence evolution of PXR were a very hot topic [Jones et al., 2000] until the first crystal structures came out in 2001 [Watkins et al., 2001], finally elucidating some features of this receptor.

After that first structures, several others were deposited in the PDB and to date we count 20 depositions of this human LBD both in the *apo* and *holo* forms, in complex with different ligands. The PXR-LBD is an α-helical sandwich, reproducing the typical NR fold, composed of three layers: $\alpha 1/\alpha 3$, $\alpha 4/\alpha 5/\alpha 8$, and $\alpha 7/\alpha 10$ (Figure 3.1). Unlike other NRs, PXR has the small β-sheet expanded to a five-stranded antiparallel β-sheet and the $\alpha 6$ helix is often converted to a loop. The latter characteristic is thought to be responsible for the accommodation of different ligands within the internal cavity of the domain [Watkins et al., 2001]. The αAF helix represents the AF-2 domain involved in binding of co-activators and co-repressors (see Figure 1.3 for the scheme of functional domains).

Ligand binding causes a conformational change leading to the release of co-repressors; then the exposure of the hydrophobic surface of αAF implies the recruitment of co-activators e. g. the steroid receptor co-activator, SRC-1) and the transcriptional activation [Chai et al., 2016].

The large majority of the *holo* crystal structures of human PXR include ligands of pharmaceutical interest: the St. John's wort compound hyperforin [Watkins et al., 2003b]); the antibiotic rifampicin [Chren-

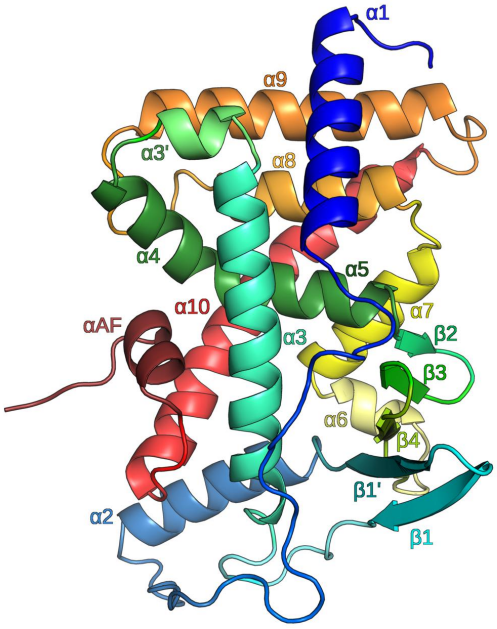

Figure 3.1 The PXR LBD structure is constituted by an α-helical sandwich. Secondary structure elements are colored from N-terminus (blue) to C-terminus (red) and labeled according to the NR LBD nomenclature proposed in Glass and Rosenfeld [2000].

cik et al., 2005]; some drug-like ligands [Xue et al., 2007, Hennessy et al., 2015, Khan et al., 2015]; the anti-HIV drug PNU-142721 [Cheng and Redinbo, 2011]. Rifampicin is widely used to treat tuberculosis at doses that induce CYP3A4 expression, and it was known that it directly binds to PXR into the ligand binding pocket before the crystal structures had been available [Jones et al., 2000]. It is interesting to note that the experimental structure could not be fully solved, but it lacked the α2 and a part of the ligand. Rifampicin is the largest ligand ever described for PXR, and probably for the NRs. Other studies [Xue et al., 2007, Hennessy et al., 2015] were aimed at searching a molecular scaffold to develop PXR antagonists, but the unique characteristics of the ligand binding cavity made antagonist design particularly difficult. Indeed PXR does not rely on the usual NR paradigm, in which both agonists and antagonists compete for the same binding site causing a different conformational change. To find antagonists of PXR it is necessary to look outside the ligand binding cavity and target the αAF surface. It is conceivable that antagonists bind αAF preventing the recruitment of co-activators; this model was confirmed by computational coupled with mutagenesis studies [Wang et al., 2006, Ekins et al., 2007].

Another interesting structure of PXR is the one able to explain the synergistic activation of PXR by the 17α-ethinylestradiol, the active substance of contraceptive pills, and the persistent organochlorine pesticide, trans-nonachlor [Delfosse et al., 2015]. It was demonstrated that

each ligand enhances the binding affinity of the other, so the binary mixture induces a substantial biological response at doses at which each chemical individually is inactive. The Authors suggested the formation of a "supramolecular ligand", which could explain the synergistic toxic effect of chemical mixtures, with implications for studies of endocrine disruption, toxicology and chemical risk assessment. This structure provideed an example of how molecular structural information can elucidate the molecular mechanism underling observable biological effects and greatly improve the understanding of biological and toxicological processes.

The most extensively studied ligand of PXR is SR12813. It was the first ligand to be co-crystallized with PXR [Watkins et al., 2001] and did not identify a unique geometry for SR12813, but proposed that SR12813 could occupy different binding sites inside the cavity. SR12813 is a cholesterol-lowering drug known to be a high-affinity ligand of PXR (dissociation constant $K_d = 41$ nM). Four X-ray depositions of PXR in complex with SR12813 have been available: the first identified three different orientations inside the ligand binding cavity [Watkins et al., 2001]; the same Authors proposed a new geometry two years later and suggested that the co-activator SRC-1 contributes to the stabilization of that single geometry [Watkins et al., 2003a]; other Authors confirmed that hypothesis finding exactly the same ligand orientation with the co-activator [Wang et al., 2008]; finally the structure of the dimeric form of PXR:RXR LBDs was determined [Wallace et al., 2013], characterized by SR12813 bound in a different orientation (although the ligand Electron Density Map is less defined compared to the previous studies).

The crystal structures of the PXR-LBD allowed identification of a buried ligand binding cavity with a remarkable volume of more than 1300 $Å^3$, that makes this pocket the largest known to date [Gallastegui et al., 2015]. The special insertion of about 60 aminoacids (compared to other NRs), the presence of a flexible loop in helix 6, and the unwinding of helix 7 are believed to be responsible for PXR ability to bind ligands of different sizes. In particular, the longer loop may contract or expand in order to accommodate different chemicals, including the big Rifampicin, inside the pocket [Chrencik et al., 2005]. The cavity is mainly hydrophobic, it is lined by only four polar residues and four charged (or potentially charged) and two of them, E321 and R410, constitute a salt-bridge that effectively neutralizes their charges. Mutagenesis studies also suggested that this salt bridge could be critical for the receptor response to ligands [Chrencik et al., 2005].

After the first structures of hPXR-LBD became available, many molecular docking studies were performed on them. One of the most important predicted the agonist binding site inside the ligand binding cavity, and the antagonist site on the αAF surface [Ekins et al., 2007]. Others not only tried to answer some fundamental questions regarding ligand binding and activation of the receptor, but also to

impact on drug-discovery programs by solving potential drug-drug interaction issues [Gao et al., 2007]. To the same aim also machine learning algorithms were used [Khandelwal et al., 2008]. Molecular modeling was also used to predict the binding of different environmental pollotants. For example, binding of some endocrine-disrupting chemicals (EDC) were investigated to analyze how they interfere with the body endocrine system. producing harmful effects. These studies were conduced in particular on bisphenol A (BPA) [Sui et al., 2012] and phthalates [Sarath Josh et al., 2016], compounds that have attracted attention during the last decades. In addition to Molecular Docking, also the QSAR approaches were used to try to predict ligand-PXR interactions [Ekins et al., 2009].

In this thesis, we tested and compared different approaches for computational modeling the ligand binding to PXR all aimed at taking into account the LBD flexibility involved in the process. To this aim, at first we exploited the experimental information available using the most updated PXR X-ray structures to set up an ensemble-docking protocol and applied it to study the binding of a set of known ligands with different structural and physico-chemical characteristics. Subsequently, we focused our attention on the controversial binding mechanism of the most studied PXR ligand, SR12813, and used advanced methods based on MD simulations to model the process and obtain mechanistic information.

3.2 METHODS

3.2.1 *Receptor and ligands preparation for docking*

We downloaded the PXR structures from the PDB [Berman et al., 2000] and removed the protein partners. We used Prime [Jacobson et al., 2002, 2004], in the Schrödinger suite, to determine the 3D structure of the unresolved aminoacids ranging from 178 to 192 (the $\alpha 1/\alpha 2$ loops). No template was given to the software, so the loops were built ab initio using the Prime Energy to select the best model. Prime energy includes all force field parameters plus the GBSA solvation terms. The models were validated using PROCHECK [Laskowski et al., 1993], and prepared using the Preparation Wizard with OPLS3 force field.

3.2.2 *Molecular Docking*

We extracted the ligands from the PDB structures and prepared them using LigPrep and Epik to get the more reliable tautomer. Rifampicin was downloaded from PubChem.

Docking was performed using Glide XP with OPLS3 force field [Friesner et al., 2004, Halgren et al., 2004]. Glide XP scales ligands vdW radii to 0.8 as default; in this case we scaled them to 0.7 on

the basis of redocking calculation trials where this scaling factor lead to better reproduction of the experimental binding geometries. The grid was centered on the SR12813 ligand in the 4j5x structure. After structural alignment of all the proteins, the grids were calculated using the same coordinates of the first one on SR12813.

For ensemble docking we used the Emodel function to select the best pose among the ones sampled in the different receptor structures. RMSD was calculated on the ligand heavy atoms after alignment of the Cα atoms of the protein structures.

3.2.3 *System preparation for MD simulations*

Crystal structures for PXR in its unbound (PDB ID: 4j5w [Wallace et al., 2013]) and SR12813-bound (PDB ID: 1ilh [Watkins et al., 2001], 1nrl [Watkins et al., 2003b], 4j5x [Wallace et al., 2013]) forms were obtained from the PDB [Berman et al., 2000], and protein partners were removed. The PXR structures have unresolved regions between the α1 and α2 helices, that was modeled using Prime [Jacobson et al., 2004] within the Schrödinger Suite, as described in section 3.2.1. Proteins were prepared with the Protein Preparation Wizard as previously described (section 3.2.1). The SR12813 ligand was parametrized using the BiKi suite [Decherchi et al., 2018] with the AM1-BCC [Jakalian et al., 2002] level of theory. Partial charges were derived using the RESP method [Bayly et al., 1993] in Antechamber [Wang et al., 2006], while a GAFF [Wang et al., 2004] parametrization was used to achieve the complete topological description of each ligand. Sensible torsion parametrization of the C=C-Ca-Ca angle was compared with QM calculations performed at the HF/6-31G* level using the Jaguar [Bochevarov et al., 2013] program in Maestro to adjust the ambiguous parametrization of GAFF.

3.2.4 *Plain MD simulation*

The plain MD simulations were performed using GROMACS 4.3 [Abraham et al., 2015]. The protein was solvated in an orthorhombic box with TIP3P water molecules [Jorgensen et al., 1983], ad neutralized with Na^+/Cl^- ions within BiKi basics [Decherchi et al., 2018]. The minimal distance between the protein and the box boundaries was set to 12 Å. The Amber ff14SB force field [Maier et al., 2015] was used for the proteins and a multistage equilibration protocol was applied to remove unfavorable contacts and provide a reliable starting point for the production runs: the system first underwent 5,000 steps of steepest descent energy minimization, and then four different consecutive equilibration steps were performed: (1) 100 ps in NVT ensemble at 100 K using a time-step of 1 fs, (2) 100 ps in NVT ensemble at 200 K and time-step increased to 2 fs, (3) 100 ps in NVT ensemble at 300 K,

and (4) 1,000 ps in NPT ensemble at 300 K. In all the stages the atoms belonging to the protein backbone (and to the ligand, if present) were restrained with a force constant of 1,000 kJ mol^{-1} nm^{-2}. Electrostatics was treated with the cut-off method for short-range interactions and with the particle mesh Ewald method (PME) [Darden et al., 1993] for long-range interactions (rlist = 1.1 nm, cut-off distance = 1.1 nm, vdW distance = 1.1 nm, PME order = 4). The constant temperature conditions were provided using the velocity rescale thermostat [Bussi et al., 2007] (coupling constant of 0.1 ps), while pressure was coupled with the Parrinello-Rahman barostat [Parrinello and Rahman, 1981] (coupling constant of 2 ps). All bonds were constrained with the LINCS algorithm [Hess et al., 1997].

3.2.5 *MD-Binding simulations*

The MD-Binding method within the BiKi suite uses an additive external force that is summed to the regular potential energy of the system to enhance the probability to observe the binding event. The bias consists in external electrostatic-like forces acting between a subset of the residues of the binding site and the ligand. The intensity of the biasing force is regulated by the adaptivity rules and gradually switches off as the process moves forward so that, after the conjectured passing of the transition state has occurred, it slowly recovers the behavior of classical unbiased MD.

In the standard protocol for MD-Binding, the possible entrances for the ligand are computed using Nanoshaper [Decherchi and Rocchia, 2013], and the ligand is positioned with a random orientation at a predetermined distance from the residues that form the entrance. In the present work, we used two selected frames extracted from the *apo* simulation, to calculate the ligand entrance with Nanoshaper for the ligand positioning. The protein attractive atoms were selected as the list of residues that belong to the internal cavity of PXR according to Nanoshaper calculations (C207, S208, L209, K210, V211, L239, H242, M243, M246, S247, F251, F281, C284, Q285, F288, W289, Y306, L308, E309, T311, A312, G313, L318, E321, M323, L324, F326, H327, I403, H407, R410, L411, F420, M425, F429); for the ligand, we used all the heavy atoms. As switch-off residue we selected the S247, with a cut-off distance of 4 Å. This residue was found to interact with the ligand in most of the experimental structures and is placed at the opposite side of both the entrances so that, when the ligand approaches this residue, the transition state for the binding process can be considered overcome. 50 independent simulation runs were launched in parallel per entrance gate, each of them lasting 20 ns, starting from the 4j5w *apo* protein conformation.

3.2.6 *Scaled MD*

In the SMD approach implemented in BiKi [Mollica et al., 2015], the potential energy terms are scaled by a scaling factor λ to reduce the energetic barrier and enhance sampling during the simulations. In the present work we used the value $\lambda = 0.5$ and applied low restraints (50 kJ mol^{-1} nm^{-2}) to the protein backbone atoms, excluding the residues around the binding site (within 8 Å from the ligand) that were unrestrained. Simulations of 30 ns were carried out for each replica in the NVT ensemble.

3.2.7 *Analysis of MD simulations*

For the comparison of binding modes, we calculated the dRMSD between the ligand-site distances in the simulated complex and the corresponding ligand-site distances in the X-ray structures:

$$\text{dRMSD} = \sqrt{\frac{\sum_i \sum_j (d_{ij}^x - d_{ij}^m)^2}{N}}$$

where x and m are the experimental and calculated complexes, respectively; d are the vectors of the distances between the ligand and the binding-site heavy atoms; i and j are the indices of the atoms; and N is the number of comparisons performed. We defined all the protein atoms within 6 Å from the ligand heavy atoms as binding-site heavy atoms. Calculations were performed with PLUMED [Tribello et al., 2014]. Using the dRMSD index, the distance calculation takes into account only the deviation of the relative position of the ligand with respect to the residues belonging to the binding site; it is a better index to evaluate the accuracy of the binding geometry than the RMSD calculated on the absolute positions of the ligand atoms, because this last one neglects the difference in the positions of protein residues in the simulated and reference geometries.

To evaluate the stability of a ligand binding conformation with SMD, we measured the time required for the ligand to reach 4 Å of RMSD (computed on heavy atoms) from the initial geometry in each replica, and we evaluated the resulting boxplot. The cut-off value of 4 Å was chosen analyzing different RMSD graphs to assure that it was a good cut-off value to discriminate when the ligand definitively left its starting conformation.

The ligand orientation within the binding site were represented in the sub-space of two selected collective variables (CVs) that correspond to the polar coordinates θ and ϕ in a reference system centered in the oxygen atom of the hydroxyl group of the ligand, relative to the vector connecting this atom and the carbon atom joining the two phosphate groups (Figure 3.2). These coordinates were computed for

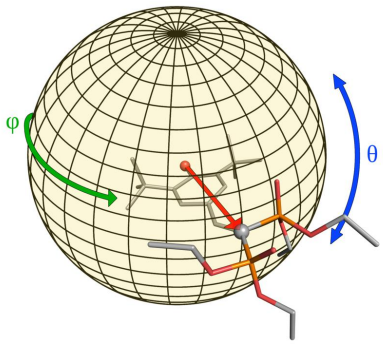

Figure 3.2 Representation of the CVs describing the ligand orientation inside the cavity. The angles θ and ϕ are computed for the vector connecting the oxygen of the hydroxyl group of the ligand and the carbon atoms between the two phosphate groups.

each conformation in the trajectory, upon alignment of the protein Cα atoms.

In this way we correctly described the ligand rotation inside the cavity, neglecting the possible translations. The conformational sampling performed in each simulation was then represented in the sub-space described by these two coordinates as a probability density surface in which the counting of each bin was reweighted to account for the different dimensions of the bins due to the spherical shape of the surface.

After structural alignment on protein Cα atoms, ligand conformations sampled during the SMD simulations starting from the final MD-Binding geometries were clustered using the GROMOS RMSD-based clustering tool applied to the ligand heavy atoms, using a cut-off value of 2 Å.

3.3 RESULTS

3.3.1 *Modeling the PXR ligand binding event by molecular docking*

Structural information on the PXR LBD in both the *apo* and *holo* forms has increased during the last years. To evaluate if this information is able to capture the conformational variability of PXR involved in ligand binding, we initially compared a standard docking protocol to an ensemble docking protocol based on the available structures. To this aim we used a set of ligands for which the binding geometry is known from X-ray crystallography (Figure 3.3).

