UNIVERSITAT DE VALÈNCIA

Study of the aryl hydrocarbon receptor mediated effects through *in silico* modeling and *in vitro* bioassays

by

Elizabeth Goya Jorge

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry.

Supervised by:

Prof. Rosa María Giner Pons

Dr. Rafael Gozalbes Botella

Dr. Stephen Jones Barigye

-Valencia, July 2020-

VNIVERSITAT D̃ VALÈNCIA

Facultat de Farmàcia

Departament de Farmacologia

ProtoQSAR
Chemoinformatics and modeling solutions
to optimize your chemicals

INFORME DE LOS DIRECTORES

La presente tesis doctoral, estructurada para su evaluación en la modalidad de compendio de publicaciones cumple con los requisitos aprobados por la Escuela de Doctorado de la Universitat de València, así como la normativa propia de la Comisión Académica del Programa de Doctorado 3154 en Química (RD 99/2011).

La doctoranda, Dña. Elizabeth Goya Jorge, ha publicado los trabajos de investigación que conforman esta tesis doctoral en revistas científicas indexadas, y en los cuales figura como primera firmante y autora principal de los mismos. Otras publicaciones de la doctoranda durante su período de tesis y manuscritos en proceso de preparación se adjuntan como anexos.

- **Goya-Jorge, E.**, Giner, R.M, Veitía, M.S.-I, Gozalbes, R., Barigye, S.J. Predictive modeling of aryl hydrocarbon receptor (AhR) agonism. Chemosphere, 2020, 256 (127068). doi: 10.1016/j.chemosphere.2020.127068. WoS, JIF: 5.778. ESI rank: 5/380 (Q1). Scopus, SJR 1.53. H Index 228 (Q1).

- **Goya-Jorge, E.**, Abdmouleh, F., Carpio, L.E., Giner, R.M., Veitía, M.S.-I. Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic activities, European Journal of Pharmaceutical Sciences, 2020, 151C (105386); doi: 10.1016/j.ejps.2020.105386. WoS, JIF: 3.616. ESI rank: 37/280 (Q1). Scopus, SJR 0.80. H Index 128 (Q1).

- **Goya-Jorge, E.**, Doan, T.Q., Scippo, M.L., Muller, M., Giner, R.M., Barigye, S.J., Gozalbes, R. Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective, SAR and QSAR in Environmental Research, 2020, 31 (3), 209-226; doi: 10.1080/1062936X.2019.1708460. WoS, JIF: 2.053. JCR rank: 23/59 (Q2). Scopus, SJR 0.49. H Index 46 (Q2).

- **Goya-Jorge, E.**, Rampal, C., Loones, N., Barigye, S.J., Carpio, L.E., Gozalbes, R., Ferroud, C., Veitía, M.S.-I, Giner, R.M. Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes, European Journal of Medicinal Chemistry, 2020, 207 (112777). doi: 10.1016/j.ejmech.2020.112777. WoS JIF: 5.572. ESI rank: 52/539 (Q1). Scopus, SJR 1.14. H Index 151 (Q1).

STATEMENT OF SUPERVISORS

As supervisors of the scientific research work presented in this Thesis, we **CERTIFY** that:

-The Ph.D. Thesis proposal entitled: "Study of the aryl hydrocarbon receptor mediated effects through *in silico* modeling and *in vitro* bioassays" submitted by the MSc. Elizabeth Goya Jorge to obtain the degree of Doctor in Chemistry has been developed for the last three years at the company ProtoQSAR and at the Department of Pharmacology, Faculty of Pharmacy of the Universitat de València with funding from a European Marie Skłodowska-Curie grant in the framework of "PROTECTED" project: Protection against Endocrine Disruptors; detection, mixtures, health effects, risk assessment and communication.

-Some of the research work presented in this Thesis was carried out in collaboration with the following academic institutions in which the Ph.D. candidate had research stays:

- Conservatoire National des Arts et Métiers, sited in Paris, France
- University of Liège sited in Liège, Belgium
- Queen's University of Belfast sited in Northern Ireland, United Kingdom

We approve the presentation of this doctoral thesis to be evaluated by the corresponding court and we signed this document on the seventh day of July of the year two thousand and twenty.
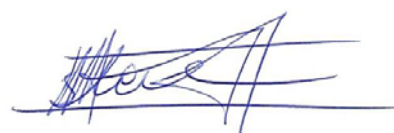
Dr. Rosa María Giner Pons	**Dr. Rafael Gozalbes Botella**	**Dr. Stephen Jones Barigye**
Full Professor	Founder and Director	Head of R&D
Department of Pharmacology	ProtoQSAR SL.	ProtoQSAR SL.
University of Valencia	Private company	Private company

"Nothing in life is to be feared, it is only to be understood"

Marie Skłodowska-Curie

Dedico esta Tesis

a la memoria de mi papá adorado

Agradecimientos

"La gratitud es el más legítimo pago al esfuerzo ajeno, es reconocer que todo lo que somos, es la suma del sudor de los demás (...) que un hombre solo no vale nada, y que la dependencia humana, además de necesaria, es hermosa"

José Martí

Agradezco ante todo a mis directores, cuya supervisión y apoyo incondicional han permitido que culmine esta importante fase de mi carrera profesional. Muchas gracias a Rafael, que me dio la oportunidad de formar parte de un proyecto de investigación espectacular, en el que he crecido como profesional y como persona. A Steph, mi "por siempre tutor" que ha supervisado cada paso desde mis estudios de pregrado, ha sido un verdadero privilegio poder contar con tu apoyo y sobre todo poder aprender de ti durante todos estos años. Y Rosa, mi madre académica, mi mentora, pero sobre todo mi más cercana familia en estos años lejos de mi país y de los míos, creo que no hay forma de agradecer con palabras todo lo que has hecho por mí, ¡muchas muchísimas gracias!

Me gustaría también dedicar mi más profundo agradecimiento a quien ha sido mi co-supervisora en la distancia, Maite: tu esmero, dedicación y cariño han hecho posible que cumpla con este sueño del que tú formas parte imprescindible y por lo que te estaré por siempre agradecida.

Agradezco entonces a la persona más importante de mi vida, a mi inspiración, mi guía, mi mejor amiga, mi apoyo, mi confidente, la mujer más fuerte y maravillosa que existe: mi mamá. ¿qué decirte que no sepas? Sin ti no lo hubiese logrado porque aún a 10 mil km estás siempre aquí, así lo siento. A mi amado Yairon, mi compañero de viaje, mi soporte diario, le debo haberme levantado de los peores momentos, ¡gracias por tu apoyo y amor incondicional!

A mi familia adorada, le agradezco infinitamente, ya que son parte imprescindible de todos mis logros, son tanto míos como son suyos. A mi hermanito querido, a mis hermanas del alma, a mis tiuchis maravillosos y a mi familia política. A mi familia extendida, esa que escogí hace 16 años y que más que amigos son mi soporte emocional aún en extremos opuestos del mundo, les agradezco también, a mis BFFs!!! sólo ellos saben cuánta falta me hacía tener +999 mensajes esperando a ser leídos y que siempre alegraron mis eternas jornadas de trabajo sin respetar nunca mucho los husos horarios.

Gracias también a mis compañeros de la empresa, a los que siguen y a los que estuvieron en algún momento durante estos tres años, en especial a Pepi cuyo cariño y soporte personal y profesional aprecio muchísimo. A Laureano, muchas gracias por sus geniales contribuciones en la modelación. Y a todos, los de hace tiempo: Eva, María J, Pepe, Julián, Joel, Guille y Sergi y los más recientes: Pablo, Pravin y Andrea, por compartir y ayudarme de una forma u otra durante este importante período de mi vida.

¡Muchísimas gracias a todos y a cada uno!

Acknowledgments

"If I have seen further than others, it is by standing upon the shoulders of giants"

Isaac Newton

I gratefully thank to my Project's coordinator Lisa Connolly and to my Project's manager Katie Austin for your closeness, dedication, and mentorship during these three years in the "PROTECTED family" and during my research visiting at Queen's University of Belfast when I learned so much and have an amazing time with my great lab-mates Mazia, Cher and Toby.

Thanks also to my advisors during my research stay at the University of Liège: Marc Müller and Marie L. Scippo and to my colleague Que for your kindness and for sharing your experience and collaborate with my research ideas. I believe that we accomplished a lot together! From my time in Liège I thank the wonderful team at GIGA research lab, especially to my kind lab-mates Gus and Ratish.

I deeply thank all the ESRs of PROTECTED project, the ones that I already mentioned, and: Melissa, Chiara, Maria, Vittoria, Clèmence, Bérénice, Solomon, Tarek, Anteneh, and Ajay with whom I have experienced incredible adventures around the world, both professional and personal. It was a great honor to be part of such an amazing team of young scientists in love with their research. I am looking forward to meeting you somewhere, someday, and nostalgically remember the years in which the study of endocrine disruption brought us all together, from 12 different countries (Cuba, Cyprus, Ethiopia, France, Italy, Nepal, Nigeria, Mexico, Morocco, Pakistan, United States, and Vietnam), with different cultures and scientific backgrounds, but with the same goal: to become Ph.D.

MANY THANKS!

THESIS INDEX

LIST OF SCIENTIFIC PUBLICATIONS

The results presented in this Thesis are published in the following scientific articles

- **Goya-Jorge, E**., Doan, T.Q., Scippo, M.L., Muller, M., Giner, R.M., Barigye, S.J, Gozalbes, R. Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective, SAR and QSAR in Environmental Research, 2020, 31 (3), 209-226; 10.1080/1062936X.2019.1708460. WoS, JIF: 2.053. JCR rank: 23/59 (Q2) in "Mathematical & Computational Biology". Scimago, SJR 0.49. H Index 46 (Q2) in Bioengineering.

- **Goya-Jorge, E**., Abdmouleh, F, Carpio, L.E., Giner, R.M., Veitía, M.S.-I. Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic activities, 2020, European Journal of Pharmaceutical Sciences, 2020, 151C (105386). 10.1016/j.ejps.2020.105386. WoS, JIF: 3.616. JCR rank (ESI): 37/280 (Q1) in Pharmacology & Toxicology. Scimago, SJR 0.80. H Index 128 (Q1) in Pharmaceutical Science.

- **Goya-Jorge, E**., Giner, R.M, Veitía, M.S.-I, Gozalbes, R., Barigye, S.J. Predictive modeling of aryl hydrocarbon receptor (AhR) agonism. Chemosphere, 2020, 256 (127068). 10.1016/j.chemosphere.2020.127068. WoS, JIF: 5.778. JCR rank (ESI): 5/380 (Q1) in "Environment/Ecology". Scimago, SJR 1.53. H Index 228 (Q1) in Chemistry (miscellaneous).

- **Goya-Jorge, E**., Rampal, C., Loones, N., Barigye, S.J., Carpio, L.E., Gozalbes, R., Ferroud, C., Veitía, M.S.-I, Giner, R.M. Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes. European Journal of Medicinal Chemistry, 2020, 207 (112777). 10.1016/j.ejmech.2020.112777. WoS, JIF: 5.572. JCR rank: 5/61 (Q1) in "Chemistry, Medicinal". Scimago, SJR 1.14. H Index 151 (Q1) in Drug Discovery.

- **Goya-Jorge, E**., de Julián-Ortiz, JV., Gozalbes, R. Revisión de los modelos computacionales que relacionan la estructura química con la disrupción del sistema endocrino [PROTECTion against Endocrine Disruption through Quantitative Structure-Activity Relationship modelling], Revista de Toxicología, 2020, 37 (1). AETOX. Scimago, SJR 0.12. H Index 10 (Q4) in Toxicology.

Research collaboration

Ricco, C., Abdmouleh, F., Riccobono, C., Guenineche, L., Martin, F., **Goya-Jorge, E**., Lagarde, N., Liagre, B., Ali, M.B., Ferroud, C., Arbi, M.E., Veitía, M.S.-I. Pegylated triarylmethanes: Synthesis, antimicrobial activity, anti-proliferative behavior and *in silico* studies. Bioorganic Chemistry 2020, 96 (103591). 10.1016/j.bioorg.2020.103591. WoS, JIF: 4.831. JCR rank: 8/57 (Q1) in "Chemistry, Organic". Scimago, SJR 0.82. H Index 55 (Q1) in Drug Discovery.

---

*Metrics last updated before the Thesis's submission in 2020 and they corresponded to the year 2019. The best quartile and category are detailed for each publication.

SCIENTIFIC DISSEMINATION

The doctoral student has presented her research work in the following scientific events

Jun 2020 (Poster Presentation). INNOVAESTIC. (Investigación en Docencia Universitaria)

May 2020 (Poster Presentation). 30th SETAC Europe, Dublin, Ireland

Feb 2020 (Oral Presentation). 5th German Pharm-Tox Summit, Leipzig, Germany

Oct 2019 (Oral Presentation). EC H2020 MSCA-ITN cluster event, Girona, Spain

Nov 2019 (Poster Presentation). Symposium EDs in aquatic environment, Valdivia, Chile

May 2019 (Oral and Poster Presentations). JRC-EC Summer School, Ispra, Italy

Jul 2018 (Oral Presentation). 11th Congress CICTA, Madrid, Spain

Sep 2018 (Participation). 11th BDS Conference, Aachen, Germany

Oct 2018 (Oral and Poster Presentations). 20th Congress ESTIV, Berlin, Germany

Apr 2018 (Oral Presentation). VI AETOX Scientific Day, Valencia, Spain

Feb 2018 (Oral Presentation). 14th Conference ChemCYS, Blankenberge, Belgium

Dec 2017 (Oral Presentation). Master Course in Toxicology, UV, Valencia, Spain

Oct 2017 (Participation). Workshop Biocides, Milan, Italy

CONFERENCE PROCEEDINGS

Goya-Jorge, E., Gozalbes, R., Barigye, S.J., Veitía, M.S.-I, Giner, R.M. Triarylmethane class of compounds as potential modulators of the Aryl Hydrocarbon Receptor (AhR): synthesis and *in vitro* evaluation. Abstracts of the 86th Annual Meeting of the German Society for Experimental and Clinical Pharmacology and Toxicology (DGPT). With contribution of the Arbeitsgemeinschaft für Angewandte Humanpharmakologie e. V. (AGAH). Naunyn-Schmiedeberg's Arch Pharmacol 393, 1-97 (2020), 10.1007/s00210-020-01828-y.

Tolosa, J., **Goya-Jorge, E**., Gozalbes, R., de Julián-Ortiz, J.V. Herramientas computacionales para la evaluación de la ecotoxicidad de biocidas. Actas de las VI Jornadas de Formación en Toxicología, AETOX, 19th April 2018. Valencia, Spain. Revista de Toxicología, Vol. 35, Núm-1 (2018).

THESIS OUTLINE

The aryl hydrocarbon receptor (AhR) is a cytoplasmatic sensor of diverse endogenous and exogenous substances. In a toxicological context, the former known as "dioxin receptor" has been investigated as a xenobiotic chemoreceptor and due to its roles in mediating carcinogenesis, endocrine disruption, among other immunological, hepatic, cardiovascular, and dermal toxicity mechanisms. The deep physiological implications of AhR in cellular proliferation, adhesion, division, differentiation, as well as in the reproductive, immunological and cardiovascular homeostasis have opened a new field of research in order to harness AhR's pharmacological potential. Hence, AhR has become a therapeutic target of inflammatory, infectious, malignant, and immunological conditions. Toxicological and pharmacological fields could benefit from discovering novel AhR modulators to elucidate further on the chemical-biological implications of this crucial transcription factor. In this Thesis, the following objective was proposed in order to contribute to such understanding.

General Objective: Evaluate diverse chemical compounds as modulators of AhR by means of *in silico* and *in vitro* methods.

The general objective was concretized in specific aims distributed in the five Chapters of this Thesis as follow:

Chapter 1. Review the literature on AhR mediated effects and the existing theoretical and experimental methods employed to study the structural and functional aspects of the receptor.

Chapter 2. Develop and experimentally validate QSAR models to predict the AhR agonist activity of chemical compounds.

Chapter 3. Analyze the dual effects of a set of benzothiazoles as AhR modulators and antimicrobial agents.

Chapter 4. Evaluate a novel set of triarylmethane compounds as AhR ligands.

Chapter 5. Study the AhR antagonism discovered in potentially harmful substances using computational methods.

Chapter 1

CHAPTER 1. Introduction

Summary

In this first Chapter, some general definitions of the research topic addressed in this Thesis, and derived from a comprehensive bibliographic review, are introduced. The signaling pathways and molecular consequences of AhR modulation are addressed, as well as the beneficial and harmful biological consequences related to the receptor's activity. Finally, different alternative methods (*in vitro* and *in silico*) to investigate the potential effects caused by chemical compounds through AhR modulation are defined, and some examples available in the literature are mentioned.

Highlights

- The transcription factor AhR is regulates multiple signaling networks.
- AhR is a key receptor in physiological conditions but it also mediates several adverse health effects.
- AhR is a promiscuous receptor with the capacity to accommodate diverse endogenous and exogenous ligands.
- Luciferase reporter gene assays are important *in vitro* techniques widely used to explore the effects of chemical compounds on AhR expression.
- Some *in silico* studies on AhR and its ligands have been reported although important knowledge gaps remain.

Table of Contents Chapter 1

Index of Figures

Glossary

AhR	Aryl hydrocarbon Receptor
AhRR	AhR Transcriptional Repressor
ANN	Artificial Neural Network
AR	Androgen Receptor
ARNT	Aryl hydrocarbon Receptor Nuclear Translocator
CALUX	Chemically Activated LUciferase eXpression
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Similarity Indices Analysis
CYP1A1	Cytochrome P450 (family 1, subfamily A, polypeptide 1)
ECHA	European Chemicals Agency
EFSA	European Food Safety Authority
EPA	Environmental Protection Agency
ESR	Estrogen Receptor
FDA	Food and Drug Administration
FICZ	5,11-dihydroindolo[3,2-b] carbazole-12-carbaldehyde
HepG2	Human Hepatoma cell line
HIF	Hypoxia Inducible Factor
HSP90	90-kDa heat shock protein
kNN	k-Nearest Neighbors
LBD	Ligand Binding Domain
NES	Nuclear Export Signals
NLS	Nuclear Localization Sequences
NPC	Nuclear Pore Complex
OECD	Organization for Economic Co-Operation and Development
PAS	Periodic (PER)-ARNT-Single minded (SIM)
PBDE	Polybrominated diphenyl ethers
PCB	Polychlorinated biphenyl
PCDD	Polychlorinated dibenzo-p-dioxins
PCDF	Polychlorinated dibenzofurans
PLIP	Protein-Ligand Interaction Profiler
PLS	Partial Least Squares
QSAR	Quantitative Structure-Activity Relationship
RBA	Relative Binding Affinity
RF	Random Forest
RGA	Reporter Gene Assays
SVM	Support Vector Machine
TAD	Transactivation Domain
TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
WHO	World Health Organization
XAP2	Termed X-associated protein 2
XRE	Xenobiotic Response Elements

1. Aryl hydrocarbon Receptor

1.1. Functional domains

The transcription factor AhR was originally discovered in hepatocytes in the 70s. Half-century later, multiple studies have confirmed that AhR signaling affects virtually every animal cell-type and organ in vertebrates and many invertebrates [1]. The highest AhR expression has been reported in the liver, kidney, thymus, lung, spleen, and placenta. Moreover, AhR is constitutively allocated in the nucleus of aggressive malignancies where its target genes are widely expressed *i.e.* skin tumors, non-small cells lung cancer, or glioblastoma [2]. Figure 1 is a schematic representation of the AhR functional domains and its binding sites.
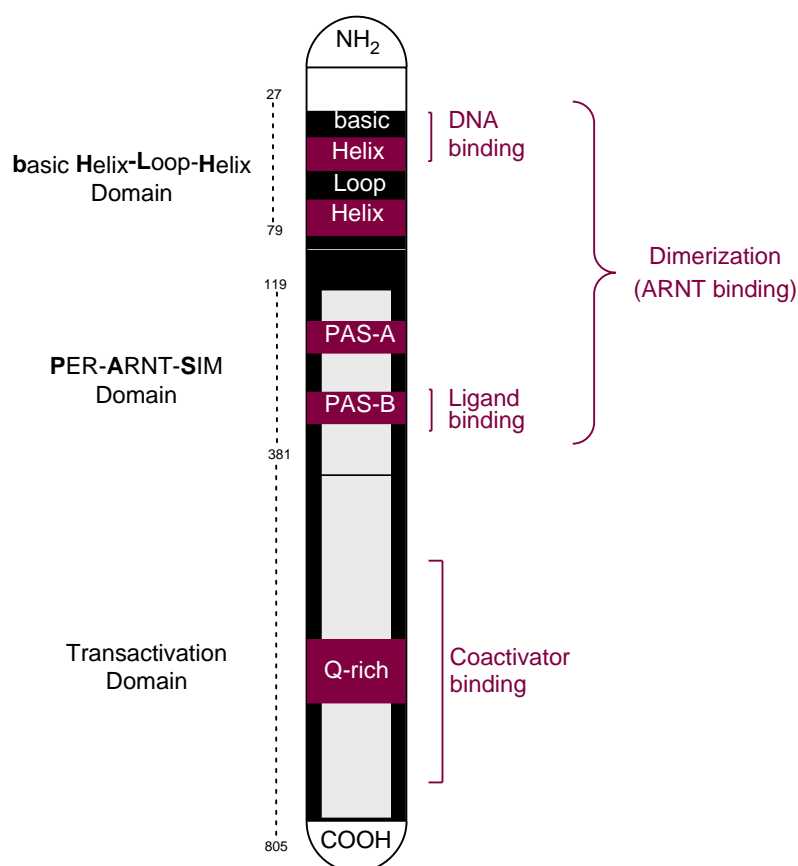
Figure 1. Domain organization of the AhR and its binding sites.

The AhR and its nuclear translocator (ARNT) belong to the bHLH-PAS superfamily of transcription factors, acronym for basic Helix-Loop-Helix and Periodic (PER)-ARNT-Single minded (**S**IM). Protein subfamilies may be stratified in two classes, with Class I including bHLH-PAS proteins such as the AhR, hypoxia-inducible factors (HIF), clock circadian

regulators (CLOCK), and neuronal PAS protein (NPAS), while Class II includes the ARNT subfamily of proteins located inside the nucleus and can heterodimerize with the cytosolic Class I members such as AhR. Both AhR and ARNT include an (N-) terminal bHLH domain, two PAS domains named A and B, and a transactivation domain (TAD) carboxy (C-) terminal. The PAS A domain controls dimerization through its connector helix and strengthens the binding to a distinct DNA motif in which bHLH plays the main role. With a three-loop insertion difference from PAS A, the PAS B is the AhR ligand binding domain (LBD) [3]. In the TAD, the Q-rich (Glutamine-rich) subdomain is the key region for transcriptional activation of xenobiotic response elements (XRE) in the DNA [4]. The binding to coactivators during the transcriptional process regulates the wide diversity and tissue-specific AhR effects [5].

1.2.　Cytoplasmatic complex and signaling pathways

The cytoplasmatic inactive form of AhR (monomer) is found associated with two molecules of 90-kDa heat shock protein (HSP90) chaperon, a co-chaperone p23, a termed X-associated protein 2 (XAP2) and a protein kinase SRC [6]. While one of the HSP90 interacts with both the bHLH and PAS regions, the other unit is only bound to the PAS domain. The HSP90 is a folding machinery found also in steroid receptors such as estrogen, progesterone, and glucocorticoid receptors. The resistance to salt treatment and the absence of molybdate led to suggestions that HSP90/AhR binding may be the most stable known for that chaperone. However, important inter-species differences have been identified [7,8]. The co-chaperone p23 is a commonly found protein in HSP90 complexes where it contributes to the ligand binding and to the client's release from HSP90 [9,10]. More recently, a protecting role of p23 against AhR's degradation independent of HSP90's functions has been acknowledged [11]. Meanwhile, XAP2 presents a discrete domain organization with a tetratricopeptide repeat motif whose carboxyl-terminal domain is responsible for the interaction with AhR protein. In the AhR core complex, XAP2 positively or negatively modulates the receptor's sensitivity to ligands and the posterior transcriptional responses [12–14]. The c-SRC tyrosine kinase contributes to ligand recognition and indirectly ensures the transcriptional process [15]. Figure 2 represents some of the molecular consequences upon ligand binding to the AhR complex.
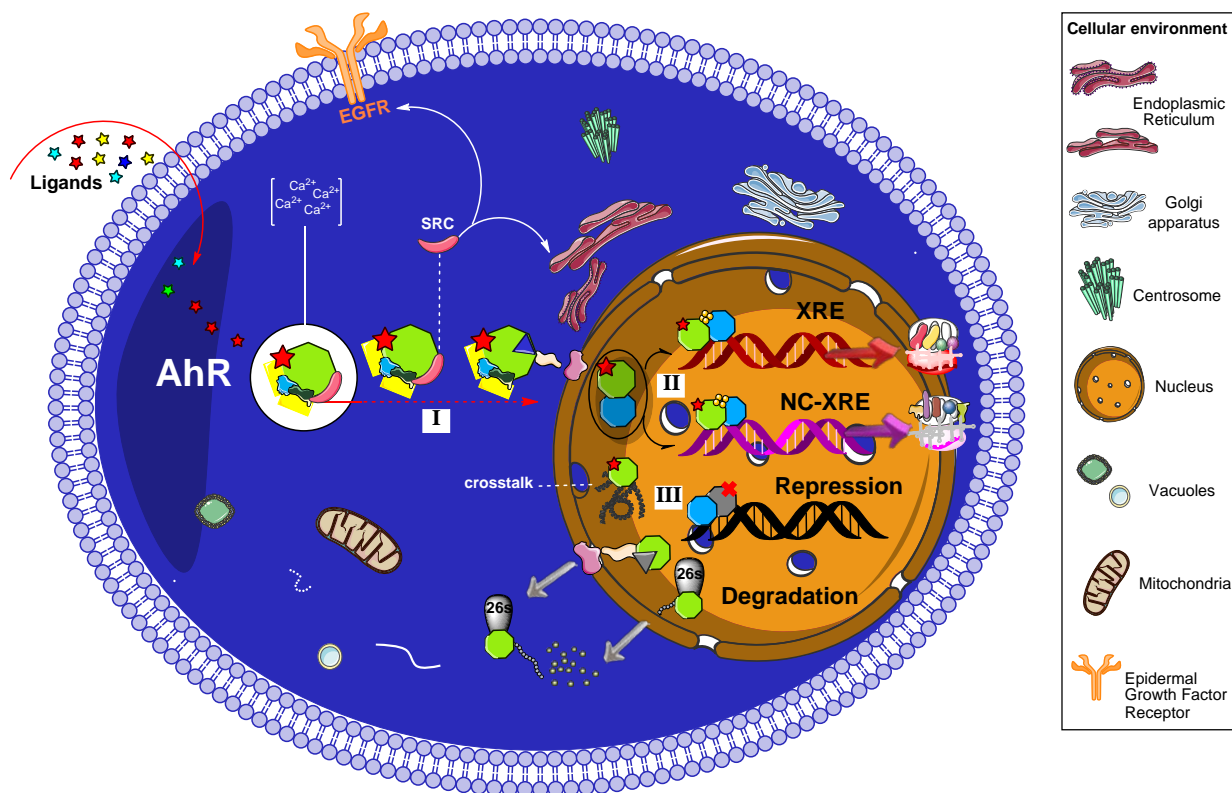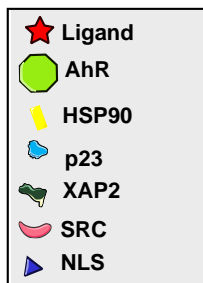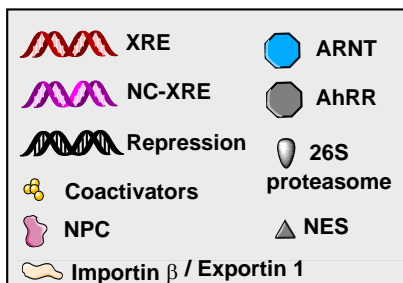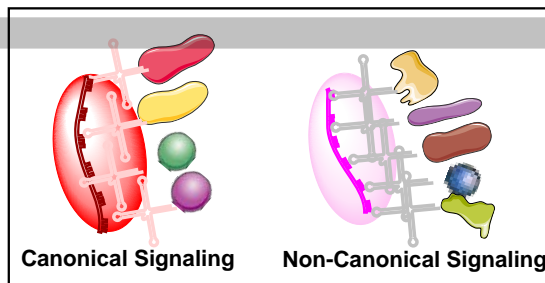
Figure 2. Schematic representation of AhR-mediated pathways. Ligands bind to the cytoplasmatic AhR, where it is forming a complex with HSP90, p23, XAP2, and SRC. **(I)** Agonist ligands promote conformational changes and the exposure of NLS, allowing the nuclear translocation mediated by importin β through the NPC. **(II)** Once in the nucleus, the heterodimer AhR-ARNT can bind to different sequences in the DNA and trigger the synthesis of proteins (*e.g.* CYP1A1, UGT1A6), which is additionally regulated by coactivators. **(III)** Crosstalk interactions with other signaling pathways are acknowledged (*e.g.* ESRs, AR). Moreover, the Ah receptor can be repressed by AhRR, and its degradation is mainly mediated by 26S proteasome.

Genomic and non-genomic pathways are suggested for AhR signaling as represented in Figure 2. The genomic induction of target genes is the fundamental and best-known mechanism of AhR-mediated signaling. The genomic pathway (I) starts once AhR agonists bind to the LBD causing conformational changes in the PAS A that trigger the exposure of nuclear localization sequences (NLS). The SRC is then released from the AhR-chaperone complex to the cytosol, triggering phosphorylation processes. The entrance to the nucleus is facilitated by importins which recognize the NLS. Specifically, importin β is the direct mediator that facilitates the passage to the cytoplasmic face of the nuclear pore complex (NPC) [16–18].

After the AhR-chaperone complex is translocated to the nucleus, the receptor heterodimerizes with ARNT losing the chaperone proteins in the process and forming the AhR/ARNT complex (II). Depending on the target genes involved in the subsequent transcriptional activity triggered by the AhR/ARNT, the genomic signaling is classified as canonical when it is mediated by XRE or else as non-canonical or no consensus XRE (NC-XRE) responses (Figure 2).

The canonical signaling starts when the AhR/ARNT complex binds to the XRE identified by the DNA consensus motif sequence 5'-TNGCGTG-3' [19,20]. The NC-XREs, sometimes referred to as AhR Responsive Elements-II, are recognized by the consensus promoter region 5'-CATG{N6}C[T|A]TG-3' [21]. A scaffold is formed upon binding to AhR core regions inducing the recruitment of multiple coactivators that ultimately regulate the transcriptional process [5]. In addition, similar to other PAS proteins involved in critical signaling functions, AhR expression is controlled by a transcriptional repressor (AhRR) with a similar N-terminal region sequence. Thus, the AhRR is an AhR competitor in that it may form the heterodimeric AhRR/ARNT complex, which binds to the XRE, and consequently repress the transcription. The AhRR effects mediate key processes such as inflammation and tumor growth that has led to its consideration as a drug target [22,23]. The repression pathway is also controlled by the receptor itself and its transcriptional outcomes forming a regulatory circuit [24].

The prototypical canonical target gene identified for AhR determines the induction of cytochrome P450 from family 1, subfamily A, polypeptide 1 (CYP1A1) xenobiotic metabolizing enzyme [25,26]. At the same time, AhR regulates the basal *CYP1A1* expression through elements such as the so-called special protein family, and particularly the specificity protein 1 (Sp1) [27]. Moreover, AhR signaling through XRE has been associated with other family members such as CYP1A2, CYP1B1 as well as phase II enzymes such as the UGT1A6

[28,29]. The understanding of the AhR complexity has been further expanded with the discovery of multiple nonclassical target genes in the past decades [30]. Some of the most important NC-XRE pathways are related to the tumor suppressor Kruppel-like factor 6 (KLF6) and the recruitment of flanking target genes such as the plasminogen activator inhibitor (*PAI*)-1 and the p21^{cip1} protein [31].

Among the non-genomic mechanisms that AhR activation could trigger is the above commented phosphorylation induced by the c-SRC of multiple key protein substrates as the epidermal growth factor receptor (EGFR) [32,33]. Furthermore, c-SRC, the mitogen-activated protein kinases (MAPKs), reactive oxygen species (ROS) and AhR-dependent paths have been recently associated with the apoptosis of neuronal cells through the stress of the endoplasmic reticulum [34]. The proteasome-mediated degradation of ubiquitinated proteins is another important non-genomic pathway of AhR activation, which has been linked to the inflammatory process [35,36]. Thus, AhR can act as a ligand-dependent ubiquitin protein ligase (E3) promoting the proteolysis of other transcription factors such as the estrogen receptor (ESR) and androgen receptor (AR) [37,38]. Furthermore, AhR controls the levels of intracellular calcium and the induction of enzymes such as the cytosolic phospholipase A2 (cPLA2) and cyclooxygenase 2 (COX2) [39].

On the other hand, AhR modulation intervenes in the signaling networks of many other crucial cellular entities [40]. The crossroad with estrogen receptors (ESRα and ESRβ) is among the best studied and it has been shown to lead to transcriptional activation of estrogen-responsive gene promoters and the consequent estrogenic effects [41]. Besides, AhR has proven to interact with the canonical Wnt/β-catenin pathway possibly mediated by some other entities such as R-spondin 1 [42], which is an important element in cellular proliferation as well as in sexual organ's development and differentiation [43,44]. Such an interaction associates the receptor with the regulation of organogenesis and other key biological processes such as cell polarity and migration attributed to the glycoproteins Wnts [45]. In addition, AhR crosstalk interactions include effects on the transforming growth factor-β/Bone morphogenetic protein (TGFβ/BMP) pathway, which is a target of drugs currently in phase I or phase II clinical trials for the treatment of ovarian cancer, hepatocellular carcinoma, glioblastoma and melanoma [46].

Lastly, the downregulation of AhR is described as a rapid crucial attenuation of AhR effects on gene transcription. The proteolytic degradation is a terminal step mainly ubiquitin-mediated

through the 26S proteasome. It could be triggered if the receptor complex has a misfolded conformation prior to entry to the nucleus. Otherwise, after the signaling processes, AhR is ubiquitinated and degraded inside the nucleus or transferred to the cytoplasm via the Exportin 1 (or CRM-1) that recognizes the nuclear export signals (NES). In the cytosol AhR is targeted and degraded by 26S proteasome [47,48].

1.3. Physiological effects

The ligand, cell, tissue-specific, and pleiotropic effects of AhR-mediated activity are still puzzling and, in many cases, poorly understood [6]. Human diet and gut microbiome are tremendous sources of AhR modulators that trigger complex and fundamental processes in health and diseases [49,50]. It is believed that at least the downstream effects of dietary AhR inducers fit the renowned quote of the father of toxicology Paracelsus that "*Sola dosis facit venenum*" (the dose makes the poison) [51]. However, countless low-dosage activities are probably still underestimated and not fully understood by current science. Overall, AhR pathways intricacies have led to consider that many of its adverse health consequences could be an exacerbated cellular response to its endogenous roles [40,52]. The truth is that these contradictory statements remain to be clarified on this fascinating transcription factor [53].

The proven AhR physiological functions, particularly important in barrier organs as the gut or the skin, are nowadays driving a new line of research of the formerly known as "dioxin receptor" [54]. Toxicologists have been for decades interested in AhR-dependent signaling as a key intermediate step in the toxic responses to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and derivatives [53]. Immunological, hepatic, and endocrine disruptions, as well as carcinogenic effects, are attributed to AhR-mediated activity [30]. The major AhR physiological functions have captivated a wide range of experts, which has led to propose the receptor as a diagnostic marker and therapeutic target in multiple human diseases [55]. Some of the main harmful, beneficial and still unknown roles of AhR are briefly summarized in Figure 3.
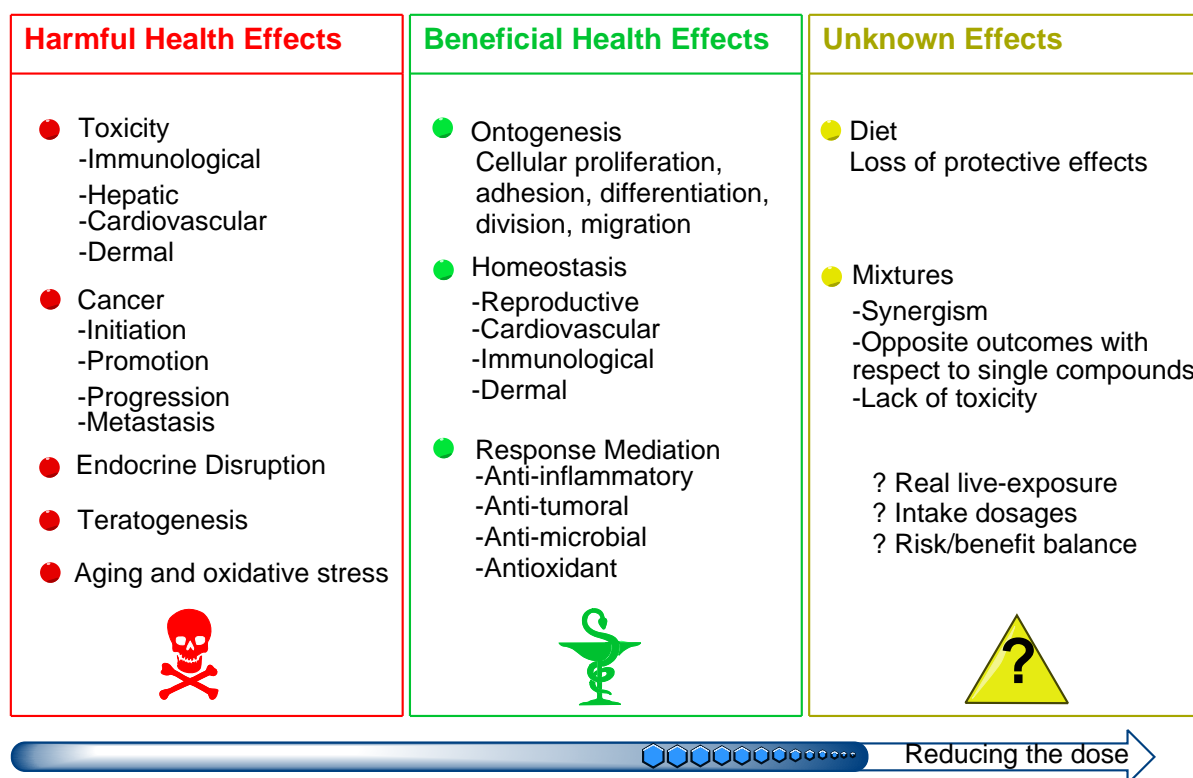
Figure 3. Some harmful, beneficial and unknown effects associated with AhR.

AhR integrates toxicological and crucial physiological functions as represented in Figure 3. Low doses, mixtures, and the general risk/benefit balance are among the main open questions to be further investigated. In general, AhR is associated with key processes such as development, cell cycle regulation, immunity, resilience against stress, homeostasis, and metabolism [56]. Therefore, potential therapeutic uses of AhR modulation include maintenance of lung health [57], control of inflammatory process in vascular tissues [58], treatment of liver and cystic fibrosis [59,60], control of the antioxidant response [61], and regulation of neural functions in both vertebrates and invertebrates [62]. Moreover, the most significant pharmacological applications of targeting AhR are probably in the treatment of several cancer types, where drugs such as Phortress (NSC 710305) have been recommended in the chemotherapy of CYP1A1-positive tumors [2,63,64]. In addition, important inflammatory and immunological conditions could be modulated through AhR expression and particularly those affecting gut and intestinal tissues [49,50,65]. Hence, promising drug candidates such as NPD-0414-2 and NPD-0414-24 have been recently suggested in the therapy of colitis [66].

Unfortunately, the intrinsic biological implications of AhR have been little exploited in pharmacological contexts due to concerns on its roles in mediating pollutants such as biphenyls

and polychlorinated dibenzodioxins [67]. However, toxicological research also bequeaths crucial leads currently exploited in therapeutics research [54].

Less understood than AhR activation, the antagonism upon the receptor's expression is particularly important as a detoxification system and as a therapeutic perspective in cancer. Certainly, dietary AhR antagonists are suggested to have multiple health benefits [68]. Moreover, several environmental contaminants and highly toxic substances and mixtures can also antagonize the receptor's effects [69,70]. Therefore, the tricky and contradictory outcomes of AhR and the implications that it may have in the toxicological or pharmacological contexts indistinctly should not be generalized under any circumstances. Both the activation and the inhibition of this multifunctional chemical sensor could be beneficial and harmful for human health [54].

2. Endogenous and exogenous ligands

Numerous and structurally diverse chemical entities have been identified as modulators of the Ah receptor. Furthermore, there are endogenous and exogenous (man-made and dietetic) sources of AhR ligands identified [71,72]. Prototypical examples are provided in Figure 4 and Figure 5, respectively.

A wide structural diversity of AhR agonist and/or antagonist modulators have been acknowledged, as illustrated in Figure 4 and Figure 5. Indeed, the receptor has been considered a "promiscuous" sensor of chemical entities [72]. However, a comprehensive understanding of the structural and functional profile of the AhR has been limited by the unavailability of an experimentally determined AhR protein structure co-crystallized with the corresponding functional domains. Consequently, studies aimed at elucidating the interaction modes involved in the AhR signal transduction pathway have relied on homologous systems [73]. Discovering chemical entities acting as AhR agonists and antagonists contribute to a better understanding of the molecular interactions as well as the toxicological and therapeutic roles of the receptor [74]. In toxicology, this has helped to identify potential high concern substances and environmental contaminants, particularly persistent organic pollutants whose exacerbated effects could trigger the adverse outcomes associated with AhR transcriptional activity [75]. Meanwhile, the identification of AhR modulators as drug candidates from naturally occurring compounds and through *de novo* synthetic procedures for the treatment of malignancies, immunological and inflammatory conditions is an active research field [76,77].

FICZ
172922-91-7

Indirubin
479-41-4

Bilirubin
635-65-4

Equilenin
517-09-9

Lipoxin A4
89663-86-5

Serotonin
50-67-9

Prostaglandin G2
51982-36-6

Haemin
16009-13-5

Kynurenine
343-65-7
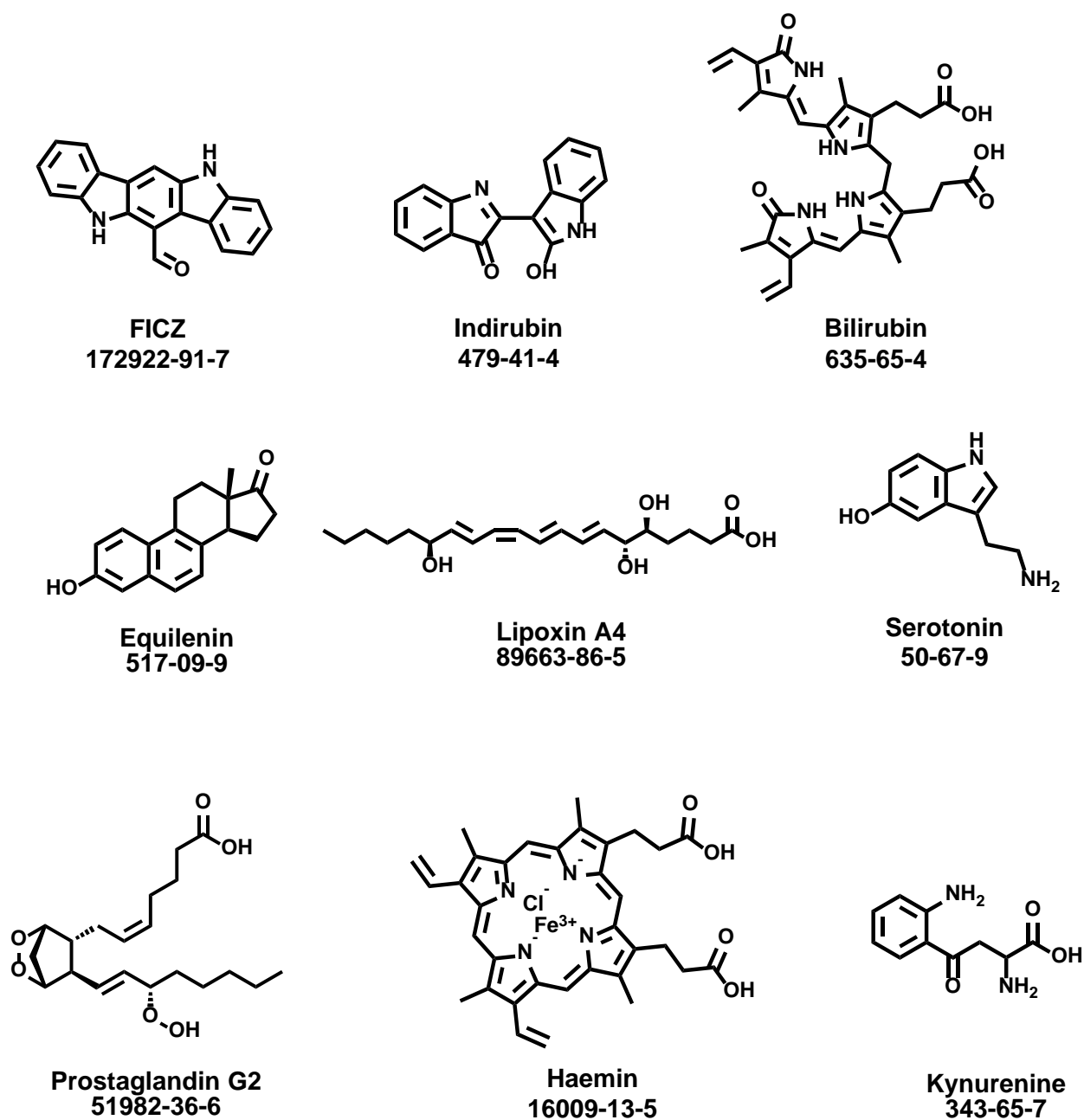
Figure 4. Representative examples of AhR endogenous ligands

SYNTHETIC

TCDD
1746-01-6

alpha-naphtoflavone
604-59-1

Phortress
741241-36-1

Benzo[A]Pyrene
50-32-8

CH223191
301326-22-7

Omeprazole
73590-58-6

NATURAL

Curcumin
458-37-7

Berberine
2086-83-1

Resveratrol
501-36-0

Diosmin
520-27-4

Harmane
486-84-0

Quercetin
117-39-5
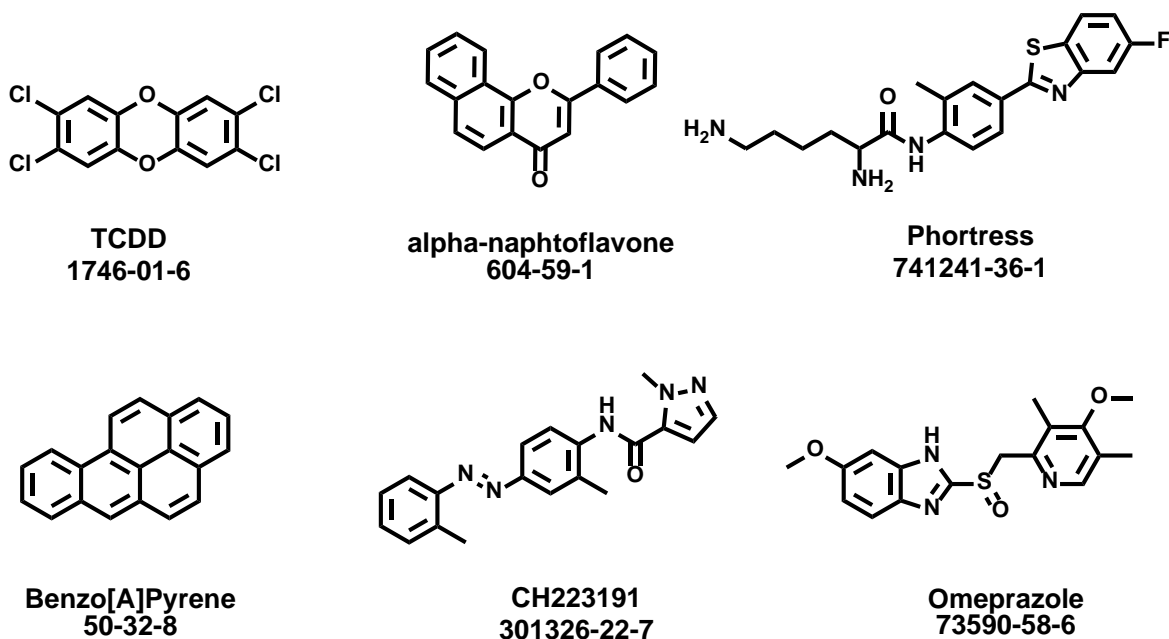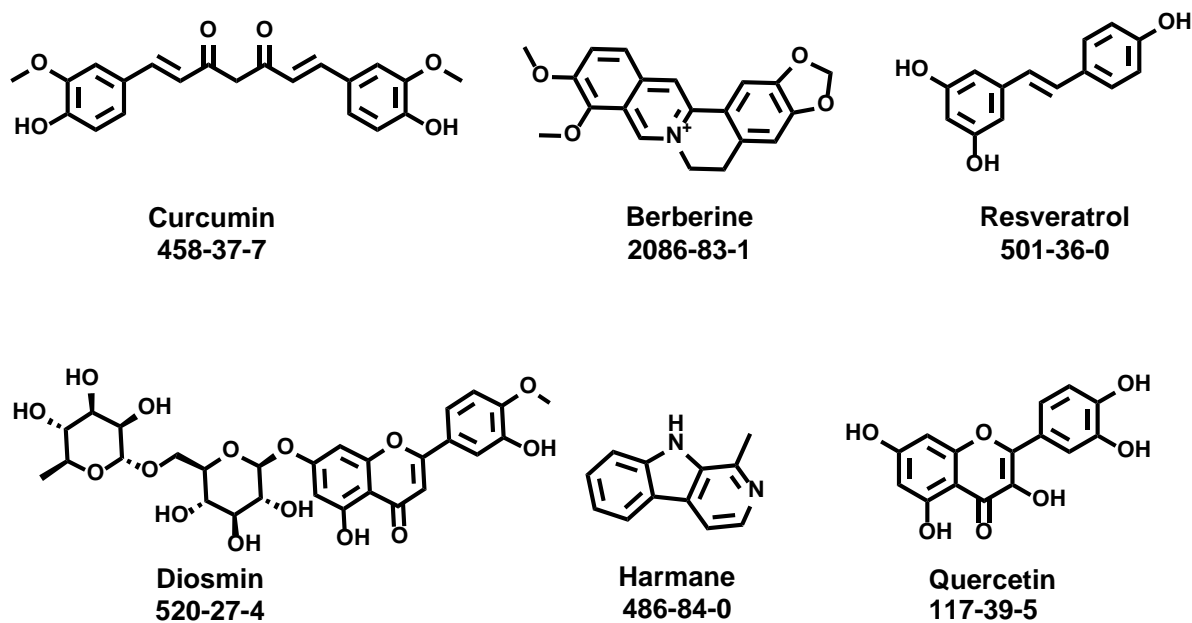
Figure 5. Examples of synthetic and natural occurring AhR exogenous ligands

3. Non-animal methods in chemical biology

Although the analysis of the chemistry of biologically assayed compounds was not usually a priority in experimental studies, this trend has gradually changed. Thus, the volume of available information relative to the molecules causing biological effects has significantly increased. Such analysis, sometimes called a "chemocentric" view of the biological data (as shown in Figure 6) has improved and brought closer the chemical and biological fields of research [78,79].
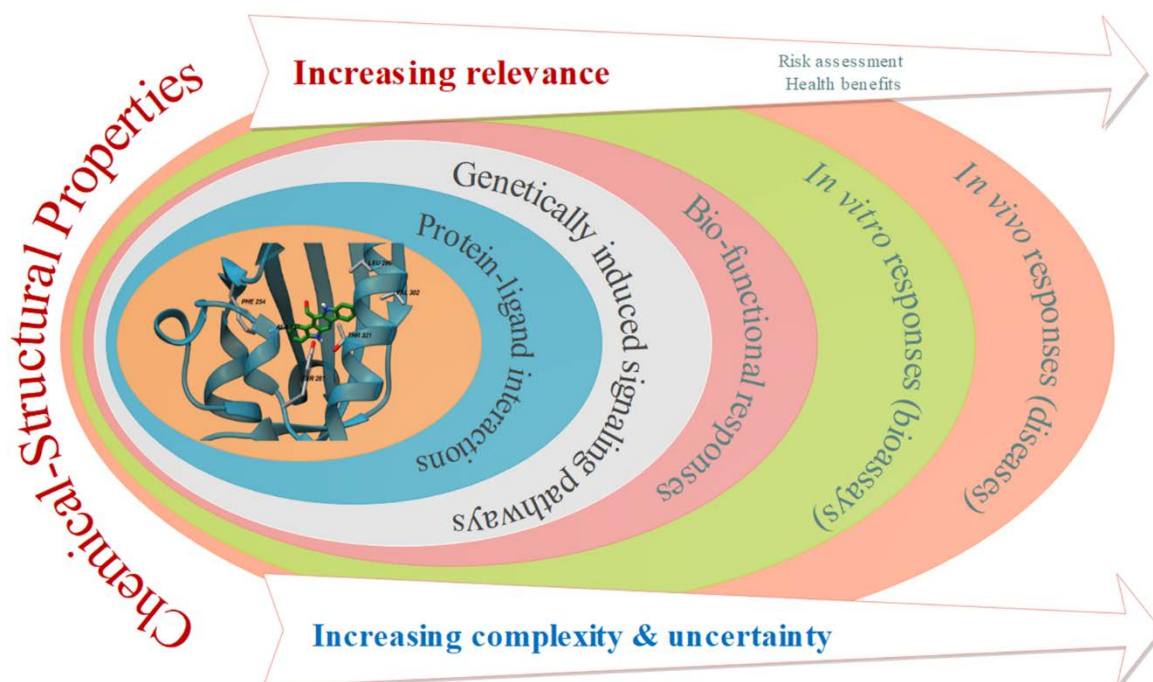
Figure 6. Chemocentric view of biological data (Modified sketch of a graph presented by Alexander Tropsha [80]). Protein-ligand (AhR-FICZ) is shown [81].

As represented in Figure 6, the chemical-structural properties and the interactions with the protein targets could represent key or "core" elements from the lowest to the highest levels of complexity in biological determinations [80].

In chemical biology, the so-called alternative methods to animal experimentation have become a fundamental research arena gaining importance and popularity [82,83]. These methods embody the 3Rs (Replacement, Reduction, and Refinement) guidelines for optimizing, reducing and ideally substituting the use of animal models in toxicology, pharmacology, agricultural and environmental research [84]. Important initiatives aimed at promoting the complete substitution of animals in life science have been launched [85]. However, to date, the applicability of non-

animal methods is still limited by knowledge gaps in multiple areas of research. Furthermore, *in vivo* testing remains crucial in fields such as veterinary and human medicine [86].

Beyond the ethical implications and economic appeal, *in vitro* assays and *in silico* modeling possess unique advantages in chemical-biological research. As major screening tools, they can provide predictive results and mechanistic interpretations of the alleged modes of action of chemical compounds in biological contexts [87]. The main contributions of *in silico* methods have been in the identification of target receptors and lead compounds [88]. Meanwhile, the *in vitro* bioassays have aided in simulating physiological environments in order to measure and analyze bioactivities with high precision [89,90].

The high success rates of *in vitro/in silico* methods in extrapolating to real contexts have contributed to their acceptance in the scientific community and stakeholders. Hence, most of the regulatory agencies and organizations (*i.e.* ECHA, EPA, OECD, FDA, EFSA, WHO), strongly support the use of alternative methods prior to any animal experimentation. Moreover, in the European Union, animal testing has been banned for the cosmetic industry, and globally, the risk assessment and hazard identification of chemicals is mainly studied through combinations of *in silico* and *in vitro* methods [91,92].

3.1. *In vitro* luciferase reporter gene assays

The *in vitro* ("in the glass") methods emerged from basic research in biochemistry and microbiology [93], and nowadays they are widely used in all fields of life science and environmental research [94]. There are multiple *in vitro* methods, that range from the study of biologically relevant reactions (*e.g.* antioxidant bioactivity assays), to the assessment of interactions (*e.g.* binding experiments), and culturing procedures [95,96]. The *in vitro* culture methods allow the extension of the 'unit of life' paradigm by isolating cells and tissues from the species-specific influences of an intact organism [97]. There are culture procedures based on subcellular systems, cell lines, multicellular systems, and organ isolated bioassays [98–100].

The reporter gene assays (RGA) are among the most popular techniques employed to study transcription factors such as nuclear receptors (*e.g.* ESRs, AR) and the AhR. The RGAs are *in vitro* bioassays that can be used to indirectly identify either induction or blocking of signaling pathways. Furthermore, the superior complexity and informativeness of RGAs with respect to

other *in vitro* techniques, make their results on the transactivity induced by chemicals, very reproducible and suitable for computational modeling [101].

The RGA methods involve the transfection of cell lines and the measurement of the receptor's transactivation mediated by signaling proteins such as luciferases [102]. The generic name of luciferase has been given to enzymes that catalyze the oxidation of substrates and thus inducing, the production of light. There are different types of luminous organisms in which bioluminescent reactions occur. However, only few luciferases have been sequenced and characterized [103]. Figure 7 shows some representative examples of luciferases and the bioluminescent organisms where they are produced.
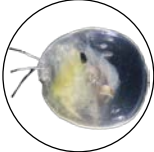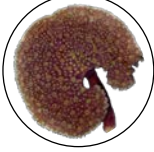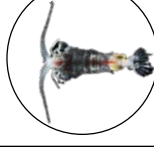
Bioluminescent organism		size	emission	secreted
	Photinus pyralis "firefly" insect	61 KDA	Glow	✖
	Vargula hilgendorfii "sea-firefly" crustacean	62 KDA	Glow	✔
	Renilla reniformis "sea pansy" anthozoa	36 KDA	Flash	✖
	Gaussia princeps crustacean	20 KDA	Flash	✔

Figure 7. Examples of luciferases and the bioluminescent organisms where they are produced. The size of the luciferases, their type of emission, and whether they are or not secreted to the culture medium, are shown.

The intracellular firefly luciferase from the North American firefly *Photinus pyralis* and the Renilla luciferase, from the sea pansy *Renilla reniformis*, are among the best known and most commonly used as reporters [104,105]. Cloned secreted luciferases from copepods such as *Gaussia princeps* and *Metridia longa*, are likewise increasingly being utilized as bioluminescent reporters in biomedical research. The firefly luciferase catalyzes the oxidation of luciferin and

copepod luciferases catalyze the coelenterazine oxidation that produces blue light emission [106]. These substrates and their signals are notoriously different, as represented in Figure 8. While the oxidation of the luciferin leads to a luminescence stable with time although not very intense, the coelenterazine's oxidation produces a flash blue light emission significantly high but detectable only during few minutes.
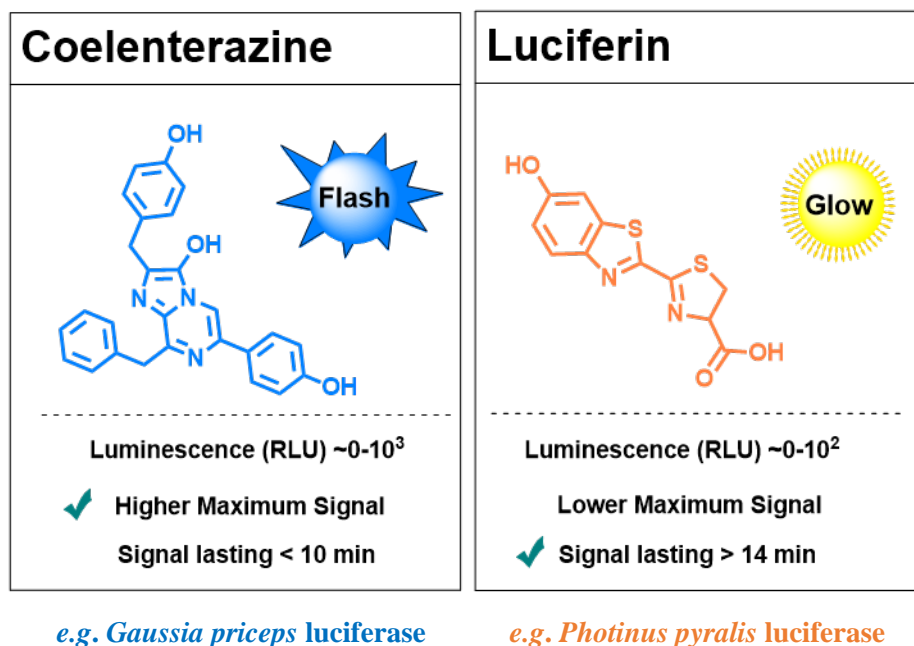
Figure 8. Differences between coelenterazine and luciferin substrates of luciferases.

Among the most popular RGAs are the Chemically Activated LUciferase eXpression (CALUX) bioassays, developed by the BioDetection Systems company and widely used in the study of dioxin and dioxin-like chemical compounds [107]. The DR-CALUX® assay uses the rat hepatoma H4IIE cell line with an incorporated firefly luciferase gene coupled to XRE as a reporter gene of the AhR transactivation [108]. However, the main disadvantage of the assays based on Firefly and Renilla luciferases is the necessity of cell lysis procedures to measure the bioluminescence, which could cause noticeable interferences in the results. A feasible alternative to overcome that disadvantage is the use of secreted luciferases [109].

In this sense, a novel secreted luciferase named "Lucia" that is produced via a synthetic gene designed from marine copepods has been engineered by the InvivoGen company. Some of the advantages claimed for this innovative luciferase compared to others are the intensity of the bioluminescence and the stable expression, which provide reliability and biological significance to the RGAs. Moreover, the Lucia luciferase's one-step detection reagent (named

QUANTI-Luc™) of contains coelenterazine and stabilizing products that limit the autoxidation of this substrate [103]. The human hepatoma cell strain (HepG2) is one of the engineered cell lines (HepG2-Lucia™ AhR) available for the detection and screening of AhR ligands in the cell culture supernatant [110].

3.2.   *In silico* **methods**

Computational chemistry is one of the most rapidly advancing scientific areas, permeating a wide array of disciplines such as nanotechnology, material sciences, analytical chemistry, molecular biology, environmental sciences, toxicology, and drug discovery, among others [111]. The pseudo-Latin expression *in silico* methods (derived from "*in silicon*" in reference to the constitution of computer chips) was first used around the 1980s to semantically associate computational methods with the *in vivo*, *in vitro*, and *in situ* procedures. Nowadays, it is a recognized terminology particularly used to allude to computational models for studying biological processes and properties [112].

Computer-aided chemical research follows two major approaches: structure-based methods and ligand-based methods. The structure-based methods require information on the target structure (*i.e.* biomolecule, receptor), while the ligand-based methods rely exclusively on the information provided by the ligand (*i.e.* compounds with demonstrated chemical biological properties) [113]. Some examples of *in silico* methods applied according to the available information are represented in Figure 9.

Figure 9. Examples of *in silico* methods that can be used according to the known information.

As shown in Figure 9, computational combinatorial chemistry methods can be useful in designing chemical libraries for drug discovery and risk assessment of chemicals, particularly in the absence of structural information on the target and possible ligands [114]. The so-called "de novo" design approaches are useful in multiple contexts to suggest new scaffolds to be synthesized and experimentally tested. Besides, they are especially important for generating structures for virgin targets (those for which no active compounds are known) [115]. If the ligand has been identified or both the ligand and the receptor are known, various *in silico* methodologies can be applied. Three of the most important ones in the toxicological and pharmacological contexts are detailed below (*i.e.* QSAR modeling, Pharmacophore mapping, and Molecular Docking).

3.2.1. QSAR Modeling

The main steps of QSAR modeling are, first, to parameterize the chemical structure, and then, to establish the relationship between these parameters and the studied physicochemical, pharmacological or toxicological activity (*e.g.* and antibacterial efficacy, toxicity index, solubility). The QSAR methods are based on the understanding that the steric, geometric, and electronic properties of the molecules, as calculated or measured by the so-called molecular descriptors, ultimately determine their behavior in biological environments, as well as the physicochemical properties attributable to them [116]. Therefore, the three fundamental elements of QSAR studies are the representation of the molecular structure, the definition of molecular descriptors, and the statistical correlations obtained when the activity is modeled as a function of these descriptors. The development of novel approaches for similarity analysis and clustering, as well as the state of the art statistical algorithms and artificial intelligence methods have contributed to the advancement of QSAR modeling [117]. Furthermore, numerous and interdisciplinary applications in diverse fields such as agronomy, food science, pharmacology, and toxicology have been ascribed to QSAR models [88].

3.2.1.1. Conceptualization of QSAR methodology

Crum-Brown and Fraser introduced in the late 19th century the "first formulation of the quantitative structure-activity relationships", in which they proposed the biological activity (Φ) as a function of the chemical constitution (C), *i.e.* $\Phi = f(C)$. In this sense, variations in one (ΔC) are reflected in the other ($\Delta\Phi$) [118]. However, the first QSAR models as such were not

proposed until a half-century later, by Corwin Hansch, Robert M. Muir, and Toshio Fujita, among others [119–121].

The molecular descriptors are defined as "*the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment*" [122]. Hence, the numeric data generated through the calculation of molecular descriptors codify the chemical information of each studied compound. This meaningful transformation of the molecular structure is the backbone of QSAR modeling [116]. Meanwhile, the evaluation of activities, for example through *in vitro* and *in vivo* bioassays, generates experimental measurements (categorical or continuous) that are statistically associated with the predictor variables (*i.e.* the molecular descriptors), as represented in Figure 10.
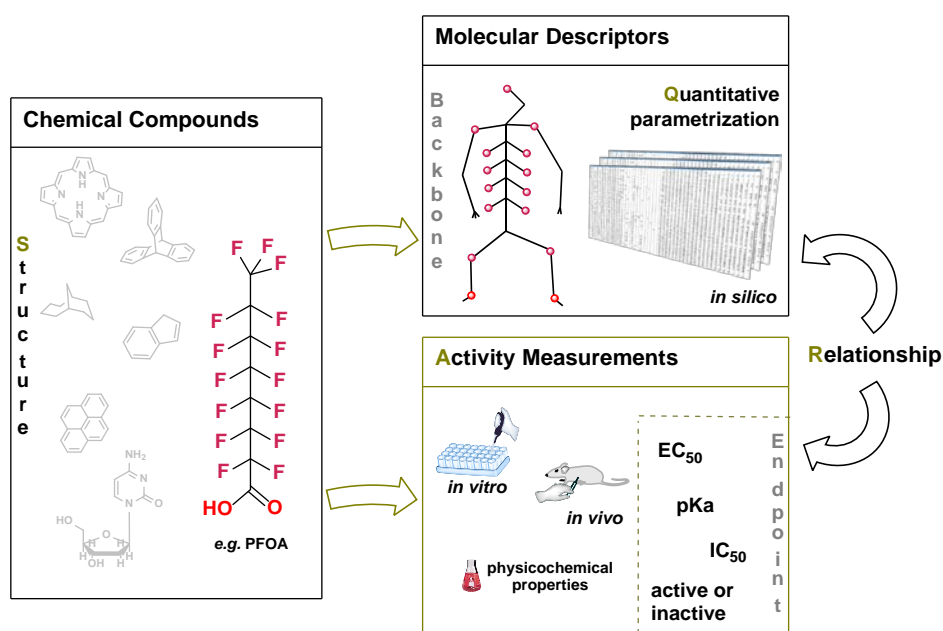
Figure 10. Schematic representation of the elements involved in a QSAR model

Thousands of molecular descriptors have been defined to date derived from the analysis of numerous theories. In general, descriptors may be stratified into two groups a) experimental measurements (*e.g.* polarizability, log P, molar refractivity), and b) descriptors derived from the symbolic representation of the molecules, also known as theoretical descriptors. Furthermore, the theoretical descriptors are classified into 0-6D descriptors depending on the considered structural information. The 0D descriptors are derived from the chemical formula of the compounds (*e.g.* atomic weight, number of atoms). The 1D descriptors are lists of

structural fragments of the molecules, (*e.g.* count of functional groups or substituents of interest). The 2D descriptors are derived from the bidimensional representation of the molecules, while the 3D descriptors provide information on spatial (configurational and conformational) descriptors. The 4D descriptors additionally include information about the interactions with the target receptor. The dimension added to 5D descriptors is related to stereodynamic information, while 6D descriptors consider the solvation functions of the molecules [123,124].

The relationships between the molecular features and measured biological properties have been modeled using different methodologies. The selection of the adequate method will depend on the available information and the ultimate goal of the study. A simplistic representation of a QSAR model can be a linear function as in Equation 1. However, much more complex algorithms have been employed in building of QSARs (*e.g.* Decision Trees, Bayesian Networks, Artificial Neural Networks, Support Vector Machines, etc.) [125].

$$Activity = a(descriptor_1) + b(descriptor_2) + (...) + c$$

(1)

3.2.1.2. QSAR workflow

Although the construction of QSAR models is a dynamic and interactive process, there are some systematic steps followed in the modeling workflow (as shown in Figure 11).

The construction and curation of the databases is the first common step in all QSAR methodologies. A suitable database must not contain duplicates, salts, mixtures, counterions, organometallic or inorganic compounds. Additionally, structure validation procedures as well as normalization of chemotypes and tautomeric forms is recommended. Meanwhile, those activity results retrieved from literature with doubtful experimental determination or contradictory effects should be excluded from the analysis to minimize potential sources of error. The analysis of the size and balance of the database is also key to design reliable QSAR models. Indeed, statistical methods are sometimes inefficient when processing large datasets, while QSARs based on small databases (*i.e.* less than 40 compounds) may result in rather narrow applicability domains. Moreover, the number of compounds in different categories or activity classes sometimes is imbalanced, which could bias the classification model and lead to incorrect predictions of the minority class [126].
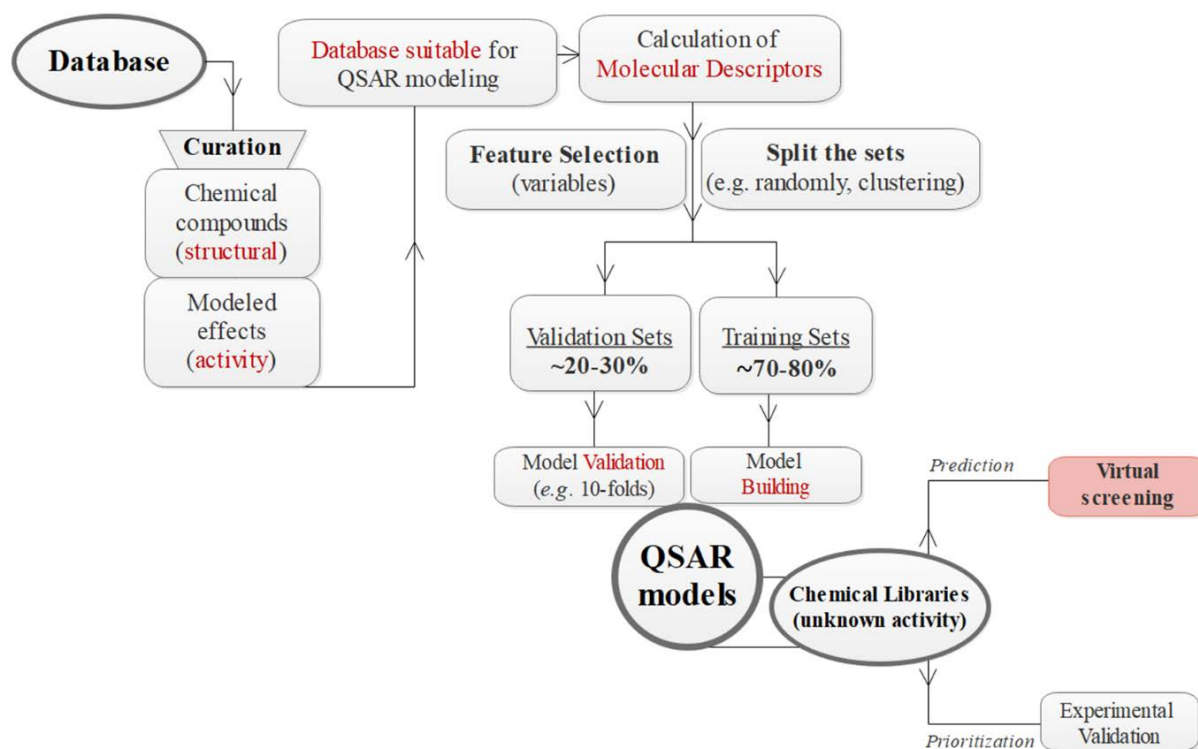
Figure 11. General flowchart of quantitative structure-activity relationship (QSAR) modeling

The calculation of molecular descriptors is followed by the selection of those descriptors more relevant to the target activity, which is commonly performed by means of a variable selection procedure (*e.g.* forward stepwise, genetic algorithms) [127]. The training and validation sets could be randomly assigned, although similarity analysis guarantees the rational distribution of objects and provides means to establish reliable QSAR models [128]. Different classes of modeling techniques, categorical, or quantitative, linear or nonlinear, could be used to describe meaningful relationships between structures and their corresponding effects (*e.g.* PLS, kNN, RF, ANN, SVM). However, the validation of QSAR models is crucial to assure their predictive capacity.

The validation procedure of QSARs comprises an evaluation of the model's capacity to generate accurate predictions. When such validation is conducted using compounds that were not included in the model building, it is denominated as external validation [129]. In this sense, the division of the original dataset in several training sets and validation sets (*e.g.* 10-folds division) has been highly recommended in the framework of the QSAR methodology [130].

The main objective of QSAR modeling is to virtually screen molecules with unknown activity to predict their behavior. These predictions could narrow down large databases, for instance,

to prioritize potential drugs or to identify health-risk chemicals. Ideally, the computational predictions should be confirmed by prospective experimental validation of some of the screened chemical compounds [130].

3.2.1.3. OECD principles of validation

When developed under standardized conditions, QSAR methods are recognized with regulatory purposes by entities such as the Organization for Economic Co-operation and Development (OECD), the European Food Safety Authority (EFSA) and the European Chemicals Agency (ECHA). Five OECD principles (Figure 12) are suggested to allow the use of QSAR models in the regulatory assessment of chemical safety. Nowadays, they constitute a helpful guide to validate QSAR models for a variety of applications.
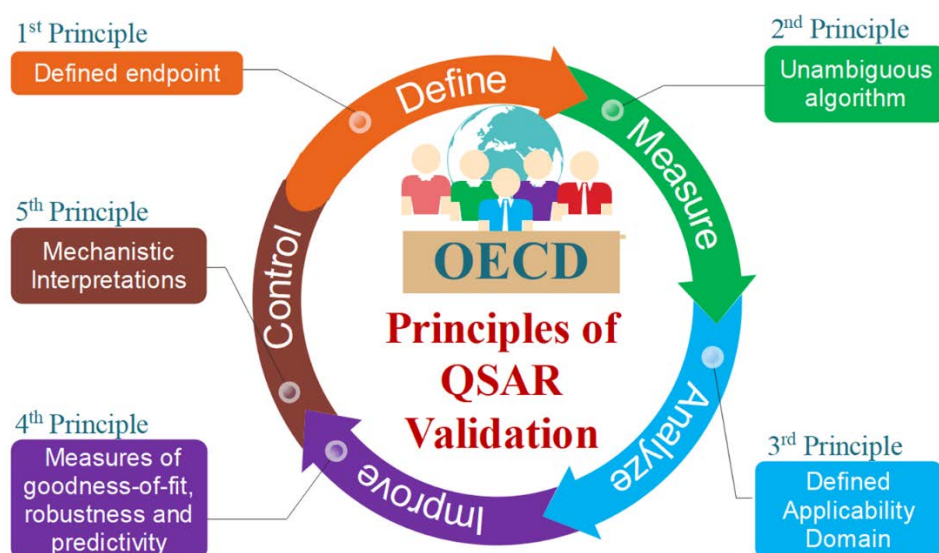
Figure 12. The five validation principles of QSAR methods according to the OECD guidelines represented as a DMAIC (Define, Measure, Analyze, Improve, and Control) process.

For the first principle, the guide indicates that complete information on the modeled property or activity should be provided in order to judge if the use of the selected endpoint is appropriate. The second principle seeks transparency in the built statistical models to ensure that they are reproducible by others.

The predictions made by QSAR models are reliable only in a "domain" where they can be applied. Therefore, to avoid making false assumptions based on unreliable predictions, the third regulatory requirement indicates that this limited space should be adequately delineated. The

fourth OECD principle evokes the need for appropriate measures of goodness-of-fit, robustness, and predictivity (*i.e.* internal and external validation). Lastly, when possible, a mechanistic interpretation is highly recommended to add credibility and acceptance to the predictions [131]. Although the mechanistic interpretation of the models is not mandatory, it is highly demanded by stakeholders. Some authors have criticized this demand highlighting that it has sometimes biased the final and most important goal of QSAR studies, which is to design models with high predictive power [130,132].

3.2.1.4. QSAR studies of AhR modulation

QSARs for the capacity of chemicals to modulate the receptor's effects have been suggested since the 80s. The work by Safe *et al*. (1985) on the structure-function relationship of polychlorinated biphenyls (PCBs), was among the pioneering studies in the field. Using a congeneric set of PCBs, they modeled the induction potencies (EC_{50}) and the binding affinities to AhR. A high correlation was obtained between the physicochemical parameters of the PCBs and their effects as AhR ligands. Thus, steric, electronic, hydrophobic, and hydrogen bonding factors of toxic haloaromatics demonstrated to be important in facilitating the interaction with the active binding state of the receptor [133]. Furthermore, QSAR models have been applied to study AhR binding affinities of polychlorinated dibenzofurans (PCDFs) using electronic descriptors based on the softness, electronegativity, and electrophilicity. Moderate to satisfactory performance was obtained with the proposed linear model [134]. The toxic potential of PCDFs through AhR bound has been further detailed by some others and using methodologies such as Heuristic Method (HM) and Support Vector Machine (SVM) [135,136].

The relative binding affinity (RBA) to the AhR is among the most widely used endpoints, particularly in the study of polybrominated diphenyl ethers (PBDEs). Indeed, different QSAR methods such as comparative molecular field analysis (CoMFA) and comparative similarity indices analysis (CoMSIA), as well as multiple regression and SVM methods, have been employed to model the RBA of PBDEs yielding satisfactory correlations [137–139].