The backbone structure of hPXR is well conserved among the different depositions (the RMSD is under 3 Å on Cα), but also the internal sidechains do not vary much their conformations (RMSD on complete internal residues is under 2 Å). The loop modeling involved 14 residues that connect α1 and α2; the modeled part is shown in Figure 3.4. We

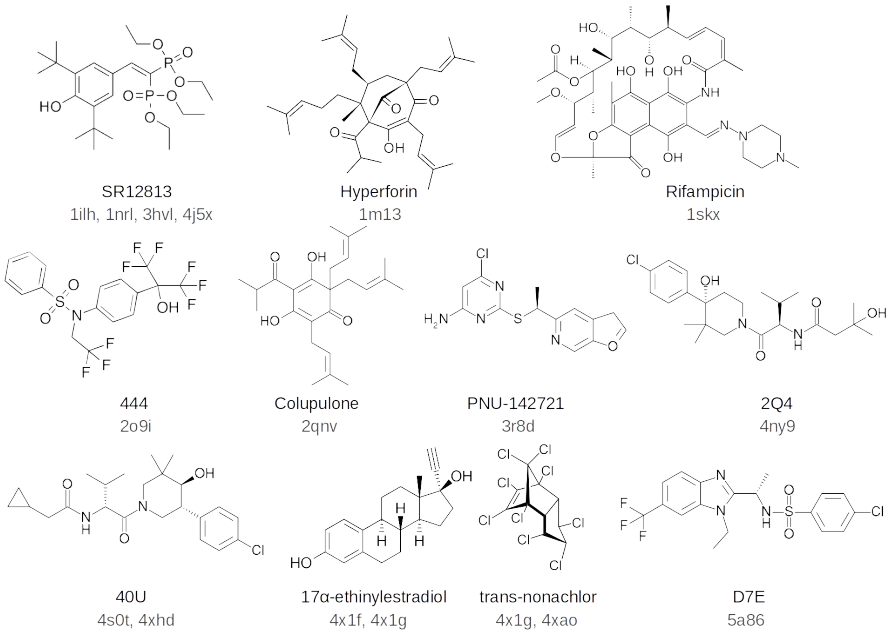

Figure 3.3 The co-crystallized ligands used in the docking and ensemble-docking studies. It is reported the common name (or the PDB id of the ligand) and below the PDB code for the structures which contain the depicted ligand.

did not observe a significant convergence of the loop in a single conformational state; this is in agreement with the experimental evidence that this highly flexible region could not be resolved. The 3D position of this insertion, however, does not affect the internal volume of the ligand binding cavity (Figure 3.4), which is instead influenced by few sidechains e. g. L209 and H407). Volumes range from about 1000 $Å^3$ in the *apo* structure 4j5w, to about 1500 $Å^3$ in 5a86.

To obtain a spectrum of the conformational characteristics of the PXR LBD as wide as possible, we chose five representative structures, among all the depositions, that are the most up to date and include complexes with different ligands: 4j5w is in the *apo* form, 4j5x is complexed with the very flexible ligand SR12813, 4s0t is the structure that has most of the residues solved and it is in complex with 40U (a inhibitor of GCR-CCR1), 4x1g has both the hormone and the pesticide bound into the cavity, and finally 5a86 has the greatest volume and is complexed with 40U (an antagonist of S1P1 receptor).

The shape of the internal cavities (Figure 3.4) differ only in some details, but maintain a certain similarity, especially if we compare 4j5w and 4j5x (obtained by the same laboratories under the same experimental conditions), despite 4j5w is in the *apo* form and 4j5x is bound to SR12813.

Molecular docking calculations were performed with a standard protocol for all the available co-crystallized ligands shown in Figure 3.3 using the *apo* 4j5w receptor structure to evaluate if the description of

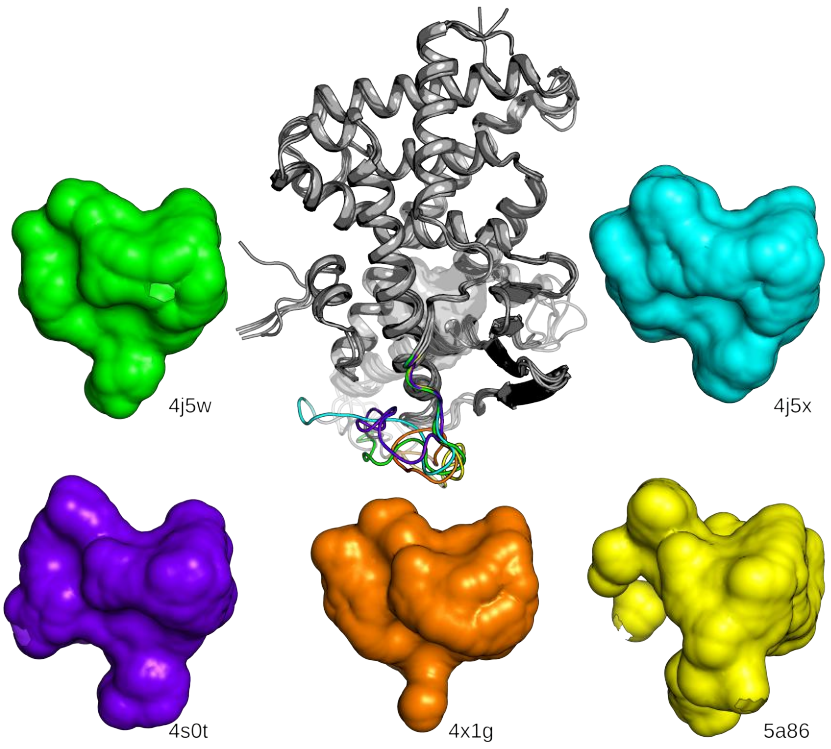

Figure 3.4 The structure of PXR is shown as gray cartoons and the modeled loops are colored according to the model; the ligand binding cavities are shown as solvent accessible surface calculated using CASTp.

the ligand binding cavity provided by a single structure is sufficient to describe such different binding modes.

Docking poses and crystallographic geometries are compared in Figure 3.5. For both Rifampicin and Hyperforin the software did not obtain any poses; but docking results were insufficient for most of the ligands. In some cases neither the binding site could be reproduced, e. g. 2Q4, 40U, 17α-ethinylestradiol and D7E; in other cases the binding site was identified, but the pose was not correct, e. g. Colupulone and trans-nonachlor; finally docking to the *apo* structure worked well with two ligands, 444 and PNU-142721, that have smaller molecular size than the other ligands. It is possible that the biggest ligands could not find their ideal site because the *apo* cavity is too small and rigid to accommodate them at the best.

The case of SR12813 is different because it is co-crystallized in six different geometries; we compared the obtained pose for this ligand with the six depositions in Figure 3.6. It is difficult to evaluate how good were these results because of the multiple reference geometries. Docking tended to reproduce the 1ilh B/1nrl/3hvl geometries, which are all very similar. The docked pose inverts the molecular "head" and "tail" if compared to 1ilh A/4j5x and it is rotated of 90°compared to 1ilh C. While the binding site that was mostly sampled in the crystallographic structures (1ilh B/1nrl/3hvl) was identified by docking, the parts of the molecule that are more flexible (the two phosphate substituents) were not well reproduced.

Therefore, we concluded that it is not sufficient to use only one structure of the receptor to describe binding to hPXR of diverse ligands using molecular docking.

To include more flexibility during calculations we tested the ensemble docking approach based on the previously described hPXR X-ray structures.

Ensemble docking gave better results compared to docking on a single structure; in fact for most of the ligands the crystallographic geometry was reproduced (Figure 3.7). It is interesting to note that we found three poses for Hyperforin, for which we did not sample any one using the standard docking protocol. On the contrary, we could not obtain any poses for Rifampicin, but it is known that to accommodate this ligand inside the binding cavity it is necessary a massive rearrangement of the secondary structure elements. We can observe that the best poses were obtained when the ligand re-docked to the experimental structure to which it was co-crystallized, if present in the considered ensemble. This was the case of 40U in 4s0t, 17α-ethinylestradiol and trans-nonachlor in 4x1g, and D7E in 5a86; all these poses perfectly reproduced the experimental binding geometries. It could be concluded that the "memory" of the original ligand is maintained into the receptor used for docking.

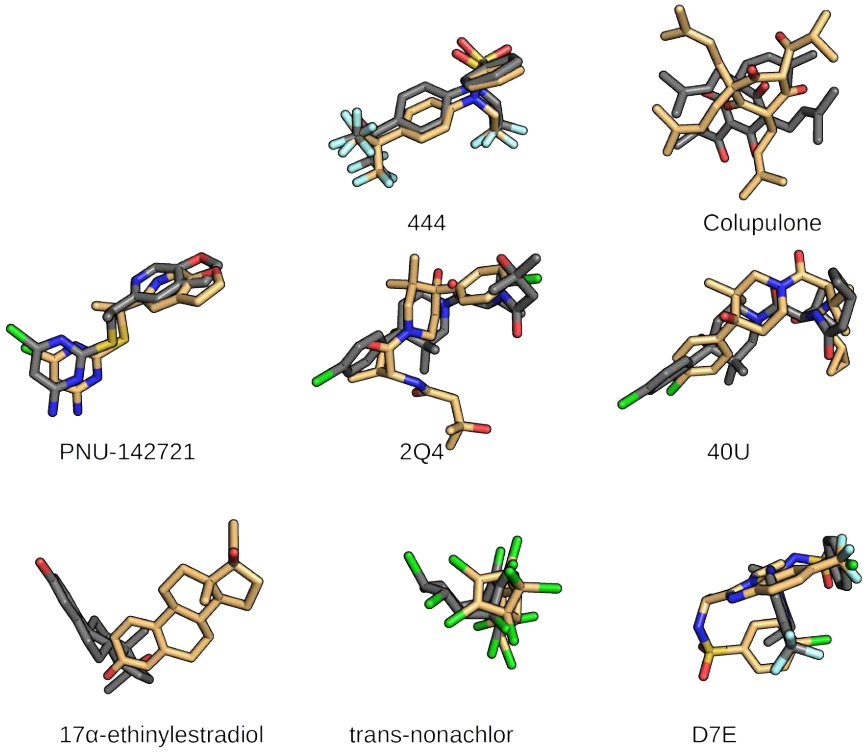

Figure 3.5 Docking to the *apo* structure of hPXR (4j5w). The crystallographic geometries of the ligands are colored in gray and the docked pose is colored in light orange.

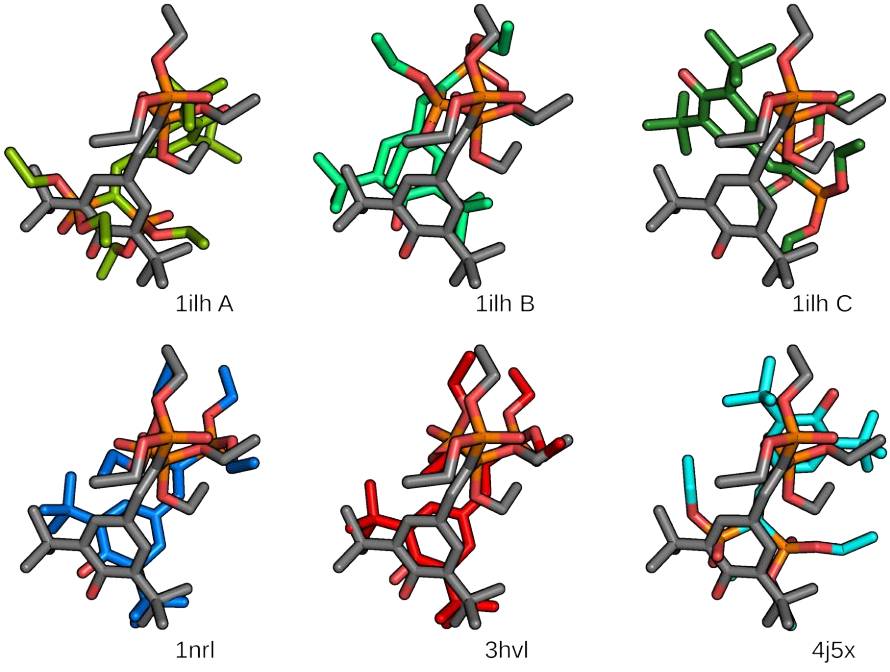

Figure 3.6 Docking of SR12813 to the *apo* structure of hPXR (4j5w). The crystallographic geometries are colored according to the deposition (labeled), the docked pose is colored in gray.

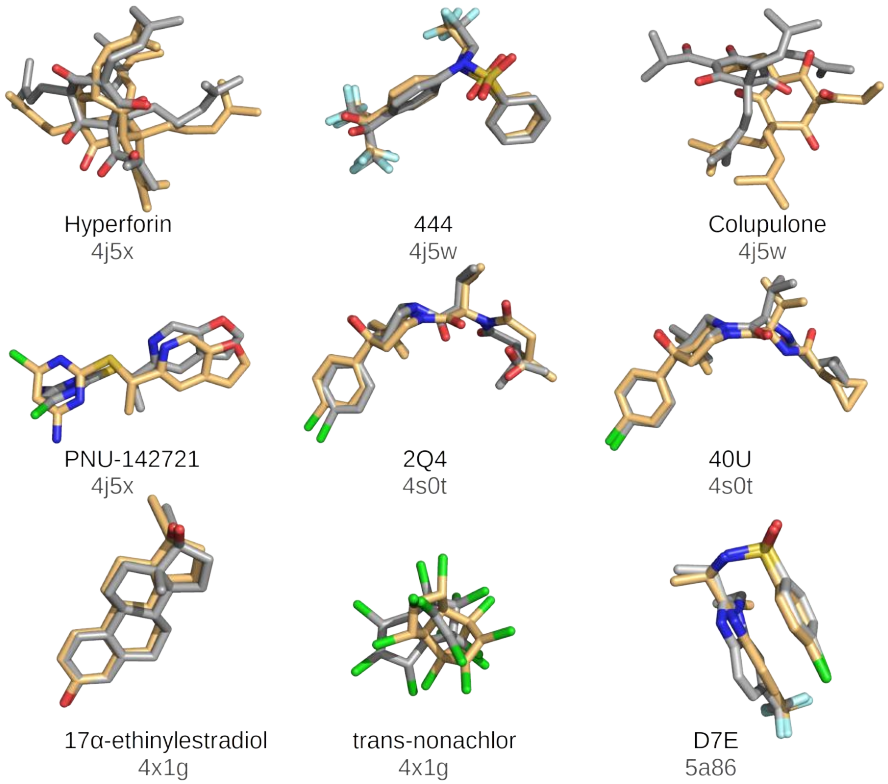

Figure 3.7 Ensemble docking on hPXR. The crystallographic geometries of the ligands are colored in gray and the docked poses are colored in light orange.

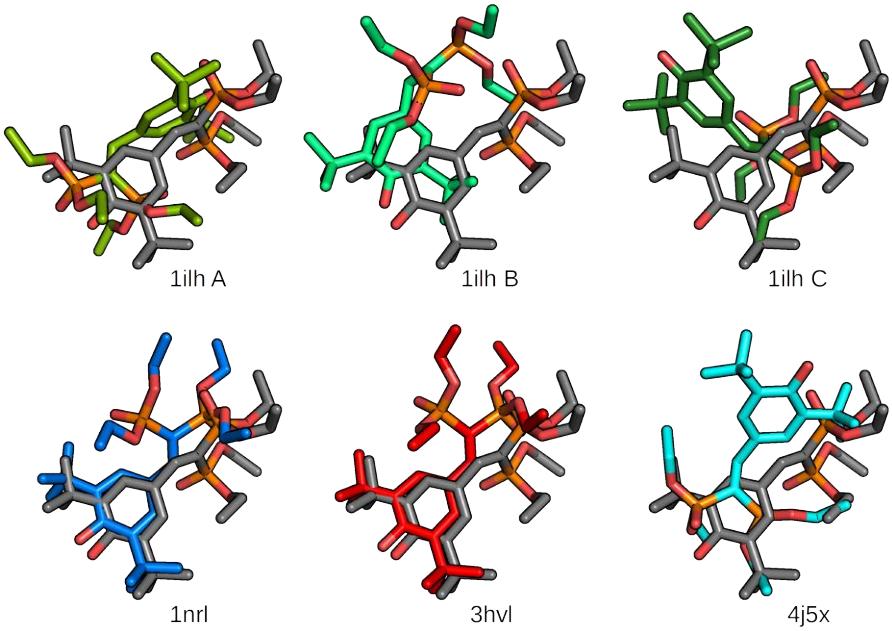

Figure 3.8 Ensemble docking of SR12813. The crystallographic geometries are colored according to the deposition (labeled), the docked pose is colored in gray.

The ligand for which we did not include the corresponding receptor structure docked into other protein conformations; e. g. 2Q4 is very similar to 40U and docked inside 4s0t, 444 and Colupulone found the best pose inside the *apo* structure 4j5w, while Hyperforin and PNU-142721 docked inside 4j5x. The observed preference of these latter ligands for the 4j5x and 4j5w structures among those available in the ensemble might be a clue that these give a better description of PXR in its biological environment, because they include the dimerization partner RXR. The inability of finding the experimental geometries of the very flexible Hyperforin and Colupulone could be explained with limitations of the docking sampling phase. Treatment of trans-nonachlor binding, for which the right binding site was identified but not the correct orientation of the chlorine atoms, probably posed some difficulties in the treatment of halogen atoms by the force field.

The results for SR12813 remained essentially unchanged compared to those obtained with the standard docking protocol in the *apo* 4j5w structure. In fact, the best pose for this ligand was found in the 4j5x structure, originally in complex with SR12813 (Figure 3.8), that does not differ much compared to the one found in 4j5w (Figure 3.6).

In conclusion, the approach of ensemble docking solved many of the problems that arose in docking to a single *apo* structure. This protocol seemed particularly suitable for docking to PXR of small molecules with a few conformational degrees of freedom and few stereocenters. But whenever the interest is focused on ligands with many stereocenters, with high flexibility or high molecular weight, it shows many limitations.

It is conceivable that the hPXR, a promiscuous protein which can bind very diverse molecules in a number of different orientations, undergoes large conformational rearrangements of the LBD during binding; therefore it represents a very demanding study case that requires an explicit treatment of the protein flexibility involved in the ligand binding process.

This work was accomplished also thanks to Dr. Lara Callea.

3.3.2 *Exploring the PXR ligand binding mechanism using advanced MD*

In the recent years, novel methods have been proposed for a fully dynamic description of the ligand binding event based on MD [De Vivo et al., 2016]. Given that the sampling issue is pivotal for the description of these slow processes, enhanced sampling methods are usually employed. Among these Steered MD [Colizzi et al., 2010] and metadynamics [Gervasio et al., 2005, Limongelli et al., 2010] were used to simulate the drug-binding events, with the ligand moving into and/or out of the binding pocket.