3.2.2. Pharmacophore/Toxicophore Mapping

The prelude to the computer-aided pharmacophoric models were the simple structure-activity relationships, mainly addressed in the medicinal chemistry field [140]. The advances of conformational chemistry in the late 1950s enabled the suggestion of the first pharmacophore

models [141]. Nowadays the modeling of pharmacophores is a well-developed area in computational chemistry [142].

The Pharmacophore/Toxicophore models allow for an intuitive interpretation of the structure-activity relationships. Their main applications have been to identify hit compounds, to assist lead optimization, and to predict side effects, interactions, and metabolism sites [143,144]. Therefore, they constitute particularly important tools in drug discovery and in toxicology prediction.

3.2.2.1. Pharmacophore Conceptualization

The term pharmacophore was defined in 1998 by the International Union of Pure and Applied Chemistry (IUPAC) as an "*ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response*" [145].

The pharmacophore concept has been sometimes misused in medicinal chemistry to refer to functional groups or molecular fragments. As a purely abstract concept, the pharmacophore only describes the molecular interactions (*e.g.* steric, electronic) needed for the optimal interaction between a set of compounds and a defined target structure. Consequently, the pharmacophore can be considered as the largest common denominator shared by a set of active molecules [146]. A more inclusive definition was recently proposed by Osman F. Güner and Phillip Bowen that "*a pharmacophore is the pattern of features of a molecule that is responsible for a biological effect*" [147]. The concept of pharmacophore has been also extrapolated to the toxicological field, describing the ensemble of features that leads to harmful biological responses as "Toxicophore" [148].

The two main steps in the building of pharmacophore hypotheses are, 1) to determine the equivalent pharmacophoric features in a diverse set of active compounds, and, 2) based on these, determine the pharmacophore geometry. Some common pharmacophoric groups include hydrogen (bond) donors, hydrogen acceptors, electron acceptors, electron donors, hydrophobic groups, and polar bonds. Once equivalent pharmacophoric groups have been identified, the pharmacophore geometry could be determined using several approaches (*e.g.* ensemble distance geometry, distance mapping in torsional space, orientation map in distance space, and genetic algorithm-based search) [149].

3.2.2.2. Pharmacophore mapping workflow

The pharmacophore hypotheses could be developed following two different approaches, ligand-based and structure-based modeling. In ligand-based pharmacophore modeling, the hypotheses are estimated by using data set of compounds with known activity (*i.e.* experimentally determined). On the other hand, in the structure-based pharmacophore models, the protein-ligand complex is studied to elucidate the ensemble of relevant features involved in their interaction. A generic workflow for pharmacophore mapping is shown in Figure 13.
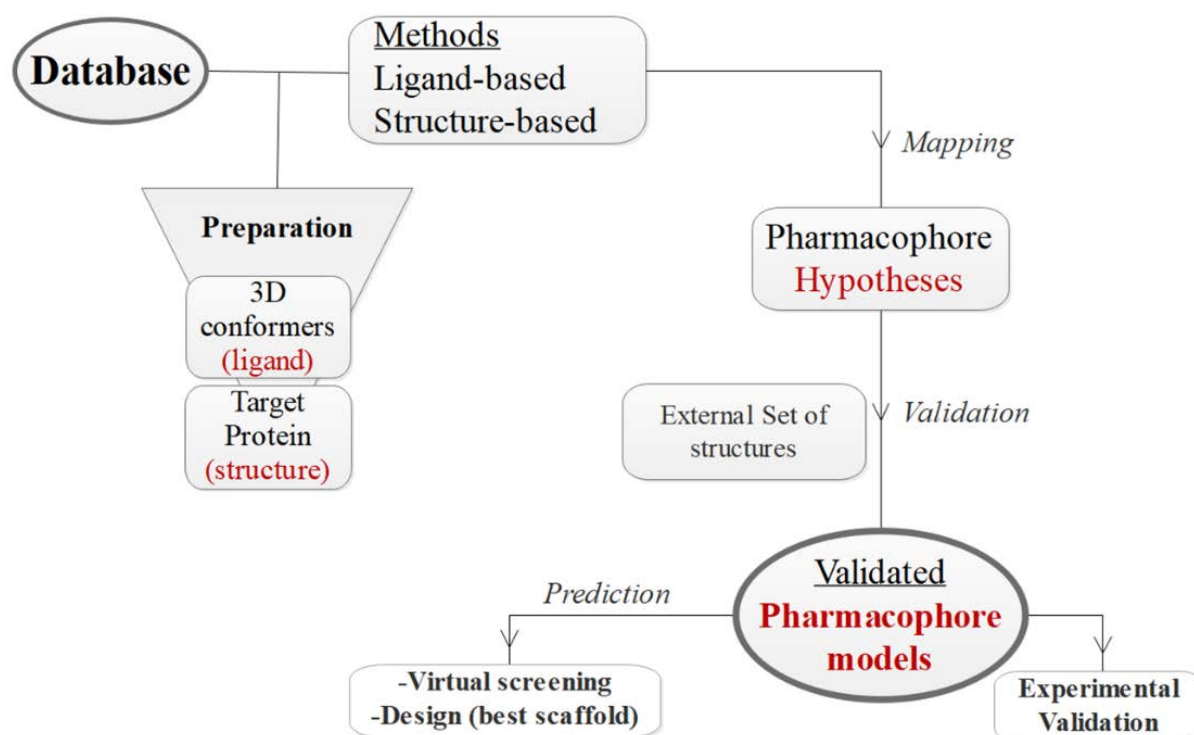
Figure 13. Flowchart of Pharmacophore mapping

As shown in Figure 13, preparation stages are required for the mapping of Pharmacophore hypotheses. Both the ligands and the target protein need to be preprocessed in the case of structure-based pharmacophores, and only the 3D conformers of studied compounds are prepared in ligand-based pharmacophores. In ligand-based pharmacophore mapping, the conformational space of each compound in the training set is explored to build an alignment that determines the common features to be included in the pharmacophore model. The sampling of the small molecules conformational space can be performed using several algorithms (*e.g.* Monte Carlo, systematic torsional grids, genetic algorithms). On the other hand, molecular alignment methods include the conventional least-squares fitting functions that superimpose the features or molecular fragments and some other novel approaches based, on stochastic

proximity embedding or fuzzy pattern recognition [150]. Structure-based pharmacophore methods commonly use "molecular probes" (*e.g.* functional groups, fragments) in order to map the interactions occurring in the receptor-binding site. The discovered "hot spots" are transformed into pharmacophore features to generate the final models. Some examples of tools used for structure-based Pharmacophore mapping are DrugScore, GRID, PHASE, and HS-Pharm [143].

Once all possible pharmacophore hypotheses are generated from congeneric or diverse series of ligands with known experimental activity outcomes, the best hypotheses are selected for further external validation. Thus, it is expected that poor alignment would be observed between the pharmacophoric features and known inactive compounds (as displayed in the example of Figure 14 [151]). Decoy structures are often used as potential non-active molecules [152]. Decoys are structures that share physicochemical properties with active ligands, while topologically distinct from them. Therefore, decoys are unlikely to bind to the target receptor. A widely used benchmarking dataset that provides decoys is DUD•E [153].
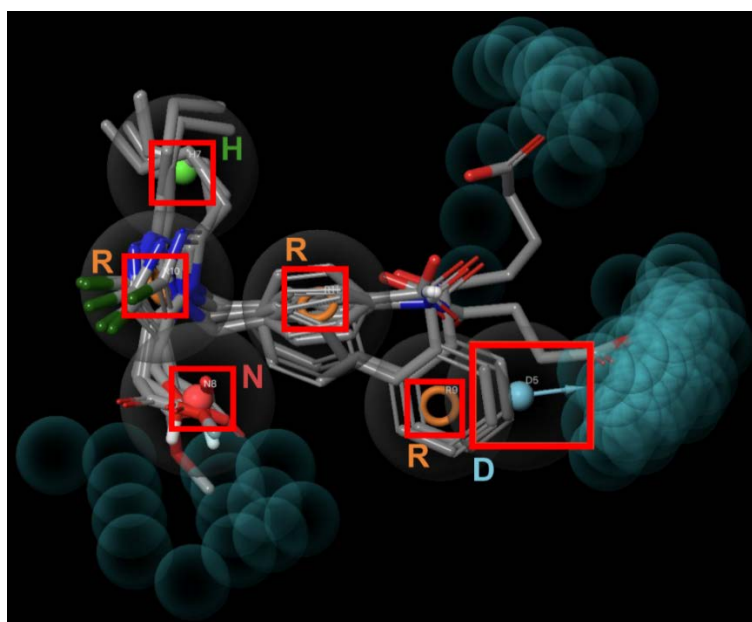
Figure 14. Example of a DHNRRR ligand-based pharmacophore model generated with PHASE. (**D**): Donor, (**H**): Hydrophobic, (**N**): Negative Ionic, (**R**): Aromatic Ring. The inactive ligands are represented with the pharmacophore features, demonstrating a poor alignment (*e.g.* feature **D**) [151].

In order to select the most optimum pharmacophore models for predictive purposes, several enrichment metrics are used (*e.g.* Receiver Operator Characteristic area under the curve (ROC), Robust Initial Enhancement (RIE), Enrichment Factor (EF)) [154]. Meanwhile, the experimental verification of the pharmacophore geometry is recommended as a final step to validate the

suggested hypothesis. Such experimental validation has been conducted, for instance, by analyzing binding affinities of comparable compounds, or else through biophysical determinations (*e.g.* binding conformations) [149].

3.2.2.3. AhR pharmacophore modeling

Structure-based pharmacophore or toxicophore models have not been employed for modeling on AhR-ligand interaction profiles to date due to the lack of a fully crystalized AhR protein (including the ligand-binding domain). Only a few and preliminary ligand-based approaches have been suggested [155]. Moreover, mainly studies on binding capacities and often with limited distinctions between agonist and antagonist ligands are reported in the literature [156].

The pharmacophore image of several dioxin-like compounds (*i.e.* polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs) and co-planar PCBs) was mapped using electrostatic potentials by Koyano *et al*. In that study, the ligand-based pharmacophore was explored prior to CoMFA modeling. One of the main observations was that in PCDD, PCDF and PCB congeners, specific patterns of chlorine substitution enhance the toxic effects of these derivatives [155].

The overexpression of P450 enzymes mediated by AhR is an unwanted effect in the design of compounds as selective inhibitors of P450 1A2. In this sense, a pharmacophore analysis of selective P450 1A2 was suggested by Liu *et al*. through a superimposition over P450 1A1. A lead compound (7,8-pyrano-flavone) and several flavone derivatives were designed as selective P450 1A2 inhibitors based in this pharmacophore mapping [156].

3.2.3. Molecular Docking

Docking methods arose from the renowned 'lock and key' theory. Since the 1980s, molecular docking has been considered a useful and efficient tool particularly important in the drug discovery pipeline [157]. Computational docking for small molecules (ligands) is based on the analysis of the supramolecular complexes formed upon ligands binding to protein targets [158].

In most docking studies the main goal is to screen databases for molecules with favorable interaction profiles with binding sites (targets) of interest in macromolecules. Thereby, the expected outcomes are potential lead compounds based on binding scores and possibly an understanding of the relevant interactions between the ligands and the pocket residues [159].

3.2.3.1. Conceptualization of the Molecular Docking methodology

Molecular docking is a structure-based approach in which possible interactions between a target macromolecule (*i.e.* protein, nucleic acids) and different biochemical entities (*e.g.* small molecules, nucleic acids, proteins), are analyzed [160]. Molecular shapes and physicochemical complementarities are two fundamental elements in docking methods, since they both contribute to the fit of ligands into the binding pocket of the macromolecules (*i.e.* ligand-binding domain of the receptor) [161]. More than a hundred thousand 3D structures of proteins have been resolved to date, in most cases through X-ray diffraction. However, in absence of an experimentally determined 3D structure, *in silico* homology modeling may be employed to construct a model of the target protein based on its sequence and 3D structures of similar (*i.e.* homologous) protein [161].

The key components of the molecular docking process are the search algorithm and the scoring function. The search algorithm explores the conformational space in which the ligand and the target could interact. Hence, the computational cost of the calculations is significantly higher with an increase in the size of the molecules. Some examples of search algorithms are fast shape matching, particle swarm optimization, simulated annealing, and evolutionary algorithms [161,162]. Meanwhile, the scoring functions evaluate the ligand-receptor interactions and establish a rank corresponding to the approximated binding affinity. These scoring functions may be categorized into three main types: a) knowledge-based scoring function (*e.g.* DrugScore, Astex statistical potential, PMF), b) empirical scoring (*e.g.* PLP, CHEMPLP, Glide SP/XP), and c) force field-based scoring functions (*e.g.* DOCK, AutoDOCK, GOLDscore).

3.2.3.2. Molecular Docking Simulations workflow

The workflow of molecular docking-based virtual screening of ligand compounds is briefly summarized in Figure 15.
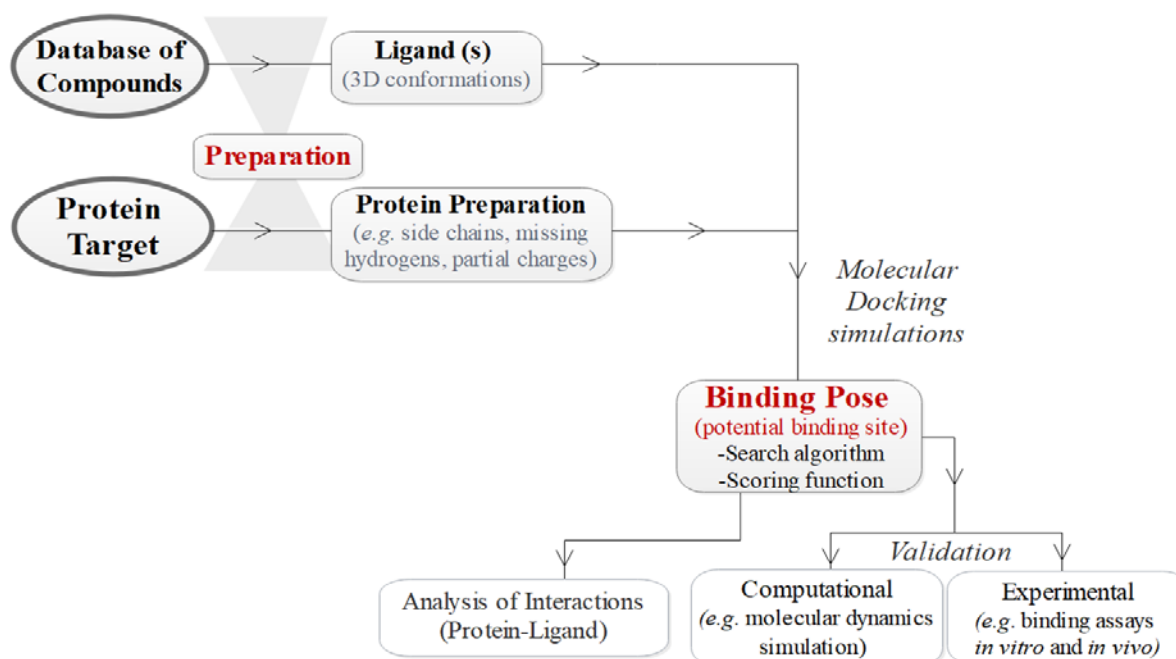
Figure 15. Flowchart of molecular docking analysis

In general, three main stages could be identified in molecular docking. First, is the preparation step of both the protein receptor and potential binders. Second, an exploration and scoring of the different interaction modes is performed to select the most favorable binding poses. Finally, post-docking evaluations are conducted to validate the results and to analyze the predicted interactions in the protein-ligand complex [161].

Docking outcomes should be carefully analyzed since the suggested binding modes might be questionable in the absence of additional validation. Furthermore, in many cases, the binding affinity may be overestimated and important issues such as the selection of a wrong binding site or the use of unsuitable databases of ligand compounds, are frequently identified [163]. The analysis of the potential binding interactions between the pocket residues of the target protein and the ligands is an important application of molecular docking. An example of an interaction diagram between a ligand and binding site residues is shown in Figure 16. Hence, the elucidation of protein-ligand interaction patterns (*e.g.* pi-Staking, H-bond) could provide key leads on strategies for designing bioactive molecules as well as valuable insights on the possible mechanisms of action [164].
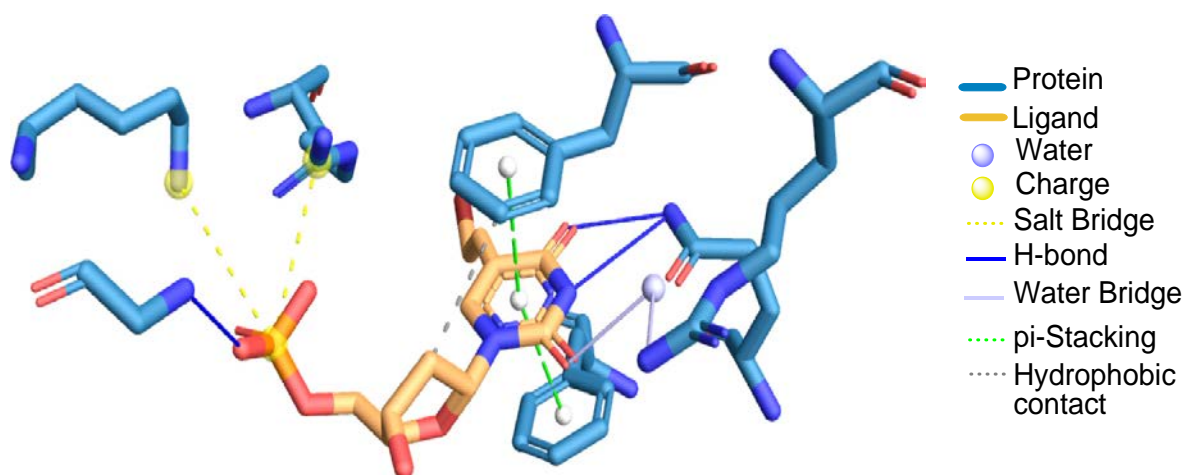
Figure 16. Example of interactions generated with Protein-Ligand Interaction Profiler (PLIP): *Varicella zoster* virus thymidine kinase (PDB: 1OSN) binding the antiherpes drug brivudine-monophosphate. The binding is dominated by a double pi-Stacking and polar interactions at the terminal regions of the ligand [164].

3.2.3.3. AhR Molecular Docking

Docking studies focused on AhR have used different templates of PAS domains of structurally and functionally related proteins that shares a high degree of sequence identity with AhR. Some examples of template used are the hypoxia-inducible factor 2α (HIF-2α), the human ARNT, and the human PAS kinase (hPASK) [165]. Using these and some other templates, several authors have proposed homology models for the AhR LBD. Furthermore, molecular dynamics simulations have corroborated some of the suggested homologous reported in the literature [165,166].

Docking analyses between PCBs, PCDDs, and PCDFs with AhR have concluded that H-bonding and hydrophobic interactions are characteristic interactions that govern the binding affinity of these ligands for the receptor [167]. Some others have suggested differential patterns of interaction between agonists and antagonists; for example, while in antagonist ligands predominated H-bonds, the hydrophobic interactions were identified as the most important ones in the case of agonist ligands [168].

Meanwhile, due to the ligand promiscuity of AhR, the ligand selective modulation caused by different compounds, has been studied through site-directed mutagenesis. In this study, the residues presumably involved in ligand specific AhR activation were suggested. Besides, the

analysis provided insights on the differential agonist/antagonist modes of ligand binding [169]. Recently, specific residues within AhR binding cavity were predicted in a novel homologous model of AhR LBD that includes some new flexibility of the receptor. These authors used a diverse set of AhR ligands, and site-direct mutagenesis as validation [170].

4. Concluding Remarks Chapter 1

Based on a comprehensive literature revision presented in this introductory chapter, the following concluding remarks are noteworthy:

- The AhR is a chemical sensor with cytosolic location in inactive state. The main physiological functions of AhR are the transcription of responsive genes associated with endocrine homeostasis, oxidative balance, circadian rhythms, and diverse metabolic, immunological, and inflammatory processes.
- Several structurally diverse substances from endogenous and exogenous origins are recognized as agonist and/or antagonist modulators of the Ah receptor.
- The promiscuity of AhR to accommodate ligands, make the selectivity and function-specific modulation of the receptor a difficult task.
- In chemical biology, diverse alternative methods to animal experimentation are used to identify AhR ligands.
- Among the most popular *in vitro* bioassays are the measurements of AhR transactivation through cell-based reporter gene methods.
- *In silico* techniques such as QSAR modeling, Pharmacophore mapping, and Molecular Docking Simulations should be further explored in the study of AhR modulation.
- Toxicological and Pharmacological fields could benefit from a better understanding of the structural determinants of ligands causing the activation or inhibition of AhR mediated effects.

5. References

1.   Nebert, D. W. Aryl hydrocarbon receptor (AHR): "pioneer member" of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of "sensors" of foreign and endogenous signals. *Prog. Lipid Res.* **67**, 38–57 (2017).

2.   Murray, I. A., Patterson, A. D. & Perdew, G. H. Aryl hydrocarbon receptor ligands in cancer: Friend and foe. *Nature Reviews Cancer* **14**, 801–814 (2014).

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

4.   Kumar, M. B., Ramadoss, P., Reen, R. K., Vanden Heuvel, J. P. & Perdew, G. H. The Q-rich Subdomain of the Human Ah Receptor Transactivation Domain Is Required for Dioxin-mediated Transcriptional Activity. *J. Biol. Chem.* **276**, 42302–42310 (2001).

5.   Hankinson, O. Role of coactivators in transcriptional activation by the aryl hydrocarbon receptor. *Arch. Biochem. Biophys.* **433**, 379–386 (2005).

6.   Rothhammer, V. & Quintana, F. J. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. *Nat. Rev. Immunol.* **19**, 184–197 (2019).

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

8.   Young, J. C., Moarefi, I. & Ulrich Hartl, F. Hsp90: A specialized but essential protein-folding tool. *J. Cell Biol.* **154**, 267–273 (2001).

9.   Cox, M. B. & Miller C.A., I. I. I. The p23 co-chaperone facilitates dioxin receptor signaling in a yeast model system. *Toxicol. Lett.* **129**, 13–21 (2002).

10.   Shetty, P. V, Bhagwat, B. Y. & Chan, W. K. p23 enhances the formation of the aryl hydrocarbon receptor-DNA complex. *Biochem. Pharmacol.* **65**, 941–948 (2003).

11.   Pappas, B. *et al.* p23 protects the human aryl hydrocarbon receptor from degradation via a heat shock protein 90-independent mechanism. *Biochem. Pharmacol.* **152**, 34–44 (2018).

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

13.   Carver, L. A., Lapres, J. J., Jain, S., Dunham, E. E. & Bradfield, C. A. Characterization of the Ah receptor-associated protein, ARA9. *J. Biol. Chem.* **273**, 33580–33587 (1998).

14.   Meyer, B. K., Petrulis, J. R. & Perdew, G. H. Aryl hydrocarbon (Ah) receptor levels are selectively modulated by hsp90-associated immunophilin homolog XAP2. *Cell Stress Chaperones* **5**, 243–254 (2000).

15.   Larigot, L., Juricek, L., Dairou, J. & Coumoul, X. AhR signaling pathways and regulatory functions. *Biochim. Open* **7**, 1–9 (2018).

16.   Petrulis, J. R., Kusnadi, A., Ramadoss, P., Hollingshead, B. & Perdew, G. H. The hsp90 co-chaperone XAP2 alters importin β recognition of the bipartite nuclear localization signal of the Ah receptor and represses transcriptional activity. *J. Biol. Chem.* **278**, 2677–2685 (2003).

17.   Esser, C. & Rannug, A. The aryl hydrocarbon receptor in barrier organ physiology, immunology, and toxicology. *Pharmacol. Rev.* **67**, 259–279 (2015).

18.   Quintana, F. J. & Sherr, D. H. Aryl hydrocarbon receptor control of adaptive immunity. *Pharmacol. Rev.* **65**, 1148–1161 (2013).

19.   Swanson, H. I., Tullis, K. & Denison, M. S. Binding of transformed Ah receptor complex to a dioxin responsive transcriptional enhancer: Evidence for two distinct heteromeric DNA-binding forms. *Biochemistry* **32**, 12841–12849 (1993).

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

21.   Dolciami, D. *et al.* Targeting Aryl hydrocarbon receptor for next-generation immunotherapies: Selective modulators (SAhRMs) versus rapidly metabolized ligands (RMAhRLs). *Eur. J. Med. Chem.* **185**, (2020).

22.   Hahn, M. E., Allan, L. L. & Sherr, D. H. Regulation of constitutive and inducible AHR signaling: Complex interactions involving the AHR repressor. *Biochem. Pharmacol.* **77**, 485–497 (2009).

23.   Vogel, C. F. A. *et al.* A protective role of aryl hydrocarbon receptor repressor in inflammation and tumor growth. *Cancers (Basel).* **11**, 1–17 (2019).

24.   Mimura, J., Ema, M., Sogawa, K. & Fujii-Kuriyama, Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes Dev.* **13**, 20–25 (1999).

25.   Kress, S., Reichert, J. & Schwarz, M. Functional analysis of the human cytochrome P4501A1 (CYP1A1) gene enhancer. *Eur. J. Biochem.* **258**, 803–812 (1998).

26.   Quattrochi, L. C. & Tukey, R. H. Nuclear uptake of the Ah (dioxin) receptor in response to omeprazole:

Transcriptional activation of the human CYP1A1 gene. *Mol. Pharmacol.* **43**, 504–508 (1993).

27.   Ye, W. *et al.* AhR regulates the expression of human cytochrome P450 1A1 (CYP1A1) by recruiting Sp1. *FEBS J.* **286**, 4215–4231 (2019).

28.   Auyeung, D. J., Kessler, F. K. & Ritter, J. K. Mechanism of rat UDP-glucuronosyltransferase 1A6 induction by oltipraz: Evidence for a contribution of the AryL hydrocarbon receptor pathway. *Mol. Pharmacol.* **63**, 119–127 (2003).

29.   Hankinson, O. The role of AHR-inducible cytochrome P450s in metabolism of polyunsaturated fatty acids. *Drug Metab. Rev.* **48**, 342–350 (2016).

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

31.   Wilson, S. R., Joshi, A. D. & Elferink, C. J. The tumor suppressor kruppel-like factor 6 is a novel aryl hydrocarbon receptor DNA binding partner. *J. Pharmacol. Exp. Ther.* **345**, 419–429 (2013).

32.   Biscardi, J. S. *et al.* C-Src-mediated phosphorylation of the epidermal growth factor receptor on Tyr845 and Tyr1101 is associated with modulation of receptor function. *J. Biol. Chem.* **274**, 8335–8343 (1999).

33.   Haarmann-Stemmann, T., Bothe, H. & Abel, J. Growth factors, cytokines and their receptors as downstream targets of arylhydrocarbon receptor (AhR) signaling pathways. *Biochem. Pharmacol.* **77**, 508–520 (2009).

34.   Yu, A. R. *et al.* Alpha-naphthoflavone induces apoptosis through endoplasmic reticulum stress via c-Src-, ROS-, MAPKs-, and arylhydrocarbon receptor-dependent pathways in HT22 hippocampal neuronal cells. *Neurotoxicology* **71**, 39–51 (2019).

35.   Domínguez-Acosta, O. *et al.* Activation of aryl hydrocarbon receptor regulates the LPS/IFNγ-induced inflammatory response by inducing ubiquitin-proteosomal and lysosomal degradation of RelA/p65. *Biochem. Pharmacol.* **155**, 141–149 (2018).

36.   Mejía-García, A. *et al.* Activation of AHR mediates the ubiquitination and proteasome degradation of c-Fos through the induction of Ubcm4 gene expression. *Toxicology* **337**, 47–57 (2015).

37.   Ohtake, F., Fujii-Kuriyama, Y. & Kato, S. AhR acts as an E3 ubiquitin ligase to modulate steroid receptor functions. *Biochem. Pharmacol.* **77**, 474–484 (2009).

38.   Ohtake, F. *et al.* Dioxin receptor is a ligand-dependent E3 ubiquitin ligase. *Nature* **446**, 562–566 (2007).

39.   Matsumura, F. The significance of the nongenomic pathway in mediating inflammatory signaling of the dioxin-activated Ah receptor to cause toxic effects. *Biochem. Pharmacol.* **77**, 608–626 (2009).

40.   Roman, Á. C., Carvajal-Gonzalez, J. M., Merino, J. M., Mulero-Navarro, S. & Fernández-Salguero, P. M. The aryl hydrocarbon receptor in the crossroad of signalling networks with therapeutic value. *Pharmacol. Ther.* **185**, 50–63 (2018).

41.   Ohtake, F. *et al.* Modulation of oestrogen receptor signalling by association with the activated dioxin receptor. *Nature* **423**, 545–550 (2003).

42.   Mathew, L. K., Sengupta, S. S., LaDu, J., Andreasen, E. A. & Tanguay, R. L. Crosstalk between AHR and Wnt signaling through R-Spondin1 impairs tissue regeneration in zebrafish. *FASEB J.* **22**, 3087–3096 (2008).

43.   Chassot, A. A. *et al.* WNT4 and RSPO1 together are required for cell proliferation in the early mouse gonad. *Development* **139**, 4461–4472 (2012).

44.   Tomaselli, S. *et al.* Human RSPO1/R-spondin1 is expressed during early ovary development and augments β-catenin signaling. *PLoS One* **6**, (2011).

45.   Komiya, Y. & Habas, R. Wnt signal transduction pathways. *Organogenesis* **4**, 68–75 (2008).

46.   Neuzillet, C. *et al.* Targeting the TGFβ pathway for cancer therapy. *Pharmacol. Ther.* **147**, 22–31 (2015).

47.   Pollenz, R. S. The mechanism of AH receptor protein down-regulation (degradation) and its impact on AH receptor-mediated gene regulation. *Chem. Biol. Interact.* **141**, 41–61 (2002).

48.   Pollenz, R. S. & Barbour, E. R. Analysis of the complex relationship between nuclear export and aryl hydrocarbon receptor-mediated gene regulation. *Mol. Cell. Biol.* **20**, 6095–6104 (2000).

49.   Lamas, B., Natividad, J. M. & Sokol, H. Aryl hydrocarbon receptor and intestinal immunity review-article. *Mucosal Immunol.* **11**, 1024–1038 (2018).

50.   Gao, J. *et al.* Impact of the Gut Microbiota on Intestinal Immunity Mediated by Tryptophan Metabolism. *Front. Cell. Infect. Microbiol.* **8**, 1–22 (2018).

51.   Esser, C. *et al.* Old receptor, new tricks—The ever-expanding universe of aryl hydrocarbon receptor functions. Report from the 4th AHR meeting, 29–31 August 2018 in Paris, France. *Int. J. Mol. Sci.* **19**, (2018).

52.   Mitchell, K. A. & Elferink, C. J. Timing is everything: Consequences of transient and sustained AhR activity. *Biochem. Pharmacol.* **77**, 947–956 (2009).

53.   Okey, A. B. An aryl hydrocarbon receptor odyssey to the shores of toxicology: The Deichmann Lecture,

International Congress of Toxicology-XI. *Toxicol. Sci.* **98**, 5–38 (2007).

54. Bock, K. W. From TCDD-mediated toxicity to searches of physiologic AHR functions. *Biochem. Pharmacol.* **155**, 419–424 (2018).

55. Mulero-Navarro, S. & Fernandez-Salguero, P. M. New trends in Aryl hydrocarbon receptor biology. *Front. Cell Dev. Biol.* **4**, 1–14 (2016).

56. Mescher, M. & Haarmann-Stemmann, T. Modulation of CYP1A1 metabolism: From adverse health effects to chemoprevention and therapeutic options. *Pharmacol. Ther.* **187**, 71–87 (2018).

57. Guerrina, N., Traboulsi, H., Eidelman, D. H. & Baglole, C. J. The aryl hydrocarbon receptor and the maintenance of lung health. *Int. J. Mol. Sci.* **19**, (2018).

58. Bock, K. W. Human AHR functions in vascular tissue: Pro- and anti-inflammatory responses of AHR agonists in atherosclerosis. *Biochem. Pharmacol.* **159**, 116–120 (2019).

59. Duval, C., Blanc, E. & Coumoul, X. Aryl hydrocarbon receptor and liver fibrosis. *Curr. Opin. Toxicol.* **8**, 8–13 (2018).

60. Puccetti, M. *et al.* Towards targeting the aryl hydrocarbon receptor in cystic fibrosis. *Mediators Inflamm.* **2018**, (2018).

61. Dietrich, C. Antioxidant Functions of the Aryl Hydrocarbon Receptor. *Stem Cells Int.* **2016**, (2016).

62. Juricek, L. & Coumoul, X. The Aryl Hydrocarbon Receptor and the Nervous System. *Int. J. Mol. Sci.* **19**, 2504 (2018).

63. Bradshaw, T. & Westwell, A. The Development of the Antitumour Benzothiazole Prodrug, Phortress, as a Clinical Candidate. *Curr. Med. Chem.* **11**, 1009–1021 (2005).

64. Safe, S., Cheng, Y. & Jin, U. H. The aryl hydrocarbon receptor (AhR) as a drug target for cancer chemotherapy. *Curr. Opin. Toxicol.* **1**, 24–29 (2017).

65. Esser, C., Rannug, A. & Stockinger, B. The aryl hydrocarbon receptor in immunity. *Trends Immunol.* **30**, 447–454 (2009).

66. Marafini, I. *et al.* NPD-0414-2 and NPD-0414-24, two chemical entities designed as aryl hydrocarbon receptor (AHR) ligands, inhibit gut inflammatory signals. *Front. Pharmacol.* **10**, 1–9 (2019).

67. McMillan, B. J. & Bradfield, C. A. The aryl hydrocarbon receptor sans xenobiotics: Endogenous function in genetic model systems. *Mol. Pharmacol.* **72**, 487–498 (2007).

68. Yang, T. *et al.* Dietary natural flavonoids treating cancer by targeting aryl hydrocarbon receptor. *Crit. Rev. Toxicol.* **49**, 445–460 (2019).

69. Goya-Jorge, E. *et al.* Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective. *SAR QSAR Environ. Res.* **31**, 209–226 (2020).

70. Doan, T. Q. *et al.* A mixture of persistent organic pollutants relevant for human exposure inhibits the transactivation activity of the aryl hydrocarbon receptor in vitro. *Environ. Pollut.* **254**, (2019).

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

72. Stejskalova, L., Dvorak, Z. & Pavek, P. Endogenous and exogenous ligands of aryl hydrocarbon receptor: current state of art. *Curr. Drug Metab.* **12**, 198–212 (2011).

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

74. Seok, S.-H. *et al.* Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex. *Proc. Natl. Acad. Sci.* **114**, 5431–5436 (2017).

75. Zhou, H. *et al.* Toxicology mechanism of the persistent organic pollutants (POPs) in fish through AhR pathway. *Toxicol. Mech. Methods* **20**, 279–286 (2010).

76. Guyot, E., Chevallier, A., Barouki, R. & Coumoul, X. The AhR twist: Ligand-dependent AhR signaling and pharmaco-toxicological implications. *Drug Discov. Today* **18**, 479–486 (2013).

77. Neavin, D. R., Liu, D., Ray, B. & Weinshilboum, R. M. The role of the aryl hydrocarbon receptor (AHR) in immune and inflammatory diseases. *Int. J. Mol. Sci.* **19**, (2018).

78. Duran-Frigola, M., Rossell, D. & Aloy, P. A chemo-centric view of human health and disease. *Nat. Commun.* **5**, 1–11 (2014).

79. Hajjo, R., Setola, V., Roth, B. L. & Tropsha, A. Chemocentric informatics approach to drug discovery: Identification and experimental validation of selective estrogen receptor modulators as ligands of 5-hydroxytryptamine-6 receptors and as potential cognition enhancers. *J. Med. Chem.* **55**, 5704–5719 (2012).

80. Tropsha, A. Best practices for developing predictive QSAR models. (2010). Available at: http://infochim.u-strasbg.fr/CS3_2010/OralPDF/Tropsha_CS3_2010.

81. Goya-Jorge, E., Abdmouleh, F., Carpio, L. E., Giner, R. M. & Sylla-Iyarreta Veitía, M. Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic

activities. *Eur. J. Pharm. Sci.* **151**, 105386 (2020).

82. Doke, S. K. & Dhawale, S. C. Alternatives to animal testing: A review. *Saudi Pharm. J.* **23**, 223–229 (2015).

83. Illing, P. *et al.* *Alternatives To Animal Testing*. (The Royal Society of Chemistry, 2006). doi:10.1039/9781847552457

84. Cronin, M. *Non-Animal Approaches - The way forward. (Report on a European Commission Scientific Conference)*. (2017). doi:10.2779/373944

85. Osakwe, O. Preclinical In Vitro Studies: Development and Applicability. in *Social Aspects of Drug Discovery, Development and Commercialization* (eds. Osakwe, O. & Rizvi Development and Commercialization, S. A. A. B. T.-S. A. of D. D.) 129–148 (Academic Press, 2016). doi:10.1016/B978-0-12-802220-7.00006-5

86. Stratton, C. W. In vitro susceptibility testing versus in vivo effectiveness. *Med. Clin. North Am.* **90**, 1077–1088 (2006).

87. Kandárová, H. & Letaáiová, S. Alternative methods in toxicology: Pre-validated and validated methods. *Interdiscip. Toxicol.* **4**, 107–113 (2011).

88. Willems, H., De Cesco, S. & Svensson, F. Computational Chemistry on a Budget – Supporting Drug Discovery with Limited Resources. *J. Med. Chem.* (2020). doi:10.1021/acs.jmedchem.9b02126

89. Johnson, C. I., Argyle, D. J. & Clements, D. N. In vitro models for the study of osteoarthritis. *Vet. J.* **209**, 40–49 (2016).

90. Eskes, C. *et al.* Good cell culture practices &in vitro toxicology. *Toxicol. Vitr.* **45**, 272–277 (2017).

91. Tourneix, F. *et al.* Skin sensitisation testing in practice: Applying a stacking meta model to cosmetic ingredients. *Toxicol. Vitr.* **66**, 104831 (2020).

92. de Ávila, R. I., Lindstedt, M. & Valadares, M. C. The 21st Century movement within the area of skin sensitization assessment: From the animal context towards current human-relevant in vitro solutions. *Regul. Toxicol. Pharmacol.* **108**, 104445 (2019).

93. Balouiri, M., Sadiki, M. & Ibnsouda, S. K. Methods for in vitro evaluating antimicrobial activity: A review. *J. Pharm. Anal.* **6**, 71–79 (2016).

94. Rehberger, K., Kropf, C. & Segner, H. In vitro or not in vitro: a short journey through a long history. *Environ. Sci. Eur.* **30**, (2018).

95. Tan, J. B. L. & Lim, Y. Y. Critical analysis of current methods for assessing the in vitro antioxidant and antibacterial activity of plant extracts. *Food Chem.* **172**, 814–822 (2015).

96. Pandey, P., Hasnain, S. & Ahmad, S. Protein-DNA Interactions. in (eds. Ranganathan, S., Gribskov, M., Nakai, K. & Schönbach, C. B. T.-E. of B. and C. B.) 142–154 (Academic Press, 2019). doi:https://doi.org/10.1016/B978-0-12-809633-8.20217-3

97. Rauws, A. G., De Waal, E. J. & Van Der Laan, J. W. Sense and non-sense in toxicity assessment of medicinal products. in *Advances in Drug Research* (eds. Testa, B. & Meyer, U. A.) **30**, 15-7234 1 (Academic Press, 1997).

98. Dyer, S. D. *et al.* In vitro biotransformation of surfactants in fish. Part II - Alcohol ethoxylate (C16EO8) and alcohol ethoxylate sulfate (C14EO2S) to estimate bioconcentration potential. *Chemosphere* **76**, 989–998 (2009).

99. Michelini, E., Cevenini, L., Mezzanotte, L., Coppa, A. & Roda, A. Cell-based assays: fuelling drug discovery. *Anal. Bioanal. Chem.* **398**, 227–238 (2010).

100. Shao, Y. & Fu, J. Synthetic human embryology: towards a quantitative future. *Curr. Opin. Genet. Dev.* **63**, 30–35 (2020).

101. Kovarich, S., Papa, E. & Gramatica, P. QSAR classification models for the prediction of endocrine disrupting activity of brominated flame retardants. *J. Hazard. Mater.* **190**, 106–112 (2011).

102. Connolly, L., Ropstad, E. & Verhaegen, S. In vitro bioassays for the study of endocrine-disrupting food additives and contaminants. *TrAC - Trends Anal. Chem.* **30**, 227–238 (2011).

103. InvivoGen. *Introducing Lucia: a new secreted luciferase*. (2012).

104. Bøtter-Jensen, L. Luminescence techniques: Instrumentation and methods. *Radiat. Meas.* **27**, 749–768 (1997).

105. Marques, S. M. & Esteves Da Silva, J. C. G. Firefly bioluminescence: A mechanistic approach of luciferase catalyzed reactions. *IUBMB Life* **61**, 6–17 (2009).

106. Markova, S. V., Larionova, M. D. & Vysotski, E. S. Shining Light on the Secreted Luciferases of Marine Copepods: Current Knowledge and Applications. *Photochem. Photobiol.* **95**, 705–721 (2019).

107. Sciuto, S. *et al.* Dioxin-like compounds in lake fish species: Evaluation by DR-CALUX bioassay. *J. Food Prot.* **81**, 842–847 (2018).

108. Aarts, J. M. M. J. G. *et al.* Species-specific antagonism of Ah receptor action by 2,2',5,5'-tetrachloro-and 2,2',3,3',4,4'-hexachlorobiphenyl. *Eur. J. Pharmacol.* **293**, 463–474 (1995).

109.   Takenaka, Y. *et al.* Evolution of bioluminescence in marine planktonic copepods. *Mol. Biol. Evol.* **29**, 1669–1681 (2012).

110.   InvivoGen. *HepG2-Lucia^{TM} AhR cells. AhR hepatoma reporter cells.* (2019).

111.   Dykstra, C. E., Frenking, G., Kim, K. S. & Scuseria, G. E. Computing technologies, theories, and algorithms. The making of 40 years and more of theoretical and computational chemistry. in *Theory and Applications of Computational Chemistry* (eds. Dykstra, C. E., Frenking, G., Kim, K. S. & Scuseria, G. E. B. T.-T. and A. of C. C.) 1–7 (Elsevier, 2005). doi:10.1016/B978-044451719-7/50044-5

112.   Roncaglioni, A., Toropov, A. A., Toropova, A. P. & Benfenati, E. In silico methods to predict drug toxicity. *Curr. Opin. Pharmacol.* **13**, 802–806 (2013).

113.   Taha, M. O. Mixing Pharmacophore Modeling and Classical QSAR Analysis as Powerful Tool for Lead Discovery. in *Virtual Screening* 3–17 (InTech, 2012). doi:10.5772/20993

114.   Liu, R., Li, X. & Lam, K. S. Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **38**, 117–126 (2017).

115.   Todorov, N. P., Alberts, I. L. & Dean, P. M. 4.13 - De Novo Design. in *Comprehensive Medicinal Chemistry II* (eds. Taylor, J. B. & Triggle, D. J.) 283–305 (Elsevier, 2007). doi:https://doi.org/10.1016/B0-08-045044-X/00255-8

116.   Todeschini, R. & Consonni, V. *Molecular Descriptors for Chemoinformatics*. (WILEY-VCH, 2009).

117.   Roy, K. *et al.* Chemical Information and Descriptors. in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* 47–80 (Elsevier, 2015). doi:10.1016/B978-0-12-801505-6.00002-8

118.   Brown, A. C. & Fraser, T. R. On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. *J. Anat. Physiol.* **2**, 224–242 (1868).

119.   Free, S. M. & Wilson, J. W. A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.* **7**, 395–399 (1964).

120.   Muir, R. M. & Hansch, C. Chemical Constitution as Related to Growth Regulator Action. *Annu. Rev. Plant Physiol.* **6**, 157–176 (1955).

121.   Hansch, C. & Fujita, T. p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **86**, 1616–1626 (1964).

122.   Todeschini, R. & Consonni, V. *Handbook of Molecular Descriptors*. (WILEY-VCH, 2000). doi:10.1002/9783527613106

123.   Sahoo, S., Adhikari, C., Kuanar, M. & Mishra, B. K. A Short Review of the Generation of Molecular Descriptors and Their Applications in Quantitative Structure Property/Activity Relationships. *Curr. Comput. Aided. Drug Des.* **12**, 181–205 (2016).

124.   Damale, M., Harke, S., Kalam Khan, F., Shinde, D. & Sangshetti, J. Recent Advances in Multidimensional QSAR (4D-6D): A Critical Review. *Mini-Reviews Med. Chem.* **14**, 35–55 (2014).

125.   Esposito, E. X., Hopfinger, A. J. & Madura, J. D. Methods for Applying the Quantitative Structure–Activity Relationship Paradigm. in *Chemoinformatics. Methods in Molecular Biology^{TM}* (ed. Bajorath, J.) 131–213 (2004). doi:10.1385/1-59259-802-1:131

126.   Tropsha, A. Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inform.* **29**, 476–488 (2010).

127.   Gonzalez, M., Teran, C., Saiz-Urra, L. & Teijeira, M. Variable Selection Methods in QSAR: An Overview. *Curr. Top. Med. Chem.* **8**, 1606–1627 (2008).

128.   Golbraikh, A. *et al.* Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided. Mol. Des.* **17**, 241–253 (2003).

129.   Golbraikh, A. & Tropsha, A. Beware of q2 ! *J. Mol. Graph. Model.* **20**, 269–276 (2002).

130.   Tropsha, A. & Golbraikh, A. Predictive QSAR Modeling Workflow, Model Applicability Domains, and Virtual Screening. *Curr. Pharm. Des.* **13**, 3494–3504 (2007).

131.   OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*. 154 (OECD Publishing, 2007). doi:10.1787/9789264085442-en

132.   Estrada, E. & Patlewicz, G. On the Usefulness of Graph-theoretic Descriptors in Predicting Theoretical Parameters. Phototoxicity of Polycyclic Aromatic Hydrocarbons (PAHs). *Croat. Chem. Acta* **77**, 203–211 (2004).

133.   Safe, S. *et al.* PCBs: Structure-function relationships and mechanism of action. *Environ. Health Perspect.* **60**, 47–56 (1985).

134.   Arulmozhiraja, S. & Morita, M. Structure-Activity Relationships for the Toxicity of Polychlorinated Dibenzofurans: Approach through Density Functional Theory-Based Descriptors. *Chem. Res. Toxicol.* **17**, 348–356 (2004).

135.   Clare, B. W. QSAR of aromatic substances: Toxicity of polychlorodibenzofurans. *J. Mol. Struct.*

THEOCHEM **763**, 205–213 (2006).

136. Luan, F. *et al.* QSAR study of polychlorinated dibenzodioxins, dibenzofurans, and biphenyls using the heuristic method and support vector machine. *QSAR Comb. Sci.* **25**, 46–55 (2006).

137. Gu, C. *et al.* Improved 3D-QSAR analyzes for the predictive toxicology of polybrominated diphenyl ethers with CoMFA/CoMSIA and DFT. *Ecotoxicol. Environ. Saf.* **73**, 1470–1479 (2010).

138. Wang, Y. *et al.* Quantitative structure-activity relationship for prediction of the toxicity of polybrominated diphenyl ether (PBDE) congeners. *Chemosphere* **64**, 515–524 (2006).

139. Zheng, G., Xiao, M. & Lu, X. Quantitative structure - Activity relationships study on the Ah receptor binding affinities of polybrominated diphenyl ethers using a support vector machine. *QSAR Comb. Sci.* **26**, 536–541 (2007).

140. Woods, D. D. The Relation of p-aminobenzoic Acid to the Mechanism of the Action of Sulphanilamide. *Br. J. Exp. Pathol.* **21**, 74–90 (1940).

141. Beckett, A. H. Stereochemical factors in biological activity. *Fortschr. Arzneimittelforsch.* **1**, 455–530 (1959).

142. Güner, O. History and Evolution of the Pharmacophore Concept in Computer-Aided Drug Design. *Curr. Top. Med. Chem.* **2**, 1321–1332 (2005).

143. Caporuscio, F. & Tafi, A. Pharmacophore Modelling: A Forty Year Old Approach and its Modern Synergies. *Curr. Med. Chem.* **18**, 2543–2553 (2011).

144. Schuster, D. Pharmacophore Models for Toxicology Prediction. in *Computational Toxicology* (ed. Sean, E.) 121–144 (John Wiley & Sons, Ltd, 2018). doi:10.1002/9781119282594.ch5

145. Wermuth, C. G., Ganellin, C. R., Lindberg, P. & Mitscher, L. a. Glossary for Chemists of Terms Used in Medicinal Chemistry. *Pure Appl. Chem.* **70**, 1129–1143 (1998).

146. Wermuth, C. G. Pharmacophores: Historical Perspective and Viewpoint from a Medicinal Chemist. in *Pharmacophores and Pharmacophore Searches* (eds. Langer, T., Hoffmann, R. D., Mannhold, R., Kubinyi, H. & Folkers, G.) 3–13 (WILEY-VCH, 2006).

147. Güner, O. F. & Bowen, J. P. Setting the record straight: The origin of the pharmacophore concept. *J. Chem. Inf. Model.* **54**, 1269–1283 (2014).

148. Williams, D. P. Toxicophores: Investigations in drug safety. *Toxicology* **226**, 1–11 (2006).

149. Ghose, A. K. & Wendoloski, J. J. Pharmacophore Modelling: Methods, Experimental Verification and Applications. in *3D QSAR in Drug Design. Volume 2. Ligand-Protein Interactions and Molecular Similarity* (eds. Kubinyi, H., Folkers, G. & Martin, Y. C.) 253–271 (Kluwer Academic Publishers, 1998).

150. Sheng-Yong, Y. Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discov. Today* **15**, 444–450 (2010).

151. Schrödinger. Tutorial: Ligand-Based Virtual Screening Using Phase. 1–20 (2019).

152. Réau, M., Langenfeld, F., Zagury, J. F., Lagarde, N. & Montes, M. Decoys selection in benchmarking datasets: Overview and perspectives. *Front. Pharmacol.* **9**, (2018).

153. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **55**, 6582–6594 (2012).

154. Truchon, J. F. & Bayly, C. I. Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **47**, 488–508 (2007).

155. Koyano, K., Nakano, T. & Kaminuma, T. Electrostatic potentials and CoMFA analysis of toxicity of dioxins. *Chem-Bio Informatics J.* **1**, 60–72 (2001).

156. Liu, J. *et al.* A Ligand-Based Drug Design. Discovery of 4-Trifluoromethyl-7,8-pyranocoumarin as a Selective Inhibitor of Human Cytochrome P450 1A2. *J. Med. Chem.* **58**, 6481–6493 (2015).

157. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161**, 269–288 (1982).

158. Roy, K., Kar, S. & Das, R. N. Chapter 10 - Other Related Techniques. in *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment* (eds. Roy, K., Kar, S. & Das, R. N.) 357–425 (Academic Press, 2015). doi:https://doi.org/10.1016/B978-0-12-801505-6.00010-7

159. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **7**, 146–157 (2012).

160. Meza Menchaca, T., Juárez-Portilla, C. & Zepeda, R. C. Past, Present, and Future of Molecular Docking. in *Drug Discovery and Development - New Advances* (InTech, 2020). doi:10.5772/intechopen.90921

161. Khanna, V., Ranganathan, S. & Petrovsky, N. Rational Structure-Based Drug Design. in *Encyclopedia of Bioinformatics and Computational Biology* (eds. Ranganathan, S., Gribskov, M., Nakai, K. & Schönbach, C.) 585–600 (Academic Press, 2019). doi:https://doi.org/10.1016/B978-0-12-809633-8.20275-6

162. Heberle, G. & de Azevedo, W. F. Bio-Inspired Algorithms Applied to Molecular Docking Simulations. *Current Medicinal Chemistry* **18**, 1339–1352 (2011).

163. Chen, Y. C. Beware of docking! *Trends Pharmacol. Sci.* **36**, 78–95 (2015).

164. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: Fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–W447 (2015).
165. Pandini, A., Denison, M. S., Song, Y., Soshilov, A. A. & Bonati, L. Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. *Biochemistry* **46**, 696–708 (2007).
166. Cao, F. *et al.* Molecular docking, molecular dynamics simulation, and structure-based 3D-QSAR studies on the aryl hydrocarbon receptor agonistic activity of hydroxylated polychlorinated biphenyls. *Environ. Toxicol. Pharmacol.* **36**, 626–635 (2013).
167. Li, F. *et al.* Docking and 3D-QSAR studies on the Ah receptor binding affinities of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). *Environ. Toxicol. Pharmacol.* **32**, 478–485 (2011).
168. Gadhwal, M. K., Patil, S., D'mello, P. & Joshi, A. Homology modeling of aryl hydrocarbon receptor and docking of agonists and antagonists. *Int. J. Pharm. Pharm. Sci.* **5**, 76–81 (2013).
169. Soshilov, A. A. & Denison, M. S. Ligand Promiscuity of Aryl Hydrocarbon Receptor Agonists and Antagonists Revealed by Site-Directed Mutagenesis. *Mol. Cell. Biol.* **34**, 1707–1719 (2014).
170. Tagliabue, S. G., Faber, S. C., Motta, S., Denison, M. S. & Bonati, L. Modeling the binding of diverse ligands within the Ah receptor ligand binding domain. *Sci. Rep.* **9**, 1–14 (2019).

Chapter 2

CHAPTER 2. Predictive modeling of AhR agonism

- **Publication Title:** "Predictive modeling of aryl hydrocarbon receptor (AhR) agonism". (https://doi.org/10.1016/j.chemosphere.2020.127068).
- **Authors**: Elizabeth Goya-Jorge, Rosa M. Giner, Maité Sylla-Iyarreta Veitía, Rafael Gozalbes, Stephen J. Barigye.

Summary

This second Chapter presents results of the combination of *in silico* and *in vitro* approaches to study the activation of AhR mediated effects. Classification QSAR models were built and validated, demonstrating adequate robustness and predictivity. A majority vote-based ensemble of the built QSAR models is proposed to predict the AhR agonism. The prospective *in vitro* experimental validation of the models included the evaluation of 40 compounds from natural and synthetic origins. The general correspondence between the ensemble predictions and the *in vitro* results suggests that the constructed ensemble may be useful in predicting the AhR agonistic activity.

Highlights

- Classification models for AhR agonism were built using a large and balanced dataset.
- A 10-fold external validation demonstrated the robustness and predictiveness of built QSARs.
- *In vitro* experimental validation corroborated the predictivity of the built ensemble classifier.
- The benzothiazoles were the most prominent group among the AhR agonists.
- The built QSARs may be useful in guiding the screening of AhR agonists.
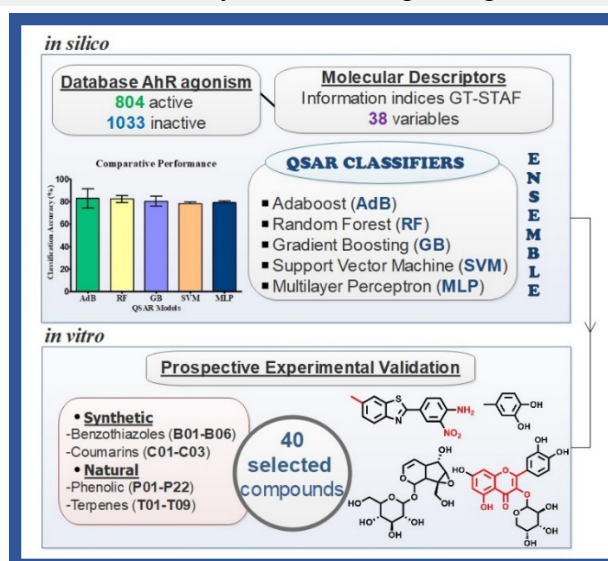
Predictive modeling of aryl hydrocarbon receptor (AhR) agonism

Elizabeth Goya-Jorge [a, b], Rosa M. Giner [b], Maité Sylla-Iyarreta Veitía [c], Rafael Gozalbes [a], Stephen J. Barigye [a, *]

[a] ProtoQSAR SL. CEEI (Centro Europeo de Empresas Innovadoras) Parque Tecnológico de Valencia, Av. Benjamin Franklin 12, 46980, Paterna, Valencia, Spain
[b] Departament de Farmacologia, Facultat de Farmàcia, Universitat de València, Av. Vicente Andrés Estellés s/n, 46100, Burjassot, Valencia, Spain
[c] Equipe de Chimie Moléculaire du Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (EA 7528), Conservatoire National des Arts et Métiers (Cnam), 2 Rue Conté, HESAM Université, 75003, Paris, France

HIGHLIGHTS

- Classification models for AhR agonism were built using a large and balanced dataset.
- A 10-fold external validation demonstrated the robustness and predictiveness of built QSARs.
- *In vitro* experimental validation corroborated the predictivity of the built ensemble classifier.
- The benzothiazoles were the most prominent group among the AhR agonists.
- The built QSARs may be useful in guiding the screening of AhR agonists.

GRAPHICAL ABSTRACT
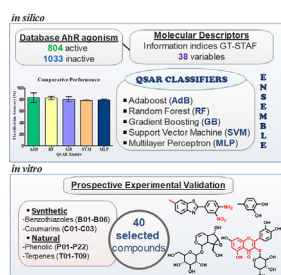
ARTICLE INFO

ABSTRACT

The aryl hydrocarbon receptor (AhR) plays a key role in the regulation of gene expression in metabolic machinery and detoxification systems. In the recent years, this receptor has attracted interest as a therapeutic target for immunological, oncogenic and inflammatory conditions. In the present report, *in silico* and *in vitro* approaches were combined to study the activation of the AhR. To this end, a large database of chemical compounds with known AhR agonistic activity was employed to build 5 classifiers based on the Adaboost (AdB), Gradient Boosting (GB), Random Forest (RF), Multilayer Perceptron (MLP) and Support Vector Machine (SVM) algorithms, respectively. The built classifiers were examined, following a 10-fold external validation procedure, demonstrating adequate robustness and predictivity. These models were integrated into a majority vote based ensemble, subsequently used to screen an in-house library of compounds from which 40 compounds were selected for prospective *in vitro* experimental validation. The general correspondence between the ensemble predictions and the *in vitro* results suggests that the constructed ensemble may be useful in predicting the AhR agonistic activity, both in a toxicological and pharmacological context. A preliminary structure-activity analysis of the evaluated compounds revealed that all structures bearing a benzothiazole moiety induced AhR expression while diverse activity profiles were exhibited by phenolic derivatives.

© 2020 Elsevier Ltd. All rights reserved.

---

1. Introduction

The aryl hydrocarbon receptor (AhR) is a basic Helix-Loop-Helix member of the Per-Arnt-Sim (bHLH/PAS) family of receptors known to act as a cytoplasmatic sensor and as a promiscuous transcription factor with crucial roles in the regulation of gene expression (Mescher and Haarmann-Stemmann, 2018; Nebert, 2017). Studies on AhR-mediated effects in normal physiological development and physiopathological conditions have improved the understanding of cellular responses to diverse chemical agents (Fujii-Kuriyama and Mimura, 2007; Kawajiri and Fuji-Kuriyama, 2016). The highest levels of AhR transcriptional activity are shown in barrier tissues and the liver, where the receptor has been related to metabolic machinery and detoxification systems. AhR activators include a diverse array of environmental toxicants, as well as endogenous ligands mainly derived from L-tryptophan metabolism (Bock, 2018; Esser et al., 2018).

AhR agonism is a long-known toxicological mechanism of dioxins and related compounds, whose exposure and human health risk have been extensively assessed (Gies et al., 2007). Reproductive and immunological disruption, as well as carcinogenic effects, are attributed to AhR transcriptional activity (Rothhammer and Quintana, 2019). However, the complexity of AhR pathways has led to consider that many of its adverse health consequences could be exacerbated cellular responses of its endogenous roles (Mitchell and Elferink, 2009; Roman et al., 2018).

The AhR is an emerging pharmacological target in intestinal immunity and inflammation (Lamas et al., 2018; O'Donnell et al., 2010), with particular interest for immune responses where Th17 and dendritic cells have been observed to express high levels of AhR (Esser et al., 2009), as well as in inflammatory lung diseases such as cystic fibrosis (Guerrina et al., 2018; Puccetti et al., 2018). Furthermore, the applicability of AhR ligands in chemotherapy has been extensively studied and anticancer prodrugs have been suggested as clinical candidates acting on the selective overexpression of AhR-mediated transcription of cytochrome P450 1A1 (CYP1A1) (Murray et al., 2014; Nandekar and Sangamwar, 2012).

Important *in silico* efforts have been made aimed at providing better understanding of the AhR binding modes, the different activation and inhibitory patterns, as well as the potential outcomes of AhR modulation (Chitrala et al., 2018; Gadaleta et al., 2018; Goya-Jorge et al., 2020). Unfortunately, most of the predictive models reported to date have been based on AhR binding affinities, and thus do not distinguish between agonist and antagonist effects, as well as their corresponding biological responses (Giani Tagliabue et al., 2019; Larsson et al., 2018). A handful of *in silico* studies on AhR agonism may be found in the literature but these have been based on small congeneric dataset of compounds (Ghorbanzadeh et al., 2014), with the exception of a recent study in which a large imbalanced database was employed (Klimenko et al., 2019).

In the present report, the AhR activation was modeled using a large and balanced dataset of chemical compounds and Quantitative Structure Activity-Relationship (QSAR) methods based on different machine learning algorithms. The models' predictivity and robustness was assessed using internal and external validation workflows, followed by prospective *in vitro* experimental validation for a set of chemical compounds of natural and synthetic origin.

2. Materials and methods

2.1. In silico

2.1.1. Database of AhR agonistic activity

A dataset of chemical compounds evaluated for their AhR agonistic profile in the TOX21 Program was retrieved from the PubChem bioassay repository (Wang et al., 2017). The AhR agonistic capacity of these chemicals was evaluated based on a quantitative high-through screening (qHTS) assay using a HepG2 (human hepatocellular carcinoma) cell model stably transfected with an AhR-responsive firefly luciferase reporter gene plasmid (Huang et al., 2016). This dataset was curated based on standard data curation protocols provided by ChemAxon Ltd software. Consequently, duplicated compounds, mixtures, salts containing organic polyatomic counterions, inorganic substances or metal complexes were all discarded. The obtained dataset comprised 7 times more inactive than active AhR agonists. Therefore, it was balanced using a structural similarity filter based on hybridization fingerprints; a Tanimoto coefficient cutoff of 0.5 was applied (Steinbeck et al., 2003).

2.1.2. Structural parametrization and variable selection

The chemical structural characteristics of the molecules comprising the compiled dataset were codified using the information indices (IFIs) implemented in the Graph Theoretical Thermodynamic STAte Functions (GT-STAF) module of the freely available software for Molecular Descriptor Computations ToMoCOMD-CARDD (acronym for Topological Molecular Computational Design-Computer Aided Rational Drug Design) (Barigye and Marrero-Ponce, 2016; Martinez Lopez et al., 2015). The following configuration was considered: Shannon's and mutual entropy IFIs were selected as the molecular descriptor types, Pauling's electronegativity, polarizability, AlogP, TPSA and the electrotopological state were employed for the descriptor weighting scheme. Moreover, a series of aggregation operators as a generalization for the sum of atomic contributions were employed and these include the norms (summation, Euclidean distance), means (geometric, arithmetic, quadratic and harmonic means), statistic aggregation operators (variance, skewness, kurtosis, range and percentile 50) and the classical algorithms which include autocorrelation, gravitational, Kier-Hall connectivity and the electrotopological states molecular indices. All these descriptors were computed for the connected subgraphs (CS) and substructure fingerprint (SS) molecular fragmentation approaches, respectively. Details on the definition of all GT-STAF indices are provided elsewhere (Barigye and Marrero-Ponce, 2016). Constant variables and those with x/x square correlation coefficients greater than 0.8 were discarded. The filtered data matrix was posteriorly employed for the QSAR modeling. Fig. 1 illustrates the general workflow followed to build the QSAR classification models in the present report.

2.1.3. QSAR modeling

For the model building, the data matrix was split into training (75%) and test (25%) sets, respectively. A mutual information rank-based filter was applied to the training dataset to select the variables that best discriminate AhR agonists (actives) from non AhR-agonists (inactives). Subsequently, several machine learning algorithms coupled with the genetic algorithm (GA) feature selection strategy were employed to build classification models for AhR agonism. For the GA configuration, the population size was set to 100, the crossover and mutation probabilities were set to 0.5 and 0.2, respectively, while the F1 score was employed as the optimization function. The best classification models were subject to a 10-fold test set validation and the average performance parameters reported (refer to Fig. 1). Lastly, in-house libraries of compounds were virtually screening and a set of potentially active AhR agonists were selected for *in vitro* prospective validation.
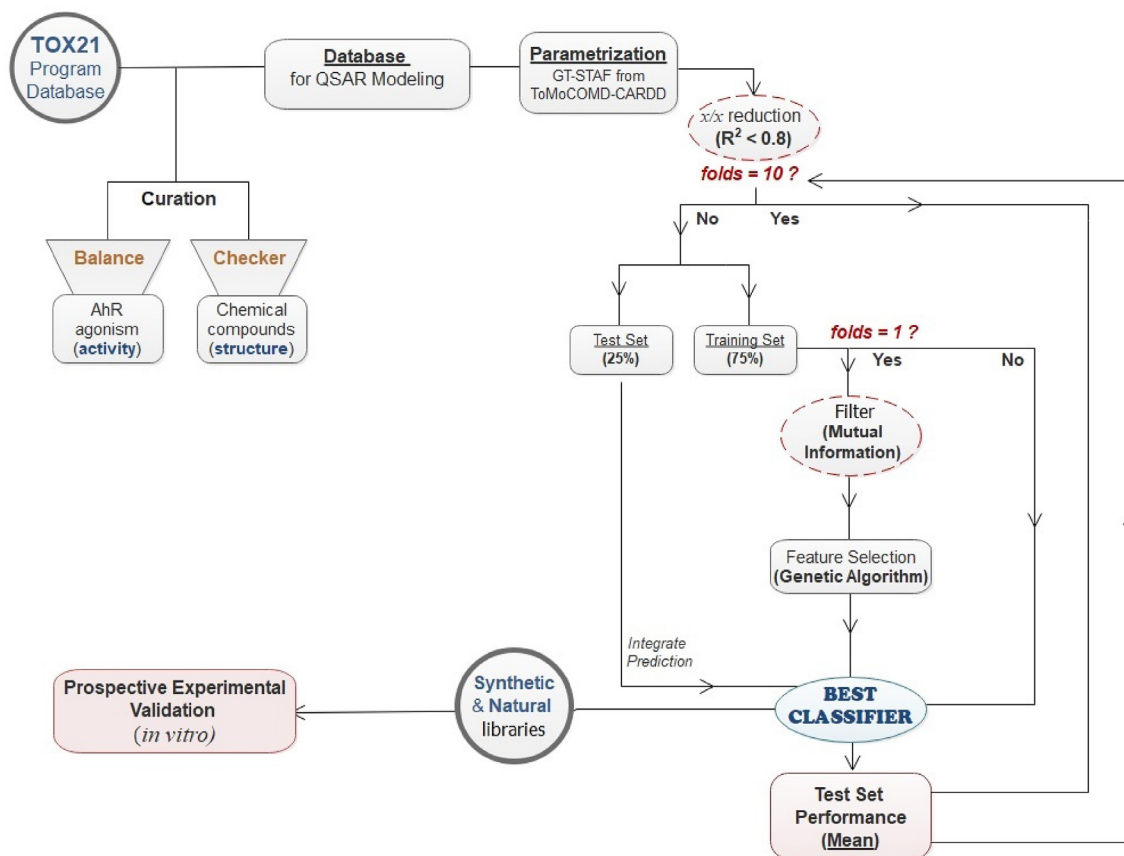
Fig. 1. Workflow employed in the development and validation of QSAR models for aryl hydrocarbon receptor (AhR) agonism.

2.2. In vitro

2.2.1. AhR transcriptional activity

The AhR-mediated transcriptional activity for the set of synthetic and natural compounds was assayed in the AhR-HepG2 Lucia™ cell line (InvivoGen). The cells were maintained in Minimum Essential Medium containing non-essential amino acids (MEM-NEAA, Gibco ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS), penicillin (100 U/mL), streptomycin (100 μg/mL) in a humidified 5% CO_2 atmosphere at 37 °C. Normocin (0.1 mg/mL) and Zeocin (0.2 mg/mL) were added to the culture medium after the third passage.

The compounds cytotoxicity was studied by the MTT assay (Mosmann, 1983). Briefly, AhR-HepG2 cells were exposed to a range of concentrations (1-100 μM) of the tested compounds for 24 h at 37 °C. Then, 100 μL/well of MTT solution (0.5 mg/mL) were added and the plates were incubated until blue deposits were visible. The absorbance was measured at 490 nm in a microplate reader (VICTORx3, PerkinElmer Inc., USA).

The induction of AhR transcriptional activity was evaluated in a reporter gene assay. The AhR-HepG2 cells suspended at 2.0×10^5 cells/mL were seeded into 96-well microplates (200 μL/well), incubated for 24 h at 37 °C, and then, exposure to different concentrations (5-100 μM) of the tested compound (10 μL/well) during 24 h. A volume of 20 μL/well of supernatant were transferred to white sterile and flat-bottom 96-well microplates (Corning). Finally, 50 μL/well of the QUANTI-Luc™ assay reagent (Invivogen) were added and the luminescence immediately measured in a microplate reader (VICTORx3, PerkinElmer Inc., USA). The AhR agonist compound 5,11-Dihydro-indolo[3,2-b]

carbazole-6-carboxaldehyde (FICZ) assayed from 1 to 18 μM was used as a positive control (PC). Compounds showing a maximum response relative to FICZ (RPC_{max}) ≥10% were considered active as agonist of AhR expression (OECD, 2016).

3. Results and discussion

3.1. Data set building and feature selection

A large and diverse dataset of compounds with known AhR agonistic profile was retrieved from TOX21, curated and balanced for subsequent in silico modeling. The final database comprised of 1837 chemical compounds of which 804 were active and 1033 inactive (this dataset is provided as Supporting Information SI-1). A total of 4139 IFIs remained after the zero variance and x/x dimensionality reduction procedure.

3.2. Classification models for AhR agonism

Classification models for AhR agonist activation based on the Adaboost (AdB), Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM) and Multilayer Perceptron (MLP) algorithms were obtained. The configuration parameters and the validation metrics obtained for the five models are presented in Table 1.

As shown in Table 1, the built AhR agonistic activity classification models yielded good external validation performance i.e. accuracies ≥0.75, true active rates (sensitivity) ≥ 0.71, true inactive rates (specificity) ≥ 0.77, active predictive values (precision) ≥ 0.70, and Matthew's correlation coefficients ≥ 0.48, respectively. Therefore, it

Table 1

Evaluation metrics of the classification models for AhR agonism.

QSAR Models [a]	Accuracy [g]		Sensitivity [h]		Specificity [i]		Precision [j]		MCC [k]	
	Train ▪	Ext.V ▪▪	Train ▪	Ext.V ▪▪	Train ▪	Ext.V ▪▪	Train ▪	Ext.V ▪▪	Train ▪	Ext.V ▪▪
AdB [b]	0.91	0.75	0.90	0.71	0.93	0.77	0.91	0.70	0.82	0.48
RF [c]	0.86	0.79	0.87	0.76	0.84	0.82	0.81	0.77	0.71	0.58
GB [d]	0.85	0.76	0.82	0.73	0.87	0.78	0.83	0.72	0.69	0.51
SVM [e]	0.80	0.77	0.77	0.72	0.82	0.81	0.77	0.75	0.59	0.53
MLP [f]	0.80	0.78	0.75	0.72	0.85	0.82	0.80	0.76	0.60	0.55

[a] Classification models for the AhR agonistic activity using 38 molecular descriptors and 1837 chemical structures (804 active and 1033 inactive) (**SI-1**). The evaluation metrics were calculated considering: TA = true active, FI = false active, TI = true inactive, FI = false inactive. ▪***Train.*** Metrics obtained in the training of the models. ▪▪***Ext.V.*** average classification metrics obtained for the 10-fold external validation.

[b] Adaboost (**AdB**) classifier: number of estimators = 250.

[c] Random Forest (**RF**) classifier: max depth = 6, number of estimators = 150.

[d] Gradient Boosting (**GB**) classifier: number of estimators = 150.

[e] Support-Vector Machine (**SVM**): radial basis function kernel (RBF), C = 14.0, gamma = 0.07.

[f] Multilayer Perceptron: **MLP** 38-100-2, activation function: relu.

[g] **Accuracy** = (TA + TI) ÷ (TA + FI + FA + TI).

[h] **Sensitivity** = TA ÷ (TA + FI).

[i] **Specificity** = TI ÷ (TI + FA).

[j] **Precision** = TA ÷ (TA + FA).

[k] Matthews Correlation Coefficient (**MCC**) = (TA × TI − FA × FI) ÷ ((TA + FA) × (TA + FI) × (TI + FA) × (TI + FI)) ^{½}.

may be deduced that the built QSAR models possess adequate robustness and predictive capacity (Cherkasov et al., 2014).

In a recently published study, composite QSAR models were developed for a large imbalanced dataset of AhR agonism using more than eight thousand structural keys transformed in factors and ensemble QSARs based on Partial Logistic Regression (PLR) (Klimenko et al., 2019). As shown in Fig. 2, no significant differences ($p > 0.05$) were observed between the balanced accuracy in the training and validation sets of the best model reported by Klimenko et al., (2019) when compared with the classification accuracy of the five QSAR models proposed herein.

3.3. Prospective experimental validation

The best classifiers of AhR agonism were prospectively validated by virtually screening in-house libraries of synthetic and natural chemical compounds. A total of 40 compounds were selected for

Comparative Performance

Fig. 2. Comparative Performance of the built QSAR models and a reference method, i.e. Partial Logistic Regression (PLR$^\phi$) model from the literature named "QSAR 2:1" by Klimenko et al., (2019). The training performance and the 10-fold cross-validation of all QSAR models (represented as a range) were compared in a One-way ANOVA following Turkey's multiple comparison post-test. No significant differences ($p > 0.05$) were identified.

in vitro experimental evaluation of their AhR agonistic activity profile (details provided as SI-2). These compounds included 9 of synthetic (S) origin i.e. 6 benzothiazoles (**B01-B06**) and 3 coumarins (**C01-C03**), and 31 of natural (N) origin *i.e.* 22 phenolic (**P01-P22**, nine of which are flavonoids) and nine terpenoids (**T01-T09**, mainly triterpenes derivatives). The predictions of AhR agonism for each individual QSAR model for the 40 compounds are shown in Table 2, while the structural identity, the consensus (majority vote) prediction based on the five classifiers base classifiers and the obtained *in vitro* results are presented in Table 3.

The cytotoxicity of the 40 compounds was evaluated by MTT assay. The non-cytotoxic concentration to perform the subsequent luciferase reporter gene assay of AhR-mediated transcriptional activity was established as one which guarantees over 85% cell viability.

As shown in Table 3, the selected exposure doses were much higher for the natural occurring compounds (12.5−100 μM) than for the synthetic derivatives (5-10 μM). The response percentage relative to the positive control used (FICZ) was considered as the threshold to identify active AhR agonists from the inactives. Based on the maximum response induced at the highest dose for each tested compound, 14 of the tested chemicals were classified as active AhR-agonists due to their RPC$_{max}$ ≥ 10% (OECD, 2016), as reported in Table 3.

The benzothiazole class of compounds was the most prominent group of AhR agonists, consistent with other reports in the literature (Bradshaw and Westwell, 2005). The most potent AhR agonist identified herein was the synthetic methyl benzothiazole **B03**, with an 86.44% of activation relative to the endogenous agonist FICZ. Thus, the disubstituted benzene ring with 2-nitroaniline significantly improves the activity when compared to the analogous methyl benzothiazoles monosubstituted with ρ-nitro benzene (**B01**) or ρ-aniline (**B02**). Variations in the substituents of the benzene ring fused with the thiazole (**B04** and **B05**) resulted in an approximately 4- and 3-fold decrease of the activity when compared with the methyl derivatives **B03** and **B02**, respectively. The only pyridine benzothiazole tested (**B06**) showed a remarkable induction of AhR (73.55%).

On the other hand, the best AhR agonist identified among the

Table 2
Predictions of the QSAR models built for the AhR agonism of 40 selected chemical compounds (light blue: active, dark blue: inactive).

ID [a]	B01	B02	B03	B04	B05	B06	C01	C02	C03	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	T01	T02	T03	T04	T05	T06	T07	T08	T09
AdB																																								
RF																																								
GB																																								
SVM																																								
MLP																																								

[a] **ID**: identification code of the 40 compounds, benzothiazoles (**B01-B06**), coumarin (**C01-C03**), phenolics (**P01-P22**), triterpenes (**T01-T09**).

natural compounds, was the small phenolic molecule 4-methylcatechol (**P02**), that exhibited activity in the same order as the benzothiazole **B06**. Interestingly, slight variations in the simple phenolic ring led to a partial (**P01**) or complete loss of activity even in cases where the catechol function was present (**P08**). Some of these (**P06**, **P03**) showed an RPC$_{max}$ close to the 10% threshold, but most of the substituted simple phenols were inactive, particularly the phenolic acids and their derivatives (**P04**, **P05**, **P07-P09**). Diferuloylmethane curcumin (**P10**) approximated the threshold of active AhR agonists, contrary to previous reports (Ciolino et al.,

Table 3
Summary of the AhR agonist activity evaluated *in vitro* and predicted *in silico* (ensemble) for 40 selected chemical compounds.