Given the limitations observed for both standard docking and ensemble-docking approaches in the study of the SR12813 binding to

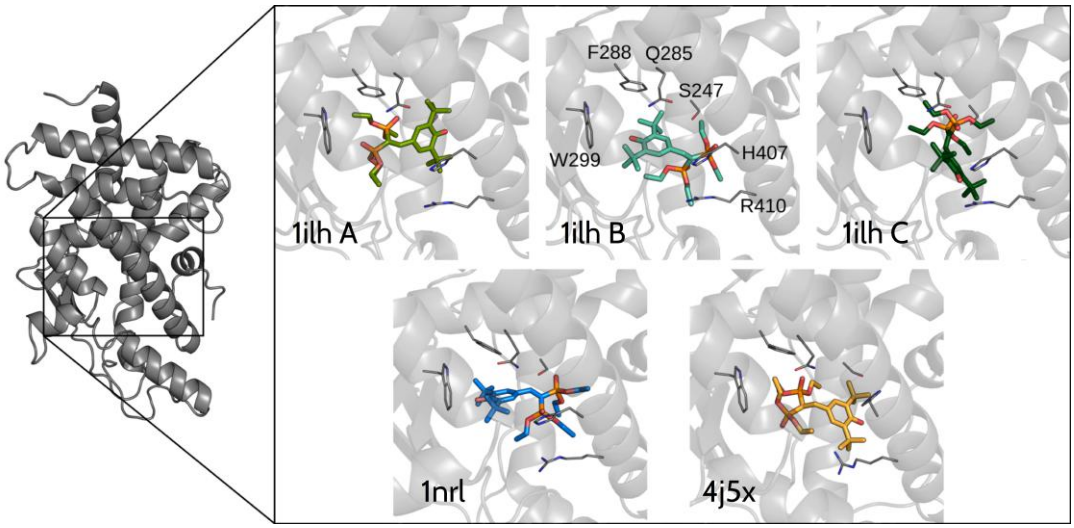

Figure 3.9 SR12813 is co-crystallized in different binding modes depending on the PDB deposition.

the hPXR (section 3.3.1), we tested one of these new approaches. We were aimed both at elucidating the controversial description of the SR12813 orientation into the binding cavity and at investigating the ligand entrance path inside the LBD to shed light into the mechanism underling the activation of PXR by ligands.

The buried nature of the cavity suggests significant conformational rearrangements of the LBD upon ligand entrance. Moreover, the presence of different binding modes implies a high degree of flexibility and plasticity of the domain during the binding process. Therefore, we propose to explicitly include the dynamic description of the binding event using recent MD-based tools implemented in the BiKi suite [Limongelli et al., 2013]. In particular, we employed the MD-Binding method [Decherchi et al., 2018] to analyze the binding mechanism and gain insights into the ligand entrance pathway. Moreover, we propose the use of the scaled MD (SMD) approach [Spitaleri et al., 2018, Tsujishita et al., 1993] to extensively sample the conformational space available to the PXR-ligand complex, thus allowing elucidation of the SR12813 dynamic behavior within the binding cavity. The characteristics of the system required the development of specific methodological approaches that may be insightful also for investigation of other ligand binding processes.

Among the X-ray structures for the PXR-LBD in complex with SR12813, 1nrl and 3hvl include SRC-1, 4j5x both SRC-1 and RXR, while 1ilh does not present any co-crystallized partners. The two protein partners bind in different regions: RXR dimerizes with the $\alpha 9/\alpha 10$ PXR helices, while SRC-1 binds to the αAF helix. The 1ilh deposition proposes three different ligand binding modes: 1ilh A, 1ilh B, and 1ilh C. In Figure 3.9

all the five different binding geometries are shown; as already pointed out, 1nrl and 3hvl are extremely similar (Figure 3.6).

Overall the residues within the cavity of the different depositions lie in similar geometries (pairwise RMSD on binding site heavy atoms: 1ilh-1nrl = 1.10 Å, 1ilh-4j5x = 1.40 Å, 1nrl-4j5x = 1.23 Å), with the ligand contacting the same set of residues, but with different orientations. In both the 1ilh A and 4j5x structures the ligand phosphate groups are directed toward the W299 residue, while the hydroxyl group orientations are slightly different: in 1ilh A it is involved in a H-bond with the S247, while in 4j5x this interaction is absent and the hydroxyl group is shifted toward the α10 helix. In the 1ilh B structure the ligand orientation is opposite to the previous ones, with a phosphate group forming a H-bond with H407, and the hydroxyl group pointing toward the W299 residue. In 1nrl the ligand maintains the same interactions, with the addition of another H-bond between S247 and the second phosphate group. Finally, in 1ilh C the phosphate groups establish H-bonds with S247 and Q285, while the hydroxyl group points toward the R410. The 3hvl structure presents a ligand orientation identical to that in 1nrl; for this reason, only the latter deposition, with better resolution and few missing residues, was retained for the analysis.

Firstly we investigated the relative stability of the experimental ligand binding geometries using the SMD approach [Sinko et al., 2013, Mollica et al., 2016]. In SMD simulations, the potential energy of the system is scaled, thus lowering the energy barriers and facilitating the barrier-crossing events. This method was proposed for the study of the whole ligand unbinding process and the prediction of kinetics constant [Mollica et al., 2015, Decherchi et al., 2015], but has already been used to evaluate the stability of different binding poses [Tsujishita et al., 1993, Decherchi and Rocchia, 2013].

All the simulations were carried out in absence of protein partners, and we measured the time that was necessary to reach a value of ligand RMSD of 4 Å from the crystallographic geometry. These results are shown as boxplots in Figure 3.10, values are representative of 23 replicas.

The binding mode in 1nrl resulted the most stable, in fact almost half of the replicas never reached the cut-off limit of 4 Å. Also 4j5x is quite stable, but in this case only one replica never reached the cut-off. The other 1ilh structures were all not very stable and drifted away from the initial crystallographic geometries. Anyway we never observed a solvated state of the ligands, neither for these unstable structures, this means that the high RMSD values can be ascribed to internal ligand reorientations. d in the simulation time.

The ligand binding site of PXR is buried like the one of AhR, consequently they do not present channels connecting the internal cavity and the solvent. For this reason the path of entrance is debatable for

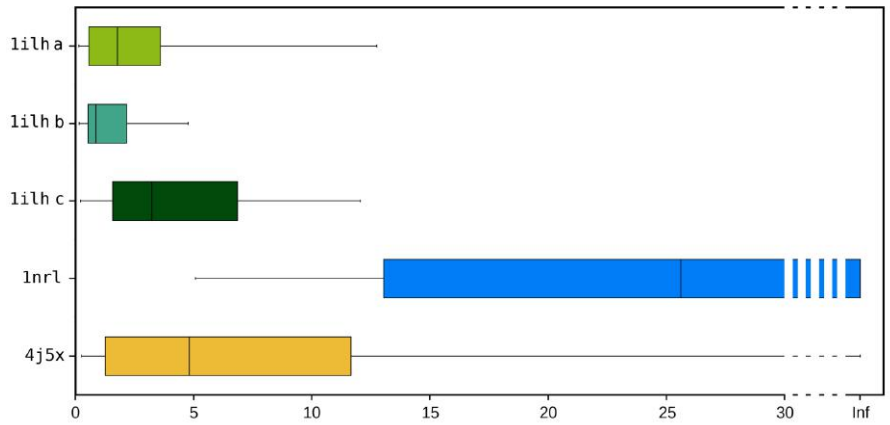

Figure 3.10 Time required to leave the ligand crystallographic geometry.

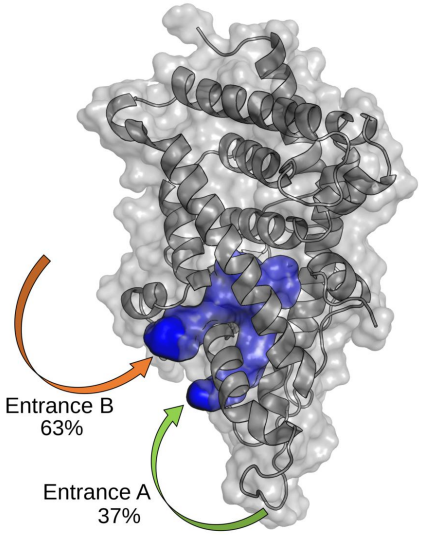

Figure 3.11 Possible entrances of water molecules into the PXR cavity. Solid and blue surface represents the solvent-accessible volumes, entrance B is preferred over entrance A.

both cases. Having available high quality structures, we computationally simulated the entrance of the ligand toward the cavity. We used the MD-Binding method that uses an additive external force to enhance sampling of the binding event.

Before performing MD-Binding, we used 100 ns of plain MD simulation of the *apo* protein structure (PDB ID: 4j5w) in explicit solvent to characterize the possible channels through which water molecules were most frequently exchanged. We identified two possible entrances connecting the bulk solvent and the binding pocket (Figure 3.11).

Entrance A is defined as the channel formed between $\alpha 2$ and $\alpha 3$ and entrance B the one between $\alpha 2$ and $\alpha 6$. We observed that the residues of both the entrances are not involved in the interaction interfaces with RXR and SRC-1. We further evaluated both entrances as possible entrances for the ligand.

The simulation campaign comprehends 50 replicas of 20 ns for each entrance, all started from the *apo* structure. Only 8 simulations starting

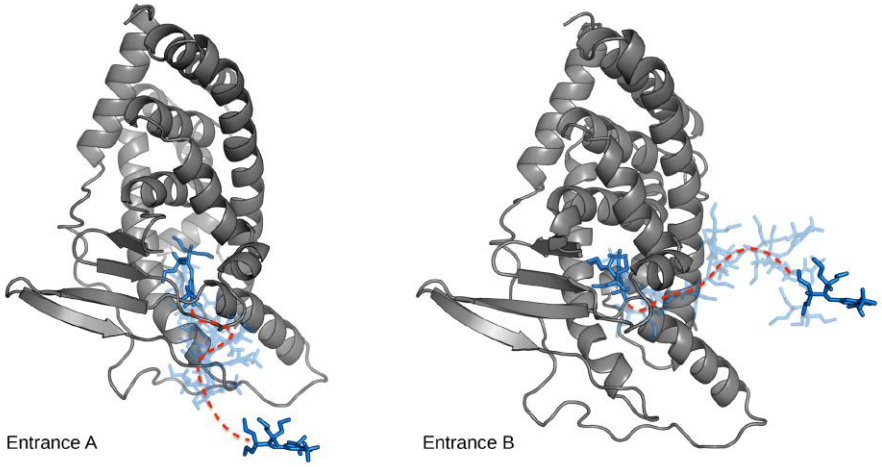

Figure 3.12 Two possible path of entrance for the ligand. Protein is represented as gray cartoons, the first and last frames in the ligand trajectory as solid blue sticks, and the trajectory as transparent sticks.

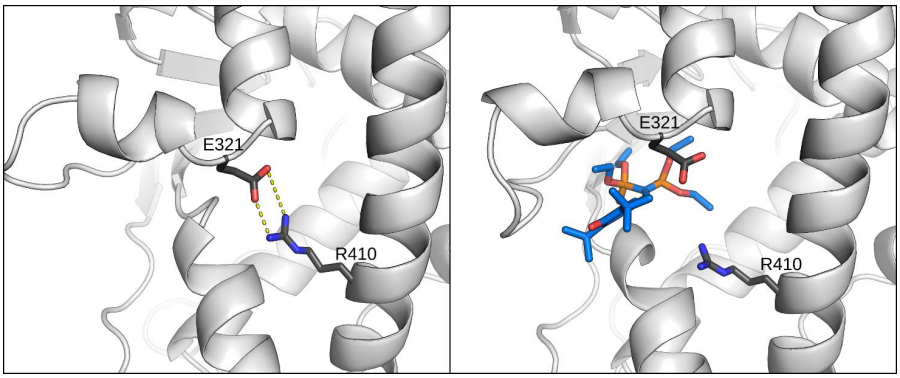

Figure 3.13 The breakdown of the salt-bridge between E321 and R410 upon ligand entrance. The steric hindrance of the ligand clearly breaks the interaction.

from point A, against the 42 starting from point B, reached the binding site. Despite the proximity of the two entrances, all simulations starting from point A passed through entrance A and those starting from point B through entrance B. Examples of the binding paths associated to the two entrances are shown in Figure 3.12.

The results confirmed the preference of SR12813 for entrance B, with 84% of simulations overcoming the energetic barrier and reaching the buried binding site. Interestingly, we found that during most of the simulations of the path through the B entrance, the ligand causes a shift of the α6 helix (RMSD on Cα atoms of the α6 helix ranging from 4 to 8 Å), thus producing the disruption of a salt-bridge between the E321 and R410 residues (see Figure 3.13). It was reported not only that this α6 is often converted into loop and thus particularly unstable, but also that the mutation of these residues (E321 and R410) alters the basal activity of PXR, highlighting their relevance in the binding

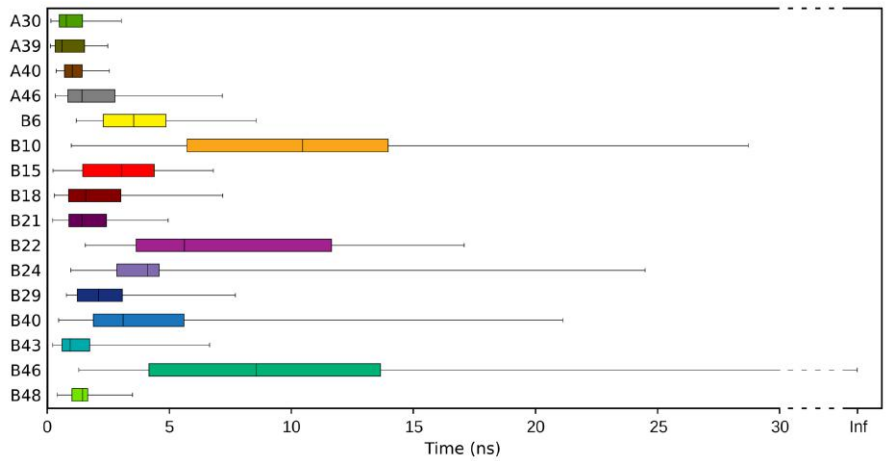

Figure 3.14 Simulation time required to leave the initial MD-Binding poses.

process [Watkins et al., 2001]. This conformational change was not found in the simulations of water entrance; therefore the salt-bridge acts as a gate for SR12813 and its breaking may be interpreted as the rate-determining step for the recognition process.

Despite the high success rate in overcoming the energetic barrier for ligand binding obtained by MD-Binding simulations, the final geometries were highly heterogeneous. The distance RMSD (dRMSD) values, calculated on the last frames show that only few replicas reached a geometry similar to one of the five crystallographic structures. While only 1ilh A was reproduced with high accuracy (dRMSD value for the replica B22: 1.1 Å) and acceptable results were obtained for 1ilh B and 4j5x (dRMSD < 2 Å), none of the replicas approached 1ilh C and 1nrl, despite the latter structure was predicted as the most stable one by SMD (Figure 3.10).

Therefore, to enhance sampling of the ligand conformational changes inside the binding cavity, we performed SMD simulations, instead of plain MD, starting from the final frames of MD-Binding. We selected the frames in which the distance between ligand and the switch-off residue was smaller than 5.5 Å(obtaining 16 different starting points) and performed 20 replicas of 30 ns collecting a total of 9.6 μs of simulations. None of these starting binding geometries showed high stability and most of the simulations rapidly drifted away (Figure 3.14; as a comparison, see the 1nrl stability in Figure 3.10). Interestingly, one of the most stable replicas was the B22 was very close to the 1ilh A X-ray structure.

To investigate the conformational space explored by the SMD simulations, we designed two CVs (θ and ϕ) representing the ligand orientation within the cavity in polar coordinates (see Figure 3.2 in the Methods section). The conformational sub-space explored by the ligand during the SMD simulations is well represented by the probabil-

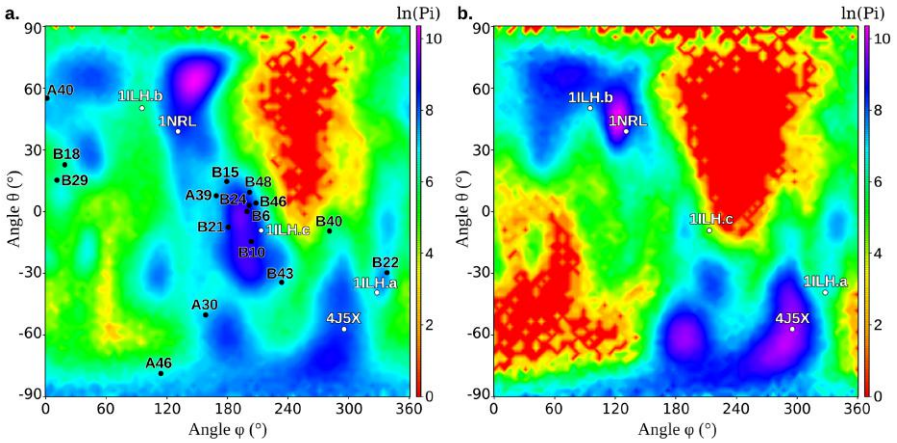

Figure 3.15 The two CVs define a sub-space of conformational state probabilities. **a.** Simulations started from MD-Binding poses (black dots); **b.** simulations started from the X-ray structures (white dots)

.

ity distribution map depending on these variables, that highlights the most sampled regions (Figure 3.15a).

Most of the MD-Binding poses are traceble to a ligand orientation similar to that in the X-ray 1ilh C, with the phosphate groups toward the F288 and S247 residues in the internal region of the cavity, and the hydroxyl group pointing toward R410 at the entrance B. Moreover, it confirmed that the B22 simulation approached the 1ilh A geometry, with the phosphate groups directed toward W299 and the hydroxyl group toward the αAF helix. The remaining MD-Binding poses were not close to any experimental structures.