ID [a]	Origin [b]	Structure & CAS number [c]	[μM] [d] Min	Max	RPC$_{max}$ (%) ±SEM [e]	*in vitro* AhR agonism [f]	*in silico* Ensemble prediction [g]
B01	S	CAS 488722-57-2	5.0	10	16.03 ± 0.21	**Active**	TRUE
B02	S	CAS 92-36-4	5.0	10	66.34 ± 0.64	**Active**	TRUE
B03	S	CAS 1448780-07-1	5.0	10	86.44 ± 0.84	**Active**	TRUE
B04	S	CAS 132064-21-2	5.0	10	22.32 ± 0.32	**Active**	TRUE
B05	S	CAS 488722-59-4	5.0	10	20.61 ± 0.42	**Active**	TRUE
B06	S	CAS 1448780-06-0	5.0	10	73.55 ± 0.65	**Active**	TRUE
C01	S	CAS 81-81-2	5.0	10	6.45 ± 0.02	Inactive	TRUE
C02	S	CAS 109936-39-2	5.0	10	6.77 ± 0.02	Inactive	FALSE
C03	S	CAS 2230787-75-2	5.0	10	6.95 ± 0.04	Inactive	TRUE
P01	N	CAS 87-66-1	50	100	15.53 ± 0.29	**Active**	TRUE
P02	N		50	100	79.35 ± 0.58	**Active**	FALSE

(*continued on next page*)

Table 3 (*continued*)

ID [a]	Origin [b]	Structure & CAS number [c]	[μM] [d]		RPC$_{max}$ (%) ±SEM [e]	*in vitro* AhR agonism [f]	*in silico* Ensemble prediction [g]
			Min	Max			
P03	N	CAS 452-86-8	50	100	9.09 ± 0.08	Inactive	TRUE
P04	N	CAS 2478-38-8	50	100	7.19 ± 0.04	Inactive	TRUE
P05	N	CAS 119-36-8	50	100	7.36 ± 0.03	Inactive	TRUE
P06	N	CAS 530-57-4	50	100	9.40 ± 0.21	Inactive	TRUE
P07	N	CAS 499-75-2	50	100	8.11 ± 0.05	Inactive	TRUE
P08	N	CAS 7400-08-0	50	100	7.14 ± 0.04	Inactive	TRUE
P09	N	CAS 331-39-5	50	100	7.62 ± 0.11	Inactive	TRUE
P10	N	CAS 1135-24-6	50	100	9.97 ± 0.24	Inactive	TRUE
P11	N	CAS 458-37-7	50	100	6.57 ± 0.14	Inactive	TRUE
P12	N	CAS 476-66-4	12.5	25	<5.0	Inactive	FALSE
P13	N	CAS 1401-55-4	50	100	4.17 ± 0.08	Inactive	TRUE
P14	N	CAS 989-51-5	12.5	25	<5.0	Inactive	TRUE
		CAS 2957-21-3					

Table 3 (*continued*)

ID [a]	Origin [b]	Structure & CAS number [c]	[μM] [d]		RPC_max (%) ±SEM [e]	*in vitro* AhR agonism [f]	*in silico* Ensemble prediction [g]
			Min	Max			
P15	N	CAS 520-33-2	50	100	14.48 ± 0.28	**Active**	TRUE
P16	N	CAS 10236-47-2	50	100	7.84 ± 0.06	Inactive	FALSE
P17	N	CAS 520-26-3	50	100	13.59 ± 0.13	**Active**	FALSE
P18	N	CAS 520-36-5	50	100	10.30 ± 0.20	**Active**	TRUE
P19	N	CAS 529-53-3	50	100	22.32 ± 0.27	**Active**	FALSE
P20	N	CAS 30370-87-7	50	100	30.76 ± 0.72	**Active**	TRUE
P21	N	CAS 153-18-4	50	100	7.47 ± 0.05	Inactive	TRUE
P22	N	CAS 446-72-0	12.5	25	<5.0	Inactive	TRUE
T01	N	CAS 2415-24-9	50	100	16.05 ± 0.09	**Active**	FALSE
T02	N	CAS 545-48-2	12.5	25	<5.0	Inactive	TRUE
T03	N	CAS 508-02-1	12.5	25	<5.0	Inactive	FALSE

Table 3 (*continued*)

ID [a]	Origin [b]	Structure & CAS number [c]	[µM] [d]		RPC$_{max}$ (%) ±SEM [e]	*in vitro* AhR agonism [f]	*in silico* Ensemble prediction [g]
			Min	Max			
T04	N	CAS 73039-13-1	12.5	25	<5.0	Inactive	TRUE
T05	N	CAS 638-95-9	50	100	7.32 ± 0.02	Inactive	TRUE
T06	N	CAS 545-47-1	12.5	25	<5.0	Inactive	TRUE
T07	N	CAS 473-98-3	50	100	9.62 ± 0.03	Inactive	FALSE
T08	N	CAS 472-15-1	12.5	25	<5.0	Inactive	TRUE
T09	N	CAS 474-62-4	12.5	25	<5.0	Inactive	TRUE

[a] **ID:** Identification code, benzothiazoles (**B01-B06**), coumarin (**C01-C03**), phenolics (**P01-P22**), triterpenes (**T01-T09**).

[b] **origin**: **S** (Synthetic) or **N** (Natural).

[c] CAS registration number below each structural representation (further details of the compounds provided as **SI-2**).

[d] **[µM]**: *in vitro* assayed concentrations (**Max**: maximum and **Min**: minimum).

[e] **RPCmax (%)**: Percentage of the maximum response relative to the positive control (agonist control FICZ) ± **SEM**.

[f] *in vitro* **AhR agonism**: based on the threshold of RPC≥10% for AhR agonists, the classification of active/inactive agonist compounds is shown.

[g] *in silico* **Ensemble Prediction**: from the predictive results of the five QSAR models built, a consensual activity prediction is suggested and compared with the experimental results, showing TRUE when there is an agreement between the *in vitro vs. in silico* results or else, FALSE.

1998), probably due to its predominant suppressive effects of AhR-mediated transcription (Nakai et al., 2017; Nishiumi et al., 2007). Surprisingly, ellagic acid (**P11**) and epigallocatechin gallate (**P12**) did not exhibit any AhR agonism despite the multiple catechol moieties in their structures.

The flavonoids (**P14-P22**) were the class of phenolic compounds with the highest number of active compounds. Their relative AhR agonistic activity was as follows: quercetin 3-arabinoside (**P20**) > scutellarein (**P19**) > hesperetin (**P15**) ≈ heperidin (**P17**) ≈ apigenin (**P18**). No differences were observed in the activity displayed by the heteroside flavonoids and their corresponding aglycones in neither flavanones nor flavones (**P16** *vs*. **P14** or **P20** *vs*. **P18**, respectively). However, in the case of flavonols, the presence of an additional glycoside unit (**P20** vs **P21**) resulted in the loss of the agonistic activity. Lastly, the isoflavone genistein (**P22**) was found inactive, probably due to other modulatory effects on AhR as some of its analogous (Chan et al., 2003).

Finally, in the terpenoids group of compounds (**T01-T09**), the only active compound was the iridoid catalpol (**T01**), while betulin (**T07**) approximated the activity threshold. Most of the assayed

terpenes were cytotoxic at exposure concentrations higher than 25 µM.

As may be observed in Table 3, the prospective *in vitro* experimentation validated the predictiveness and applicability of the built classification models for AhR agonist activity. Indeed, the ensemble of five QSAR models built in this work, correctly predicted the activity profile of 31 of the 40 evaluated chemicals, yielding a 77.5% of classification accuracy.

4. Conclusions

Five classification models for predicting the AhR antagonistic activity of chemical compounds were built and validated, demonstrating adequate robustness and predictive power. These classifiers comprised an ensemble model, which was used to screen an in-house repository of synthetic and natural occurring compounds. Consequently, a set of 40 diverse chemicals was selected for experimental validation. The satisfactory *in vitro* collaboration of the predicted AhR agonistic activity supports the utility of the built ensemble classifier in guiding the prioritization of chemical

compounds with desired AhR agonistic activity profiles both in the in toxicological and pharmacological contexts. It is important to highlight that all the synthetic and most of the natural occurring compounds tested *in vitro* herein have not been previously evaluated for their AhR agonistic activity and is thus reported here for the first time.

Author contributions

All authors contributed to the drafting and revision of the article and approved the final version presented.

Declaration of competing interest

The authors declare no conflict of interest.

CRediT authorship contribution statement

Elizabeth Goya-Jorge: Investigation, Formal analysis, Validation, Data curation, Writing - original draft. **Rosa M. Giner:** Supervision, Methodology, Writing - review & editing. **Maité Sylla-lyarreta Veitía:** Supervision, Resources. **Rafael Gozalbes:** Supervision, Funding acquisition, Project administration. **Stephen J. Barigye:** Supervision, Conceptualization, Software, Writing - review & editing.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2020.127068.

References

Barigye, S.J., Marrero-Ponce, Y., 2016. Digital communication and chemical structure codification. In: Meyers, R.A. (Ed.), Encyclopedia of Complexity and Systems Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–28. https://doi.org/10.1007/978-3-642-27737-5_625-2.

Bock, K.W., 2018. From TCDD-mediated toxicity to searches of physiologic AHR functions. Biochem. Pharmacol. 155, 419–424. https://doi.org/10.1016/j.bcp.2018.07.032.

Bradshaw, T., Westwell, A., 2005. The development of the antitumour benzothiazole prodrug, phortress, as a clinical candidate. Curr. Med. Chem. 11, 1009–1021. https://doi.org/10.2174/0929867043455530.

Chan, H.Y., Wang, H., Leung, L.K., 2003. The red clover (Trifolium pratense) isoflavone biochanin A modulates the biotransformation pathways of 7, 12-dimethylbenz[a]anthracene. Br. J. Nutr. 90, 87–92. https://doi.org/10.1079/BJN2003868.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'Min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., 2014. QSAR modeling: where have you been? Where are you going to? J. Med. Chem. 57, 4977–5010. https://doi.org/10.1021/jm4004285.

Chitrala, K.N., Yang, X., Nagarkatti, P., Nagarkatti, M., 2018. Comparative analysis of interactions between aryl hydrocarbon receptor ligand binding domain with its ligands: a computational study. BMC Struct. Biol. 18 https://doi.org/10.1186/s12900-018-0095-2.

Ciolino, H.P., Daschner, P.J., Wang, T.T.Y., Yeh, G.C., 1998. Effect of curcumin on the aryl hydrocarbon receptor and cytochrome P450 1A1 in MCF-7 human breast carcinoma cells. Biochem. Pharmacol. 56, 197–206. https://doi.org/10.1016/S0006-2952(98)00143-9.

Esser, C., Lawrence, B.P., Sherr, D.H., Perdew, G.H., Puga, A., Barouki, R., Coumoul, X., 2018. Old receptor, new tricks—the ever-expanding universe of aryl hydrocarbon receptor functions. In: Report from the 4th AHR Meeting, 29–31 August 2018 in Paris, France. Int. J. Mol. Sci., vol. 19. https://doi.org/10.3390/ijms19113603.

Esser, C., Rannug, A., Stockinger, B., 2009. The aryl hydrocarbon receptor in immunity. Trends Immunol. 30, 447–454. https://doi.org/10.1016/j.it.2009.06.005.

Fujii-Kuriyama, Y., Mimura, J., 2007. Transcriptional roles of AhR in expression of biological effects induced by endocrine disruptors. Pure Appl. Chem. 75, 1819–1826. https://doi.org/10.1351/pac200375111819.

Gadaleta, D., Manganelli, S., Roncaglioni, A., Toma, C., Benfenati, E., Mombelli, E., 2018. QSAR modeling of ToxCast assays relevant to the molecular initiating events of AOPs leading to hepatic steatosis. J. Chem. Inf. Model. 58, 1501–1517. https://doi.org/10.1021/acs.jcim.8b00297.

Ghorbanzadeh, M., Van Ede, K.I., Larsson, M., Van Duursen, M.B.M., Poellinger, L., Lücke-Johansson, S., Machala, M., Pěnčíková, K., Vondráček, J., Van Den Berg, M., Denison, M.S., Ringsted, T., Andersson, P.L., 2014. In vitro and in silico derived relative effect potencies of ah-receptor-mediated effects by PCDD/Fs and PCBs in rat, mouse, and Guinea pig CALUX cell lines. Chem. Res. Toxicol. 27, 1120–1132. https://doi.org/10.1021/tx5001255.

Giani Tagliabue, S., Faber, S.C., Motta, S., Denison, M.S., Bonati, L., 2019. Modeling the binding of diverse ligands within the Ah receptor ligand binding domain. Sci. Rep. 9, 1–14. https://doi.org/10.1038/s41598-019-47138-z.

Gies, A., Neumeier, G., Rappolder, M., Konietzka, R., 2007. Risk assessment of dioxins and dioxin-like PCBs in food − comments by the German federal environmental agency. Chemosphere 67, S344–S349. https://doi.org/10.1016/j.chemosphere.2006.05.128.

Goya-Jorge, E., Doan, T.Q., Scippo, M.L., Muller, M., Giner, R.M., Barigye, S.J., Gozalbes, R., 2020. Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective. SAR QSAR Environ. Res. 31, 209–226. https://doi.org/10.1080/1062936X.2019.1708460.

Guerrina, N., Traboulsi, H., Eidelman, D.H., Baglole, C.J., 2018. The aryl hydrocarbon receptor and the maintenance of lung health. Int. J. Mol. Sci. 19, 3882. https://doi.org/10.3390/ijms19123882.

Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S.A., Attene-ramos, M., Zhao, T., Austin, C.P., Simeonov, A., 2016. Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat. Commun. 7, 1–10. https://doi.org/10.1038/ncomms10425.

Kawajiri, K., Fuji-Kuriyama, Y., 2016. The aryl hydrocarbon receptor: a multifunctional chemical sensor for host defense and homeostatic maintenance. Exp. Anim. 66, 75–89. https://doi.org/10.1538/expanim.16-0092.

Klimenko, K., Rosenberg, S.A., Dybdahl, M., Wedebye, E.B., Nikolov, N.G., 2019. QSAR modelling of a large imbalanced aryl hydrocarbon activation dataset by rational and random sampling and screening of 80,086 REACH pre-registered and/or registered substances. PloS One 14, 1–21. https://doi.org/10.1371/journal.pone.0213848.

Lamas, B., Natividad, J.M., Sokol, H., 2018. Aryl hydrocarbon receptor and intestinal immunity review-article. Mucosal Immunol. 11, 1024–1038. https://doi.org/10.1038/s41385-018-0019-2.

Larsson, M., Fraccalvieri, D., Andersson, C.D., Bonati, L., Linusson, A., Andersson, P.L., 2018. Identification of potential aryl hydrocarbon receptor ligands by virtual screening of industrial chemicals. Environ. Sci. Pollut. Res. 25, 2436–2449. https://doi.org/10.1007/s11356-017-0437-9.

Martinez Lopez, Y., Marrero-Ponce, Y., Barigye, S.J., Martinez Santiago, O., 2015. Software for Molecular Descriptor Computations: TOMOCOMD-CARDD (TOpological MOlecular COMputer Design - Computed-Aided Rational Drug Design).

Mescher, M., Haarmann-Stemmann, T., 2018. Modulation of CYP1A1 metabolism: from adverse health effects to chemoprevention and therapeutic options. Pharmacol. Ther. 187, 71–87. https://doi.org/10.1016/j.pharmthera.2018.02.012.

Mitchell, K.A., Elferink, C.J., 2009. Timing is everything: consequences of transient and sustained AhR activity. Biochem. Pharmacol. 77, 947–956. https://doi.org/10.1016/j.bcp.2008.10.028.

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J. Immunol. Methods 65, 55–63. https://doi.org/10.1016/0022-1759(83)90303-4.

Murray, I.A., Patterson, A.D., Perdew, G.H., 2014. Aryl hydrocarbon receptor ligands in cancer: friend and foe. Nat. Rev. Canc. 14, 801–814. https://doi.org/10.1038/nrc3846.

Nakai, R., Fukuda, S., Kawase, M., Yamashita, Y., Ashida, H., 2017. Curcumin and its derivatives inhibit 2,3,7,8,-tetrachloro-dibenzo-p-dioxin-induced expression of drug metabolizing enzymes through aryl hydrocarbon receptor-mediated pathway. Biosci. Biotechnol. Biochem. 82, 616–628. https://doi.org/10.1080/09168451.2017.1386086.

Nandekar, P.P., Sangamwar, A.T., 2012. Cytochrome P450 1A1-mediated anticancer drug discovery: in silico findings. Expet Opin. Drug Discov. 7, 771–789. https://doi.org/10.1517/17460441.2012.698260.

Nebert, D.W., 2017. Aryl hydrocarbon receptor (AHR): "pioneer member" of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of "sensors" of foreign and endogenous signals. Prog. Lipid Res. 67, 38–57. https://doi.org/10.1016/j.plipres.2017.06.001.

Nishiumi, S., Yoshida, K. ichi, Ashida, H., 2007. Curcumin suppresses the transformation of an aryl hydrocarbon receptor through its phosphorylation. Arch. Biochem. Biophys. 466, 267–273. https://doi.org/10.1016/j.abb.2007.08.007.

O'Donnell, E.F., Saili, K.S., Koch, D.C., Kopparapu, P.R., Farrer, D., Bisson, W.H., Mathew, L.K., Sengupta, S., Kerkvliet, N.I., Tanguay, R.L., Kolluri, S.K., 2010. The anti-inflammatory drug leflunomide is an agonist of the aryl hydrocarbon receptor. PloS One 5. https://doi.org/10.1371/journal.pone.0013128.

OECD, 2016. Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation in Vitro Assays to Detect Estrogen Receptor Agonists and Antagonists. https://doi.org/10.1787/9789264265295-en.

Puccetti, M., Paolicelli, G., Oikonomou, V., De Luca, A., Renga, G., Borghi, M., Pariano, M., Stincardini, C., Scaringi, L., Giovagnoli, S., Ricci, M., Romani, L., Zelante, T., 2018. Towards targeting the aryl hydrocarbon receptor in cystic fibrosis. Mediat. Inflamm. 1601486. https://doi.org/10.1155/2018/1601486, 2018.

Roman, Á.C., Carvajal-Gonzalez, J.M., Merino, J.M., Mulero-Navarro, S., Fernández-Salguero, P.M., 2018. The aryl hydrocarbon receptor in the crossroad of signalling networks with therapeutic value. Pharmacol. Ther. 185, 50–63. https://doi.org/10.1016/j.pharmthera.2017.12.003.

Rothhammer, V., Quintana, F.J., 2019. The aryl hydrocarbon receptor: an environmental sensor integrating immune responses in health and disease. Nat. Rev. Immunol. 19, 184–197. https://doi.org/10.1038/s41577-019-0125-8.

Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E., Willighagen, E., 2003. The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bioinformatics. J. Chem. Inf. Comput. Sci. 43, 493–500. https://doi.org/10.1021/ci025584y.

Wang, Y., Cheng, T., Bryant, S.H., 2017. PubChem BioAssay: a decade's development toward open high-throughput screening data sharing. SLAS Discov 22, 655–666. https://doi.org/10.1177/2472555216685069.

Chapter 3

CHAPTER 3. Study of benzothiazoles as AhR ligands

- **Publication Title**: "Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic activities".
  (https://doi.org/10.1016/j.ejps.2020.105386).
- **Authors**: Elizabeth Goya-Jorge, Fatma Abdmouleh, Laureano E. Carpio, Rosa M. Giner, Maité Sylla-Iyarreta Veitía.

Summary

This third Chapter presents the analysis of a set of 16 synthetically produced benzothiazoles as dual AhR modulators and antimicrobial agents. *In vitro* methods were used to evaluate the biological activities. Furthermore, *in silico* predictions revealed a good ADMET profile and druglikeness for the most prominent benzothiazole, as well as similarities in binding to AhR compared to the endogenous agonist FICZ.

Highlights

- Sixteen functionalized benzothiazoles were evaluated as antimicrobial agents and as AhR modulators.
- The benzothiazoles showed noticeable antimicrobial effects against Gram-positive and Gram-negative pathogens and against the yeast *C. albicans*.
- Six benzothiazoles exhibited significant AhR agonist effects in a cell-based reporter gene assay.
- Structure-activity relationship analysis exposed some relevant contributions of the substituents to the studied biological effects.
- Compound **12** (4-(6-methyl-1,3-benzothiazol-2-yl)-2-nitroaniline) displayed promising biocide activity and AhR agonism as well as an adequate ADMET profile and binding similarities with the agonist control FICZ.
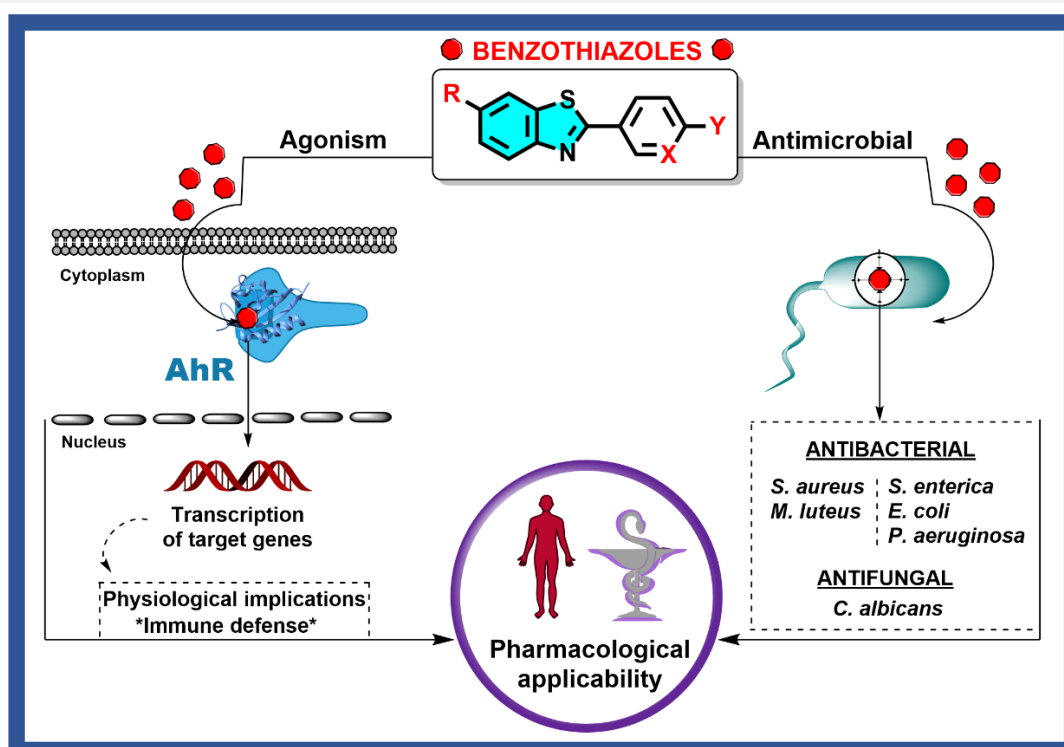
Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic activities

Elizabeth Goya-Jorge[a,b,1], Fatma Abdmouleh[c,d,1], Laureano E. Carpio[b], Rosa M. Giner[a], Maité Sylla-Iyarreta Veitía[c,*]

[a] Departament de Farmacologia, Facultat de Farmàcia, Universitat de València. Av. Vicente Andrés Estellés, s/n, 46100 Burjassot, Valencia, Spain
[b] ProtoQSAR SL. CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia, Av. Benjamin Franklin 12, 46980 Paterna, Valencia, Spain
[c] Equipe de Chimie Moléculaire du Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (EA 7528), Conservatoire National des Arts et Métiers (Cnam), 2 rue Conté, 75003, HESAM Université, Paris, France.
[d] Laboratoire de Biotechnologie Microbienne et d'Ingénierie des Enzymes (LBMIE). Centre de Biotechnologie de Sfax, Université de Sfax, Route de Sidi Mansour Km 6, BP 1177, 3018, Sfax, Tunisie.

ABSTRACT

Benzothiazole is a privileged scaffold in medicinal chemistry present in diverse bioactive compounds with multiple pharmacological applications such as analgesic, anticonvulsant, antidiabetic, anti-inflammatory, anticancer and radioactive amyloidal imagining agents. We reported in this work the study of sixteen functionalized 2-aryl and 2-pyridinylbenzothiazoles as antimicrobial agents and as aryl hydrocarbon receptor (AhR) modulators. The antimicrobial activity against Gram-positive (*S. aureus* and *M. luteus*) and Gram-negative (*P. aeruginosa, S. enterica* and *E. coli*) pathogens yielded MIC ranging from 3.13 to 50 µg/mL and against the yeast *C. albicans*, the benzothiazoles displayed MIC from 12.5 to 100 µg/mL. All compounds showed promising antibiofilm activity against *S. aureus* and *P. aeruginosa*. The arylbenzothiazole **12** displayed the greatest biofilm eradication in *S. aureus* (74%) subsequently verified by fluorescence microscopy. The ability of benzothiazoles to modulate AhR expression was evaluated in a cell-based reporter gene assay. Six benzothiazoles (**7, 8-10, 12, 13**) induced a significant AhR-mediated transcription and interestingly compound **12** was also the strongest AhR-agonist identified. Structure-activity relationships are suggested herein for the AhR-agonism and antibiofilm activities. Furthermore, *in silico* predictions revealed a good ADMET profile and druglikeness for the arylbenzothiazole **12** as well as binding similarities to AhR compared with the endogenous agonist FICZ.

1. Introduction

The interaction with microorganisms is fundamental to produce nutrients and to suppress the pathogen colonization in the mucous membranes of healthy hosts. Microbial communities are playing important roles in organ development, metabolism, and immune homeostasis (Sommer and Bäckhed, 2013). However, the diseases triggered by infectious agents have caused most of the biggest health tragedies of mankind. The coronavirus SARS-CoV-2 responsible for the COVID-19 pandemic is a current example (Shereen et al., 2020). The preservation of health against pathogenic colonization is determined by the immune response as well as the adequate biotransformation of toxicants in the host cells (Janeway et al., 2001). The transformation and appearance of new pathogens and the so-

called emerging diseases, as well as the adaptive resistance to pharmacological treatments, mean that all efforts are insufficient for the control and/or eradication of infections.

The aryl hydrocarbon receptor (AhR) is a long known chemical sensor highly expressed in the liver and in barrier organs (Esser and Rannug, 2015). Despite the extensive association of AhR activation with dioxin-like toxicity during several decades, nowadays is well known that agonist modulations upon AhR can lead to physiotherapeutical benefits in multiple pathological conditions (Hu et al., 2007; Zhao et al., 2019). AhR activity regulates, among others, the immune system functions, through direct or indirect paths, and the differentiation of diverse immune cell types (Esser et al., 2009). Important roles of AhR are suggested in infective disease tolerance and in defense pathways. As an example, the influence of AhR in primary LPS

ADMET	absorption, distribution, metabolism, and excretion - toxicity	FA	fusidic acid
		FICZ	5,11-dihydroindolo[3,2-b]carbazole-12-carbaldehyde
AhR	aryl hydrocarbon receptor	Lv	levofloxacin
AhR-HepG2, Lucia™	human hepatoma cell line stably transfected to express AhR	MBC	Minimal Bactericidal Concentration
		MIC	Minimal Inhibitory Concentration
BT	benzothiazole	MTT	3-(4,5-dimethyl thiazol-2-yl)-2,5-diphenyl tetrazolium bromide
CH223191	2-methyl-2H-pyrazole-3-carboxylic acid		
CYP	cytochrome P450	OD	optical density
Eff	efficiency		

responsiveness and in fact the induction of endotoxin tolerance has been demonstrated using a *Streptococcus*-induced multifocal septic arthritis model. The endotoxin-tolerant state LPS triggered was found to protect mice against immunopathology in Gram-positive infections, steering the contribution of AhR to host fitness (Bessede et al., 2014) . Moreover, AhR plays a key role in innate defense against bacteria. Bacterial pigmented virulence factors such as the phenazines from *P. aeruginosa* and the naphthoquinone phthiocol from *M. tuberculosis*, have been characterized, as ligands of AhR. During ligand binding, AhR activation leads to degradation of virulence factor and regulation of cytokine and chemokine production. Therefore, AhR recognition of these bacterial pigments is reported as a prominent alternative path to control the antibacterial response. It has also been shown that AhR activation regulates inflammatory leukocyte recruitment to the infected lung and control of bacterial replication demonstrating that AhR plays a central role in defense against both acute and chronic bacterial infection (Moura-Alves et al., 2014). Besides its protective effect against intestinal pathogenic bacteria, AhR is also involved in resistance to fungi. It has been described that some indole-type AhR ligands dramatically inhibit *Candida albicans* biofilm formation, as well as attachment to intestinal epithelial cells suggesting that AhR activation may be a potential therapeutic target to combat bacterial and fungal

intestinal infection (Oh et al., 2012).

Benzothiazole (BT) is a heterocyclic structure bearing a benzene ring fused with a five-membered ring containing nitrogen and sulfur atoms. The BT nucleus is considered a privileged structure particularly important in medicinal chemistry due to the wide range of pharmacological applications ascribed to its derivatives (Ali and Siddiqui, 2013; Bondock et al., 2010; Choudhary et al., 2017; Gill et al., 2015). Anticancer, antimicrobial, and anti-inflammatory are among the most extensively reported clinical uses of this versatile fused heterocyclic scaffold (Kamal et al., 2015). Furthermore, BT skeleton is present in imaging agents and as radiotracers of positron emission tomography (PET) used to diagnose neurological diseases and in biocidal agents employed as industrial chemicals (Mathis et al., 2003; Muthusubramanian et al., 2001). Some examples of widely used benzothiazoles are represented in Figure 1.

In 2013, encouraged by the high interest in [^{11}C] PIB for diagnosing Alzheimer's disease, we developed a new series of 2-arylbenzothiazoles and 2-pyridinylbenzothiazole derivatives by Szuki-Miyaura coupling reaction. The one-step protocol was performed from commercially available reactants to obtain the desired radiopharmaceutical precursors in high yields under thermal conditions or microwave activation, Scheme 1 (Bort et al., 2013).
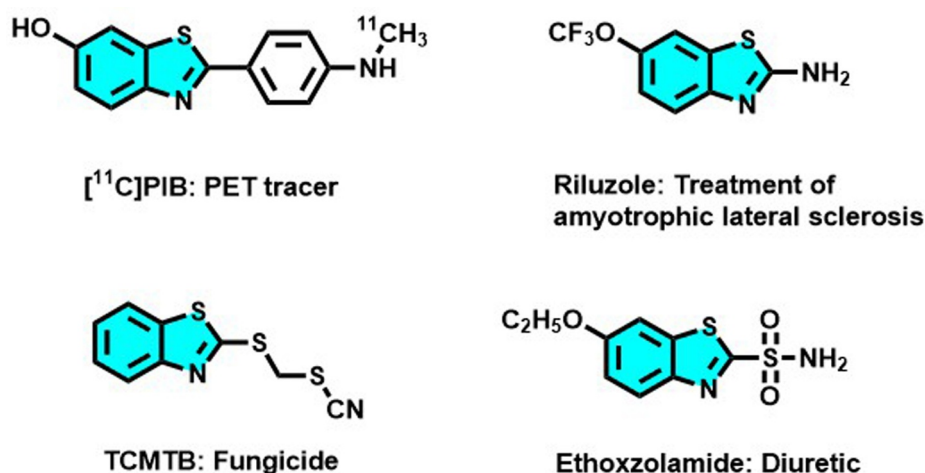
Figure 1. Examples of marketed benzothiazoles and their main applications (Hergesheimer et al., 2015; Mathis et al., 2003; Moyer and Ford, 1958; Muthusubramanian et al., 2001)
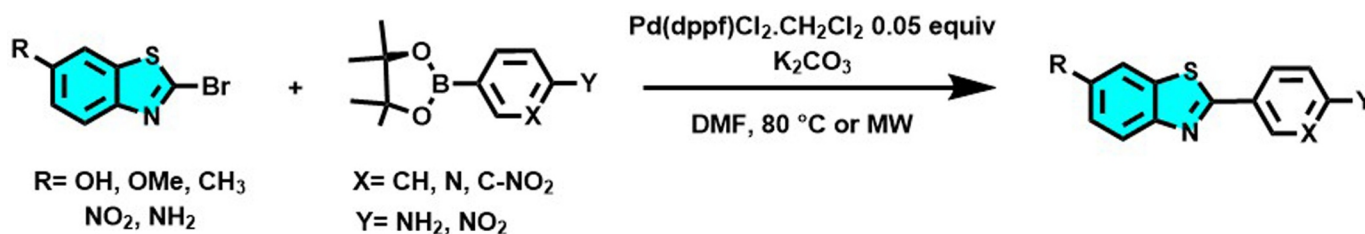
Scheme 1. Synthesis of a new series of arylbenzothiazoles, potential precursors of the radiochemical tracer [^{11}C]PIB, by Suzuki-Miyaura coupling reaction (Bort et al., 2013).

Inspired by the therapeutic potentialities of the benzothiazole scaffold and taking into account the aforementioned pharmacological interest of combining antimicrobial and AhR-modulatory activities we investigate the antimicrobial potential of a series of sixteen arylbenzothiazoles, previously synthesized in our laboratory, against Gram positive and Gram negative pathogens and the opportunist yeast *C. albicans*. The AhR agonist/antagonist behavior of benzothiazoles was studied using an AhR reporter assay preceded by a viability study. The structure-activity relationships considerations between biocidal and AhR modulatory activities were also discussed herein.

2. Experimental Section

The thirteen 2-arylbenzothiazoles and the three 2-pyridinylbenzothiazoles biologically evaluated in this work were previously prepared by our team from the corresponding 2-bromobenzothiazoles (1 eq.) and phenylboronic acid pinacol ester (1.2 eq.) in anhydrous DMF, and in presence of K_2CO_3 (6.0 eq.) and Pd (dppf)Cl_2.CH_2Cl_2 (0.05 eq.). The reactions were performed under argon bubbling at 80 ˚C or under microwave irradiation and monitoring by TLC or by GC-MS. After treatment, the compounds were isolated by filtration as colored powders and the structures were checked by comparison of their NMR, IR, and HRMS data. The detailed synthetic procedures and characterization can be found in the original research article (Bort et al., 2013).

2.1. Biological studies

2.1.1. Antibacterial and antifungal activities

The *in vitro* antibacterial and antifungal activities of the studied benzothiazoles (**1**-**16**) were evaluated by using a microplate broth dilution assay according to the National Committee for Clinical Laboratory Standard (NCCLS. 2002). They were screened against the two Gram-positive pathogens *Staphylococcus aureus* (ATCC 9144) and *Micrococcus luteus* (LB14110) and the three Gram-negative *Salmonella enterica* (NCTC 6017), *Pseudomonas aeruginosa* (ATCC 9027), and *Escherichia coli* (ATCC 8739), and the yeast *Candida albicans* (ATCC 2091) by using microplate broth dilution assay according to the National Committee for Clinical Laboratory Standard (NCCLS. 2002). Serial dilutions of compounds, dissolved in DMSO, were prepared in LB broth medium (1% Bactotryptone, 0.5% Yeast extract, 0.5% NaCl).

A volume of 100 μL of the tested compounds (final concentration ranging from 200-1.15 μg/mL) and 100 μL of bacterial suspension (10^7 CFU/mL) was dispensed into each well of 96-well plates (Corning) and incubated at 37 C for 20 h. Levofloxacin and fusidic acid were used as positive controls. After the incubation period, the Minimal Inhibitory Concentrations (MICs) were recorded as the lowest concentrations of the compounds at which no visible growth occurred. The 3-(4,5-dimethyl thiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) reagent was added to each well to evaluate the microbial growth, indicating live cells when a color transformation from yellow to dark blue was observed. The Minimal Bactericidal Concentrations (MBCs) were determined by sampling one loopful from each well and cultured on plate count agar at 30 C for 24 h. The lowest concentration of compounds that resulted in microbial death was reported as their MBCs.

The optical density (OD) was measured at 600 nm using a microplate reader (Varioskan, Thermofisher). The antimicrobial activity was expressed as the inhibition percentage when compared with microbial culture untreated.

2.1.2. Antibiofilm activity

Biofilm formation in *S. aureus* and *P. aeruginosa* was determined spectrophotometrically by using the crystal violet test in 96-well plates as previously described (Ricco et al., 2020). The tested compounds (**1**-**16**) were assayed at their MIC against *S. aureus* and *P. aeruginosa*, respectively, to determine the biofilm anti-adhesion activity. In addition, the biofilm eradication activity of benzothiazoles was evaluated at 200 μg/mL. A volume of 100 μL of each BT/well was delivered into the plates. Then, 100 μL/well of a fresh overnight cultured bacterial suspension diluted was added until obtaining a final OD of 0.1 at 600 nm. After 24 h incubation at 37 ºC, the wells were emptied, gently rinsed twice with PBS, and the plates dried at 60 ºC for 45 min. The biofilm was stained with 150 μL of crystal violet solution (0.2%) for 15 min at room temperature. After staining, plates were rinsed with water and 200 μL/well of glacial acetic acid were added to dissolve the crystals. After 1 h incubation at room temperature, the OD of the adherent biofilm was measured at 570 nm using a microplate reader (Varioskan, Thermofisher). The antibiofilm activity was expressed as the inhibition percentage when compared with the negative control.

The antibiofilm activity against *S. aureus* of compound 4-(6-methyl-1,3-benzothiazol-2-yl)-2-nitroaniline (**12**) was also determined by fluorescence microscopy images (OLYMPUS Fluorescent microscope BX50 equipped with a digital camera OLYMPUS DP70). The biofilms grown on glass pieces (Ø 10 mm) and placed in 24-well polystyrene plates were treated with **12** at 50 μg/mL (MIC value against *S. aureus*) for the anti-adhesion test and at 200 μg/mL for the eradication test. Non-treated wells containing the LB medium served as control. The bacterial inoculation was adjusted to an OD 600 nm of 0.1. Plates were incubated at 37 ˚C for 24 h. The wells were carefully emptied, the glass slides washed with PBS, and 500 μL/well of acridine orange (0.1% in PBS) were added. A 40× objective using U-MWB2 filter with excitation at 460–490 nm and emission at 520 nm was used for the visualization.

2.1.3. Cell viability evaluation

Cell Line characteristics and maintenance. AhR-HepG2 Lucia™ cell line engineered from the human HepG2 hepatoma cell line was obtained from InvivoGen stably transfected to express the endogenous Ah receptor and thereby screening potential AhR ligands. Lucia luciferase reporter gene is coupled with the human *Cyp1a1*, while the Lucia luciferase reporter protein is readily measurable in the cell culture supernatant. According to the provider's recommendations, AhR-HepG2 cells were maintained in Minimum Essential Medium containing nonessential amino acids (MEM-NEAA, Gibco ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS), penicillin (100 U/mL), streptomycin (100 μg/mL) in a humidified 5% CO_2 atmosphere at 37 ˚C. Normocin (0.1 mg/mL) and Zeocin (0.2 mg/mL) from Invivogen were added to the culture medium after the third passage.

Cell viability bioassay. The reduction caused by dehydrogenases and other reducing agents present in metabolically active cells was evaluated following the colorimetric MTT assay (Mosmann, 1983), with the aim of avoiding misinterpretations caused by cell damage in AhR activity tests.

Briefly, AhR-HepG2 cells were trypsinized, seeded on 96-well plates at a density of 2.0×10^5 cells/mL, and cultured overnight in MEM-NEAA. Next, cells were treated during 24 h with 10 μL/well of 5 μM and 10 μM concentrations of the sixteen benzothiazoles, or else with the positive control ligands which were 5,11-dihydroindolo[3,2-b]carbazole-12-carbaldehyde (FICZ) for agonism, and 2-methyl-2H-pyrazole-3-carboxylic acid (CH223191) for antagonism. A volume of 100 μL/well of MTT reagent (0.5 mg/mL) was added to the emptied plates containing the cells attached. Approximately 3 h of incubation at 37 ˚C allowed the transformation of the yellow MTT to the water-insoluble violet-blue formazan that was dissolved by adding 100 μL/well of DMSO. The OD was measured by reading the absorbance at 490 nm using a microplate reader (VICTORx3, PerkinElmer Inc., USA).

2.1.4. AhR reporter assays

AhR agonist and antagonist bioassay. To measure the effects on AhR transcriptional activity, AhR-HepG2 cells were trypsinized, passed through 18-gauge (18G) needles, and a suspension of 2.0×10^5 cells/mL was seeded into 96-wells microplates (200 μL/well) in a MEM-NEAA medium without Normocin nor Zeocin. After 24 h of incubation at 37

°C, cells were exposed to the tested compounds during another 24 h. The treatment of cells in agonistic mode was conducted by adding 10 μL/well of 5 μM and 10 μM of the benzothiazoles. In antagonistic mode 10 μL/well of the benzothiazoles (or CH223191) were added followed by 10 μL/well of the EC_{50} determined for the known AhR agonist FICZ used as positive control.

A volume of 20 μL/well of the supernatant of treated cells was transferred to white sterile and flat-bottom 96-wells microplates (Corning). After 50 μL/well of the QUANTI-Luc™ assay reagent (Invivogen) was added and the luminescence immediately measured in a microplate reader (VICTORx3, PerkinElmer Inc., USA). Results were expressed as fold response and as efficiency (Eff) percentage relative to the activity displayed by the positive control FICZ.

2.2. Computational studies

ADMET profile and druglikeness. Physicochemical, biopharmaceutical, and toxicological properties of **12** were predicted using the software ADMET Predictor™ v9.5 (Simulations Plus, Inc., Lancaster, CA, USA) (Ghosh et al., 2016).

Molecular docking simulations. Molecular docking analysis was performed with Autodock Vina (Trott and Olson, 2009) as implemented in YASARA (Krieger and Vriend, 2014). The crystallized protein structure was obtained from the Protein Data Bank (PDB ID 3F1O). Benzothiazole **12** and the known agonist FICZ were the ligands used. Simulations were performed for the entire target structure making stiff the protein and flexible the ligand compounds. The interactions of the best protein/ligand complexes were predicted using the Protein-Ligand

Interaction Profiler web server (Salentin et al., 2015). Molecular graphics and analyses were performed with UCSF Chimera (Pettersen et al., 2004).

3. Results and Discussion

3.1. Biological Evaluation

3.1.1. Antimicrobial activity

A set of BTs was screened for antibacterial activity against Gram-positive (*M. luteus* and *S. aureus*) and Gram-negative (*E. coli*, *P. aeruginosa*, and *S. enterica*) pathogen strains and the yeast *C. albicans*. The Minimal Inhibitory Concentration (MIC) and the Minimal Bactericidal Concentration (MBC) for the 16 tested benzothiazoles and for levofloxacin (Lv) and fusidic acid (FA), used as positive controls, are reported in Table 1. As shown, levofloxacin had a notable potency against all the pathogens with MICs and MBCs lower than 1.52 μg/mL as expected from this broad-spectrum fluoroquinolone (Keam et al., 2005). Meanwhile, the MICs for fusidic acid ranging from 12.5-25 μg/mL and the MBCs were 100 μg/mL in Gram positive bacteria and higher in the other microorganisms studied.

The functionalized 2-aryl and 2-pyridinylbenzothiazoles assayed showed in general good antimicrobial potency. The MBC/MIC ratio (≥4) indicated predominant bacteriostatic effects. The antibacterial activity did not reveal important differences in the potency of the benzothiazoles against Gram-positive and Gram-negative bacteria, although Gram-positive strains seem to be more sensitive.

The inhibition potency against *M. luteus* was the most relevant.

Table 1
Minimal inhibitory concentration (MIC) and Minimal bactericidal concentration (MBC) values exhibited by the BTs and the positive controls levofloxacin and fusidic acid.

ID	R	X	Y	[c][a]	Gram (+) *M. luteus*	*S. aureus*	Gram (-) *E. coli*	*P. aeruginosa*	*S. enterica*	Yeast *C. albicans*
1	OH	CH	NH_2	MIC	[6.25-12.5]	[6.25-12.5]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]
				MBC	100	100	>100	>100	100	>100
2	OH	N	NH_2	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[25-50]	[12.5-25]	[12.5-25]
				MBC	100	>100	>100	>100	>100	>100
3	OH	CH	NO_2	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[25-50]	[25-50]	[25-50]
				MBC	100	>100	>100	>100	>100	>100
4	OH	CNO_2	NH_2	MIC	[12.5-25]	[12.5-25]	[25-50]	[50-100]	[25-50]	[50-100]
				MBC	>100	>100	>100	>100	>100	>100
5	OCH_3	CH	NH_2	MIC	[6.25-12.5]	[25-50]	[25-50]	[25-50]	[25-50]	[25-50]
				MBC	>100	>100	>100	>100	>100	>100
6	OCH_3	N	NH_2	MIC	[3.13-6.25]	[25-50]	[25-50]	[25-50]	[25-50]	[25-50]
				MBC	>100	>100	>100	>100	>100	>100
7	CH_3	CH	NO_2	MIC	[6.25-12.5]	[25-50]	[12.5-25]	[12.5-25]	[25-50]	[25-50]
				MBC	>100	>100	>100	>100	>100	>100
8	OCH_3	CNO_2	NH_2	MIC	[6.25-12.5]	[25-50]	[12.5-25]	[12.5-25]	[25-50]	[25-50]
				MBC	>100	>100	>100	>100	>100	>100
9	CH_3	N	NH_2	MIC	[12.5-25]	[25-50]	[12.5-25]	[25-50]	[50-100]	[50-100]
				MBC	>100	>100	>100	>100	>100	>100
10	CH_3	CH	NH_2	MIC	[12.5-25]	[25-50]	[12.5-25]	[25-50]	[50-100]	[50-100]
				MBC	>100	>100	>100	>100	>100	>100
11	OCH_3	CH	NO_2	MIC	[6.25-12.5]	[25-50]	[12.5-25]	[25-50]	[50-100]	[50-100]
				MBC	>100	>100	100	>100	>100	>100
12	CH_3	CNO_2	NH_2	MIC	[6.25-12.5]	[12.5-25]	[12.5-25]	[25-50]	[12.5-25]	[50-100]
				MBC	>100	>100	100	>100	50	100
13	NO_2	CH	NH_2	MIC	[6.25-12.5]	[12.5-25]	[25-50]	[12.5-25]	[25-50]	[25-50]
				MBC	>100	>100	100	>100	50	100
14	NO_2	CH	NO_2	MIC	[6.25-12.5]	[12.5-25]	[25-50]	[25-50]	[12.5-25]	[25-50]
				MBC	>100	>100	>100	>100	>100	>100
15	NO_2	CNO_2	NH_2	MIC	[6.25-12.5]	[12.5-25]	[25-50]	[25-50]	[25-50]	[25-50]
				MBC	>100	>100	>100	>100	>100	>100
16	NH_2	CH	NH_2	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[25-50]	[25-50]	[25-50]
				MBC	100	>100	>100	100	100	100
Levofloxacin	MIC	<1.52	<1.52	<1.52	<1.52	<1.52	<1.52			
	MBC	<1.52	<1.52	<1.52	<1.52	<1.52	<1.52			
Fusidic acid	MIC	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]	[12.5-25]			
	MBC	100	100	>100	>100	>100	>100			

[a] [c] Concentration values are expressed in μg/mL.

Hence, against *M. luteus* the same MICs as FA were estimated for compounds **2-4, 9, 10** and **16**, and the majority of benzothiazoles (**5, 7, 8, 11-15**) showed MICs below that of FA, ranging between 6.25-12.5 µg/mL except for **6** (3.13-6.25 µg/mL). Against *S. aureus* **1** showed the strongest potency with MIC between 6.25-12.5 µg/mL, while half of the compounds (**2-4, 12-16**) exhibited the same MIC as FA (12.5-25 µg/mL). Against *E. coli, P. aeruginosa* and *S. enterica*, compounds **2, 7, 8,** and **12** displayed the higher potency with MICs similar to FA in two bacteria indistinctly, while **1** showed the same MIC as FA for the three Gram-negative strains. Only **12** and **13** displayed MBCs of 50 µg/mL for *S. enterica*. The antifungal MICs for the 2-arylbenzothiazole **1** and its 2-pyridinyl analogous **2** were comparable with FA (12.5-25 µg/mL). The rest of the benzothiazoles displayed higher inhibitory concentrations against the opportunistic pathogenic yeast *C. albicans* ranging from 25-100 µg/mL.

These results suggest that the concurrent functionalization of a hydroxyl group and an amino group as substituents in R and Y positions, respectively, has a positive influence on the antibacterial activity against Gram-positive pathogens (compound **1**) and on the antifungal activity (compounds **1** and **2**). When the hydroxyl function of compound **1** is replaced by a methoxy (compound **5**) or a nitro function

(compound **13**), no difference in activity was observed against *M. luteus* (MIC values between 6.25-12.5). However, the antimicrobial activity decreases for most pathogens. The introduction of an amino group (compound **16**) has no positive influence on the activity. Therefore, a hydroxyl group may be required to exhibit notable antimicrobial effects.

3.1.2. Antibiofilm activity

The biocidal effects of the BTs were evaluated through their ability to inhibit the biofilm formation in *S. aureus* and *P. aeruginosa*. The early biofilm adherence was studied according to the minimum inhibitory concentration corresponding to each compound. All compounds exhibited antibiofilm effects in both strains. The inhibition percentages were greater than 20% for *P. aeruginosa* and above 50% for *S. aureus*, as shown in Figure 2 a) and b), respectively.

The 16 functionalized benzothiazoles exhibited good biocidal effects against Gram-negative *P. aeruginosa*, showing inhibition percentages at their MIC greater 76%. Compounds **5, 7,** and **12** with 85% of anti-adhesive activity and benzothiazole **9** (87%) showed the most prominent anti-biofilm activity on *P. aeruginosa*, significantly higher (p < 0.001) than both positive controls (Lv and FA) used. Unexpectedly,
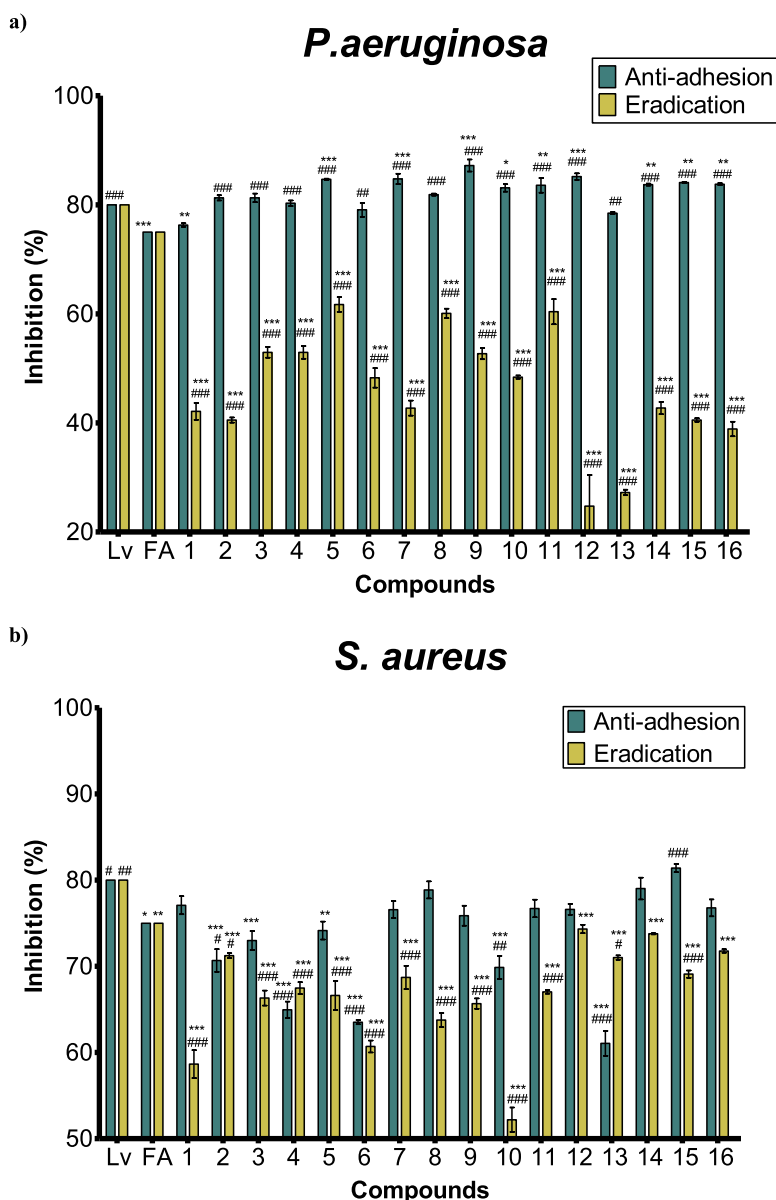
Figure 2. Inhibition percentage of the antibiofilm activity against a) *P. aeruginosa* b) *S. aureus* of levofloxacin (Lv), fusidic acid (FA), and benzothiazoles (**1-16**). Bars represent the effect at MICs (anti-adhesive)-green or at 200 µg/mL (eradication)-yellow of each BT. The percentage is the OD measured at 570 nm of treated bacteria relative to the negative control (untreated). Means ± SD for at least three independent experiments (n = 3) are shown. The levels of significance were determined using one-way ANOVA followed by Dunnett's post-test when compared to Lv: ***p < 0.001, **p < 0.01 and *p < 0.05, or to FA ###p < 0.001, ##p < 0.01 and #p < 0.05 in the anti-adhesion or eradication tests, indistinctly.

most of the compounds exhibited eradication percentages of *P. aeruginosa* lower than those obtained for *S. aureus*, displaying **5** the greater value recorded of 62%.

Meanwhile, against *S. aureus*, the studied benzothiazoles **1, 7-9, 11, 12, 14, 16** displayed inhibition percentages equivalent to FA at their MIC in the anti-adhesion test as shown in Figure 2 a). In the same, compounds **14** (79%) and **15** (81%) exhibited a biocidal effect similar to levofloxacin (80%) against *S. aureus*, significantly greater than FA. In the eradication test, the antibiofilm effect of compound **12** (74%) against *S. aureus* was the most important of the set, being less than Lv but not significantly different from FA (75%).

The biocidal activity of **12** over *S. aureus* was confirmed by fluorescence microscopy (Figure 3). The images of the acridine orange staining treated slides with compound **12** confirmed the ability of this benzothiazole to alter the surface characteristics of bacterial cells and at both anti-adhesive (b) and eradication (c) concentrations when compared to untreated bacteria (a).

3.1.3. Cell viability evaluation

Effects on AhR-HepG2 proliferation caused by the studied 2-aryl and 2-pyridinylbenzothiazoles were determined *in vitro* by the MTT bioassay. The obtained cell viability expressed as a percentage of non-treated cells is represented in Figure 4.

The treatment up to 10 μM of exposure concentration with all the compounds allowed the maintenance of cell proliferation above 85%. Therefore, none of the 16 BTs evaluated was considered cytotoxic on the AhR-HepG2 cell line.

3.1.4. AhR reporter assays

The modulation of AhR-mediated expression caused by the set of BTs was evaluated in a cell-based method using AhR-HepG2 cells. In a previous study, we have reported a preliminary *in silico/in vitro* screening of the AhR agonism of diverse compounds, including some of the studied benzothiazoles herein analyzed (Goya-Jorge et al., 2020). The prominent activity exhibited by these derivatives motivated to elucidate further on the AhR effects caused by BT class of compounds as presented herein. Thus, agonistic and antagonistic bioassays were conducted *in vitro* for a greater number of benzothiazole derivatives, aiming to analyze the substituents' influence in AhR-mediated activity.

AhR antagonism. The EC_{50} of the agonist compound FICZ was 9.06 μM in this cell line, as we previously reported (Goya-Jorge et al., 2020). The antagonist bioassays were developed by co-exposure the EC_{50} of FICZ with each tested BT or else, with the known AhR antagonist CH223191. The dose-response curve obtained for CH223191 is presented in Figure 5.

The dose-response curve obtained for the control CH223191 to validate the competitive antagonist test was consistent with some other results reported in the literature (Mohammadi-Bardbori et al., 2019). However, none of the 16 benzothiazoles assayed exhibited any
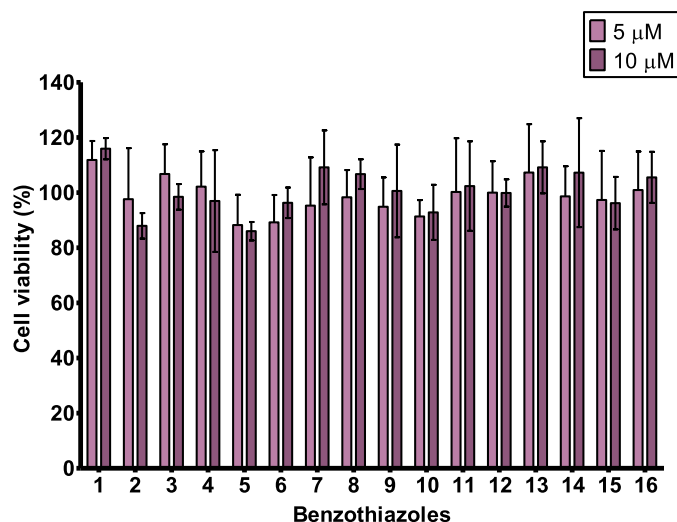
Figure 4. Viability percentages of cells exposed to the BTs by MTT assay. Compounds **1-16** were assayed at 5 μM and 10 μM. Bars represent the mean percentage ± SD from at least four independent experiments (n = 4). Compounds that reduce cell viability below 85% were considered cytotoxic. No significant differences were observed from vehicle control using one-way ANOVA followed by Dunnett's post-test (p < 0.05).
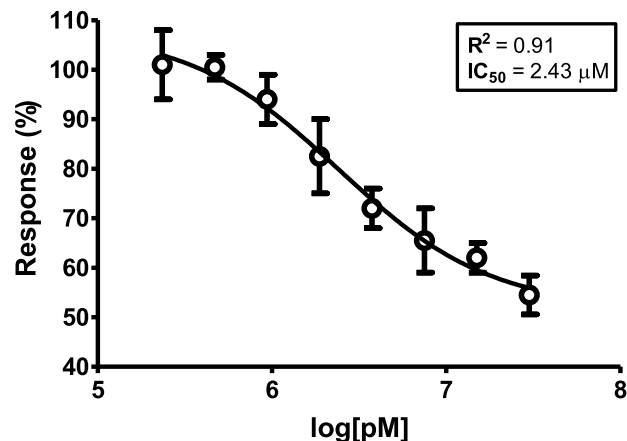
Figure 5. Dose-response curve of compound CH223191 used as positive control in the AhR antagonist assay. The dosage is represented as the logarithm of the concentration expressed in pM (1.0×10^{-12} M), while the effect is expressed as % of FICZ [EC_{50}] ± SEM. The R^2 and the IC_{50} (μM) estimated from the curve are informed.
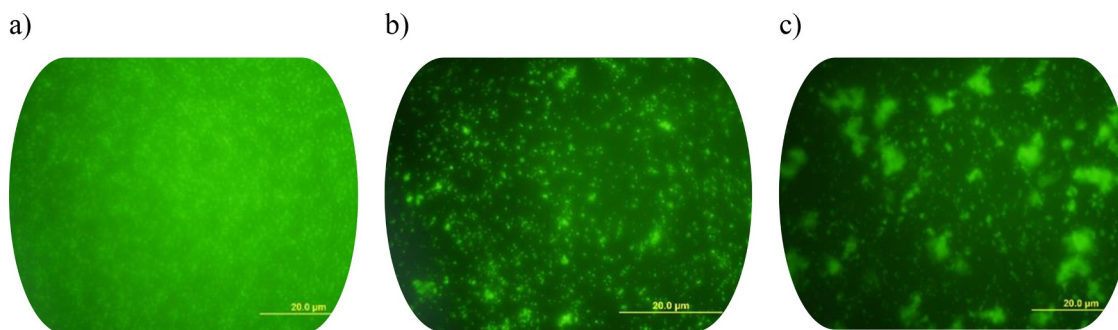
Figure 3. Fluorescence microscopy images (9× 40) of *S.aureus* biofilm formation. Green areas represent the formed biofilm and black areas indicate necrotic zones. a) Non-treated biofilm, b) Biofilm treated with compound **12** at 50 μg/mL (MIC) *i.e.* anti-adhesion activity, c) Biofilm treated with compound **12** at 200 μg/mL *i.e.* eradication activity.
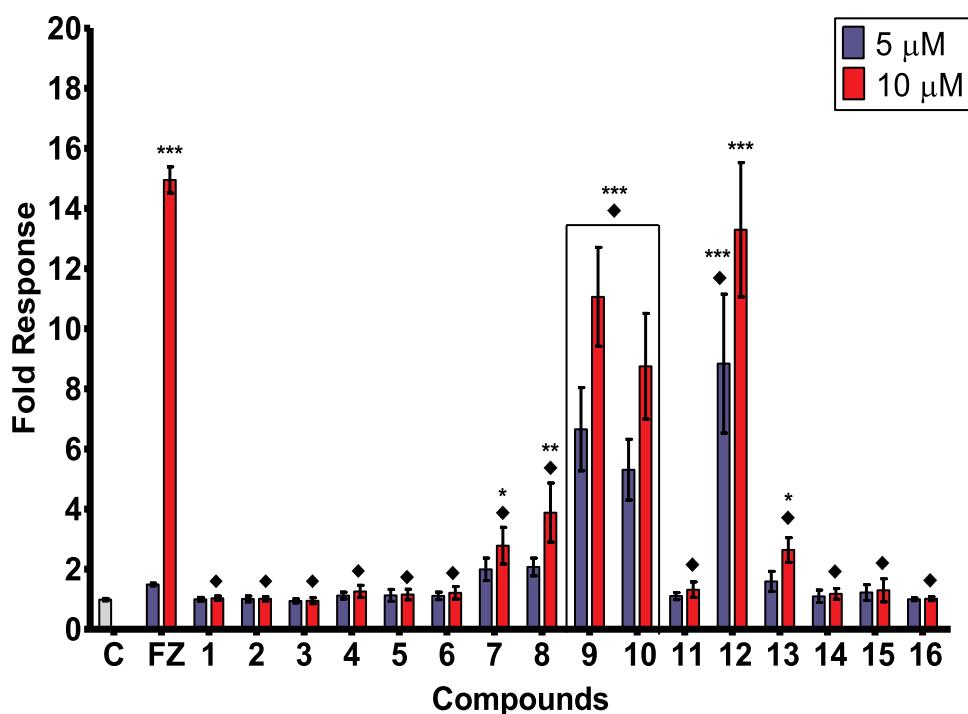
Figure 6. AhR agonistic activity (fold response) induced in AhR-HepG2 cells by the set of BTs (**1-16**). Bars represent means ± SD from at least four individual experiments (n = 4). The levels of significance were determined using one-way ANOVA followed by Dunnett's post-test when compared to vehicle control (**C**): ***$p < 0.001$, **$p < 0.01$ and *$p < 0.05$, and when compared with FICZ's effect ◆$p < 0.001$ at the same exposure concentration.

blockage capacity of AhR transcriptional activity, by the contrary, most of them displayed potential synergism when co-exposure with FICZ.

AhR agonism. Results obtained in the AhR agonist bioassay are represented in Figure 6 for the two assayed concentrations of the benzothiazoles, for the positive control FICZ (FZ) and for the vehicle control (C) of non-treated cells.

Overall, the set of BT derivatives exhibited a remarkable induction of AhR-mediated expression as displayed in Figure 6. Compounds **9, 10**, and **12** at both concentrations tested exhibited a very significant AhR agonist activity (p < 0.001) compared with the vehicle, while the positive control FICZ induced a significant activity only at the highest dosage. Moreover, compounds **7, 8**, and **13** showed a notable response when compared with non-treated cells.

The induction of AhR transcriptional activity of benzothiazole **12** at 10 µM was comparable with that displayed by the agonist FICZ, while the rest of the tested compounds exhibited effects significantly lower.

3.2. SAR considerations: antibiofilm activity vs. AhR agonism

The structural differences across the set of BTs and their influence on the antibiofilm activity and the AhR agonism were analyzed and summarized in Table 2.

As observed in Table 2, the AhR agonism was represented as a percentage of the activity displayed by FICZ (agonist efficiency). All benzothiazoles at 5 µM reached percentages higher than 50% of agonist efficiency. Compounds **7, 8**, and **13** exhibited a similar induction (~100%) of AhR transcriptional activity than the positive control at 5 µM, while **9, 10**, and **12** displayed AhR agonist efficiencies more than three times stronger than FICZ. At 10 µM, which is slightly above the EC$_{50}$ determined for FICZ (9.06 µM), only the benzothiazole **12** induced an effect comparable (89%) with the positive control followed by compounds **9** (74%) > **10** (66%) > **8** (22%) > **13** (21%). The rest of BTs induced below 20% of FICZ's response at the highest exposure concentration. In the anti-adhesion tests where the MIC of each compound was assayed against *S. aureus* and *P. aeruginosa*, the greatest inhibition percentages in the biofilm formation for *P. aeruginosa* were identified for **9** (87%) followed by **12** = **7** = **5** (85%). Compounds **12** (74%) and **14** (73%) showed the strongest eradication of *S. aureus* followed by **16** (72%) > **13** = **2** (71%). Therefore, based on the effects

displayed in both antibacterial and AhR agonist effects the following structure-activity considerations stand out:

1. AhR agonism does not seem to be determined by belonging to the 2-aryl or to the 2-pyridinylbenzothiazole subsets, noticeable when compared **1** *vs.* **2** or **5** *vs.* **6**, all of them below 10% of efficiency at 10 µM, and also when compared the strong inducers of AhR agonistic activity **9** *vs.* **10** whose efficiencies were no significantly different. Moreover, the stronger inducer of AhR-mediated effects was the arylbenzothiazole **12** followed by the pyridinylbenzothiazole **9**. Similarly, the biofilm formation of *S. aureus* and *P. aeruginosa* was not crucially determined by the aminophenyl or the aminopyridine substituent in position 2 of the benzothiazole, however, the aminopyridinyl derivative **6** exhibited lower inhibitory percentages in both anti-adherence and eradication tests when compared with its phenyl analogous **5**.

2. The *p,m*-disubstitutions with amino and nitro groups of 2-phenyl benzothiazoles, positively contributed to the AhR agonism and to the biocidal activity when compared **12** *vs.* **10**. In **8** the disubstitution significantly contributed to the AhR agonism compared to **5**, but non important differences were observed regarding the anti-adhesive or eradication effects. Lastly, the *p,m*-disubstitution positively contributed to the antibiofilm activity of **15** when compared with **13**, contrary to the AhR agonism, which was much higher in the monosubstituted **13** than in **15**.

3. The occurrence of AhR-mediated transcription was significantly higher for methyl substituents in the benzothiazole skeleton compared with all the corresponding analogous hydroxy, methoxy, nitro and amino for both 2-phenyl and 2-pyridinyl subseries (**10** *vs.* **1, 5, 13, 16**; **12** *vs.* **4, 8, 15**; **9** *vs.* **2, 6**; **7** *vs.* **3, 11, 14**).

3.3. Computational studies

3.3.1. ADMET and druglikeness profile

Benzothiazole **12** was selected by its attractive antimicrobial and AhR agonist activities to predict *in silico* its ADMET and druglikeness profile. Physicochemical and pharmacokinetic properties, as well as potential metabolic reactions predicted for **12**, are presented in Table 3.

As reported in Table 3, **12** showed acceptable physicochemical

Table 2
Summary of the AhR agonist efficiency and the biocidal activity exhibited by the sixteen benzothiazole derivatives.

BT	STRUCTURE	AhR agonist activity % Efficiency	Antibiofilm activity a. ☐ *S. aureus* b. ☐ *P. aeruginosa* % Anti-adhesion	% Eradication
01	HO–benzothiazole–C6H4–NH2	(5 μM / 10 μM)	a, b	a, b
02	HO–benzothiazole–pyridine(N)–NH2	(5 μM / 10 μM)	a, b	a, b
03	HO–benzothiazole–C6H4–NO2	(5 μM / 10 μM)	a, b	a, b
04	HO–benzothiazole–C6H4(NH2)(NO2)	(5 μM / 10 μM)	a, b	a, b
05	O–benzothiazole–C6H4–NH2	(5 μM / 10 μM)	a, b	a, b
06	O–benzothiazole–pyridine(N)–NH2	(5 μM / 10 μM)	a, b	a, b
07	CH3–benzothiazole–C6H4–NO2	(5 μM / 10 μM)	a, b	a, b
08	O–benzothiazole–C6H4(NH2)(NO2)	(5 μM / 10 μM)	a, b	a, b
09	CH3–benzothiazole–pyridine(N)–NH2	(5 μM / 10 μM)	a, b	a, b
10	CH3–benzothiazole–C6H4–NH2	(5 μM / 10 μM)	a, b	a, b
11	O–benzothiazole–C6H4–NO2	(5 μM / 10 μM)	a, b	a, b

parameters with not excessive lipophilicity (log P = 4.23), low solubility (Sw), dissociation (pKa) and diffusion in water (DiffC) standards as well as a not overly large structure as demonstrated by the molal volume (MolVol) of 266 cm^3/mol. High passive permeability (predicted Madin-Darby Canine Kidney (MDCK) model) and jejunal permeability (Peff) were predicted for **12**. The skin and cornea permeabilities were normal while the blood-brain barrier (BBB) penetration was identified as high.

In addition, Lipinski's rule of 5 revealed that **12** complied with molecular weight < 500, partition coefficient (logP) < 5, hydrogen bond donor ≤ 5, and hydrogen bond acceptors ≤ 10, suggesting favorable oral bioavailability for this arylbenzothiazole (data no shown).

Concerning the pharmacokinetic properties, only 1.06% was predicted as unlikely to be bond to blood plasma proteins in human (hum_fup%) with a blood to plasma ratio (RBP) of 0.12. The Volume of distribution (Vd) at steady state in humans was estimated as adequate and the Extended Clearance Classification System (ECCS) assessed a Class 2 to **12**. Such classification places metabolism as a primary clearance mechanism which is very common for drug-like compounds (Varma et al., 2015). The clearance parameters (CL_Metb, CL_Renal, CL_Uptake) corroborated the critical role of metabolism and the non-relevant contributions of the renal elimination nor the hepatic uptake for the studied molecule.

Compound **12** was predicted to be a substrate for several Phase I metabolizing enzymes of cytochrome P450 (CYP). The metabolic rate constants K$_m$ and V$_{max}$ were above 23 μM and below 8 nmol/min/nmol

Table 3

ADME-related physicochemical parameters, pharmacokinetic properties, and metabolic reactions predicted for benzothiazole **12**.

PHYSICOCHEMICAL										
logP[a]	pKa[b]	Sw[c]	DiffC[d]	MolVol[e]	MDCK[f]	Peff[g]	Skin[h]	Cornea[i]	BBB[j]	logBB[k]
4.23	3.73	4.79	8.29	266	5.45	4.90	1.25	1.48	High	0.15
PHARMACOKINETIC										
Hum_fup[l]	RBP[m]	V_d[n]	ECCS[o]	CL_Metb[p]	CL_Renal[q]	CL_Uptake[r]				
1.06%	0.12	0.31	Class 2	Yes (99%)	No (95%)	No (99%)				
METABOLISM										
Phase 1. Oxidation[s]	**CYP1A2**	**CYP2A6**	**CYP2B6**	**CYP2C8**	**CYP2C9**	**CYP2C19**	**CYP2D6**	**CYP2E1**	**CYP3A4**	
Substrate[t]	Yes	Yes	Yes	Yes	No	No	Yes	No	Yes	
Confidence[u]	88%	60%	55%	79%	81%	98%	82%	97%	83%	
Km[v]	23.02	-	-	-	-	-	63.02	-	55.73	
Vmax[w]	7.10	-	-	-	-	-	3.48	-	0.41	
CLint[x]	16.04	-	-	-	-	-	0.44	-	0.82	
Phase 2. Glucuronidation[y]	**UGT1A1**	**UGT1A3**	**UGT1A4**	**UGT1A6**	**UGT1A8**	**UGT1A9**	**UGT1A10**	**UGT2B7**	**UGT2B15**	
Substrate[t]	Yes	No	No	No	Yes	Yes	No	No	No	
Confidence[u]	90%	51%	72%	75%	87%	88%	86%	56%	98%	

[a] logP: octanol-water partition coefficient (lipophilicity)

[b] pKa: dissociation constant using submodels to predict protic ionization microconstants for all identified ionizable groups in a molecule.

[c] Sw (mg/mL $\times 10^{-4}$): native water solubility.

[d] DiffC (cm^2/s$\times 10^6$): molecular diffusion coefficient in water

[e] MolVol (cm^3/mol): molal volume at the normal boiling point

[f] MDCK (cm^2/s$\times 10^5$): apparent MDCK COS permeability

[g] Peff (cm^2/s$\times 10^4$): human effective jejunal permeability

[h] Skin (cm^2/s$\times 10^5$): permeability through human skin

[i] Cornea (cm^2/s$\times 10^5$): permeability through rabbit cornea

[j] BBB: qualitative likelihood (High/Low) of crossing the blood-brain barrier (98% confidence)

[k] logBB: logarithm of the brain/blood partition coefficient.

[l] Hum_fup: percent unbound to blood plasma proteins in human.

[m] RBP: blood-to-plasma concentration ratio in human.

[n] V_d (L/kg): volume of distribution in humans at steady state.

[o] ECCS Class: Extended Clearance Classification System (ECCS) assignment

[p] CL_Metab: predicts whether or not metabolism will be critical to clearance

[q] CL_Renal: predicts whether or not renal elimination will be critical to clearance.

[r] CL_Uptake: predicts whether or not hepatic uptake will be critical to clearance.

[s] Most common chemical reaction of Phase 1 of metabolism (oxidation) by Cytochromes P450 (CYP) enzymes

[t] Substrate: substrate classification models (yes/no) for human CYPxxx or UGTxxx

[u] Confidence of predictions

[v] Km (μM): estimated Michaelis-Menten Km constant for predicted sites of metabolism by human CYPxxx

[w] Vmax: [nmol/min/nmol enzyme]: estimated Michaelis-Menten V_{max} constant for predicted sites of metabolism by CYPxxx

[x] CLint (μL/min/mg human liver microsomes (HLM) protein): estimated intrinsic clearance for predicted sites of metabolism by CYPxxx

[y] Most common chemical reaction of Phase 2 of metabolism (glucuronidation) by uridine 5′-diphospho-glucuronosyltransferase (UGT)

Table 4

Human and environmental toxicological parameters predicted for benzothiazole **12**

HUMAN TOXICITY								
MaxRTD[a]	Plipidosis[b]	Developmental and genetic toxicity		Human liver adverse effects				
		Repro_Tox[c]	Mutagenesis[d]	AlkPhos[e]	GGT[f]	LDH[g]	AST[h]	ALT[i]
3.16	Nontoxic	Nontoxic	Negative	Normal	Normal	Normal	Elevated	Elevated
96%	99%	69%	99%	65%	97%	74%	85%	86%
ECOTOXICITY								
Bioconcn[j]	Biodegradn[k]	Daphnia_LC$_{50}$[l]	Minnow_LC$_{50}$[m]	Estrogenic[m]				
68.75	No (95%)	0.11	7.53	Nontoxic (98%)				

[n] Fathead minnow (*Pimephales promelas*) toxicity

[a] MaxRTD (mg/kg/day): qualitative assessment of the Maximum Recommended Therapeutic Dose administered as an oral dose.

[b] Plipidosis: qualitative estimation of potential for causing phospholipidosis.

[c] Repro_Tox: qualitative estimation of reproductive / developmental toxicity, (confidence %).

[d] Mutagenesis: classification model for the mutagenicity of pure compounds in *S. typhimurium* strain TA102 or *E. coli* strain WP2 uvrA, (confidence%).

[e] AlkPhos: human liver adverse effect as the likelihood of causing elevation in the levels of Alkaline Phosphatase enzyme, (confidence %).

[f] GGT: human liver adverse effect as the likelihood of causing elevation in the levels of gamma-glutamyl transferase (GGT) enzyme, (confidence %).

[g] LDH: human liver adverse effect as the likelihood of causing elevation in the levels of lactate dehydrogenase (LDH) enzyme, (confidence %).

[h] AST: human liver adverse effect as the likelihood of causing elevation in the levels of serum glutamic oxaloacetic transaminase (SGOT) enzyme, (confidence %).

[i] ALT: human liver adverse effect as the likelihood of causing elevation in the levels of serum glutamic pyruvic transaminase (SGTP) enzyme, (confidence %).

[j] Bioconcn (Cfish/Cwater): bioconcentration factor - partition coefficient between fish tissues and environmental water at steady state.

[k] Biodegradn: likelihood of biodegradation in the environment expressed as relative biological oxygen demand.

[l] Daphnia_LC$_{50}$ (mg/L): LC$_{50}$ for *Daphnia magna* (water flea) lethal toxicity after 48 hours of exposure.

[m] Estrogenic: classification model for predicting antiestrogen activity in rats (confidence %).

a)

- Binding Energy : 9.950 Kcal/mol

b)

- Binding Energy : 10.418 Kcal/mol
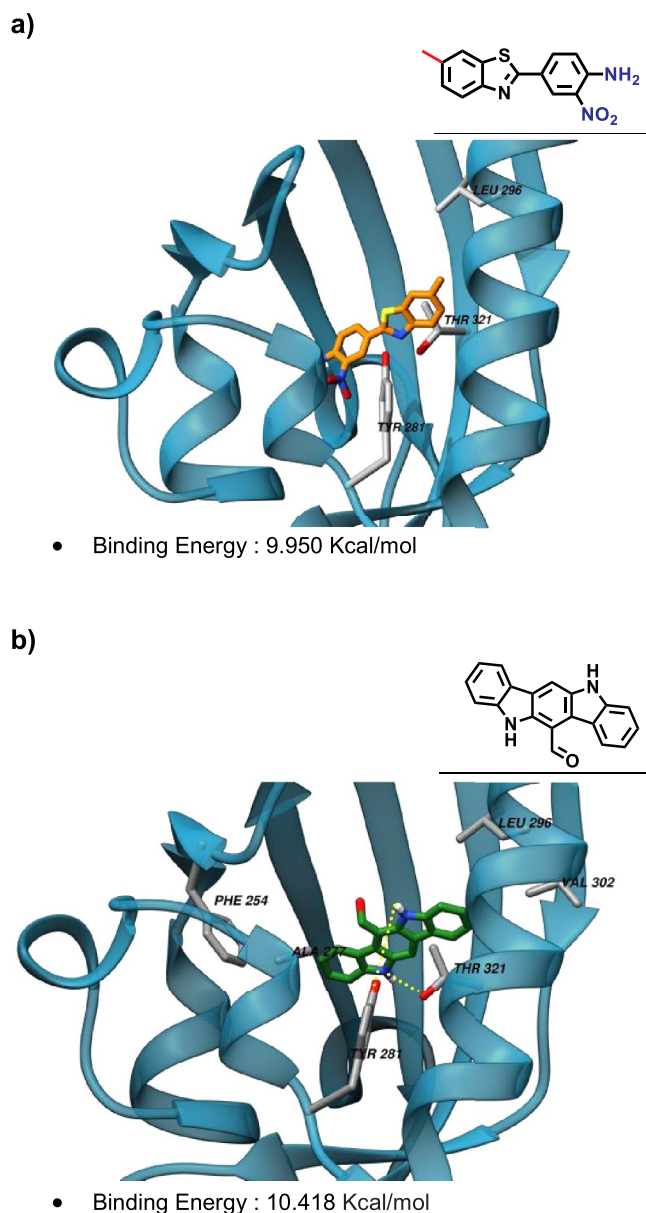
Figure 7. Molecular docking analyses for a) Benzothiazole **12** and b) **FICZ** with AhR-Ligand Binding Domain (PBD ID: 3F1O). <u>Left side:</u> the structure of the protein is represented as transparent blue ribbons and the best pose obtained for **12** and FICZ are displayed as sticks. The residues involved in hydrophobic interactions are labeled. Hydrogen bond interactions are represented in b) as yellow dot lines. <u>Right side:</u> the binding pocket residues are listed, in bold are highlight the matching between the two ligands.

enzyme, respectively. For CYP1A2, the lowest K_m and the highest V_{max} and intrinsic clearance (CLint) were predicted. Comparable K_m values were estimated for CYP2D6 and CYP3A4 (63.02 and 55.73 μM) and, while a higher V_{max} was predicted for CYP2D6 (3.48 *vs.* 0.41 nmol/min/nmol enzyme), a greater clearance was predicted for CYP3A4 (0.82 *vs.* 0.44 μL/min/mg human liver microsomes protein).