The similarity between the two maps in Figure 3.15 indicates that the method well sampled the whole conformational sub-space. The main difference was observed in the region nearby 1ilh C, that was highly sampled by SMD starting from the MD-Binding poses (Figure 3.15a), while it was poorly sampled starting from any X-ray structures (Figure 3.15b). This finding can however be explained considering the high number of MD-Binding poses falling in that region, representing similar ligand orientations. Another difference concerns the most sampled region in the two maps: while in SMD simulations starting from the X-ray structures it is close to 1nrl structure, in those starting from the MD-Binding poses it is shifted of about 20°toward higher θ values.

We extracted a limited number of poses from the whole ensemble representing the most sampled regions of the map using cluster analysis. The seven most populated clusters, out of the 227 obtained, well represent the regions with high conformational probability in the CVs sub-space (Figure 3.16).

In particular, the three most populated clusters contain ligand orientations similar to those observed in three experimental structures

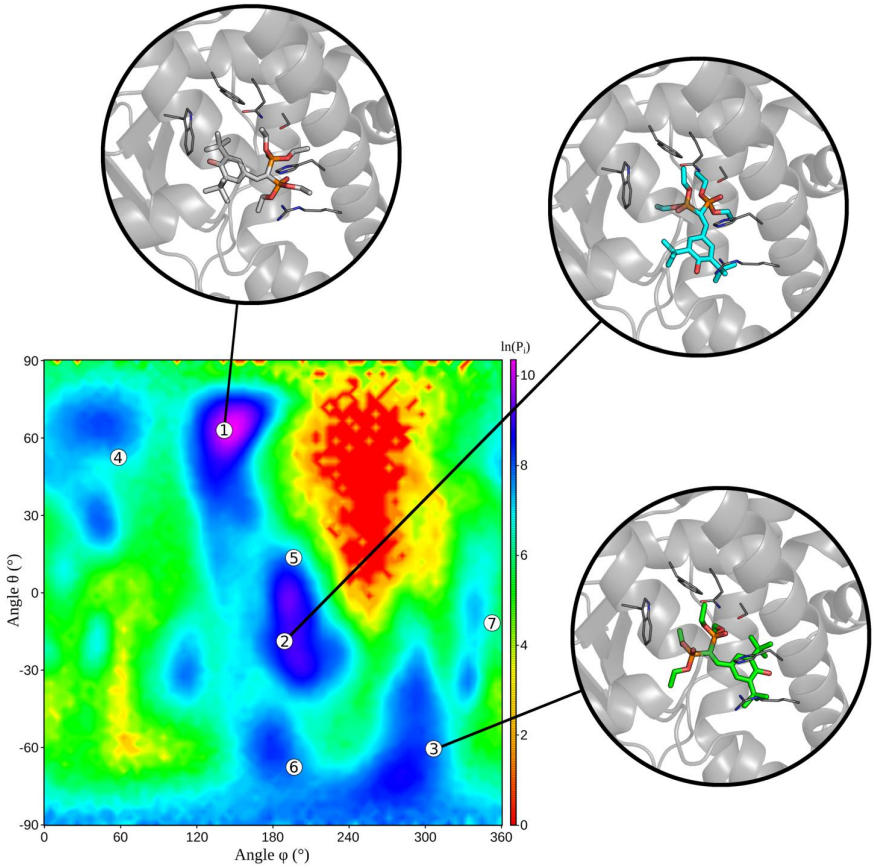

Figure 3.16 Centroids of the seven most populated clusters projected onto the conformational state probability surface obtained by SMD simulations starting from the MD-Binding poses. For the three most populated cluster, the 3D structure of the cluster centroid is reported.

(1nrl, 1ilh C, and 4j5x, respectively). Despite the dRMSD from the X-ray structures registered for the centroid of the most populated clusters were not very low, at least one conformation close to each of the X-ray ones (dRMSD < 1.5 Å) was found within the whole set of cluster centroids.

In conclusion, the investigation of the binding mechanism of the most studied PXR ligand, SR12813, here reported allowed elucidation of a number of important methodological issues. The intrinsic flexibility of the domain involved in ligand binding and the buried nature of the cavity, as well as the lack of information about the ligand path of entrance, make this system a challenging task. To obtain a mechanistic description of the association process, we suggest the use of MD-Binding method to simulate the ligand entrance into the cavity and SMD simulations to extend the sampling of the bound states and to evaluate their relative stability.

The combination of different methods can produce a complete picture of the binding event. Overall, the obtained results provide several insights into the binding mechanism of SR12813. The ligand preferentially enters the binding cavity through the B entrance, between the $\alpha 2$ and $\alpha 6$ helices. We speculate that the rate-determining step of the binding process is the breakdown of the salt bridge between E321 and R410 by the ligand. Moreover we can consider the 1ilh C geometry as a relative minimum in the free energy surface, but the 1nrl one as the most energy stable (in addition to our SMD calculations, this particular ligand orientation was the only one found in two PDB depositions, obtained from two different experimental groups).

It was proposed that the two protein partners co-crystallized with PXR in some of the available structures might play a role in the stabilization of a particular conformation of the SR12813-PXR complex [Watkins et al., 2003b]. In fact, the 1nrl/3hvl binding geometry was found in presence of SRC-1 and the 4j5x geometry corresponds to PXR bound to both SRC-1 and RXR, while the most miscellaneous deposition 1ilh, that depicts a more flexible image of the SR12813 binding, is associated to the PXR LBD crystallized without protein partners.

Our simulations were based on PXR LBD structures with no partners, and accordingly they describe the ligand binding event as characterized by high flexibility and plasticity of the protein domain. Our results provide further insights in addition to the hypotheses based on experimental data. They support the presence of multiple pre-existing conformational states of the SR12813-bound LBD also in absence of the protein partners. This could be considered as a particular case of conformational-selection, in which RXR or SRC-1 can select one among these different metastable states of the complex thus shifting the dynamic population equilibrium toward a specific bound state. The analysis of the multiple accessible states detected in our simulations can explain the different binding modes observed by X-ray crystallog-

raphy for the SR12813 ligand and support the hypothesis that the 1ilh binding geometries are kinetically favored states, while 1nrl represent the most thermodynamically favored state.

The approaches here proposed to study ligand binding to PXR effectively treated the dynamics of this system during binding and shed light on some of the unresolved mechanistic issues. On the basis of this positive outcome, these methods appear suitable for analyzing the mechanistic features of other ligand binding processes involving promiscuous protein domains.

This work was published in Motta et al. [2018a].

4

CONCLUSIONS

During my PhD, I applied molecular modeling methods to face environmental relevant problems. These methods were used to describe biological systems and processes at a molecular level, in particular ligand binding to xenobiotic receptors.

Xenobiotic receptors are highly promiscuous, because they have to bind a large variety of chemicals in order to eliminate toxic compounds, which can be very diverse. Promiscuity is often achieved increasing protein flexibility and plasticity.

In case of AhR, we modeled the LBD by homology modeling and managed to include flexibility during both the docking step, by the ensemble docking approach using different homology models, and the post-docking step, by a short MD simulation (section 2.3.5). This new protocol allowed the identification of three groups of ligands, each binding in a specific site inside the cavity. The differences observed in the ligand-protein interactions could result in differential effects downstream in the AhR signaling pathway, thus these findings could help to explain the toxicity of some agonists.

In case of PXR, given that many experimental structures are available, we included flexibility using the ensemble docking approach with different X-ray structures (section 3.3.1). The comparison between docking and ensemble docking results confirmed that it is necessary to include a variety of protein conformations, but that often these structures are biased toward the native co-crystallized ligand geometry. In contrast, the explicit inclusion of flexibility using MD-based methods with enhanced sampling exhibited no initial bias and lead us to identify the entrance path of ligands to the PXR binding cavity and to rationalize the multiple SR12813 binding modes (section 3.3.2).

Promiscuous receptors often display species-specificity; the modeling of these protein structures shed light on the molecular determinants of the observed different responses among different species. We built the first homology models of the invertebrate *C. elegans* and the amphibian *G. multiplicata* AhR LBDs. The first one showed a peculiar internal cavity, that is probably unable to bind any of the classical AhR ligands. The second helped in elucidating that the low-sensitivity of amphibians to TCDD arose before their divergence from the common lineage. Finally, the direct comparison between mouse/rat and human AhR LBDs highlighted the differences and similarities in binding to the different receptors of both agonists and SAhRMs.

The goal of computationally predict the potential activity of a ligand given a protein sequence or structure is clearly a big challenge. We

already pointed out some known limitations of *in silico* approaches (section 1.2), but we should also consider that function itself is not simply a product of the laws of physics and chemistry, which we can model, but primarily of those of evolution, which remain harder to predict.

Despite these limitations, we observed that homology modeling can give structural information that greatly helps in the understanding of species-specific AhRs features. Instead, the application of molecular docking on homology models remains a challenging task, with problems already pointed out. The inclusion of protein flexibility in docking calculations improved the reliability of the results, but the lack of an experimental structure of the LBD still hampers a detailed molecular description. The results obtained using the X-ray structures for PXR underline the difference between using a homology model and a crystal structure. Finally, the use of enhanced sampling techniques in MD simulations of the binding process can give, more than other methods, mechanistic explanation of the entrance path of the ligand into the binding site.

We can conclude that molecular modeling, coupled with experimental techniques, is extremely useful to give mechanistic insights about biological and toxicological events. The elucidation of these events will greatly improve the possible application of *in silico* methods in ecological and human risk assessment. Moreover, the understanding of the mechanism of action could contribute to a better estimation of interspecies translation between animal models (both *in vivo* and *in vitro*) and human, increasing the reliability of prediction based on these data.

A

APPENDIX

A.1 BINDING OF DIVERSE LIGANDS TO DIFFERENT AHRS

A.1.1 *Experimental session*

Chemicals

TCDD was obtained from Dr. Stephen Safe (Texas A&M University), [^{3}H]TCDD (14.3 Ci/mmol) was obtained from ChemSyn Laboratories (Lenexa, Kansas), and TCDF, PCB126 and DBA were from Accustandard (New Haven, CT). [^{32}P]-ATP (~6,000 Ci/mmol) was from Perkin Elmer Life & Analytical Sciences. The structures of the specific AhR ligands used in these studies are shown in Figure 2.15. 3MC, BNF, LEFL, and dimethyl sulfoxide (DMSO) were from Sigma-Aldrich (St. Louis, MO) and FICZ were from Tocris Bioscience (Minneapolis, MN) and Indirubin (IR) from AmplaChem (Carmel, IN). All chemical stocks and dilutions were prepared in DMSO. Please refer to Table A.1 for IUPAC nomenclature.

Plasmids

The mAhR and mouse ARNT expression plasmids, mβAhR/pcDNA3 and mβARNT/pcDNA3, have been previously described [Soshilov and Denison, 2014]. Point mutations of mβAhR/pcDNA3 were carried out using the QuikChange Lightning Mutagenesis Kit (Agilent Technologies) and all constructs were verified by sequencing.

Hydroxyapatite [^{3}H]TCDD Ligand Binding Assay

[^{3}H]TCDD specific binding to wild-type or mutant mAhRs synthesized *in vitro* using the Promega TNT Quick coupled transcription/translation rabbit reticulocyte lysate kit (Madison, WI) was carried out as previously described [Soshilov and Denison, 2014]]. [^{3}H]TCDD specific binding was determined by subtracting the amount of [^{3}H]TCDD bound to unprogrammed lysate (nonspecific binding) from the total amount of [^{3}H]TCDD binding to lysate containing *in vitro* expressed AhR. The amount of [^{3}H]TCDD specific binding remaining in the presence of the indicated competitor chemical was expressed as a percent of the total [^{3}H]TCDD specific binding. The relative binding affinity (IC_{50}) of each chemical was determined from concentration-dependent competitive inhibition curves obtained using [^{3}H]TCDD and in creasing concentrations of each test chemical and the mean IC_{50} value was determined using three-parameter non-linear regression.

AhR DNA Binding (Gel Retardation) Assay

Wild-type and mutant mAhRs and mARNT were synthesized *in vitro* in the presence of unlabeled L-methionine, the resulting mAhRs and mARNT translation mixtures and MEDGK (25 mM MOPS (3-(N-morpholino)propanesulfonic acid; pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, 10% (v/v) glycerol, 150 mM KCl) were mixed in a 1:1:8 (v/v/v) ratio and incubated with DMSO (1% final concentration) or the indicated concentration of TCDD or test chemical for 2 hours at room temperature. An aliquot of each incubation was mixed with [^{32}P]-labeled double-stranded oligonucleotide containing the mouse AhR DNA binding site XRE3 (a dioxin responsive element from the upstream region of the murine *cyp1a1* gene [Denison et al., 1988] and protein-DNA complexes resolved by gel retardation analysis as previously described in detail [Soshilov and Denison, 2014]. Gels were visualized using a FLA9000 Fujifilm Imager and protein-DNA complexes quantitated with Fujifilm MultiGauge software. For potency measurements, reactions were incubated with increasing concentrations of the test chemical and the chemical concentration producing half maximal AhR:ARNT:XRE complex formation (EC_{50}) determined.

Transactivation assay

Transactivation assays were performed in the Kenyon College, Gambier, OH (USA) by the Prof. Powell laboratories. COS-7 cells were transfected using Lipofectamine 2000 (Invitrogen). Cells were co-transfected with 13.4 ng of pGudLuc6.1 (reporter construct), 1.0 ng of Renilla (transfection control construct), 33.4 ng of *X. laevis* ARNT and AhR expression constructs. AhR protein expression was detected using polyclonal antibody SA-210 (1:500, Enzo) and five hours after transfection, cells were treated with TCDD, FICZ, or DMSO. Luminescence was measured 18 hours after dosing. Nonlinear regression curves of the average fractional induction values were used to calculate EC_{50} values (n=3).

DNA binding (gel retardation) assay

The DNA binding assays were performed in the Dept. of Environmental Toxicology, University of California, Davis, CA (USA) by the Prof. Denison research team.

Wild-type and mutant AhRs and ARNT were synthesized *in vitro* in the presence of unlabeled l-methionine, the resulting AhR and ARNT translation mixtures and MEDGK (25 mM MOPS (3-(N-morpholino) propanesulfonic acid; pH 7.5), 1 mM EDTA, 1 mM dithiothreitol, 10% (v/v) glycerol, 150 mM KCl) were mixed in a 1:1:8 (v/v/v) ratio and incubated with DMSO (1% final concentration) or the indicated concentration of TCDD or IR for the indicated periods of time at room temperature. An aliquot of each incubation was mixed with [^{32}P]-labeled

double-stranded oligonucleotide containing the AhR-ARNT XRE3 DNA binding site (a dioxin responsive element from the upstream region of the murine *cyp1a1* gene and protein-DNA complexes resolved by gel retardation analysis. Gels were visualized using a FLA9000 Fujifilm Imager and protein-DNA complexes quantitated with Fujifilm MultiGauge software (FujiFilm Corporation, Valhalla, NY, USA).

Transient transfection assays

The transient transfection assays were performed in the Dept. of Environmental Toxicology, University of California, Davis, CA (USA) by the Prof. Denison research team.

COS-1 cells were plated at a density of 75,000 cells/well and allowed to attach overnight in a 24-well plate. Cells were transiently transfected (per well) with the following amounts per well: 2 μL Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA), 20 ng wild-type (mβAhR/pcDNA3) or mutant AhR expression plasmids (in pcDNA3.1), 20 ng of mβARNT/pcDNA3 or pcDNA3.1(+) and 200 ng pGudLuc6.1. Twenty-four hours after transfection, cells were incubated with DMSO (0.1%, v/v), TCDD (10 nM final concentration) or IR (1 μM final concentration) for 18 to 22 h, followed by washing with PBS, lysis and measurement of luciferase activity in 50 μL aliquots.

A.1.2 *Figures and tables*

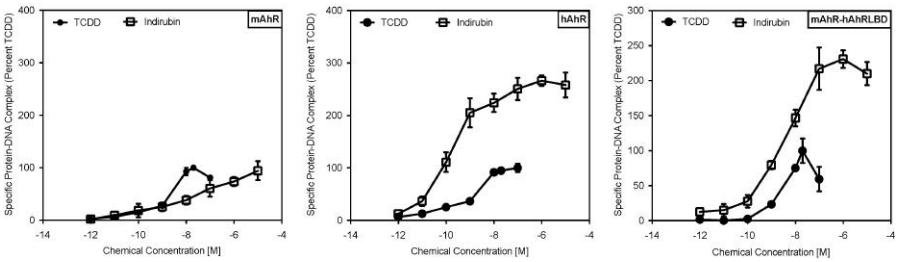

Figure A.1 The human AhR LBD plays a major role in ligand-selective activation by IR

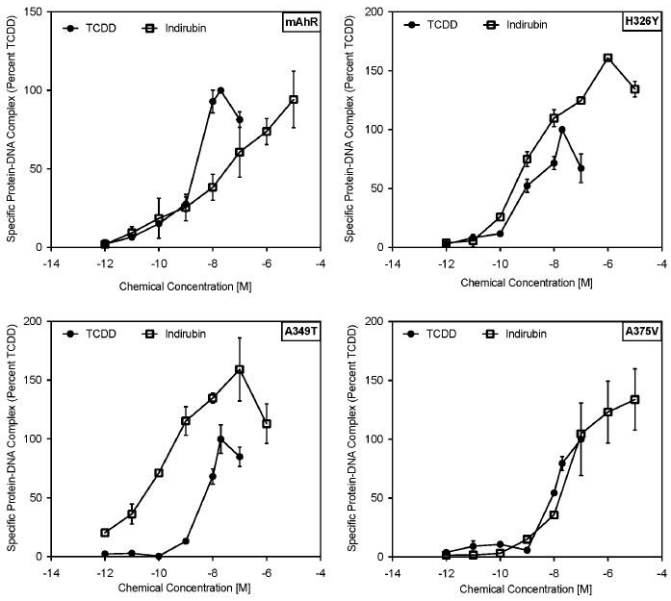

Figure A.2 Enhanced signaling of the hAhR by IR is associated with selected residues within the hAhR PASB LBD. In vitro synthesized mutant mAhRs and wild-type mARNT were incubated in the presence of DMSO (1%, v/v) or AhR agonists in DMSO (TCDD (0.001-100 nM) or IR (0.001-10,000 nM) for 2 h, followed by analysis of AhR DNA binding by gel retardation analysis. The amount of inducible protein-DNA complex at each TCDD or IR concentration was quantitated, and values normalized to the amount of complex formed with a maximal activating concentration of TCDD (20 nM)

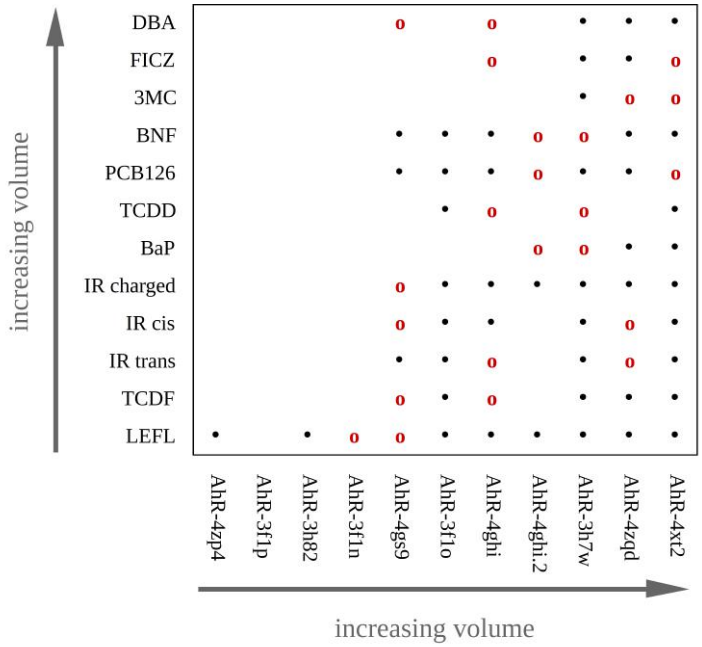

Figure A.3 Docking matrix; each sign indicates that a pose was found for that ligand in that model. Red rounds stand for the two poses chosen as representative of the binding mode variability of each ligand. Ligands and cavities are ordered by increasing occupancy or volume, respectively.