In Phase II of metabolism, **12** was predicted to be a substrate of three out of nine uridine 5′-diphospho-glucuronosyltransferase (UGT) isoforms studied. Thus, with predictive confidence above 87%, the arylbenzothiazole is probably metabolized in glucuronidation reactions mediated by UGT1A1, UGT1A8, and UGT1A9.

Finally, the global parameter ADMET_Risk was calculated in ADME Predictor software, which concisely summarizes the information from other properties and descriptors predicted to identify important

liabilities in drug candidates. The risk assessment is conducted by comparing with a reference set of 2270 commercial drugs from the World Drug Index (WDI) of which 90% have an ADMET_Risk score lower than 7. For **12**, a score of 3.75 was predicted as ADMET_Risk.

Toxicological indexes for humans and for ecosystems were also predicted for **12** (Table 4). Hence, the arylbenzothiazole was not predicted as a potential trigger of phospholipidosis nor as a reproductive or mutagenic agent. Regarding the hepatotoxicity, elevated levels of the two transaminases (AST and ALT) and normal levels of the rest of the enzymes were predicted. Therefore, **12** was not predicted to cause relevant harm to humans. Meanwhile, the ecotoxicological parameters *in silico* estimated for **12**, standing out the prediction of a poor biodegradability and an unlike probability of causing endocrine disruption via the estrogen receptor.

3.3.2. Molecular docking analysis

The binding interactions with the ligand binding domain of AhR were explored through molecular docking for **12** and for the known agonist FICZ and the outcome is presented in Figure 7 a) and b), respectively.

The binding energy with AhR, as well as the pocket residues identified in the molecular docking simulations, were similar for the arylbenzothiazole and for the ligand/agonist FICZ as represented in Figure 7. Furthermore, the THR 321 residue was predicted to interact hydrophobically with **12** and FICZ, while TYR 281 and LEU 296 seemed to establish hydrophobic interactions with **12** and hydrogen bonds with FICZ.

4. Conclusions

Ah receptor is an important chemical sensor that integrates dietary, environmental, metabolic, and microbial signals to regulate transcriptional programs in a context, cell type, and ligand-specific manner. The association of AhR with the immune system has been widely studied, particularly important in gut microbiota and barrier tissues, where crucial immune responses are related to AhR expression. The dual association of antimicrobial and AhR modulatory effects could be potentially beneficial in drug discovery. Hence, in this work, both effects were evaluated for a set of functionalized benzothiazoles yielding a promising activity for one derivative **12** as biocidal against *S. aureus* and as AhR agonist. Structure-activity analyses revealed general headlines on the substituent's contributions. Finally, computational studies with benzothiazole **12** predicted an adequate ADMET profile and a potentially similar binding to AhR when compared with the known agonist FICZ.

Abreviation

ADMET	absorption, distribution, metabolism, and excretion - toxicity
AhR	aryl hydrocarbon receptor
AhR-HepG2, Lucia™	human hepatoma cell line stably transfected to express AhR
BT	benzothiazole
CH223191	2-methyl-2H-pyrazole-3-carboxylic acid
CYP	cytochrome P450
Eff	efficiency
FA	fusidic acid
FICZ	5,11-dihydroindolo[3,2-b]carbazole-12-carbaldehyde
Lv	levofloxacin
MBC	Minimal Bactericidal Concentration
MIC	Minimal Inhibitory Concentration
MTT	3-(4,5-dimethyl thiazol-2-yl)-2,5-diphenyl tetrazolium bromide
OD	optical density

Author Contributions

All authors contributed to the drafting and revision of the article and approved the final version.

Credit Author Statement

E.G.J accomplished biological experiments to evaluated AhR expression ability, data curation, and structure-activity relationship analyses. F. A performed antibacterial, antifungal, and biocidal evaluation. L.E.C accomplished molecular docking studies. R. M. G designed and supervised biological studies. M.S-IV was responsible for chemicals supply, structure-activity and biological analysis and supervised the project. All authors contributed to the drafting and revision of the article and approved the final version presented.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

References

Ali, R., Siddiqui, N., 2013. Biological aspects of emerging benzothiazoles: A short review. J. Chem 2013. https://doi.org/10.1155/2013/345198.

Bessede, A., Gargaro, M., Pallotta, M.T., Matino, D., Servillo, G., Brunacci, C., Bicciato, S., Mazza, E.M.C., Macchiarulo, A., Vacca, C., Iannitti, R., Tissi, L., Volpi, C., Belladonna, M.L., Orabona, C., Bianchi, R., Lanz, T.V., Platten, M., Della Fazia, M.A., Piobbico, D., Zelante, T., Funakoshi, H., Nakamura, T., Gilot, D., Denison, M.S., Guillemin, G.J., Duhadaway, J.B., Prendergast, G.C., Metz, R., Geffard, M., Boon, L., Pirro, M., Iorio, A., Veyret, B., Romani, L., Grohmann, U., Fallarino, F., Puccetti, P., 2014. Aryl hydrocarbon receptor control of a disease tolerance defence pathway. Nature 511, 184–190. https://doi.org/10.1038/nature13323.

Bondock, S., Fadaly, W., Metwally, M.A., 2010. Synthesis and antimicrobial activity of some new thiazole, thiophene and pyrazole derivatives containing benzothiazole moiety. Eur. J. Med. Chem. 45, 3692–3701. https://doi.org/10.1016/j.ejmech.2010.05.018.

Bort, G., Sylla-Iyarreta Veitía, M., Ferroud, C, 2013. Straightforward synthesis of PET tracer precursors used for the early diagnosis of Alzheimers disease through Suzuki-Miyaura cross-coupling reactions. Tetrahedron 69, 7345–7353. https://doi.org/10.1016/j.tet.2013.06.085.

Choudhary, S., Jeyabalan, G., Kalra, N., 2017. A Review: Therapeutic and Biological Activity of Benzothiazole Derivatives. Int. J. Recent Adv. Sci. Technol. 4, 8–20. https://doi.org/10.30750/ijrast.432.

Esser, C., Rannug, A., 2015. The aryl hydrocarbon receptor in barrier organ physiology, immunology, and toxicology. Pharmacol. Rev. 67, 259–279. https://doi.org/10.1124/pr.114.009001.

Esser, C., Rannug, A., Stockinger, B., 2009. The aryl hydrocarbon receptor in immunity. Trends Immunol 30, 447–454. https://doi.org/10.1016/j.it.2009.06.005.

Ghosh, J., Lawless, M.S., Waldman, M., Gombar, V., Fraczkiewicz, R., 2016. Modeling ADMET. In: Benfenati, E. (Ed.), Silico Methods for Predicting Drug Toxicity. Springer, New York, New York, NY, pp. 63–83. https://doi.org/10.1007/978-1-4939-3609-0_4.

Gill, R.K., Rawal, R.K., Bariwal, J., 2015. Recent advances in the chemistry and biology of benzothiazoles. Arch. Pharm. (Weinheim). 348, 155–178. https://doi.org/10.1002/ardp.201400340.

Goya-Jorge, E., Giner, R.M., Veitía, M.S.-I., Gozalbes, R., Barigye, S.J., 2020. Predictive modeling of aryl hydrocarbon receptor (AhR) agonism. Chemosphere 256, 127068.

https://doi.org/10.1016/j.chemosphere.2020.127068.

Goya-Jorge, E., Rampal, C., Loones, N., Barigye, S.J., Carpio, L.E., Gozalbes, R., Ferroud, C., Veitía, M.S.-I., Giner, R.M., 2020. Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes. Eur. J. Med. Chem Submitted for publication.

Hergesheimer, R., Lanznaster, D., Vourc'h, P., Andres, C., Bakkouche, S., Beltran, S., Blasco, H., Corcia, P., Couratier, P., 2015. Advances in pharmacotherapy for the treatment of gout. Expert Opin. Pharmacother. 6566, 1–8. https://doi.org/10.1517/14656566.2015.997213.

Hu, W., Sorrentino, C., Denison, M.S., Kolaja, K., Fielden, M.R., 2007. Induction of Cyp1a1 is a nonspecific biomarker of aryl hydrocarbon receptor activation: Results of large scale screening of pharmaceuticals and toxicants in vivo and in vitro. Mol. Pharmacol. 71, 1475–1486. https://doi.org/10.1124/mol.106.032748.

Janeway, C., Travers, P., Walport, M., 2001. Manipulating the immune response to fight infection, in: Immunobiology: The Immune System in Health and Disease. Garland Science, New York.

Kamal, A., Syed, M.A.H., Mohammed, S.M., 2015. Therapeutic potential of benzothiazoles: A patent review (2010-2014). Expert Opin. Ther. Pat. 25, 335–349. https://doi.org/10.1517/13543776.2014.999764.

Keam, S.J., Croom, K.F., Keating, G.M., 2005. Levofloxacin. A review of its use in the treatment of bacterial infections in the United States. Drugs 63, 2769–2802. https://doi.org/10.2165/00003495-200565050-00007.

Krieger, E., Vriend, G., 2014. YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics 30, 2981–2982. https://doi.org/10.1093/bioinformatics/btu426.

Mathis, C.A., Wang, Y., Holt, D.P., Huang, G.F., Debnath, M.L., Klunk, W.E., 2003. Synthesis and evaluation of 11C-labeled 6-substituted 2-arylbenzothiazoles as amyloid imaging agents. J. Med. Chem. 46, 2740–2754. https://doi.org/10.1021/jm030026b.

Mohammadi-Bardbori, A., Omidi, M., Arabnezhad, M.R., 2019. Impact of CH223191-Induced Mitochondrial Dysfunction on Its Aryl Hydrocarbon Receptor Agonistic and Antagonistic Activities. Chem. Res. Toxicol. 32, 691–697. https://doi.org/10.1021/acs.chemrestox.8b00371.

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J. Immunol. Methods 65, 55–63. https://doi.org/10.1016/0022-1759(83)90303-4.

Moura-Alves, P., Faé, K., Houthuys, E., Dorhoi, A., Kreuchwig, A., Furkert, J., Barison, N., Diehl, A., Munder, A., Constant, P., Skrahina, T., Guhlich-Bornhof, U., Klemm, M., Koehler, A.B., Bandermann, S., Goosmann, C., Mollenkopf, H.J., Hurwitz, R., Brinkmann, V., Fillatreau, S., Daffe, M., Tümmler, B., Kolbe, M., Oschkinat, H., Krause, G., Kaufmann, S.H.E., 2014. AhR sensing of bacterial pigments regulates antibacterial defence. Nature 512, 387–392. https://doi.org/10.1038/nature13684.

Moyer, J.H., Ford, R.V, 1958. Laboratory and clinical observations on ethoxzolamide (Cardrase) as a diuretic agent. Am. J. Cardiol. 1, 497–504. https://doi.org/10.1016/0002-9149(58)90121-8.

Muthusubramanian, L., Rao, V.S.S., Mitra, R.B., 2001. Efficient synthesis of 2-(thiocyanomethylthio)benzothiazole. J. Clean. Prod. 9, 65–67. https://doi.org/10.1016/S0959-6526(00)00031-7.

Oh, S., Go, G.W., Mylonakis, E., Kim, Y., 2012. The bacterial signalling molecule indole attenuates the virulence of the fungal pathogen Candida albicans. J. Appl. Microbiol. 113, 622–628. https://doi.org/10.1111/j.1365-2672.2012.05372.x.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.

Ricco, C., Abdmouleh, F., Riccobono, C., Guenineche, L., Martin, F., Goya-Jorge, E., Lagarde, N., Liagre, B., Ali, M.Ben, Ferroud, C., Arbi, M.El, Veitía, M.S.-I., 2020. Pegylated triarylmethanes: Synthesis, antimicrobial activity, anti-proliferative behavior and in silico studies. Bioorg. Chem. 96, 103591. https://doi.org/10.1016/j.bioorg.2020.103591.

Salentin, S., Schreiber, S., Haupt, V.J., Adasme, M.F., Schroeder, M., 2015. PLIP: Fully automated protein-ligand interaction profiler. Nucleic Acids Res 43, W443–W447. https://doi.org/10.1093/nar/gkv315.

Shereen, M.A., Khan, S., Kazmi, A., Siddique, R., 2020. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J. Adv. Res. 24, 91–98. https://doi.org/10.1016/j.jare.2020.03.005.

Sommer, F., Bäckhed, F., 2013. The gut microbiota-masters of host development and physiology. Nat. Rev. Microbiol. 11, 227–238. https://doi.org/10.1038/nrmicro2974.

Trott, O., Olson, A.., 2009. Software News and Update. AutoDock Vina Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. J. Comput. Chem 31, 455–461. https://doi.org/10.1002/jcc.21334.

Varma, M.V., Steyn, S.J., Allerton, C., El-Kattan, A.F., 2015. Predicting Clearance Mechanism in Drug Discovery: Extended Clearance Classification System (ECCS). Pharm. Res. 32, 3785–3802. https://doi.org/10.1007/s11095-015-1749-4.

Zhao, H., Chen, L., Yang, T., Feng, Y.L., Vaziri, N.D., Liu, B.L., Liu, Q.Q., Guo, Y., Zhao, Y.Y., 2019. Aryl hydrocarbon receptor activation mediates kidney disease and renal cell carcinoma. J. Transl. Med. 17, 1–14. https://doi.org/10.1186/s12967-019-2054-5.

Chapter 4

CHAPTER 4. Study of triarylmethanes as AhR ligands

- **Publication Title:** "Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes" (https://doi.org/10.1016/j.ejmech.2020.112777).
- **Authors**: Elizabeth Goya-Jorge, Celine Rampal, Nicolas Loones, Stephen J. Barigye, Laureano E. Carpio, Rafael Gozalbes, Clotilde Ferroud, Maité Sylla-Iyarreta Veitía, Rosa M. Giner.

Summary

This fourth Chapter presents the synthetic procedures to obtain thirty-two novel unsymmetrical triarylmethanes, and the *in vitro* evaluation of these compounds as AhR agonists and antagonists. To our knowledge, this is the first study to assess the AhR modulatory effects in a diverse set of TAM class of compounds. The binding modes, molecular interactions, and ADME profile of the most prominent AhR agonist were analyzed *in silico*.

Highlights

- Series of novel triarylmethane compounds were straightforwardly synthesized.
- 32 triarylmethanes were screened in a cell-based bioassay as AhR modulators and 8 exhibited agonist activity.
- The introduction of heteroaromatic or naphthol substituents was favorable for AhR agonistic activity
- Adequate ADME profiles and druglikeness were predicted for the active AhR agonist compounds.
- Docking studies revealed similarities in AhR binding for the most potent TAM identified **22** (4-(pyridin-2-yl(thiophen-2-yl)methyl)phenol and the endogenous agonist ligand FICZ.
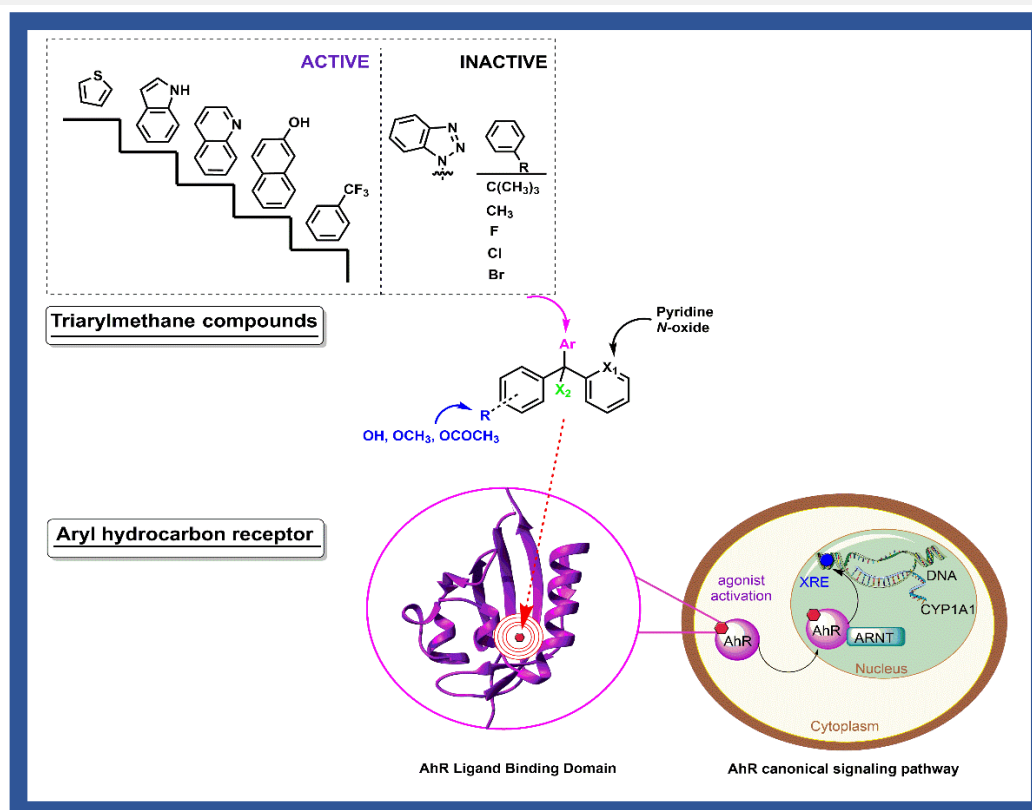
Targeting the aryl hydrocarbon receptor with a novel set of triarylmethanes☆

Elizabeth Goya-Jorge [a, b], Celine Rampal [c], Nicolas Loones [c], Stephen J. Barigye [b], Laureano E. Carpio [b], Rafael Gozalbes [b], Clotilde Ferroud [c], Maité Sylla-Iyarreta Veitía [c, *], Rosa M. Giner [a, **]

[a] Departament de Farmacologia, Facultat de Farmàcia, Universitat de València. Av. Vicente Andrés Estellés, S/n, 46100, Burjassot, Valencia, Spain
[b] ProtoQSAR SL, CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia, Av. Benjamin Franklin 12, 46980, Paterna, Valencia, Spain
[c] Equipe de Chimie Moléculaire Du Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (EA 7528), Conservatoire National des Arts et Métiers (Cnam), 2 Rue Conté, 75003, HESAM Université, Paris, France

ARTICLE INFO

ABSTRACT

The aryl hydrocarbon receptor (AhR) is a chemical sensor upregulating the transcription of responsive genes associated with endocrine homeostasis, oxidative balance and diverse metabolic, immunological and inflammatory processes, which have raised the pharmacological interest on its modulation. Herein, a novel set of 32 unsymmetrical triarylmethane (TAM) class of structures has been synthesized, characterized and their AhR transcriptional activity evaluated using a cell-based assay. Eight of the assayed TAM compounds (**14**, **15**, **18**, **19**, **21**, **22**, **25**, **28**) exhibited AhR agonism but none of them showed antagonist effects. TAMs bearing benzotrifluoride, naphthol or heteroaromatic (indole, quinoline or thiophene) rings seem to be prone to AhR activation unlike phenyl substituted or benzotriazole derivatives. A molecular docking analysis with the AhR ligand binding domain (LBD) showed similarities in the binding mode and in the interactions of the most potent TAM identified 4-(pyridin-2-yl (thiophen-2-yl)methyl)phenol (**22**) compared to the endogenous AhR agonist 5,11-dihydroindolo[3,2-b]carbazole-12-carbaldehyde (FICZ). Finally, *in silico* predictions of physicochemical and biopharmaceutical properties for the most potent agonistic compounds were performed and these exhibited acceptable druglikeness and good ADME profiles. To our knowledge, this is the first study assessing the AhR modulatory effects of unsymmetrical TAM class of compounds.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

The widely expressed and multifunctional aryl hydrocarbon receptor (AhR) protein is a ligand-activated, evolutionarily conserved and pleiotropic transcription factor. It is classified as a member of the basic helix—loop—helix (bHLH) family of receptors. The cytosolic and resting state of AhR is found in association with the chaperones heat shock protein 90 (Hsp 90), the immunophilin-like protein XAP2 (ARA9 or AIP) and p23 [1]. Although AhR is present in most tissues, its highest level of transcriptional activity is in cells of epithelial origin in the liver, kidney, lung and spleen [2].

The ligand binding domain (LBD) of AhR is allocated in the PAS-B [(PER)/AhR nuclear translocator (ARNT)/single-minded (SIM)] domain of the receptor. Once ligands arrive at the cytosolic locations of the receptor, they induce or inhibit the conformational modifications needed to prompt its nuclear translocation. If AhR is activated, the chaperone proteins are dissociated and its HLH domain forms a heterodimer with the nuclear translocator ARNT. The differential recognition of specific sequences in the promoter of downstream genes is determined by the recruitment of coactivators and corepressors, modulating thereby AhR expression. Such sequences of recognition are known as xenobiotic response elements (XRE) and they are identified by the core sequence 5′-GCGTG-3′ of the DNA [3]. Some XRE-independent mechanisms of AhR activation have been suggested on inflammatory and auto-immune conditions, particularly in selective hormone-sensitive

---

Abbreviations

ADME	absorption, distribution, metabolism, and excretion
AhR	Aryl hydrocarbon Receptor
AhR-HepG2	AhR-Lucia™ human liver carcinoma HepG2
ANOVA	(one-way) analysis of variance
bHLH	basic helix−loop−helix
BSD	bisacodyl
calcd	calculated
CH223191	2-methyl-2H-pyrazole-3-carboxylic acid
Cy	cyclohexane
CYP1A1	cytochrome P450 family 1 subfamily A polypeptide 1
DCM	dichloromethane
DCE	dichloroethane
DMSO	dimethyl sulfoxide
EC$_{50}$	half effective concentration
ER	estrogen receptor
FBS	fetal bovine serum
FCC	Flash Column Chromatography
FICZ	5,11-dihydroindolo[3,2-b]carbazole-12-carbaldehyde
GC-MS	Gas Chromatography-Mass Spectrometry
HPLC	High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectra
IC$_{50}$	half Inhibitory Concentration
LRMS	low-resolution mass spectra
m-CPBA	m-chloroperbenzoic acid
MEM	Minimum Essential Medium
MTT	3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide
NEAA	non-essential amino acids
NMR	Nuclear Magnetic Resonance
OECD	Organisation for Economic Co-operation and Development
PAS	PER, ARNT (AhR-nuclear translocator), Single-minded SIM
PBS	Phosphate Buffer Saline
PC	Positive Control
PTSA	p-toluenesulfonic acid
TAM	triarylmethane
THF	tetrahydrofuran
TLC	Thin Layer Chromatography
RPC$_{max}$	maximum response relative to the positive control
rt	room temperature
SAR	Structure-Activity Relationship
SEM	standard error of the mean
SERM	selective Estrogen Receptor modulators
XRE	xenobiotic response elements

cancer [4]. However, the main outcome of AhR expression is its canonical XRE-mediated signaling linked to the induction of xenobiotic metabolizing enzyme of the cytochrome P450 (CYP), in particular CYP1A1 from family 1, subfamily A, polypeptide 1 [5].

Several ligands have been identified as modulators of AhR including endogenous metabolites such 5,11-dihydroindolo[3,2-b]carbazole-12-carbaldehyde (FICZ) and indoxyl sulfate [6] as well as extensively used drugs such as omeprazole and leflunomide [7,8], and dietary phytocompounds such as quercetin [9]. While exact interaction patterns of different ligands upon binding with AhR still lack a completed crystallized structure of the receptor, important contributions are available for the LBD [10]. Moreover, vast studies of the toxic ligand/agonist of AhR known as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) has shed light on the activation mechanism as well as on the signaling patterns of the receptor [11,12]. The functional activity of AhR has proved to be determined by each specific ligand that binds to the LBD and that ultimately leads to dissimilar ligand- and AhR-dependent biological responses [13]. In general, aromatic or heteroaromatic hydrocarbon moieties are crucial structural determinants in all kinds of AhR modulators suggested to date [14−18].

AhR ligands are associated with key physiological processes such as proper development and metabolism, cell cycle regulation and immune defense [19]. Hence, while earlier perspectives focused on the function of AhR as xenobiotic sensor of toxicants like dioxins and polyaromatic hydrocarbons, recent suggestions placed AhR as an attractive pharmacological target [20−22]. Among the potential therapeutical uses of AhR modulation are included lung and vascular tissues health [23,24], treatment of liver and cystic fibrosis [25,26], control of the antioxidant response [27] and regulation of neural functions in both vertebrates and invertebrates [28]. Moreover, probably the most significant pharmacological applications of targeting AhR are in the treatment of several cancer types, in which the prodrug Phortress (NSC 710305) has been recommended as anticancerogenic and tumor suppressor chemotherapy for CYP1A1-positive tumors [4,29,30]. In addition, important inflammatory and immunological conditions could be modulated through AhR activation and particularly those affecting gut and intestinal tissues [31−33]. Hence, promising drug candidates such as NPD-0414-2 and NPD-0414-24 have been recently suggested in the pharmacotherapy of colitis [34]. AhR-mediated transcription converge with various nuclear receptor signaling pathway, mainly with the estrogen receptor (ER) [35]. Indeed, selective ER modulators (SERM) have been also identified as AhR ligands, which probably contributes to their therapeutical effects in postmenopausal osteoporosis and breast cancers [36]. Some SERMs identified hold the triarylmethane (TAM) skeleton [37]. Moreover, the symmetric TAM compound tris-indolyl methane was evaluated in a recent publication as a dual modulator of AhR and Pregnane X receptor (PXR) [38]. However, to the best of our knowledge, unsymmetrical TAM compounds have never been addressed as potential modulators of AhR.

The TAMs are privileged structures in medicinal chemistry [39]. Numerous TAM derivatives have found applicability in neurodegenerative diseases and vascular disorders and as anti-inflammatory, antitumoral and anti-infective agents against tuberculosis, human immunodeficiency virus and respiratory syncytial virus [40−42]. Notable examples are the TAM drug bisacodyl (BSD) and its analogs pointed out as anti-inflammatory, antimicrobial and antiproliferative agents [43,44], and the well-known antimycotic drug clotrimazole suggested in the antiproliferative and antiangiogenic pharmacotherapy [44,45].

Considering the aforementioned evidence that endows TAMs as an interesting scaffold in medicinal chemistry, added to the pharmacological relevance of targeting AhR [21], led to the hypothesis pursued herein. That is, TAM class of compounds could modulate AhR activation with potential therapeutic applicability in malignancies, immunological and inflammatory processes. Hence, novel TAMs were synthetized and their AhR-mediated transcriptional activity in AhR-HepG2 cells was assayed in vitro. The differential effects displayed by the set of compounds allowed to suggest theoretical contributions of the substituents in the AhR modulatory effects. ADME properties were predicted and the binding affinity preliminarily studied using computational methods for the most

significant AhR activators identified.

2. Results and discussion

2.1. Synthesis of triarylmethanes

The syntheses of triarylmethanes derivatives are shown in Schemes 1–4. Details about the synthetic protocol and chemical characterization of all intermediates are given in the Supplementary Information (SI-1).

The syntheses of the *p,p-N*-oxides **6a-e** and *o,p*-diarylmethylpyridines **7a-e** were carried out following the synthetic pathways represented in Scheme 1. First, synthesis of the corresponding carbinols **3a-e** was performed from 2-bromopyridine **1** and the corresponding aromatic aldehydes **2a-e**, by a bromine-lithium exchange following the procedure of Seto et al., 2004 [46] or by a bromine-magnesium exchange using isopropylmagnesium chloride in tetrahydrofuran at room temperature [47].

The key step to obtain the desired TAMs involved a regioselective Friedel-Crafts hydroxyalkylation of the corresponding carbinol **3a-e** with phenol in nitrobenzene under acidic activation [48]. The *p,p* regioisomers **4a-e** were obtained with 4 equivalents of sulfuric acid at 80 °C in a range of 33%–72% yield. The *o,p* compounds **7a-e** were obtained with 20 equivalents of catalyst at 0 °C in a range of 23%–98% yield. Acetates **5a-e** were obtained by treating the corresponding triarylmethanes derivatives with acetic anhydride in the presence of sodium hydroxide at room temperature. After workup, the desired compounds **5a-e** were isolated in a range of 76%–98% yield and pure enough to be used in the next step without any supplementary purification as suggested by the ¹H NMR analysis. *N*-oxide derivatives **6a-e** were prepared from the corresponding acetates by oxidation with *m*-chloroperbenzoic acid in dichloromethane at room temperature. After 2–3 h of reaction, *N*-oxide derivatives **6a-e** were isolated with prior purification by flash column chromatography (FCC) on silica gel with non-

optimized yields in a range of 38%–88%.

The syntheses of the benzotriazolyl triarylmethanes **10–12** were carried out following the synthetic pathways represented in Scheme 2.

First, the synthesis of the (4-methoxyphenyl) (pyridin-2-yl) methanol **9** was performed by a bromine-lithium exchange as previously described for compounds **3** from 2-bromopyridine 1 and *p*-anisaldehyde **8** in anhydrous tetrahydrofuran. Pyridylaryl-benzotriazol **10** was prepared from benzotriazole and the corresponding diarylmethanol **9** in the presence of a catalytic amount of *p*-toluenesulfonic acid in perfluorooctane (C_8F_{18}). In these conditions, the desired regioisomer **10** was obtained in 50% yield in high purity (HPLC, 95%). The regioisomer **10a** was also isolated and its characterization is described in the SI-1. An optimization of this procedure could probably improve the obtained yield.

The dimethoxylated compound **11** was synthesized by reaction with boron tribromide in dichloromethane. The reaction was conveniently carried out by mixing the reagents at 0 °C in an inert solvent and then allowing the mixture to warm up to room temperature during 6 h. Under these conditions the 4-((1H-benzo [d] [1–3]triazol-1-yl) (pyridin-2-yl)methyl)phenol **11** was obtained in 50% yield. The corresponding acetate derivative **12** was obtained by treating **11** with acetic anhydride in the presence of sodium hydroxide at room temperature. After workup and purification by FCC, the desired compound **12** was isolated in 52% yield.

The syntheses of the TAMs bearing heteroaromatic rings (naphthol, indole, quinoline or thiophene) are outlined in Scheme 3. Unsymmetrical naphthol (**14**, **15**), pyridylaryl indoles (**18,19**), and thiophene (**21**, **22**) were synthesized under acid conditions by condensation of the corresponding heterocycle with (4-methoxyphenyl) (pyridin-2-yl)methanol **9** previously obtained by a lithium-bromine exchange as described in Scheme 2. On the other hand, TAM **25** bearing a quinoline fragment was prepared from the corresponding aryl ketone **23** previously synthesized from the carbinol **9** in excellent yield (98%) *via* a base-promoted aerobic
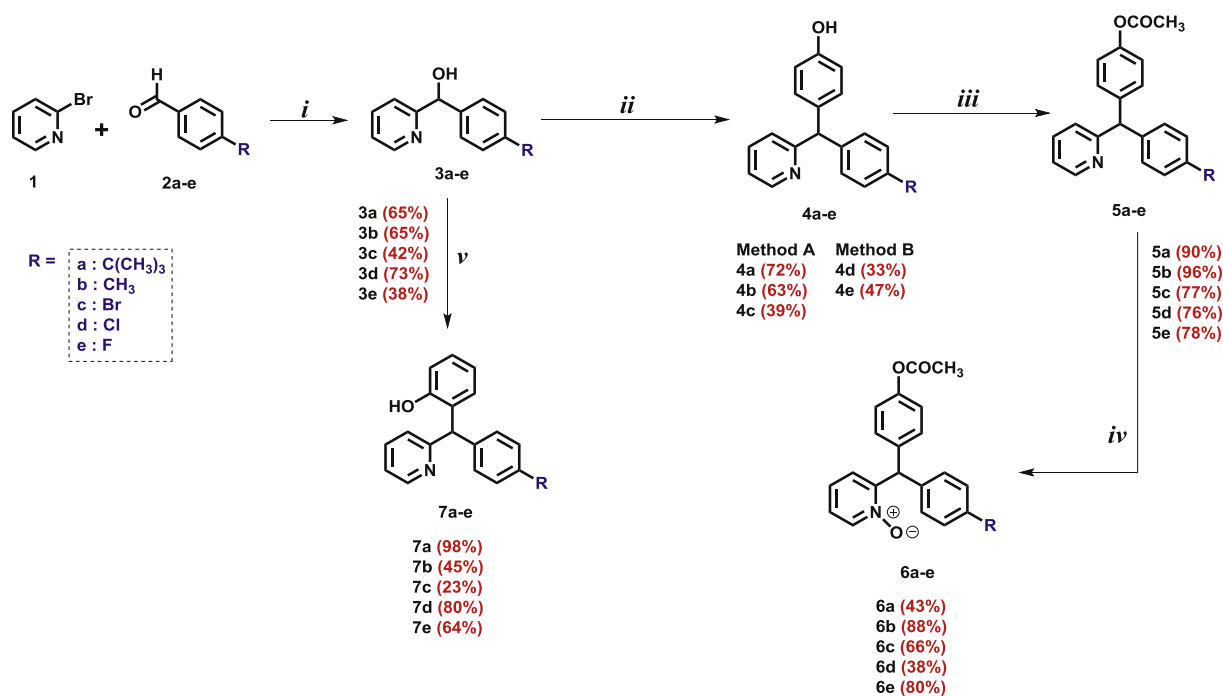
Scheme 1. Synthesis of *p,p-* and *o,p*-triarylmethanes. (*i*) *i*-PrMgCl (1 M) in 2-Me-THF, anh THF, 2 h, rt, Ar; or *n*-BuLi, anh THF, −78 °C/rt, Ar. (*ii*) phenol, H₂SO₄ (4 eq.), nitrobenzene, method A or B (A: 5 min at 80 °C, then at rt. B: from 0 °C to rt), Ar (*iii*) Ac₂O, NaOH, ≤15 h at 20 °C or 40 °C (*iv*) *m*-CPBA, anh DCM, 2 h at 20 °C. (*v*) phenol, H₂SO₄ (20 eq.), nitrobenzene at 80 °C, Ar, 5 min..
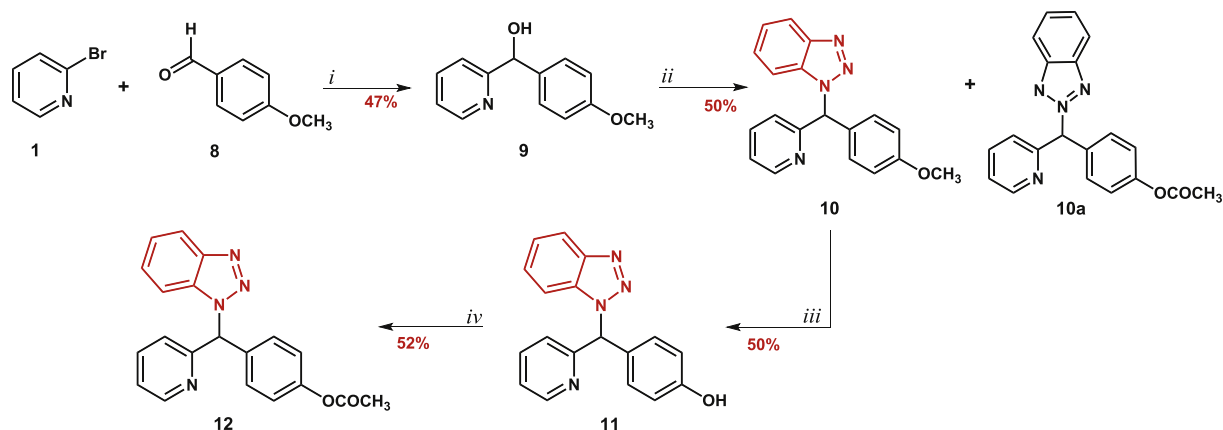
Scheme 2. Synthesis of benzotriazolyl triarylmethanes. (*i*) *n*-BuLi, anh THF, −78 °C/rt. (*ii*) benzotriazole, PTSA monohydrate, C_8F_{18}, 104 °C, 24 h (*iii*) BBr$_3$, DCM, 6 h, from 0 °C to rt., Ar (*iv*) Ac$_2$O, NaOH, 24 h, from 0 °C to rt.
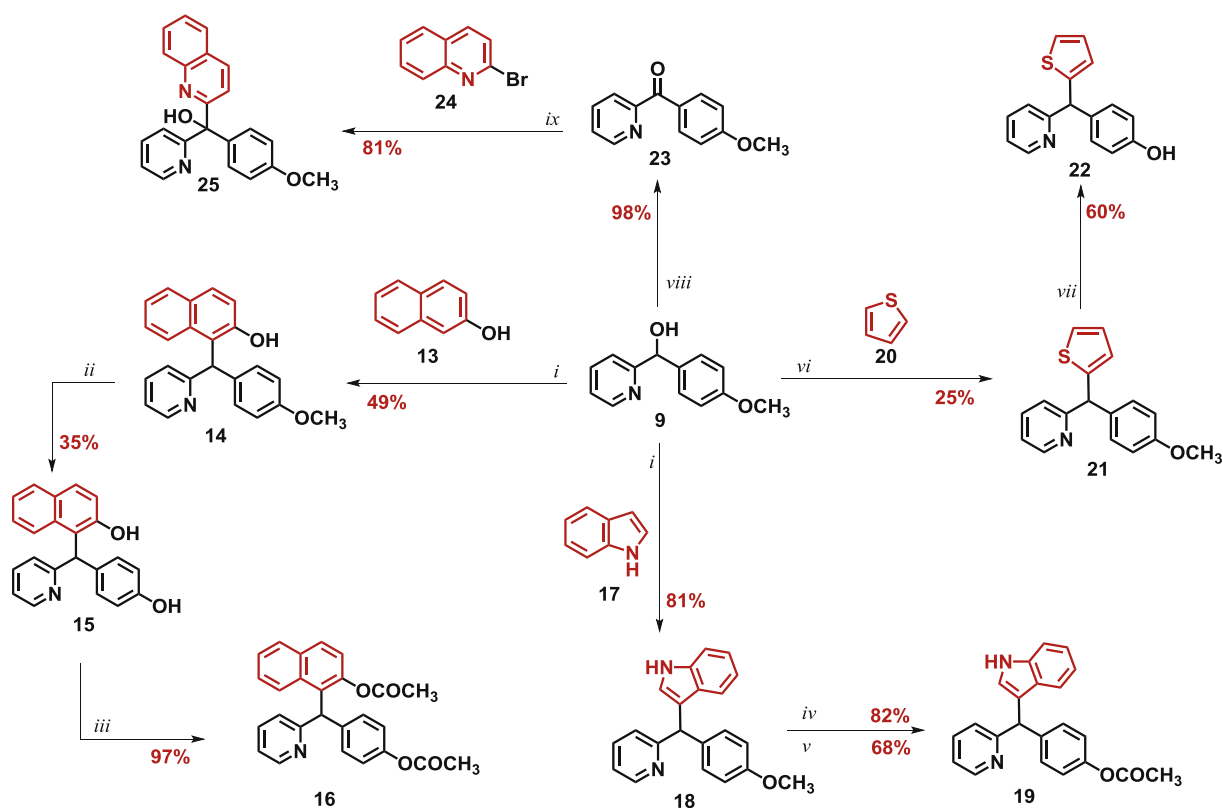
Scheme 3. Synthesis of naphthol, indole, thiophene and quinoline triarylmethanes. (*i*) NH$_2$SO$_3$H, DCE, 20 h, at 85 °C, Ar (*ii*) HI, AcOH, 5.5 h at 100 °C, Ar (*iii*) Ac$_2$O, NaOH, 24 h, from 0 °C to rt. (*iv*) BBr$_3$, DCM, 19 h, from 0 °C to rt (*v*) **18a**, Ac$_2$O, NaOH, 3.5 h, from 0 °C to rt. (*vi*) CH$_3$SO$_3$H, DCE, microwave irradiation 2 h at 80 °C (*vii*) BBr$_3$, DCM, 19 h, from 0 °C to rt., Ar (*viii*) O$_2$, NaOH, toluene at 110 °C (*ix*) prior mix of *n*-BuLi and **24** in anh THF, 1.5 h at −78 °C, Ar, then **23** in dry THF, 17 h at rt.

oxidation using air as a free and clean oxidant [49]. The other desired TAMs (**16**, **19**) were prepared using sequence series of including methoxy group deprotection followed by acylation as indicated conditions in Scheme 3.