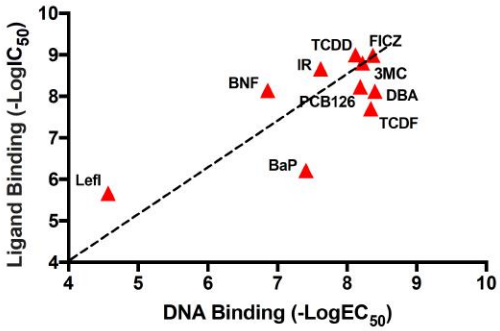

Figure A.4 Relationship between the relative affinity ($LogIC_{50}$) of each test compound to bind to the AhR and their relative potency to stimulate AhR DNA binding ($LogEC_{50}$).

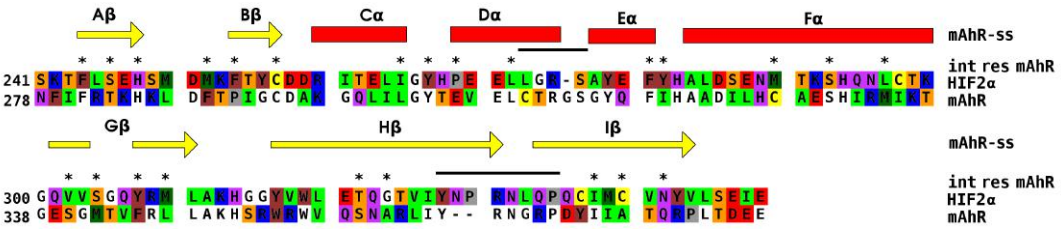

Figure A.5 Sequence alignment between HIF-2α and mAhR used for Homology Modeling. Amino acid residues are colored by physico-chemical properties, secondary structures attributed by DSSPcont to the AhR-3h7w model are reported. Internal residues are marked by asterisks and residues subjected to loop-modeling refinement (Dα-Eα: 310-314 and Hβ-Iβ: 365-370) are underlined.

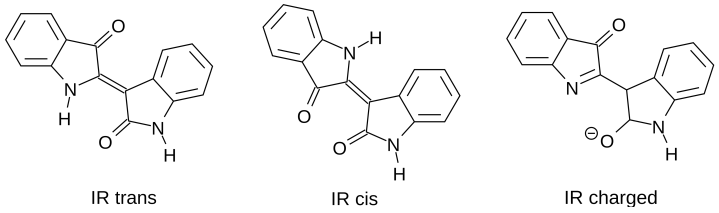

Figure A.6 The three possible forms of IR at pH = 7 found by Epik.

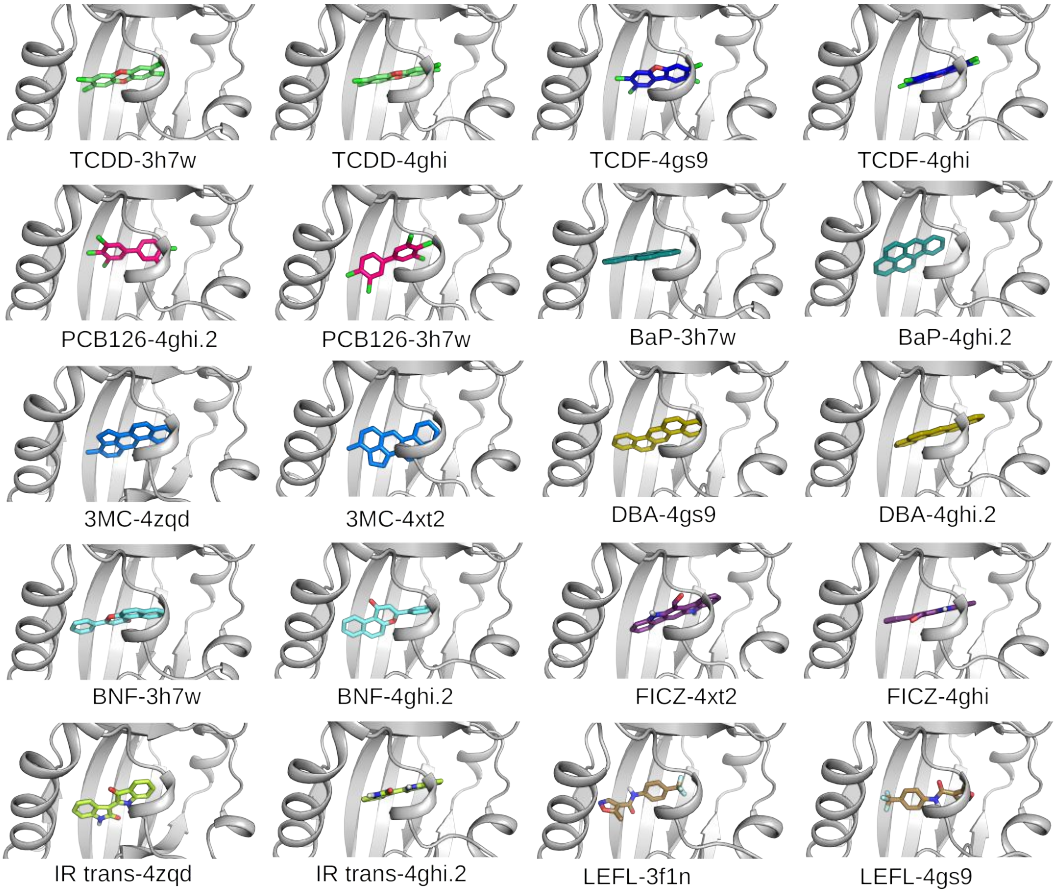

Figure A.7 Three-dimensional representation of the two representative docking poses selected for each ligand. Protein is shown as cartoons, ligands as sticks.

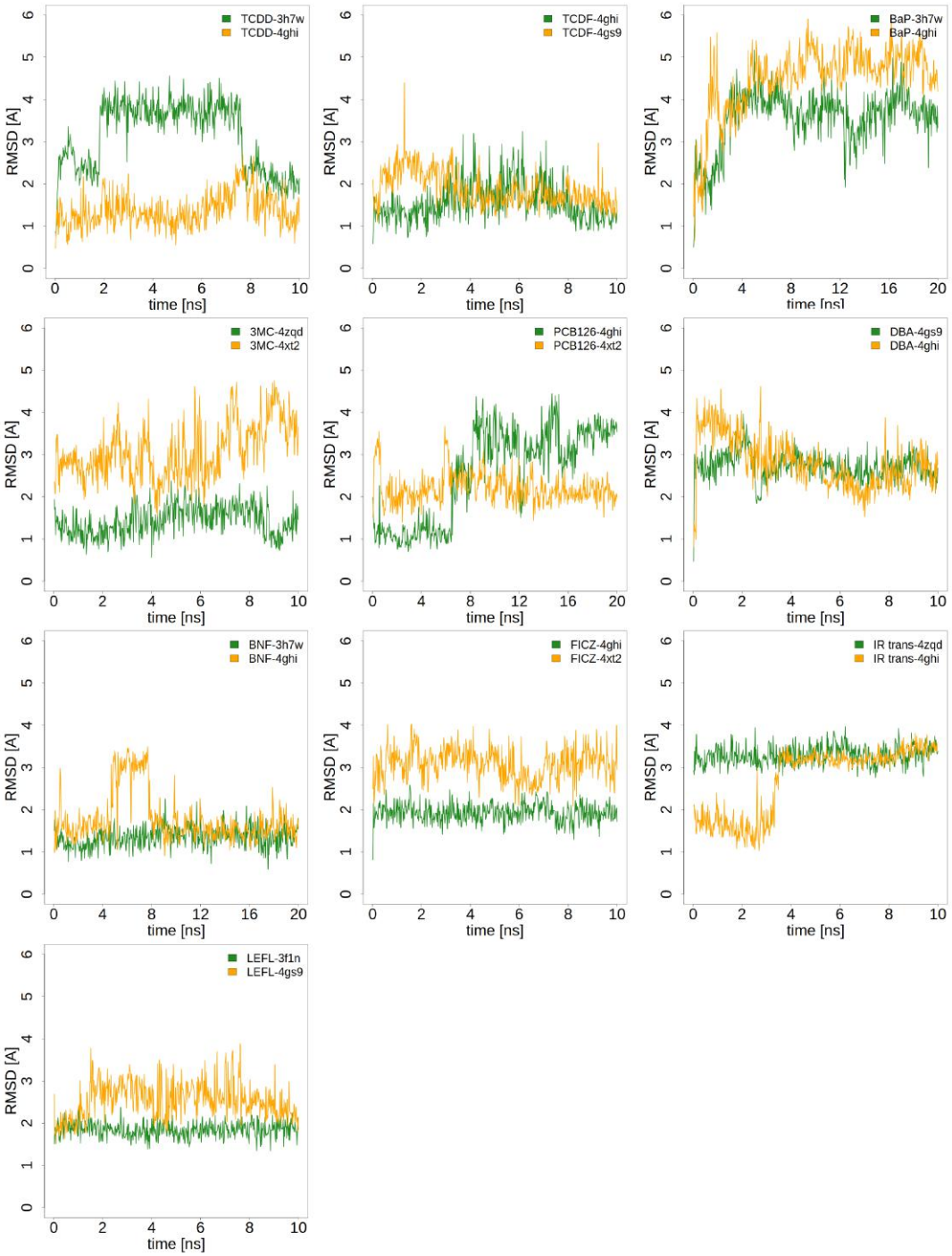

Figure A.8 RMSD graphs in the MD simulations of the two representative poses of each ligand (the pose with the lowest ΔG_{bind} is shown in green). RMSD is calculated against the initial docking pose.

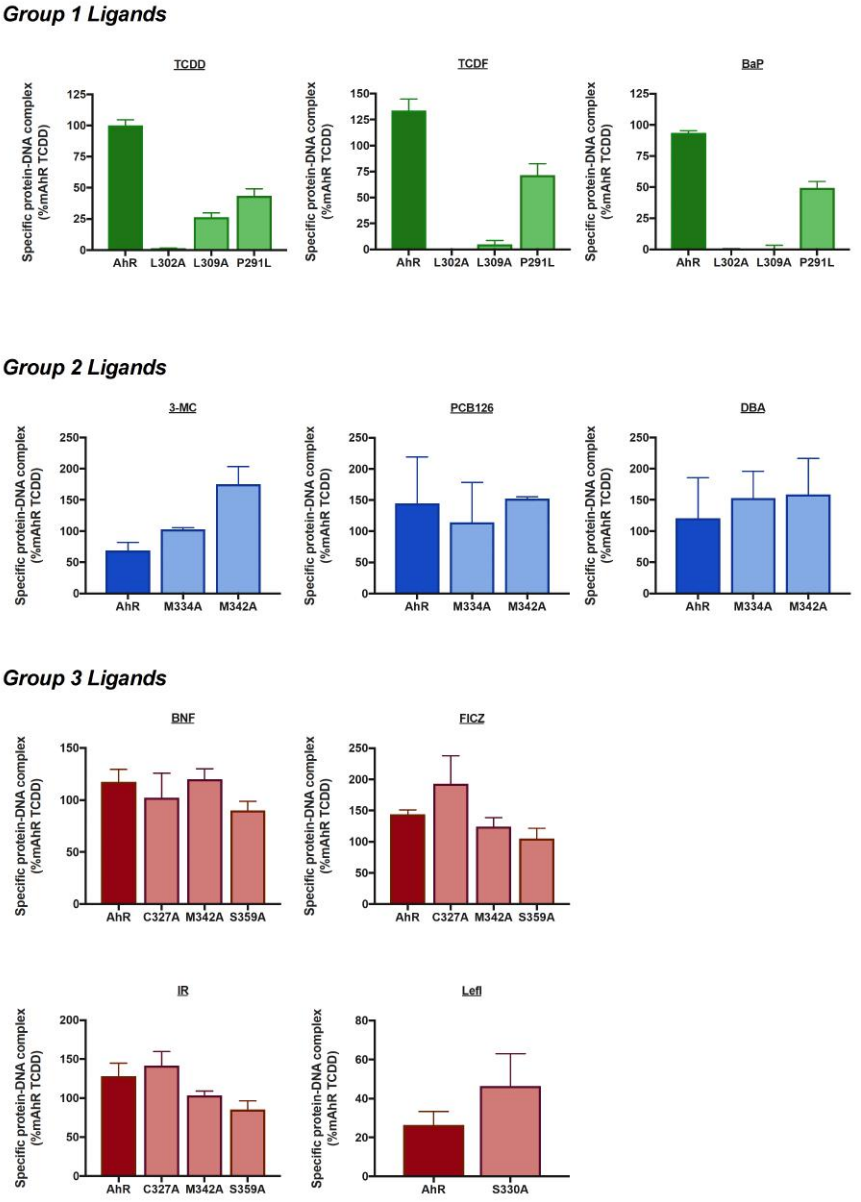

Figure A.9 Relative ability of test chemicals to stimulate AhR transformation/DNA binding of wild-type and mutant AhRs. Single concentrations of AhR ligands were used to assess AhR transformation/DNA binding measured as specific protein-DNA complex (% mAhR TCDD) by Gel Retardation Analysis as described in the Material and Methods section. Maximal concentrations of AhR ligands were used and include BNF (10 μM), BaP (10 μM), 3MC (10 μM), IR (1 μM), DBA (10 μM), FICZ (1 μM), PCB126 (100 nM), LEFL (100 μM), TCDD (20 nM), and TCDF (100 nM). Values represent the Three-parameter non-linear regression performed with nine independent reactions for mean specific protein-DNA complex (%mAhR TCDD) ± standard deviation of two independent reactions.

Table A.1 Ligands used in the study. Common names, IUPAC names and CID are reported.

Ligand id	name	IUPAC name	CID
TCDD	tetrachlorodibenzo dioxin	2,3,7,8-tetrachlorodibenzo-p-dioxin	15625
TCDF	tetrachlorodibenzofuran	2,3,7,8-tetrachlorodibenzofuran	39929
PCB126	polychlorobiphenyl	1,2,3-trichloro-5-(3,4-dichlorophenyl) benzene	63090
BaP	benzo[a]pyrene	benzo[a]pyrene	2336
3MC	3-methylcholanthrene	3-methyl-1,2-dihydrobenzo[j]aceanthrylene	1674
DBA	dibenzanthracene	naphtho[1,2-b]phenanthrene	5889
BNF	beta-naphthoflavone	3-phenylbenzo[f]chromen-1-one	2361
FICZ	6-formylindolo carbazole	5,11-dihydroindolo[3,2-b]carbazole-12carbaldehyde	1863
IR trans	indirubin	(3Z)-3-(3-oxo-1H-indol-2-ylidene)-1H-indol-2-one	5318433
LEFL	leflunomide	5-methyl-N-[4-(trifluoromethyl)phenyl]-1,2-oxazole-4-carboxamide	3899

Table A.2 Relative affinity and potency of test chemicals determined in AhR ligand binding (AhR LB) and DNA binding (AhR DNAB) assays. The relative potency of each test chemical for AhR ligand binding (IC_{50}) and transformation/DNA binding (EC_{50}) was determined from concentration-inhibition or concentration-response curves obtained using [3H]TCDD ligand binding and gel retardation analysis, respectively, as described under Materials and Methods. The mean IC_{50}/EC_{50} $\pm$ standard deviation (sd) was determined using three-parameter non-linear regression with nine independent reactions.

Ligand id	LB IC_{50} [M]	LB sd [M]	LB -LogIC$_{50}$	DNAB EC_{50} [M]	DNAB sd [M]	DNAB -LogEC$_{50}$
TCDD	1.00E-09	-	9.00	7.53E-09	1.10E-09	8.12
TCDF	2.01E-08	2.24E-09	7.70	4.55E-09	5.37E-10	8.34
PCB126	5.89E-09	3.29E-09	8.23	6.51E-09	9.45E-10	8.19
BaP	6.17E-07	8.41E-08	6.21	3.89E-08	4.65E-09	7.41
3MC	1.58E-09	1.27E-09	8.80	6.00E-09	4.87E-10	8.22
DBA	7.60E-09	4.20E-09	8.12	4.02E-09	3.80E-10	8.40
BNF	7.24E-09	2.30E-09	8.14	1.38E-07	3.06E-08	6.86
FICZ	1.02E-09	7.86E-10	8.99	4.25E-09	5.85E-10	8.37
IR	2.17E-09	1.02E-09	8.66	2.39E-08	9.90E-09	7.62
LEFL	2.19E-06	1.03E-06	5.66	2.68E-05	1.75E-06	4.57

Table A.3 Ligands in the holo X-ray structures of HIF-2α used as templates for homology modeling.

ligand id	name	IUPAC name	CID	PDB id
2XY	THS-044	N-[2-nitro-4-(trifluoromethyl)-phenyl]morpholin-4-amine	12318718	3f10
018	THS-017	2-nitro-N-(thiophen-3-ylmethyl)-4-(trifluoromethyl)aniline	43805017	3h7w
020	THS-020	N-(furan-2-ylmethyl)-2-nitro-4-(trifluoromethyl)aniline	4080866	3h82
0X3	0X3	N-(3-chloro-5-fluorophenyl)-4-nitro-2,1,3-benzoxadiazol-5-amine	70697712	4ghi
0XB	0XB	N-(3-fluorophenyl)-4-nitro-2,1,3-benzoxadiazol-5-amine	2836648	4gs9
43L	43L	(5S,7R)-5,7-bis(3-bromophenyl)-1,5,6,7-tetrahydrotetrazolo[1,5-a]pyrimidine	51551597	4xt2
0X3	0X3	N-(3-chloro-5-fluorophenyl)-4-nitro-2,1,3-benzoxadiazol-5-amine	70697712	4zqd

Table A.4 Comparison of the internal cavity volumes (calculated by CASTp) in the HIF2α template structures and in the mAhR models.

PDB id	resolution [Å]	vol [Å^3]	ligand id	model name	vol [Å^3]	DOPE score
4zp4	2.35	375	water	AhR-4zp4	417	-1146.24
3f1p	1.17	364	water	AhR-3f1p	478	-1222.75
3h82	1.50	439	020	AhR-3h82	489	-1216.72
3f1n	1.48	383	EDO	AhR-3f1n	491	-1213.07
4gs9	1.72	424	0XB	AhR-4gs9	533	-1192.83
3f10	1.59	425	2XY	AhR-3f10	568	-1186.26
4ghi	1.50	443	0X3	AhR-4ghi	575	-1160.52
4ghi	1.50	443	0X3	AhR-4ghi.2	629	-1159.28
3h7w	1.65	421	018	AhR-3h7w	689	-1283.28
4zqd	2.87	569	0X3	AhR-4zqd	712	-1153.25
4xt2	1.69	561	43L	AhR-4xt2	808	-1187.60

Table A.5 Validation scores of PROCHECK and ProSA for the mAhR models.

model	PROCHECK G-factor	ProSA Z-Score
AhR-4zp4	-0.15	-0.15
AhR-3f1p	-0.14	-0.14
AhR-3h82	-0.30	-0.30
AhR-3f1n	-0.19	-0.19
AhR-4gs9	-0.25	-0.25
AhR-3f10	-0.26	-0.26
AhR-4ghi	-0.35	-0.35
AhR-4ghi.2	-0.35	-0.35
AhR-3h7w	-0.22	-0.22
AhR-4zqd	-0.24	-0.24
AhR-4xt2	-0.36	-0.36

Table A.6 XP GlideScore for all the docking poses. The two selected representative poses are highlighted in light-gray.

ligand	model	XP Score	ligand	model	XP Score
TCDD	AhR-4ghi	-7.9	BNF	AhR-4zqd	-8.8
	AhR-3h7w	-7.3		AhR-3h7w	-8.7
	AhR-4xt2	-6.9		AhR-4gs9	-8.3
	AhR-3f10	-4.9		AhR-4xt2	-8.2
TCDF	AhR-4ghi	-8.5		AhR-4ghi.2	-7.5
	AhR-4zqd	-7.6		AhR-4ghi	-7.1
	AhR-4gs9	-7.5		AhR-3f10	-6.5
	AhR-3h7w	-7.1	FICZ	AhR-4ghi	-10.1
	AhR-3f10	-6.2		AhR-3h7w	-8.1
	AhR-4xt2	-6.1		AhR-4xt2	-7.8
PCB126	AhR-4ghi.2	-8.5		AhR-4zqd	-7.5
	AhR-4ghi	-8.1	IR trans	AhR-3h7w	-8.0
	AhR-4zqd	-7.5		AhR-4xt2	-7.5
	AhR-3h7w	-7.4		AhR-4zqd	-7.0
	AhR-4gs9	-7.4		AhR-4ghi	-6.9
	AhR-4xt2	-6.6		AhR-4gs9	-6.9
	AhR-3f10	-4.7		AhR-3f10	-6.3
BaP	AhR-3h7w	-8.8	LEFL	AhR-4ghi	-9.1
	AhR-4zqd	-7.6		AhR-3h82	-8.1
	AhR-4ghi.2	-7.4		AhR-4zqd	-7.4
	AhR-4xt2	-6.5		AhR-4gs9	-6.4
3MC	AhR-4zqd	-8.3		AhR-4ghi.2	-6.3
	AhR-3h7w	-7.9		AhR-3f10	-6.0
	AhR-4xt2	-7.5		AhR-3f1n	-5.7
DBA	AhR-3h7w	-9.2		AhR-3h7w	-5.6
	AhR-4xt2	-8.8		AhR-4xt2	-5.4
	AhR-4gs9	-8.6		AhR-4zp4	-4.9
	AhR-4zqd	-8.4			
	AhR-4ghi	-8.3			

Table A.7 MM-GBSA ΔG_{bind} mean values in the last 8 ns of MD simulations for the selected docking poses. The pose with the lowest ΔG_{bind} value is highlighted in light gray

ligand	ΔG_{bind} mean [kcal mol^{-1}]	sd [kcal mol^{-1}]
TCDD-3h7w	-36.8	1.9
TCDD-4ghi	-34.8	1.8
TCDF-4ghi	-34.0	1.8
TCDF-4gs9	-32.7	1.9
PCB126-4ghi.2	-38.0	2.0
PCB126-4xt2	-36.1	1.9
BaP-4ghi.2	-34.7	2.3
BaP-3h7w	-30.0	2.1
3MC-4zqd	-38.3	3.0
3MC-4xt2	-31.0	2.3
DBA-4gs9	-37.6	2.6
DBA-4ghi	-34.1	1.9
BNF-3h7w	-38.7	2.0
BNF-4ghi.2	-37.0	2.4
FICZ-4ghi	-40.4	2.3
FICZ-4xt2	-35.9	3
IR trans-4zqd	-39.5	3.0
IR trans-4ghi	-34.0	4.0
LEFL-3f1n	-36.6	2.0
LEFL-4gs9	-31.0	2.0

Table A.8 Relative ligand binding affinity of test chemicals for mutant AhRs. The relative affinity (IC_{50}) of each test chemical for the AhR ligand was determined from concentration-dependent inhibition curves obtained using [^{3}H]TCDD ligand binding analysis, as described under Materials and Methods. The mean IC_{50} $\pm$ standard deviation was determined using three-parameter non-linear regression with nine independent reactions.

AhR Mutant	AhR Ligand	IC_{50} [M]	SD [M]	-LogIC_{50}
P291L	TCDF	4.35E-11	2.20E-10	10.36
	BaP	6.04E-07	1.82E-07	6.22
	3MC	2.48E-07	2.93E-08	6.61
	PCB126	9.33E-09	1.76E-07	8.03
	DBA	6.72E-09	1.13E-08	8.17
M342A	BNF	9.60E-09	4.38E-09	8.02
	FICZ	2.07E-11	5.86E-11	10.68
	IR	1.26E-08	4.60E-09	7.90
	LEFL	9.03E-09	2.74E-08	8.04
	3MC	3.47E-08	1.74E-08	7.46
M334A	PCB126	7.41E-08	5.79E-09	7.13
	DBA	7.33E-09	1.56E-08	8.13
H331A	3MC	1.41E-10	3.83E-11	9.85
S330A	LEFL	6.29E-06	1.64E-06	5.20
	BNF	4.71E-08	3.75E-10	7.33
S359A	FICZ	4.10E-09	1.30E-09	8.39
	IR	9.20E-08	1.98E-09	7.04
	LEFL	non-converged	non-converged	non-converged
	DBA	1.46E-08	3.15E-09	7.83
C327A	BNF	4.05E-08	8.92E-09	7.39
	FICZ	4.32E-09	2.59E-08	8.36
	IR	7.58E-09	2.16E-08	8.12

BIBLIOGRAPHY

Abiko, Y., F.-Y. Lin, H. Lee, A. Puga, and Y. Kumagai
2016. Quinone-mediated induction of cytochrome P450 1A1 in HepG2 cells through increased interaction of aryl hydrocarbon receptor with aryl hydrocarbon receptor nuclear translocator. *J. Toxicol. Sci.*, 41(6):775–781.

Abiko, Y., A. Puga, and Y. Kumagai
2015. Covalent binding of quinones activates the Ah receptor in Hepa1c1c7 cells. *J. Toxicol. Sci.*, 40(6):873–886.

Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl
2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25.

Adcock, S. A. and J. A. McCammon
2006. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.*, 106(5):1589–1615.

Amaro, R. E., J. Baudry, J. Chodera, Ö. Demir, J. A. McCammon, Y. Miao, and J. C. Smith
2018. Ensemble Docking in Drug Discovery. *Biophys. J.*, 114(10):2271–2278.

Andreasen, E. A., M. E. Hahn, W. Heideman, R. E. Peterson, and R. L. Tanguay
2002. The zebrafish (Danio rerio) aryl hydrocarbon receptor type 1 is a novel vertebrate receptor. *Mol. Pharmacol.*, 62(2):234–249.

Ankley, G. T., R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, J. A. Serrrano, J. E. Tietge, and D. L. Villeneuve
2010. Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environ. Toxicol. Chem.*, 29(3):730–741.

Banerjee, M. and T. Chen
2013. Differential regulation of CYP3A4 promoter activity by a new class of natural product derivatives binding to pregnane X receptor. *Biochem. Pharmacol.*, 86(6):824–835.

Banerjee, M., D. Robbins, and T. Chen
2015. Targeting xenobiotic receptors PXR and CAR in human diseases. *Drug Discov. Today*, 20(5):618–628.

Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman
1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, 97(40):10269–10280.

Benson, J. M. and D. M. Shepherd
2011. Aryl hydrocarbon receptor activation by TCDD reduces inflammation associated with Crohn's disease. *Toxicol. Sci.*, 120(1):68–78.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak
1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne
2000. The Protein Data Bank. *Nucleic Acids Res.*, 28(1):235–242.

Bersten, D. C., A. E. Sullivan, D. J. Peet, and M. L. Whitelaw
2013. bHLH-PAS proteins in cancer. *Nat. Rev. Cancer*, 13(12):827–41.

Bessede, A., M. Gargaro, M. T. Pallotta, D. Matino, G. Servillo, C. Brunacci, S. Bicciato, E. M. C. Mazza, A. Macchiarulo, C. Vacca, R. Iannitti, L. Tissi, C. Volpi, M. L. Belladonna, C. Orabona, R. Bianchi, T. V. Lanz, M. Platten, M. a. Della Fazia, D. Piobbico, T. Zelante, H. Funakoshi, T. Nakamura, D. Gilot, M. S. Denison, G. J. Guillemin, J. B. DuHadaway, G. C. Prendergast, R. Metz, M. Geffard, L. Boon, M. Pirro, A. Iorio, B. Veyret, L. Romani, U. Grohmann, F. Fallarino, and P. Puccetti
2014. Aryl hydrocarbon receptor control of a disease tolerance defence pathway. *Nature*, 511(7508):184–190.

Bisson, W. H., D. C. Koch, E. F. O'Donnell, S. M. Khalil, N. I. Kerkvliet, R. L. Tanguay, R. Abagyan, and S. K. Kolluri
2009. Modeling of the aryl hydrocarbon receptor (AhR) ligand binding domain and its utility in virtual ligand screening to predict new AhR ligands. *J. Med. Chem.*, 52(18):5635–41.

Biswas, A., S. Mani, M. R. Redinbo, M. D. Krasowski, H. Li, and S. Ekins
2009. Elucidating the 'Jekyll and Hyde' nature of PXR: The case for discovering antagonists or allosteric antagonists. *Pharm. Res.*, 26(8):1807–1815.

Blumberg, B., W. Sabbagh, H. Juguilon, J. Bolado, C. M. Van Meter, E. S. Ong, and R. M. Evans
1998. SXR, a novel steroid and xenobiotic-sensing nuclear receptor. *Genes Dev.*, 12(20):3195–3205.

Bochevarov, A. D., E. Harder, T. F. Hughes, J. R. Greenwood, D. A. Braden, D. M. Philipp, D. Rinaldo, M. D. Halls, J. Zhang, and R. A. Friesner
2013. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.*, 113(18):2110–2142.

Boehr, D. D., R. Nussinov, and P. E. Wright
2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.*, 5(11):789–796.

Bonati, L., D. Corrada, S. Giani Tagliabue, and S. Motta
2017. Molecular modeling of the AhR structure and interactions can shed light on ligand-dependent activation and transformation mechanisms. *Curr. Opin. Toxicol.*, 2:42–49.

Bordogna, A., A. Pandini, and L. Bonati
2010. Predicting the accuracy of protein-ligand docking on homology models. *J. Comput. Chem.*, 32(1):81–98.

Borrelli, K. W., A. Vitalis, R. Alcantara, and V. Guallar
2005. PELE: Protein energy landscape exploration. A novel Monte Carlo based technique. *J. Chem. Theory Comput.*, 1(6):1304–1311.

Bussi, G., D. Donadio, and M. Parrinello
2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101.

Carter, P., C. A. F. Andersen, and B. Rost
2003. DSSPcont: Continuous secondary structure assignments for proteins. *Nucleic Acids Res.*, 31(13):3293–5.

Case, D. a., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods
2005. The Amber biomolecular simulation programs. *J. Comput. Chem.*, 26(16):1668–88.

Chai, S. C., M. T. Cherian, Y. M. Wang, and T. Chen
2016. Small-molecule modulators of PXR and CAR. *Biochim. Biophys. Acta - Gene Regul. Mech.*, 1859(9):1141–1154.

Changeux, J.-P. and S. Edelstein
2011. Conformational selection or induced-fit? 50 years of debate resolved. *F1000 Biol. Rep.*, 3(September):19.

Cheng, Y. and M. R. Redinbo
2011. Activation of the human nuclear xenobiotic receptor PXR by the reverse transcriptase-targeted anti-HIV drug PNU-142721. *Protein Sci.*, 20(10):1713–1719.

Chrencik, J. E., J. Orans, L. B. Moore, Y. Xue, L. Peng, J. L. Collins, G. B. Wisely, M. H. Lambert, S. A. Kliewer, and M. R. Redinbo
2005. Structural disorder in the complex of human pregnane X receptor and the macrolide antibiotic rifampicin. *Mol. Endocrinol.*, 19(5):1125–1134.

Colizzi, F., R. Perozzo, L. Scapozza, M. Recanatini, and A. Cavalli
2010. Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc.*, 132(21):7361–7371.

Collier, A., L. Orr, J. Morris, and J. Blank
2008. The effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) on the mortality and growth of two amphibian species (Xenopus laevis and Pseudacris triseriata). *Int. J. Environ. Res. Public Health*, 5(5):368–377.

Corrada, D., M. S. Denison, and L. Bonati
2017. Structural modeling of the AhR:ARNT complex in the bHLH-PASA-PASB region elucidates the key determinants of dimerization. *Mol. Biosyst.*, 13(5):981–990.

Corrada, D., A. A. Soshilov, M. S. Denison, and L. Bonati
2016. Deciphering Dimerization Modes of PAS Domains: Computational and Experimental Analyses of the AhR:ARNT Complex Reveal New Insights Into the Mechanisms of AhR Transformation. *PLoS Comput. Biol.*, 12(6).

Darden, T., D. York, and L. Pedersen
1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092.

Davis, I. W. and D. Baker
2009. RosettaLigand Docking with Full Ligand and Receptor Flexibility. *J. Mol. Biol.*, 385(2):381–392.

De Vivo, M., M. Masetti, G. Bottegoni, and A. Cavalli
2016. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.*, 59(9):4035–4061.

Decherchi, S., A. Berteotti, G. Bottegoni, W. Rocchia, and A. Cavalli
2015. The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.*, 6(1):6155.

Decherchi, S., G. Bottegoni, A. Spitaleri, W. Rocchia, and A. Cavalli
2018. BiKi Life Sciences: A New Suite for Molecular Dynamics and Related Methods in Drug Discovery. *J. Chem. Inf. Model.*, 58(2):219–224.

Decherchi, S. and W. Rocchia
2013. A general and Robust Ray-Casting-Based Algorithm for Triangulating Surfaces at the Nanoscale. *PLoS One*, 8(4):e59744.

Delfosse, V., B. Dendele, T. Huet, M. Grimaldi, A. Boulahtouf, S. Gerbal-Chaloin, B. Beucher, D. Roecklin, C. Muller, R. Rahmani, V. Cavaillès, M. Daujat-Chavanieu, V. Vivat, J.-M. Pascussi, P. Balaguer, and W. Bourguet
2015. Synergistic activation of human pregnane X receptor by binary cocktails of pharmaceutical and environmental compounds. *Nat. Commun.*, 6(1):8089.

Denison, M. S., M. J. Fisher, and J. P. Whitlock
1988. The DNA Recognition Site for the Dioxin-Ah Receptor Complex. *J. Biol. Chem.*, Pp. 17221–17224.

Denison, M. S. and S. R. Nagy
2003. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu. Rev. Pharmacol. Toxicol.*, 43:309–34.

Denison, M. S., A. A. Soshilov, G. He, D. E. DeGroot, and B. Zhao
2011. Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci.*, 124(1):1–22.

di Masi, A., E. D. Marinis, P. Ascenzi, and M. Marino
2009. Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. *Mol. Aspects Med.*, 30(5):297–343.

Doering, J. A., R. Farmahin, S. Wiseman, S. C. Beitel, S. W. Kennedy, J. P. Giesy, and M. Hecker
2015. Differences in activation of aryl hydrocarbon receptors of white sturgeon relative to lake sturgeon are predicted by identities of key amino acids in the ligand binding domain. *Environ. Sci. Technol.*, 49(7):4681–4689.

Dror, R. O., R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw
2012. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.*, 41(1):429–452.

Dundas, J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, and J. Liang
2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.*, 34(Web Server issue):W116–8.

Eckers, A., S. Jakob, C. Heiss, T. Haarmann-Stemmann, C. Goy, V. Brinkmann, M. M. Cortese-Krott, R. Sansone, C. Esser, N. Ale-Agha, J. Altschmied, N. Ventura, and J. Haendeler
2016. The aryl hydrocarbon receptor promotes aging phenotypes across species. *Sci. Rep.*, 6(1):19618.

Edgar, R. C.
2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113.

Ekins, S., C. Chang, S. Mani, M. D. Krasowski, E. J. Reschly, M. Iyer, V. Kholodovych, N. Ai, W. J. Welsh, M. Sinz, P. W. Swaan, R. Patel, and K. Bachmann
2007. Human pregnane X receptor antagonists and agonists define molecular requirements for different binding sites. *Mol. Pharmacol.*, 72(3):592–603.

Ekins, S., S. Kortagere, M. Iyer, E. J. Reschly, M. A. Lill, M. R. Redinbo, and M. D. Krasowski
2009. Challenges predicting ligand-receptor interactions of promiscuous proteins: The nuclear receptor PXR. *PLoS Comput. Biol.*, 5(12):e1000594.

Eswar, N., B. Webb, M. A. Marti-Renom, M. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, and A. Sali
2006. Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinforma.*, 15(1):5.6.1–5.6.30.

Faber, S., A. Soshilov, S. Giani Tagliabue, L. Bonati, and M. Denison
2018. Comparative In Vitro and In Silico Analysis of the Selectivity of Indirubin as a Human Ah Receptor Agonist. *Int. J. Mol. Sci.*, 19(9):2692.

Fan, H., J. J. Irwin, and A. Sali
2012. Virtual Ligand Screening Against Comparative Models of Proteins. *Methods Mol. Biol.*, 819:105–126.

Fan, H., J. J. Irwin, B. M. Webb, G. Klebe, B. K. Shoichet, and A. Sali
2009. Molecular Docking Screens Using Comparative Models of Proteins. *J. Chem. Inf. Model.*, 49(11):2512–2527.

Farmahin, R., G. E. Manning, D. Crump, D. Wu, L. J. Mundy, S. P. Jones, M. E. Hahn, S. I. Karchner, J. P. Giesy, S. J. Bursian, M. J. Zwiernik, T. B. Fredricks, and S. W. Kennedy
2013. Amino acid sequence of the ligand-binding domain of the aryl hydrocarbon receptor 1 predicts sensitivity of wild birds to effects of dioxin-like compounds. *Toxicol. Sci.*, 131(1):139–152.

Feher, M. and C. I. Williams
2009. Effect of Input Differences on the Results of Docking Calculations. *J. Chem. Inf. Model.*, 49(7):1704–1714.

Fiser, A., R. K. G. Do, and A. Sali
2000. Modeling of loops in protein structures. *Protein Sci.*, 9(9):1753–1773.

Fraccalvieri, D., A. Pandini, F. Stella, and L. Bonati
2011. Conformational and functional analysis of molecular dynamics trajectories by Self-Organising Maps. *BMC Bioinformatics*, 12(1):158.

Fraccalvieri, D., A. A. Soshilov, S. I. Karchner, D. G. Franks, A. Pandini, L. Bonati, M. E. Hahn, and M. S. Denison
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(4):714–725.

Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin
2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, 47(7):1739–49.

Friesner, R. a., R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. a. Halgren, P. C. Sanschagrin, and D. T. Mainz
2006. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.*, 49(21):6177–96.

Gallastegui, N., J. A. Mackinnon, R. J. Fletterick, and E. Estébanez-Perpiñá
2015. Advances in our structural understanding of orphan nuclear receptors. *Trends Biochem. Sci.*, 40(1):25–35.

Gao, Y. D., S. H. Olson, J. M. Balkovec, Y. Zhu, I. Royo, J. Yabut, R. Evers, E. Y. Tan, W. Tang, D. P. Hartley, and R. T. Mosley
2007. Attenuating pregnane X receptor (PXR) activation: a molecular modelling approach. *Xenobiotica.*, 37(2):124–138.

Gasiewicz, T. A. and E. C. Henry
2011. History of Research on the AhR. In *AH Recept. Biol. Toxicol.*, R. Pohjanvirta, ed., chapter 1. Wiley.

Gervasio, F. L., A. Laio, and M. Parrinello
2005. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.*, 127(8):2600–2607.

Glass, C. K. and M. G. Rosenfeld
2000. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.*, 14(2):121–41.

Goldsmith, M. R., S. D. Peterson, D. T. Chang, T. R. Transue, R. Tornero-Velez, Y.-M. Tan, and C. C. Dary
2012. Informing mechanistic toxicology with computational molecular models. *Methods Mol. Biol.*, 929:139–65.

Goodale, B. C., J. K. la Du, W. H. Bisson, D. B. Janszen, K. M. Waters, and R. L. Tanguay
2012. AHR2 mutant reveals functional diversity of aryl hydrocarbon receptors in zebrafish. *PLoS One*, 7(1):e29346.

Greenidge, P. A., C. Kramer, J. C. Mozziconacci, and W. Sherman
2014. Improving docking results via reranking of ensembles of ligand poses in multiple X-ray protein conformations with MM-GBSA. *J. Chem. Inf. Model.*, 54(10):2697–2717.

Greenwood, J. R., D. Calkins, A. P. Sullivan, and J. C. Shelley
2010. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution.

Gutierrez, M. A., S. S. Davis, A. Rosko, S. M. Nguyen, K. P. Mitchell, S. Mateen, J. Neves, T. Y. Garcia, S. Mooney, G. H. Perdew, T. D. Hubbard, D. A. Lamba, and A. Ramanathan
2016. A novel AhR ligand, 2AI, protects the retina from environmental stress. *Sci. Rep.*, 6(1):29025.

Guyot, E., A. Chevallier, R. Barouki, and X. Coumoul
2013. The AhR twist: ligand-dependent AhR signaling and pharmacotoxicological implications. *Drug Discov. Today*, 18(9-10):479–486.

Hahn, M. E.
2011. Structural and Functional Diversification of AHRs during Metazoan Evolution. In *AH Recept. Biol. Toxicol.*, R. Pohjanvirta, ed., chapter 27. Wiley.

Hahn, M. E. and S. I. Karchner
2006. Unexpected diversity of aryl hydrocarbon receptors in non mammalian vertebrates: insights from comparative genomics. *J. Exp. Zool.*, 706(May 2005):693–706.

Hahn, M. E., S. I. Karchner, and R. R. Merson
2017. Diversity as opportunity: Insights from 600 million years of AHR evolution. *Curr. Opin. Toxicol.*, 2:58–71.

Halgren, T. A., R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks
2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.*, 47(7):1750–9.

Hawkins, G. D., C. J. Cramer, and D. G. Truhlar
1996. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.*, 100(51):19824–19839.

Henkler, F., K. Stolpmann, and A. Luch
2012. Exposure to Polycyclic Aromatic Hydrocarbons: Bulky DNA Adducts and Cellular Responses. In *Mol. Clin. Environ. Toxicol.*, A. Luch, ed., volume 101 of *Experientia Supplementum*, Pp. 107–131. Basel: Springer Basel.

Hennessy, E. J., V. Oza, A. Adam, K. Byth, L. Castriotta, G. Grewal, G. A. Hamilton, V. M. Kamhi, P. Lewis, D. Li, P. Lyne, L. Öster, M. T. Rooney, J. C. Saeh, L. Sha, Q. Su, S. Wen, Y. Xue, and B. Yang
2015. Identification and Optimization of Benzimidazole Sulfonamides as Orally Bioavailable Sphingosine 1-Phosphate Receptor 1 Antagonists with in Vivo Activity. *J. Med. Chem.*, 58(17):7057–7075.

Hernandez, J. P., L. C. Mota, and W. S. Baldwin
2009. Activation of CAR and PXR by Dietary, Environmental and Occupational Chemicals Alters Drug Metabolism, Intermediary Metabolism, and Cell Proliferation. *Curr. Pharmacogenomics Person. Med.*, 7(2):81–105.

Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije
1997. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.*, 18(12):1463–1472.

Hirano, M., J.-H. Hwang, H.-J. Park, S.-M. Bak, H. Iwata, and E.-Y. Kim
2015. In Silico Analysis of the Interaction of Avian Aryl Hydrocarbon Receptors and Dioxins to Decipher Isoform-, Ligand-, and Species-Specific Activations. *Environ. Sci. Technol.*, 49(6):3795–3804.

Homeyer, N. and H. Gohlke
2012. Free energy calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area method. *Mol. Inform.*, 31(2):114–122.

Huang, N., Y. Chelliah, Y. Shan, C. A. Taylor, S.-H. Yoo, C. Partch, C. B. Green, H. Zhang, and J. S. Takahashi
2012. Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science*, 337(6091):189–94.

Hubbard, T. D., I. A. Murray, W. H. Bisson, T. S. Lahoti, K. Gowda, S. G. Amin, A. D. Patterson, and G. H. Perdew
2015. Adaptation of the human aryl hydrocarbon receptor to sense microbiota-derived indoles. *Sci. Rep.*, 5(1):12689.

Humphrey, W., A. Dalke, and K. Schulten
1996. VMD: Visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–38.

IARC
2002. Some traditional herbal medicines, some mycotoxins, naphthalene and styrene. IARC Monographs on the evaluation of carcinogenic risks to humans. Technical report.

IARC
2012. Benzo[a]pyrene. In *IARC Monogr. Eval. Carcinog. risks to humans*, Pp. 111–144.

Jacobson, M. P., R. A. Friesner, Z. Xiang, and B. Honig
2002. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.*, 320(3):597–608.

Jacobson, M. P., D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw, and R. A. Friesner
2004. A hierarchical approach to all-atom protein loop prediction. *Proteins*, 55(2):351–67.

Jakalian, A., B. L. Bush, D. B. Jack, and C. I. Bayly
2000. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, 21(2):132–146.

Jakalian, A., D. B. Jack, and C. I. Bayly
2002. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.*, 23(16):1623–1641.

Jansson, K., T. Fristedt, A. Olsson, B. Svensson, and S. Jönsson
2006. Synthesis and reactivity of laquinimod, a quinoline-3-carboxamide: Intramolecular transfer of the enol proton to a nitrogen atom as a plausible mechanism for ketene formation. *J. Org. Chem.*, 71(4):1658–1667.

Jiang, F. and S.-H. Kim
1991. "Soft docking": Matching of molecular surface cubes. *J. Mol. Biol.*, 219(1):79–102.

Jones, S. A., L. B. Moore, J. L. Shenk, G. B. Wisely, G. A. Hamilton, D. D. McKee, N. C. O. Tomkinson, E. L. LeCluyse, M. H. Lambert, T. M. Willson, S. A. Kliewer, and J. T. Moore
2000. The Pregnane X Receptor: A Promiscuous Xenobiotic Receptor That Has Diverged during Evolution. *Mol. Endocrinol.*, 14(1):27–39.

Jorgensen, W. L.
1991. Rusting of the lock and key model for protein-ligand binding. *Science (80-.).*, 254(5034):954–955.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein
1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926.

Karchner, S. I., D. G. Franks, and M. E. Hahn
2005. AHR1B, a new functional aryl hydrocarbon receptor in zebrafish: tandem arrangement of ahr1b and ahr2 genes. *Biochem. J.*, 392(Pt 1):153–61.

Kewley, R. J., M. L. Whitelaw, and A. Chapman-Smith
2004. The mammalian basic helix–loop–helix/PAS family of transcriptional regulators. *Int. J. Biochem. Cell Biol.*, 36(2):189–204.

Key, J., T. H. Scheuermann, P. C. Anderson, V. Daggett, and K. H. Gardner
2009. Principles of ligand binding within a completely buried cavity in HIF2alpha PAS-B. *J. Am. Chem. Soc.*, 131(48):17647–54.

Khan, J. A., D. M. Camac, S. Low, A. J. Tebben, D. L. Wensel, M. C. Wright, J. Su, V. Jenny, R. D. Gupta, M. Ruzanov, K. A. Russo, A. Bell, Y. An, J. W. Bryson, M. Gao, P. Gambhire, E. T. Baldwin, D. Gardner, C. L. Cavallaro, J. V. Duncia, and J. Hynes
2015. Developing adnectins that target src co-activator binding to PXR: A structural approach toward understanding promiscuity of PXR. *J. Mol. Biol.*, 427(4):924–942.

Khandelwal, A., M. D. Krasowski, E. J. Reschly, M. W. Sinz, P. W. Swaan, and S. Ekins
2008. Machine learning methods and docking for predicting human pregnane X receptor activation. *Chem. Res. Toxicol.*, 21(7):1457–1467.

Kim, I.-S., J.-H. Hwang, M. Hirano, H. Iwata, and E.-Y. Kim
2016. In vitro and in silico evaluation of transactivation potencies of avian AHR1 and AHR2 by endogenous ligands: Implications for the physiological role of avian AHR2. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.*, 187:1–9.

Kitchen, D. B., H. Decornez, J. R. Furr, and J. Bajorath
2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.*, 3(11):935–49.

Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham
2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33(12):889–97.

Koshland, D. E.
1958. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci.*, 44(2):98–104.

Krumrine, J., F. Raubacher, N. Brooijmans, and I. Kuntz
2005. Principles and Methods of Docking and Ligand Design. In *Struct. Bioinforma.*, P. E. Bourne and H. Weissig, eds., chapter 22, Pp. 441–476. John Wiley & Sons, Ltd.

Kumar, M. B. and G. H. Perdew
1999. Nuclear receptor coactivator SRC-1 interacts with the Q-rich subdomain of the AhR and modulates its transactivation potential. *Gene Expr.*, 8(5-6):273–286.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton
1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 26(2):283–291.

Laudet, V. and H. Gronemeyer
2002. *The Nuclear Receptor FactsBook*. Elsevier.

Leach, A. R.
2001. Empirical Force Field Models: Molecular Mechanics. In *Mol. Model. Princ. Appl.*, chapter 4, Pp. 165–252. Prentice Hall.

Lexa, K. W. and H. a. Carlson
2012. Protein flexibility in docking and surface mapping. *Q. Rev. Biophys.*, 45(03):301–343.

Limongelli, V., M. Bonomi, L. Marinelli, F. L. Gervasio, A. Cavalli, E. Novellino, and M. Parrinello
2010. Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc. Natl. Acad. Sci.*, 107(12):5411–5416.

Limongelli, V., M. Bonomi, and M. Parrinello
2013. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci.*, 110(16):6358–6363.

Loncharich, R. J., B. R. Brooks, and R. W. Pastor
1992. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, 32(5):523–35.

Luch, A. and W. Baird
2010. Carcinogenic Polycyclic Aromatic Hydrocarbons. In *Compr. Toxicol.*, volume 14, Pp. 85–123. Elsevier.

Ma, Q.
2011. Overview of AHR Functional Domains and the Classical AHR Signaling Pathway: Induction of Drug Metabolizing Enzymes. In *AH Recept. Biol. Toxicol.*, R. Pohjanvirta, ed., chapter 2, Pp. 33–45. Wiley.

Mackowiak, B. and H. Wang
2016. Mechanisms of xenobiotic receptor activation: Direct vs. indirect. *Biochim. Biophys. Acta - Gene Regul. Mech.*, 1859(9):1130–1140.

Madhavi Sastry, G., M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman
2013. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.*, 27(3):221–234.

Mahiout, S., S. Giani Tagliabue, A. Nasri, I. M. Omoruyi, L. Pettersson, L. Bonati, and R. Pohjanvirta
2018. In vitro toxicity and in silico docking analysis of two novel selective AH-receptor modulators. *Toxicol. Vitr.*, 52:178–188.

Mahiout, S., J. Lindén, J. Esteban, I. Sánchez-Pérez, S. Sankari, L. Pettersson, H. Håkansson, and R. Pohjanvirta
2017. Toxicological characterisation of two novel selective aryl hydrocarbon receptor modulators in Sprague-Dawley rats. *Toxicol. Appl. Pharmacol.*, 326:54–65.

Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling
2015. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, 11:3696–3713.

Mandal, P. K.
2005. Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. *J. Comp. Physiol. B.*, 175(4):221–30.

McGovern, S. L. and B. K. Shoichet
2003. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.*, 46(14):2895–2907.

Möglich, A., R. A. Ayers, and K. Moffat
2009. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure*, 17(10):1282–1294.

Mollica, L., S. Decherchi, S. R. Zia, R. Gaspari, A. Cavalli, and W. Rocchia
2015. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.*, 5(1):11539.

Mollica, L., I. Theret, M. Antoine, F. Perron-Sierra, Y. Charton, J. M. Fourquez, M. Wierzbicki, J. A. Boutin, G. Ferry, S. Decherchi, G. Bottegoni, P. Ducrot, and A. Cavalli
2016. Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein-Ligand Residence Times. *J. Med. Chem.*, 59(15):7167–7176.

Motta, S. and L. Bonati
2017. Modeling Binding with Large Conformational Changes: Key Points in Ensemble-Docking Approaches. *J. Chem. Inf. Model.*, 57(7):1563–1578.

Motta, S., L. Callea, S. Giani Tagliabue, and L. Bonati
2018a. Exploring the PXR ligand binding mechanism with advanced Molecular Dynamics methods. *Sci. Rep.*, 8(1):16207.

Motta, S., C. Minici, D. Corrada, L. Bonati, and A. Pandini
2018b. Ligand-induced perturbation of the HIF-2α:ARNT dimer dynamics. *PLOS Comput. Biol.*, 14(2):e1006021.

Motto, I., A. Bordogna, A. A. Soshilov, M. S. Denison, and L. Bonati
2011. New aryl hydrocarbon receptor homology model targeted to improve docking reliability. *J. Chem. Inf. Model.*, 51(11):2868–81.

Moura-Alves, P., K. Faé, E. Houthuys, A. Dorhoi, A. Kreuchwig, J. Furkert, N. Barison, A. Diehl, A. Munder, P. Constant, T. Skrahina, U. Guhlich-Bornhof, M. Klemm, A.-B. Koehler, S. Bandermann, C. Goosmann, H.-J. Mollenkopf, R. Hurwitz, V. Brinkmann, S. Fillatreau, M. Daffe, B. Tümmler, M. Kolbe, H. Oschkinat, G. Krause, and S. H. E. Kaufmann
2014. AhR sensing of bacterial pigments regulates antibacterial defence. *Nature*, 512(7515):387–92.

Moustakas, D. T.
2007. Application of Docking Methods to Structure-Based Drug Design. In *Comput. Struct. Approaches to Drug Discov. Ligand-Protein Interact.*, R. Stroud, ed., chapter 9, Pp. 155–180. Cambridge: Royal Society of Chemistry.

Murray, I. A., A. D. Patterson, and G. H. Perdew
2014. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nat. Rev. Cancer*, 14(12):801–814.

Murray, I. A. and G. H. Perdew
2011. Role of Chaperone Proteins in AHR Function. In *AH Recept. Biol. Toxicol.*, R. Pohjanvirta, ed., chapter 3, Pp. 47–61. Wiley.

Nakata, K., Y. Tanaka, T. Nakano, T. Adachi, H. Tanaka, T. Kaminuma, and T. Ishikawa
2006. Nuclear Receptor-Mediated Transcriptional Regulation in Phase I, II, and III Xenobiotic Metabolizing Systems. *Drug Metab. Pharmacokinet.*, 21(6):437–457.

Nguyen, L. P. and C. a. Bradfield
2008. The Search for Endogenous Activators of the Aryl Hydrocarbon Receptor. *Chem. Res. Toxicol.*, 21(1):102–116.

Nobeli, I., A. D. Favia, and J. M. Thornton
2009. Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.*, 27(2):157–167.

Nuti, R., M. Gargaro, D. Matino, D. Dolciami, U. Grohmann, P. Puccetti, F. Fallarino, and A. Macchiarulo
2014. Ligand Binding and functional selectivity of L-tryptophan metabolites at the mouse aryl hydrocarbon receptor (mAhR). *J. Chem. Inf. Model.*, 54(12):3373–3383.

Odio, C., S. A. Holzman, M. S. Denison, D. Fraccalvieri, L. Bonati, D. G. Franks, M. E. Hahn, and W. H. Powell
2013. Specific ligand binding domain residues confer low dioxin responsiveness to AHR1β of xenopus laevis. *Biochemistry*, 52(10):1746–1754.

O'Donnell, E. F., D. C. Koch, W. H. Bisson, H. S. Jang, and S. K. Kolluri
2014. The aryl hydrocarbon receptor mediates raloxifene-induced apoptosis in estrogen receptor-negative hepatoma and breast cancer cells. *Cell Death Dis*, 5:e1038.

O'Donnell, E. F., K. S. Saili, D. C. Koch, P. R. Kopparapu, D. Farrer, W. H. Bisson, L. K. Mathew, S. Sengupta, N. I. Kerkvliet, R. L. Tanguay, and S. K. Kolluri
2010. The anti-inflammatory drug leflunomide is an agonist of the aryl hydrocarbon receptor. *PLoS One*, 5(10).

Omiecinski, C. J., J. P. Vanden Heuvel, G. H. Perdew, and J. M. Peters
2011. Xenobiotic metabolism, disposition, and regulation by receptors: From biochemical phenomenon to predictors of major toxicities.

Onufriev, A., D. Bashford, and D. A. Case
2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Struct. Funct. Bioinforma.*, 55(2):383–394.

Pandini, A., M. S. Denison, Y. Song, A. A. Soshilov, and L. Bonati
2007. Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. *Biochemistry*, 46(3):696–708.

Pandini, A., A. A. Soshilov, Y. Song, J. Zhao, L. Bonati, and M. S. Denison
2009. Detection of the TCDD binding-fingerprint within the Ah receptor ligand binding domain by structurally driven mutagenesis and functional analysis. *Biochemistry*, 48(25):5972–83.

Parkinson, A. and B. W. Ogilvie
2001. Biotransformation of Xenobiotics. In *Essentials Toxicol.*, C. D. Klaassen and J. B. Watkins III, eds., chapter 6, Pp. 529–540. McGraw-Hill.

Parrinello, M. and A. Rahman
1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52(12):7182–7190.

Perdew, G., I. Murray, A. Patterson, and J. Peters
2018. Mechanisms: Xenobiotic Receptor-Mediated Toxicity. In *Compr. Toxicol.*, number September 2016, chapter 1.11, Pp. 202–228. Elsevier.

Perdew, G. H., I. a. Murray, and J. M. Peters
2010. Xenobiotic Receptor-Mediated Toxicity. In *Compr. Toxicol.*, chapter 1.17, Pp. 361–388. Newnes.

Perkins, A., J. L. Phillips, N. I. Kerkvliet, R. L. Tanguay, G. H. Perdew, S. K. Kolluri, and W. H. Bisson
2014. A Structural Switch between Agonist and Antagonist Bound Conformations for a Ligand-Optimized Model of the Human Aryl Hydrocarbon Receptor Ligand Binding Domain. *Biology (Basel).*, 3(4):645–69.

Plewczynski, D., M. Lazniewski, R. Augustyniak, and K. Ginalski
2010. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.*, 32:742–755.

Pohjanvirta, R., M. Korkalainen, I. D. Moffat, P. C. Boutros, and A. B. Okey
2011. Role of the AhR and its structure in TCDD toxicity. In *AH Recept. Biol. Toxicol.*, R. Pohjanvirta, ed., chapter 12, Pp. 181–196. Wiley.

Poland, A., D. Palen, and E. Glover
1994. Analysis of the four alleles of the murine aryl hydrocarbon receptor. *Mol. Pharmacol.*, 46(5):915–921.

Procopio, M., A. Lahm, A. Tramontano, L. Bonati, and D. Pitea
2002. A model for recognition of polychlorinated dibenzo-p-dioxins by the aryl hydrocarbon receptor. *Eur. J. Biochem.*, 269(1):13–18.

Raunio, H.
2011. In Silico Toxicology – Non-Testing Methods. *Front. Pharmacol.*, 2(June):1–8.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, S. I. Karchner, M. E. Hahn, D. Nacci, M. F. Oleksiak, D. L. Crawford, and A. Whitehead
2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science (80-.).*, 354(6317):1305 LP – 1308.

Reschly, E. J. and M. D. Krasowski
2006. Evolution and function of the NR1I nuclear hormone receptor subfamily (VDR, PXR, and CAR) with respect to metabolism of xenobiotics and endogenous compounds. *Curr. Drug Metab.*, 7(4):349–65.

Rogers, J. L., L. Bayeh, T. H. Scheuermann, J. Longgood, J. Key, J. Naidoo, L. Melito, C. Shokri, D. E. Frantz, R. K. Bruick, K. H. Gardner, J. B. MacMillan, and U. K. Tambar
2013. Development of inhibitors of the PAS-B domain of the HIF-2α transcription factor. *J. Med. Chem.*, 56(4):1739–1747.

Russell, W. M. S. and R. L. Burch
1959. *The principles of humane experimental technique*. Methuen.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen
1977. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *J. Comp. Phys.*, 23:327–41.

Safe, S.
2017. Carbidopa: a selective Ah receptor modulator (SAhRM). *Biochem. J.*, 474(22):3763–3765.

Safe, S., C. Gayathri, and I. Jutooru
2011. AHR-Active Compounds in the Human Diet. In *AH Recept. Biol. Toxicol.*, R. Pohjanvirta, ed., Pp. 331–342. Wiley.

Sakurai, S., T. Shimizu, and U. Ohto
2017. The crystal structure of the AhRR–ARNT heterodimer reveals the structural basis of the repression of AhR-mediated transcription. *J. Biol. Chem.*, 292(43):17609–17616.

Sali, A. and T. L. Blundell
1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234(3):779–815.

Salomon-Ferrer, R., D. A. Case, and R. C. Walker
2013. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 3(2):198–210.

Sarath Josh, M. K., S. Pradeep, A. K. Balan, M. N. Sreejith, and S. Benjamin
2016. Accessing the molecular interactions of phthalates and their primary metabolites with the human pregnane X receptor using in silico profiling. *J. Appl. Toxicol.*, 36(12):1599–1604.

Schectera, A., L. Birnbaumb, J. J. Ryanc, and J. D. Constabled
2006. Dioxins: An Overview. *Environ. Res.*, 101:419–428.

Scheuermann, T. H., Q. Li, H. W. Ma, J. Key, L. Zhang, R. Chen, J. A. Garcia, J. Naidoo, J. Longgood, D. E. Frantz, U. K. Tambar, K. H. Gardner, and R. K. Bruick
2013. Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nat. Chem. Biol.*, 9(4):271–276.

Scheuermann, T. H., D. Stroud, C. E. Sleet, L. Bayeh, C. Shokri, H. Wang, C. G. Caldwell, J. Longgood, J. B. MacMillan, R. K. Bruick, K. H. Gardner, and U. K. Tambar
2015. Isoform-Selective and Stereoselective Inhibition of Hypoxia Inducible Factor-2. *J. Med. Chem.*, 58(15):5930–5941.

Scheuermann, T. H., D. R. Tomchick, M. Machius, Y. Guo, R. K. Bruick, and K. H. Gardner
2009. Artificial ligand binding within the HIF2α PAS-B domain of the HIF2 transcription factor. *Proc. Natl. Acad. Sci.*, 106(2):450–455.

Schrödinger LLC
2010. The PyMOL Molecular Graphics System.

Schrödinger LLC
2015a. LigPrep.

Schrödinger LLC
2015b. MacroModel.

Schrödinger LLC
2016a. Epik.

Schrödinger LLC
2016b. Protein Preparation Wizard.

Schulte, K. W., E. Green, A. Wilz, M. Platten, and O. Daumke
2017. Structural Basis for Aryl Hydrocarbon Receptor-Mediated Gene Activation. *Structure*, 25(7):1025–1033.e3.

Seok, S.-H., W. Lee, L. Jiang, K. Molugu, A. Zheng, Y. Li, S. Park, C. A. Bradfield, and Y. Xing
2017. Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proc. Natl. Acad. Sci. U. S. A.*, 114(21):5431–5436.

Shanahan, T.
2017. Selfish Genes and Lucky Breaks: Richard Dawkins' and Stephen Jay Gould's Divergent Darwinian Agendas. In *Darwinian Tradit. Context*, Pp. 11–36. Cham: Springer International Publishing.

Shelley, J. C., A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin, and M. Uchimaya
2007. Epik: A software program for pKa prediction and protonation state generation for drug-like molecules. *J. Comput. Aided. Mol. Des.*, 21(12):681–691.

Shen, M.-Y. and A. Sali
2006. Statistical potential for assessment and prediction of protein structures. *Protein Sci.*, 15(11):2507–2524.

Sherman, W., T. Day, M. P. Jacobson, R. A. Friesner, and R. Farid
2006. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.*, 49(2):534–553.

Shoots, J., D. Fraccalvieri, D. G. Franks, M. S. Denison, M. E. Hahn, L. Bonati, and W. H. Powell
2015. An aryl hydrocarbon receptor from the salamander ambystoma mexicanum exhibits low sensitivity to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Environ. Sci. Technol.*, 49(11):6993–7001.

Sinko, W., S. Lindert, and J. A. Mccammon
2013. Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design. *Chem. Biol. Drug Des.*, 81(1):41–49.

Smith, K. J., I. A. Murray, R. Tanos, J. Tellew, A. E. Boitano, W. H. Bisson, S. K. Kolluri, M. P. Cooke, and G. H. Perdew
2011. Identification of a High-Affinity Ligand That Exhibits Complete Aryl Hydrocarbon Receptor Antagonism. *J. Pharmacol. Exp. Ther.*, 338(1):318–327.

Soshilov, A. A. and M. S. Denison
2014. Ligand promiscuity of aryl hydrocarbon receptor agonists and antagonists revealed by site-directed mutagenesis. *Mol. Cell. Biol.*, 34(9):1707–19.

Spitaleri, A., S. Decherchi, A. Cavalli, and W. Rocchia
2018. Fast Dynamic Docking Guided by Adaptive Electrostatic Bias: The MD-Binding Approach. *J. Chem. Theory Comput.*, 14(3):1727–1736.

Spyrakis, F. and C. N. Cavasotto
2015. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Arch. Biochem. Biophys.*, 583:105–19.

Srinivasan, J., J. Miller, P. A. Kollman, and D. A. Case
1998. Continuum Solvent Studies of the Stability of RNA Hairpin Loops and Helices. *J. Biomol. Struct. Dyn.*, 16(3):671–682.

Sui, Y., N. Ai, S. H. Park, J. Rios-Pilier, J. T. Perkins, W. J. Welsh, and C. Zhou
2012. Bisphenol A and its analogues activate human pregnane X receptor. *Environ. Health Perspect.*, 120(3):399–405.

Taylor, B. L. and I. B. Zhulin
1999. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.*, 63(2):479–506.

Tian, W., C. Chen, X. Lei, J. Zhao, and J. Liang
2018. CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Res.*, 46(W1):W363–W367.

Totrov, M. and R. Abagyan
2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.*, 18(2):178–184.

Tribello, G. A., M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi
2014. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.*, 185(2):604–613.

Tsujishita, H., I. Moriguchi, and S. Hirono
1993. Potential-scaled molecular dynamics and potential annealing: Effective conformational search techniques for biomolecules. *J. Phys. Chem.*, 97(17):4416–4420.

Vogel, C. F. A., E. M. Khan, P. S. C. Leung, M. E. Gershwin, W. L. W. Chang, D. Wu, T. Haarmann-Stemmann, A. Hoffmann, and M. S. Denison
2014. Cross-talk between aryl hydrocarbon receptor and the inflammatory response: A role for nuclear factor-kB. *J. Biol. Chem.*, 289(3):1866–1875.

Wallace, B. D., L. Betts, G. Talmage, R. M. Pollet, N. S. Holman, and M. R. Redinbo
2013. Structural and functional analysis of the human nuclear xenobiotic receptor PXR in complex with RXRα. *J. Mol. Biol.*, 425(14):2561–2577.

Wang, C. Y., C. W. Li, J. D. Chen, and W. J. Welsh
2006. Structural model reveals key interactions in the assembly of the pregnane X receptor/corepressor complex. *Mol. Pharmacol.*, 69(5):1513–7.

Wang, J., R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case
2004. Development and testing of a general amber force field. *J. Comput. Chem.*, 25(9):1157–1174.

Wang, W., W. W. Prosise, J. Chen, S. S. Taremi, H. V. Le, V. Madison, X. Cui, A. Thomas, K. C. Cheng, and C. A. Lesburg
2008. Construction and characterization of a fully active PXR/SRC-1 tethered protein with increased stability. *Protein Eng. Des. Sel.*, 21(7):425–433.

Warren, G. L., C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head
2006. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, 49(20):5912–5931.

Watkins, R. E., P. R. Davis-Searles, M. H. Lambert, and M. R. Redinbo
2003a. Coactivator Binding Promotes the Specific Interaction Between Ligand and the Pregnane X Receptor. *J. Mol. Biol.*, 331(4):815–828.

Watkins, R. E., J. M. Maglich, L. B. Moore, G. B. Wisely, S. M. Noble, P. R. Davis-Searles, M. H. Lambert, S. A. Kliewer, and M. R. Redinbo
2003b. 2.1 Å crystal structure of human PXR in complex with the St. John's wort compound hyperforin. *Biochemistry*, 42(6):1430–1438.

Watkins, R. E., G. B. Wisely, L. B. Moore, J. L. Collins, M. H. Lambert, S. P. Williams, T. M. Willson, S. A. Kliewer, and M. R. Redinbo
2001. The Human Nuclear Xenobiotic Receptor PXR: Structural Determinants of Directed Promiscuity. *Science (80-.).*, 292(5525):2329–2333.

Webb, B. and A. Sali
2016. Comparative Protein Structure Modeling Using MODELLER. In *Curr. Protoc. Bioinforma.*, Pp. 5.6.1–5.6.37. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Weiser, J., P. S. Shenkin, and W. C. Still
1999. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.*, 20(2):217–230.

Wiederstein, M. and M. J. Sippl
2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.*, 35:407–10.

Willson, T. M. and S. A. Kliewer
2002. PXR, CAR and Drug Metabolism. *Nat. Rev. Drug Discov.*, 1(4):259–266.

Wincent, E., N. Amini, S. Luecke, H. Glatt, J. Bergman, C. Crescenzi, A. Rannug, and U. Rannug
2009. The Suggested Physiologic Aryl Hydrocarbon Receptor Activator and Cytochrome P4501 Substrate 6-Formylindolo[3,2-b]carbazole is Present in Humans. *J. Biol. Chem.*, 284(5):2690–2696.

Wright, E. J., K. Pereira De Castro, A. D. Joshi, and C. J. Elferink
2017. Canonical and non-canonical aryl hydrocarbon receptor signaling pathways. *Curr. Opin. Toxicol.*, 2:87–92.

Wu, D., N. Potluri, J. Lu, Y. Kim, and F. Rastinejad
2015. Structural integration in hypoxia-inducible factors. *Nature*, 524(7565):303–308.

Wu, D., X. Su, N. Potluri, Y. Kim, and F. Rastinejad
2016. NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. *Elife*, 5:e18790.

Xing, Y., M. Nukaya, and K. Satyshur
2012. Identification of the Ah-receptor structural determinants for ligand preferences. *Toxicol. Sci.*, 129(1):86–97.

Xue, Y., E. Chao, W. J. Zuercher, T. M. Willson, J. L. Collins, and M. R. Redinbo
2007. Crystal structure of the PXR-T1317 complex provides a scaffold to examine the potential for receptor antagonism. *Bioorganic Med. Chem.*, 15(5):2156–2166.