The synthesis of the TAMs bearing trifluoromethyl group was carried out following the synthetic pathways represented in Scheme 4. The synthesis of the pyridin-2-yl (4-(trifluoromethyl) phenyl)methanol **26** and the corresponding arylketone **27** was performed following the same procedure described in Scheme 1. Then, TAM **28** was obtained by a halogen-metal exchange from 4-bromoanisole in 67% yield. Demethoxylation was conveniently

carried out with hydroiodic acid in acetic acid at reflux. Under these conditions the 4-(pyridin-2-yl (4-(trifluoromethyl)phenyl)methyl) phenol **29** was obtained in 82% yield. The acetate derivative **30** was obtained by treating **29** with acetic anhydride in presence of sodium hydroxide at room temperature. After workup and purification by FCC, the desired compound 4-(pyridin-2-yl (4-(trifluoromethyl)phenyl) methylphenyl acetate **30** was isolated in 87% yield.

All compounds biologically evaluated were obtained in high purity (HPLC or NMR, generally > 95%).
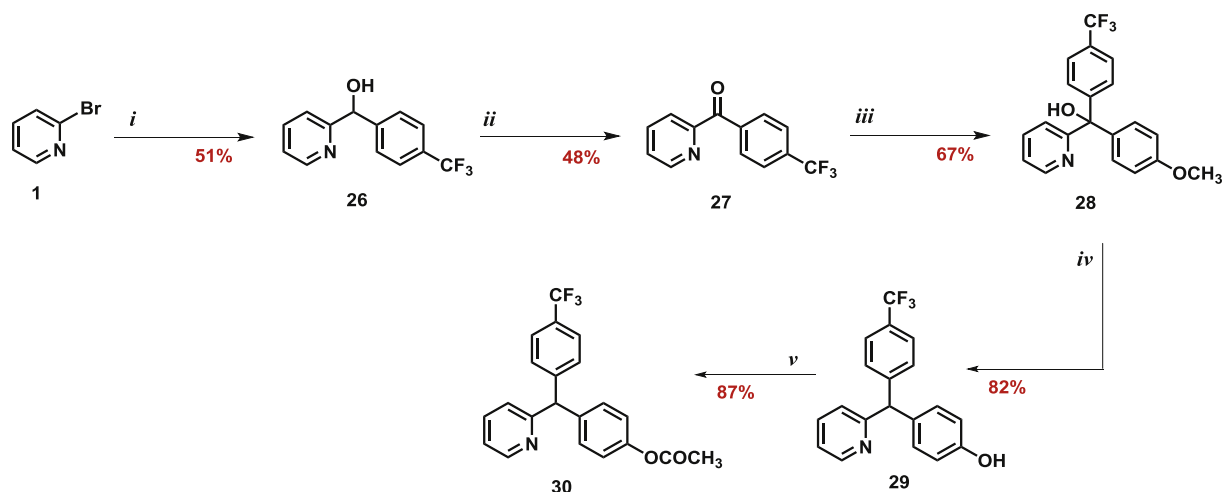
Scheme 4. Synthesis of TAMs bearing trifluoromethyl group. (*i*) prior mix of *n*-BuLi and 2-bromopyridine in anh THF at −78 °C, Ar, then 4-(trifluoromethyl)benzaldehyde, 17 h at rt. (*ii*) NaOH, O₂, toluene, reflux, 24 h (*iii*) prior mix of *n*-BuLi and 4-bromoanisole in anh THF at −78 °C, Ar and then **27** (*iv*) HI 57%, AcOH, reflux, Ar (*v*) Ac₂O, NaOH, 4 h, from 0 °C to rt.

2.2. Biological evaluation

2.2.1. Cell viability

The effects on cell viability caused by the synthetized TAMs on AhR-HepG2 cell line were determined by the 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide (MTT) assay, which constitutes a valuable method to study cell proliferation, cytotoxicity and chemosensibility *in vitro* [50]. The cell viability percentages obtained for 32 novel TAMs and the drug BSD are shown in Fig. 1. The adopted criterium considered as cytotoxic was a reduction of cell viability above 15% upon treatment with TAMs.

The eight TAMs **14**, **15**, **18**, **19**, **21**, **22**, **25** and **28** were studied in more detail due to their AhR agonist effects as will be described later. For these compounds, no cytotoxic effect was observed at the six assayed concentrations. The same conditions but longer
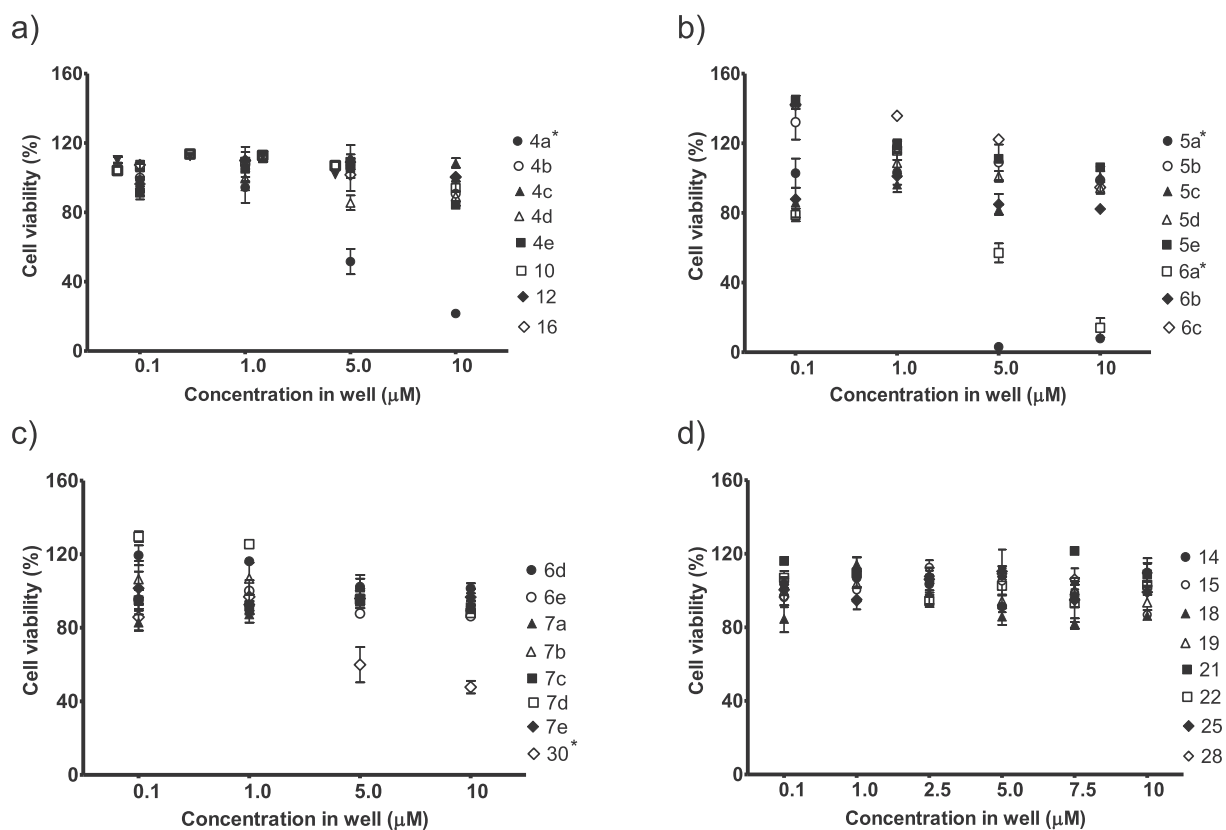
Fig. 1. Viability percentages of cells exposed to the TAMs by MTT assay. a) Compounds **4 a-e**, **10**, **12**, **16**, b) Compounds **5 a-e** and **6 a-c**, c) Compounds **6d**, **6e**, **7 a-e** and **30**, d) Compounds **14**, **15**, **18**, **19**, **21**, **22**, **25** and **28**. All compounds were assayed at 0.1 μM, 1.0 μM, 5.0 μM and 10.0 μM, and compounds in d) were also tested at 2.5 μM and 7.5 μM. Each chart represents the mean percentage ± SEM from at least three independent experiments (n = 3). Cell viability lower than 85% was considered cytotoxic. *p < 0.05 significantly different from vehicle control (using one-way ANOVA followed by Dunnett's post-test).

exposure times (48 h and 72 h) in the MTT test revealed similar results (data not shown) than those obtained after 24 h of treatment (Fig. 1d).

The TAMs **4a**, **5a**, **6a** and **30** were found to be cytotoxic over 5 μM. The strongest cytotoxicity was caused by compound **4a**, whose reduction of cell viability was above 90%, while comparable reductions (~50%) of cell viability were observed for compounds **5a**, **6a** and **30**. The concentration limits established in the AhR transcriptional activity bioassay only considered no cytotoxic concentrations of the tested TAMs.

2.2.2. AhR transcriptional activity

Standard Positive Controls. During the validation and optimization of the *in vitro* method, sigmoidal dose-response curves of the endogenous AhR agonist FICZ were obtained with concentrations from 0.01 to 18.0 μM ($R^2 = 0.99$), according to the recommendations of the AhR-HepG2 cell line provider (See SI-2). Similarly, the antagonist bioassay was validated with the inhibition curve of the AhR antagonist 2-methyl-2H-pyrazole-3-carboxylic acid (CH223191), obtained by co-exposure the EC_{50} of FICZ and concentrations from 1 to 30 μM of the CH223191 as described elsewhere [51].

The maximum induction of AhR transcription caused by FICZ was up to 30 folds at the maximum concentration tested (18 μM). From the dose-response curve of AhR agonist, the estimated EC_{50} of FICZ in the cell model was 9.06 μM. Meanwhile, in presence of FICZ concentration at the EC_{50}, the antagonist compound CH223191 reduced to half the transcriptional activity of AhR at the maximum concentration tested (30 μM). From the dose-response curve of AhR inhibition, the estimated IC_{50} of CH223191 was 2.43 μM, consistent with data reported in the literature [52].

TAM compounds. The results of AhR agonist and AhR antagonist assays are presented in Table 1 for all the studied compounds, that include the 32 novel TAMs synthetized, the commercial TAM drug BSD and the positive controls. Dose response-curves are provided as SI-2.

In the AhR reporter gene assay, the maximum effect observed corresponded to the maximum concentration tested for all TAMs except for **15** and **18**. Compounds **4a**, **5a**, **6a** and **30** were assayed at non-cytotoxic concentrations (up to 1 μM) and they were unable to induce or blockage AhR transcriptional activity in this cell model.

AhR agonist assay. The eight TAMs **14**, **15**, **18**, **19**, **21**, **22**, **25** and **28** were identified as agonists of AhR (RPC_{max}> 10%) while the rest of them and the BSD were classified as inactive AhR agonists (Table 1). The agonist effectiveness of the active compounds compared to FICZ followed the order: **18** ≈ **22** > **19** > **25** > **14** > **28** ≈ **15** > **21**. Compounds **18** and **22** were more active as agonist than FICZ at a comparable exposure concentration showing an RPC_{max} of 114.2% and 111.0%, respectively.

In cases where a dose-response curve of agonism was achieved, the half effective concentration (EC_{50}) was estimated as a measure of the potency of active compounds. As reported in Table 1, compound **22** ($EC_{50} = 13.16$ μM) was suggested as the most potent AhR agonist. Compounds **25** and **18** showed comparable half effective concentrations of 19.88 μM and 21.72 μM, respectively, while compound **19** was less potent ($EC_{50} = 27.86$ μM). The EC_{50} estimated for compounds **14** and **28** were 53.62 μM and 52.03 μM, respectively. Lastly, it was not possible to obtain a dose-response curve for compounds **15** nor **21**.

AhR antagonist assay. None of the tested compounds showed antagonist effects on AhR activation. However, additive effects and probably synergism in presence of FICZ were observed for most of the agonist compounds. Interestingly, the level of AhR transcriptional response in presence of FICZ was not proportional to the effectiveness of compounds individually tested as agonists. Thus, a

remarkable high induction during the co-exposure was registered for compounds **25** ($RPC_{max} = 272.13$%) and **28** ($RPC_{max} = 242.95$%) despite that they were not by themselves among the strongest activators of AhR. Similarly, a notable induction of AhR activation in presence of FICZ during the antagonist assay was showed by the inactive compounds **5b**, **4d** and **7a** (RPC_{max} 203.94%, 200.45% and 196.65%, respectively).

2.3. SAR considerations

The AhR-mediated transactivation induced by the TAMs allowed a comprehensive structure-activity relationship (SAR) analysis. The key structural features of the active TAMs and their AhR agonist effects expressed as fold responses are shown in Fig. 2.

X1, X2 substitution. Regarding X_1, the presence of a pyridine ring (**14**, **15**, **18**, **19**, **21**, **22**, **25**, **28**) was important for AhR agonist activity similarly to some other compounds reported *in vivo* as CYP1A1 inducers [54]. The *N*-oxidation had no influence on AhR activation, noticeable when comparing the TAMs **5a-e** *vs.* the corresponding *N*-oxides **6a-e**. None of the substituted phenyls at **Ar** showed any effect on AhR except for **28** in which the introduction of a hydroxyl group at X_2 position turned it moderately active. Compound **25** bearing a quinoline and a hydroxyl group at X_2 also exhibited significant AhR agonist induction.

R substitution. The presence of different oxygenated functional group at **R** does not appear to determine the agonist effects on AhR of the synthetized TAMs as it has been reported for other aromatic compounds in the literature [55]. Most of the active compounds in this study carry a hydroxyl or methoxy group at **R**. The methoxy substitution in most cases was a better feature to exhibit AhR agonist activity (**14** *vs.* **15** and **18** *vs.* **19**). However, the free hydroxyl group in **22** resulted in a 30-fold increase of the AhR agonism when compared with the methoxylated derivative **21**.

Ar substitution. The AhR-agonist activity was crucially influenced by the third aromatic or heteroaromatic system occupying **Ar**. Thus, the most potent AhR agonism was exhibited by TAMs with heteroaromatic moiety such as thiophene (**22**), indole (**18**, **19**) and quinoline (**25**). Otherwise, derivatives with a naphthol substituent (**14**, **15**) displayed some AhR agonist ability although considerably weaker than the rest of heteroaromatic derivatives (except for **21**) as shown in Fig. 2. Curiously, compounds bearing a benzotriazole moiety (**10**, **12**) were found to be inactive. It should be notice that even though **15** and **21** were considered agonists according to the RPC_{max} threshold (Table 1), their induced fold response was not significant compared to the vehicle control as shown in Fig. 2.

On the other hand, the introduction of heteroaromatic substituent at **Ar** were in no case harmful to cells according to the cell viability study (Fig. 1). Most of the substituted phenyl derivatives at **Ar** that caused cytotoxicity at the highest concentrations tested (**4a**, **5a**, **6a**) bear a tertbutyl functional group.

Consistent to the above results, phloroglucinol TAMs have shown better safety index and *anti*-HIV effects when bearing a heteroaromatic moiety [42]. Additionally, indole-containing chemicals have long been recognized as AhR ligands from endogenous and dietary sources, sustaining the strong agonism displayed by **18** and **19** [19,56]. Although the thiophene ring in TAM-class of compounds has been suggested as an attractive moiety for antimycobacterial activity [57], it is not commonly found in either classical or nonclassical AhR modulators identified to date [58]. Thus, to our knowledge, the agonist effects on AhR transcriptional activity of thiophene derivatives are suggested herein for the first time.

Table 1

AhR-mediated transcriptional activity of the 32 rationally designed TAMs **4**–**7** [**a**-**e**], **10**, **12**, **14**–**16**, **18**, **19**, **21**, **22**, **25**, **28**, **30**, the drug bisacodyl and the agonist (FICZ) and antagonist (CH223191) controls.

ID	Ar	R	X_1	X_2	[µM][a]	Agonist RPC (%) ±SEM[b]	EC_{50} (µM) ±SEM[c]	Antagonist RPC (%) ±SEM[b]	Activity Criteria[d]
	AhR-HepG2 transcriptional activity								
4a	Phe (4-C(CH₃)₃)	OH	N	H	1.0	6.89 ± 0.29	ND	113.44 ± 3.16	Inactive
4b	Phe (4-CH₃)	OH	N	H	10.0	7.09 ± 0.35	>100	138.98 ± 4.72	Inactive
4c	Phe (4-Br)	OH	N	H	10.0	7.71 ± 0.56	>100	168.82 ± 7.81	Inactive
4d	Phe (4-Cl)	OH	N	H	10.0	6.59 ± 0.37	>100	200.45 ± 7.41	Inactive
4e	Phe (4-F)	OH	N	H	10.0	7.79 ± 0.52	>400	150.92 ± 7.29	Inactive
5a	Phe (4-C(CH₃)₃)	OCOCH₃	N	H	1.0	5.32 ± 0.29	ND	112.56 ± 2.01	Inactive
5b	Phe (4-CH₃)	OCOCH₃	N	H	10.0	9.42 ± 0.46	>100	203.94 ± 4.05	Inactive
5c	Phe (4-Br)	OCOCH₃	N	H	10.0	6.32 ± 0.13	>1000	143.15 ± 6.44	Inactive
5d	Phe (4-Cl)	OCOCH₃	N	H	10.0	6.26 ± 0.10	>1000	128.16 ± 5.01	Inactive
5e	Phe (4-F)	OCOCH₃	N	H	10.0	8.77 ± 0.33	>100	118.36 ± 5.34	Inactive
6a	Phe (4-C(CH₃)₃)	OCOCH₃	N⁺O⁻	H	1.0	6.10 ± 0.23	ND	120.3 ± 2.55	Inactive
6b	Phe (4-CH₃)	OCOCH₃	N⁺O⁻	H	10.0	6.28 ± 0.18	>200	93.35 ± 1.09	Inactive
6c	Phe (4-Br)	OCOCH₃	N⁺O⁻	H	10.0	8.80 ± 0.52	>100	106.22 ± 3.50	Inactive
6d	Phe (4-Cl)	OCOCH₃	N⁺O⁻	H	10.0	7.78 ± 0.28	>200	78.81 ± 3.89	Inactive
6e	Phe (4-F)	OCOCH₃	N⁺O⁻	H	10.0	6.51 ± 0.31	>1000	92.41 ± 2.75	Inactive
7a	Phe (2-OH)	(CH₃)₃C	N	H	10.0	6.81 ± 0.19	>1000	196.65 ± 7.59	Inactive
7b	Phe (2-OH)	CH₃	N	H	10.0	6.93 ± 0.15	>100	114.43 ± 6.55	Inactive
7c	Phe (2-OH)	Br	N	H	10.0	7.99 ± 0.17	>1000	117.10 ± 3.45	Inactive
7d	Phe (2-OH)	Cl	N	H	10.0	7.18 ± 0.16	>1000	114.57 ± 5.72	Inactive
7e	Phe (2-OH)	F	N	H	10.0	6.89 ± 0.14	>1000	111.59 ± 3.21	Inactive
10	1H-benzotriazole	OCH₃	N	H	10.0	8.86 ± 0.26	>100	163.53 ± 7.29	Inactive
12	1H-benzotriazole	OCOCH₃	N	H	10.0	9.23 ± 0.24	>100	141.94 ± 6.44	Inactive
14	1-Naph (2-OH)	OCH₃	N	H	10.0	39.70 ± 0.76	53.62 ± 0.22	142.31 ± 2.78	**Agonist**
15	1-Naph (2-OH)	OH	N	H	5.0	28.57 ± 0.51	>50	168.86 ± 5.74	**Agonist**
16	1-Naph (2-OCOCH₃)	OCOCH₃	N	H	10.0	5.17 ± 0.25	>200	143.80 ± 1.99	Inactive
18	3-indole	OCH₃	N	H	10.0	114.20 ± 0.80	21.72 ± 0.32	182.99 ± 5.12	**Agonist**
19	3-indole	OCOCH₃	N	H	10.0	76.95 ± 0.63	27.86 ± 0.15	226.23 ± 7.45	**Agonist**
21	2-thiophene	OCH₃	N	H	10.0	10.54 ± 0.41	>50	136.57 ± 2.70	**Agonist**
22	2-thiophene	OH	N	H	10.0	111.01 ± 0.86	13.16 ± 0.08	202.27 ± 4.42	**Agonist**
25	2-quinoline	OCH₃	N	OH	10.0	53.78 ± 0.37	19.88 ± 0.08	272.13 ± 5.51	**Agonist**
28	Phe(4-CF₃)	OCH₃	N	OH	10.0	26.60 ± 0.89	52.03 ± 0.29	242.95 ± 5.42	**Agonist**
30	Phe(4-CF₃)	OCOCH₃	N	H	1.0	7.71 ± 0.22	ND	117.24 ± 3.81	Inactive
BSD	Phe(4-OCOCH₃)	OCOCH₃	N	H	10.0	6.58 ± 0.19	>1000	107.20 ± 3.03	Inactive
Control	FICZ	—			18.0	100%	9.06 ± 0.02	—	Agonist
Control	CH223191	—			30.0	—	(IC_{50})2.43 ± 0.18	54.50 ± 0.79	Antagonist

[a] Maximum concentration tested in the absence of limitations due to cytotoxicity or insolubility.

[b] Average of the percentages of the maximum response relative to the positive control (RPCmax) ± SEM of AhR agonist/antagonist activity from at least three independent experiments (n = 3).

[c] Estimated half effective concentration (EC_{50}) for agonists or half inhibitory concentration (IC_{50}) for antagonist ± SEM, all extrapolated from the dose-response curve. ND: non-determined.

[d] Activity criteria for AhR agonist (>10%) or antagonist (<70%) compounds [53].

2.4. Computational studies

2.4.1. Molecular docking

The binding to AhR of the strongest TAM agonist identified (**22**) was compared by means of molecular docking analysis with the known ligand/agonist compounds FICZ and TCDD. In the absence of a crystalized structure of AhR-LBD, the structurally related PAS-B domain of HIF2α was used for molecular docking analysis (details are provided as SI-3). The best poses obtained for the three ligands during the docking simulations are represented in Fig. 3.

The three docked ligands (**22**, FICZ and TCDD) seem to concurr in the internal cavity of the crystallized PAS-B heterodimer as expected [10]. While **22** and FICZ were predicted to bound with AhR-LBD in a similar region of the cavity, the best pose estimated for TCDD binding seems to lie in a different region as shown in Fig. 3. Details on the residues involved in the molecular docking for each ligand are provided in Supplementary Information (SI-3). Furthermore, the hydrophobic interactions in the predicted protein/ligand complex for **22** as well as for the endogenous agonist FICZ shared PHE 254 and ALA 277 residues. Π-Cation interactions with HIS 248, hydrogen bond with TYR 281 and THR 321, and halogen bond with

GLU 320 were also identified for **22**, FICZ and TCDD, respectively as represented in Fig. 3. The obtained results suggested differences in the predictive binding for these three ligands that could ultimately lead to distinct biological responses [13]. A comparison between the binding energies and interactions modes of compounds **22** and **21** did not provided plausible rationalization for the remarkable activity differences *in vitro* identified (see Section 2, SI-3).

On the other hand, AhR ligands often modulate other transcription factors, particularly nuclear receptors [59,60]. Therefore, as a preliminary off-targeting screening, the binding capacity of compound **22** in ER, Androgen Receptor (AR), Progesterone Receptor (PR) and Pregnane X Receptor (PXR) was analyzed by molecular docking. A comparison between **22** and well-known ligands of each receptor did not reveal any apparent binding resemblances. Although the binding energies of **22** were in most cases similar to those exhibited by the specific ligand for each receptor, the pocket and binding sites were different in all cases (details are provided in Section 4, SI-3).

2.4.2. Druglikeness and ADME profile

Characterizing the druglikeness as well as the bioavailability are
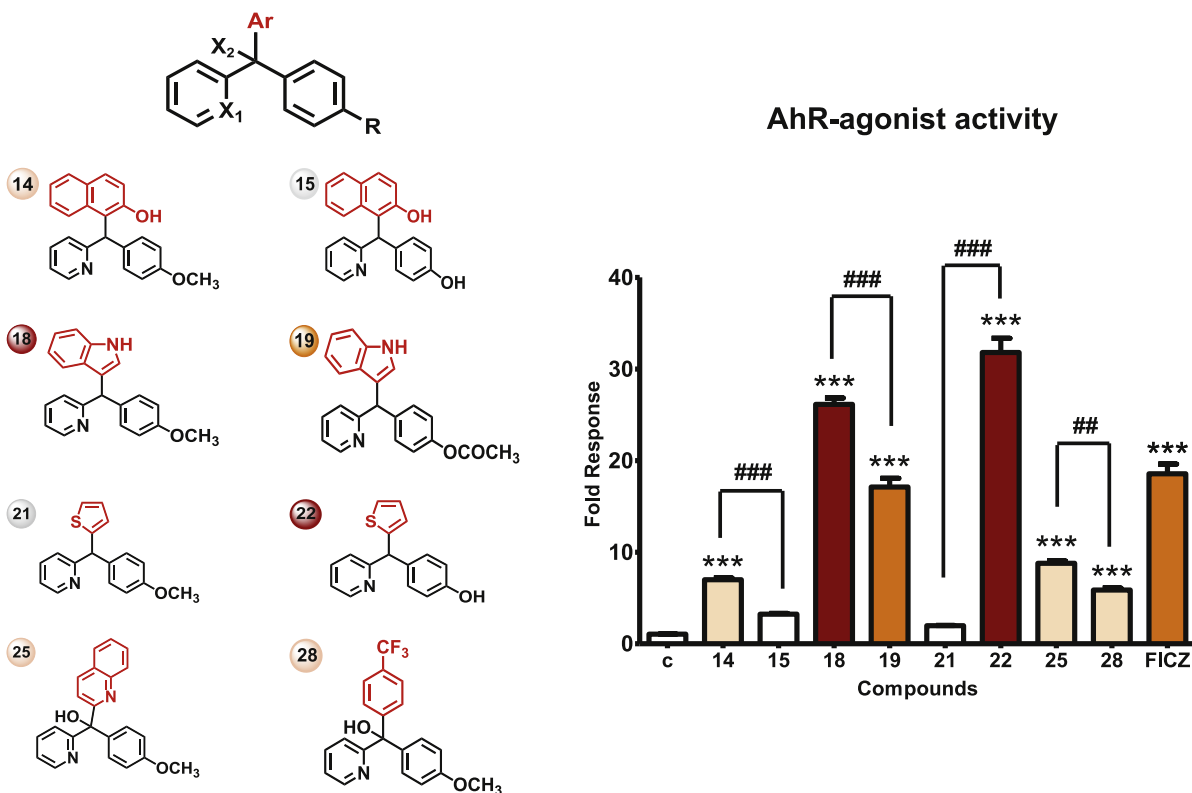
Figure 2. Maximum AhR agonist activity induced in cells by compounds **14**, **15**, **18**, **19**, **21 22**, **25** and **28**. General structure and agonist compounds are represented at the left. Data are expressed as fold responses, as compared to non-induced cells (*i.e.* vehicle control (c)). The bar chart at the right shows mean fold response ± SEM (n = 4) as: ≥25 folds (red), between 10 and 20 folds (orange), ≤10 folds and significant (light orange), <10 folds and not significant (white). The levels of significance were determined using one-way ANOVA, followed by Dunnett's post-test when compared to vehicle control (***$p < 0.001$) or by Bonferroni post-test when compared between pairs of structural analogous (##$p < 0.01$, ###$p < 0.001$).
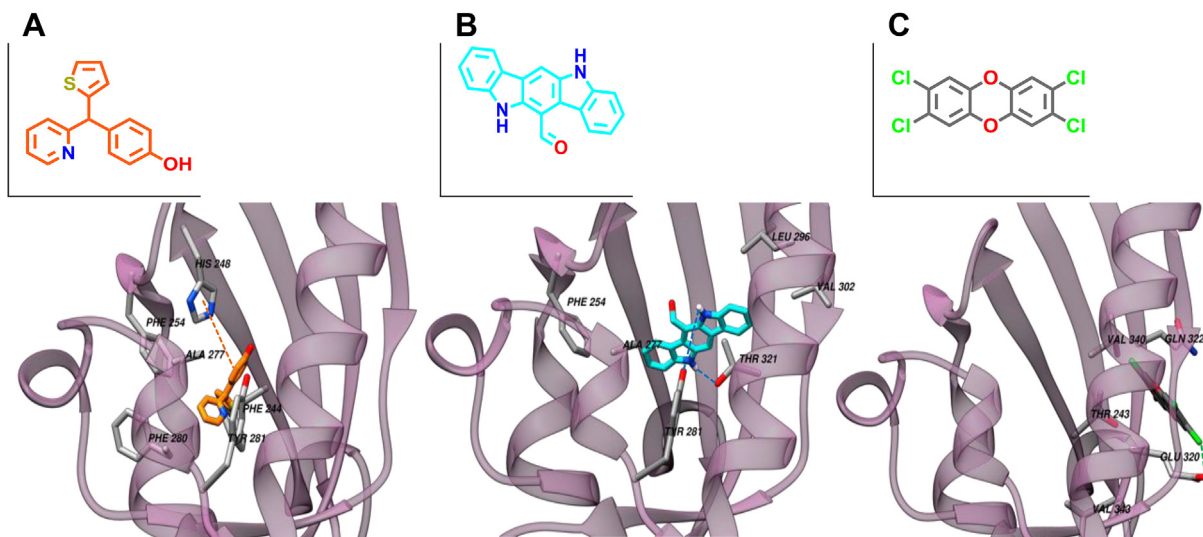
Fig. 3. Representation of molecular docking and the molecular interactions between HIF2α (PBD ID: 3F1O) and **22** (A), FICZ (B) and TCDD (C). The structure of the protein is represented as transparent lilac ribbons and the best pose obtained for **22**, FICZ and TCDD are displayed as sticks. The residues involved in hydrophobic interactions are labelled. Π-Cation (orange), hydrogen bond (blue) and halogen bond (green) interactions are represented in dot lines. (color online only). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

important steps toward the prioritization in drug discovery [61]. Therefore, physicochemical properties (Table 2) were predicted *in silico* for the six novel synthesized TAMs with the most significant AhR agonist activity. Additional properties are provided in SI-3.

The presence of reactive functional groups (#rtvFG) has been related to decomposition, reactivity and toxicity *in vivo*. Therefore, this molecular descriptor serves as an alert system of structural groups such as azo, diazo, carbonate, aluminum or silicon (full list

Table 2
Physicochemical properties and druglikeness criteria of the AhR-agonist TAMs and the drug bisacodyl.

ID	#rtvFG[a]	MW [g/mol][b]	Dipole[c]	SASA [Å²][d]	Volume [Å³][e]	TPSA[f]	Donor HB[g]	Accpt HB[h]	Polrz [Å³][i]	logP o/w[j]	logS [S: mol/dm³][k]	Lipinski's rule of five[l]
14	0	341.41	0.68	599.04	1081.50	42.35	1	2.50	38.39	5.32	−5.46	1
18	0	314.39	1.75	578.44	1034.32	37.91	1	1.75	36.98	5.36	−5.51	1
19	1	342.40	1.75	625.36	1111.64	54.98	1	3.50	40.02	4.82	−5.74	0
22	0	267.35	3.53	503.96	869.78	50.36	1	1.75	29.76	4.09	−4.32	0
25	0	342.40	6.52	615.79	1090.79	61.36	1	3.50	38.90	4.95	−5.41	0
28	0	359.35	7.33	606.32	1064.55	42.35	1	2.50	36.68	5.57	−6.12	1
BSD	2	361.40	7.66	658.56	1169.98	65.49	0	6.00	40.99	3.66	−4.74	0

[a] #rtvFG: Number of reactive functional groups in the structure of the molecule (listed in Experimental Section). Recommended values: 0−2.
[b] MW: Molecular weight of the molecule. Recommended values: 130−725 g/mol.
[c] Dipole: Computed dipole moment of the molecule. Recommended values: 1.0−12.5.
[d] SASA: Total solvent accessible surface area. Recommended values: 300−1000 Å² using a probe with a 1.4 Å radius.
[e] Volume: Total solvent-accessible volume. Recommended values: 500−2000 Å³ using a probe with a 1.4 Å radius.
[f] TPSA: Topological polar surface area.
[g] Donor HB: Estimated number of hydrogen bonds that would be donated by the solute to water molecules in an aqueous solution. Recommended values: 0−6.
[h] Accpt HB: Estimated number of hydrogen bonds that would be accepted by the solute from water molecules in an aqueous solution. Recommended values: 2-20.
[i] Polrz: Predicted polarizability. Recommended values: 13−70 Å³.
[j] logP o/w: Predicted logarithm of octanol/water partition coefficient. Recommended values: −2.0 to 6.5.
[k] logS: Predicted logarithm of solubility (S expressed in mol/dm³). Recommended values: −6.5 to 0.5.
[l] Lipinski's rule of five (druglikeness): Number of violations of Lipinski's rule of five: MW < 500, logP o/w < 5, donor HB ≤ 5, accpt HB ≤ 10. Maximum: 4 violations.

in SI-3). All the agonist TAMs were consistent the recommended criterium for drug-like compounds. They all were free from potentially reactive functional groups except for **19**. Interestingly, the TAM drug bisacodyl possessed two. The molecular weights, the computed dipole moments, the topological polar surface area (TPSA) and the total solvent accessible surface area (SASA) and volume, the polarizability as well as the estimated number of hydrogen bonds (HB) that may be donated or accepted in an aqueous solution were predicted within the recommended values for all the studied TAMs.

The octanol/water partition coefficient (logP o/w) directly influences the effects of chemical entities on biological systems, particularly their pharmacokinetics and pharmacodynamics [62]. Similarly, the aqueous solubility (logS) of a compound influences its ability to reach the site of action and produce any kind of effect. According to the predicted logP o/w and logS, all the agonist TAMs met the recommended range for druglikeness. Finally, Table 2 shows the Lipinski's rule of five as a global criterium of druglikeness that suggests the limits of some physicochemical properties as follow: MW < 500, logP o/w < 5, donor HB ≤ 5, accpt HB ≤ 10. Hence, none of the agonist TAMs showed more than one violation of such rule and only the predicted logP o/w for the TAMs **14**, **18** and **28** was slightly higher than 5.

In order to analyze the six agonist compounds in the context of rapidly metabolized AhR ligands (RMAhRLs) or Selective AhR modulators (SAhRMs), a comparative analysis was performed based on their physicochemical profiles, as proposed by Dolciami D. et al. [63]. The mean values ± SD of the molecular descriptors suggested for RMAhRLs are MW (335 ± 91), log P o/w (3.46 ± 1.10) and TPSA (65.1 ± 24.8), while those for SAhRMs are MW (307 ± 77), log P o/w (4.24 ± 2.24) and TPSA (43.7 ± 34) [63]. Comparisons performed using one-way ANOVA (p < 0.05) followed by Dunnett's post-test of the TAMs within the RMAhRLs and SAhRMs context, revealed no significant differences, as represented in Fig. 4 a) and b), respectively. Therefore, the studied TAMs cannot be classified as SAhRMs or RMAhRLs according to this criterion.

On the other hand, predictive results of some properties contributing to the ADME profile of the AhR-agonist TAMs, are shown in Table 3.

Considering the predicted parameters collectively related to oral bioavailability (Caco-2 permeability, Human Oral Absorption (HOA) and Jorgensens rule of three), it can be concluded that the AhR-agonist TAMs synthetized probably have good permeability to cross the gut-blood barrier, a high human oral absorption and only **19** and **28** violated the solubility criterium of Jorgensen's rule. Adequate binding to human serum albumin (logKhsa) as well as an appropriate number of metabolic reactions (#Metab) were predicted for all the TAMs. Lastly, the apparent capacity to cross the brain/blood barrier (logBB) was predicted as good for all the studied TAMs while the predicted skin permeability (logKp) was in acceptable limit only for the strongest AhR-agonist **22** and BSD.

3. Conclusions

TAM compounds were straightforwardly synthesized and characterized by means of efficient synthetic strategies in this work. The effects of 32 newly TAM derivatives as potential modulators of the emerging pharmacological target AhR were determined in vitro using a novel secreted luciferase assay system. The bioassays revealed an exclusive agonism of eight derivatives and a lack of antagonist activity on AhR activation across the TAM set. Hetero-aromatic or naphthol moieties crucially determined the occurrence of AhR agonism and the thiophene derivative **22** was the most potent agonist compound on AhR-mediated transcription yielding over 30-fold response, comparable to the endogenous metabolite FICZ. The structural adequacy, absence of cytotoxicity as well as druglikeness and favorable ADME profile, allow to suggest **22** as a new lead compound in the study of AhR-mediated transcription. In general, these results could provide valuable insights to design new potent AhR modulators based on the TAM scaffold.

4. Experimental Section

4.1. Chemistry

4.1.1. Materials and methods

All reagents were obtained from commercial sources unless otherwise noted and used as received. Heated experiments were conducted using thermostatically controlled oil baths. Reaction requiring anhydrous conditions were performed under an atmosphere oxygen-free in oven-dried glassware. Drying of the products was carried out under reduced pressure using a vacuum pump and/ or a desiccant heated to 40 °C in the presence of P_2O_5. All reactions were monitored by analytical thin layer chromatography (TLC) or

a)

b)
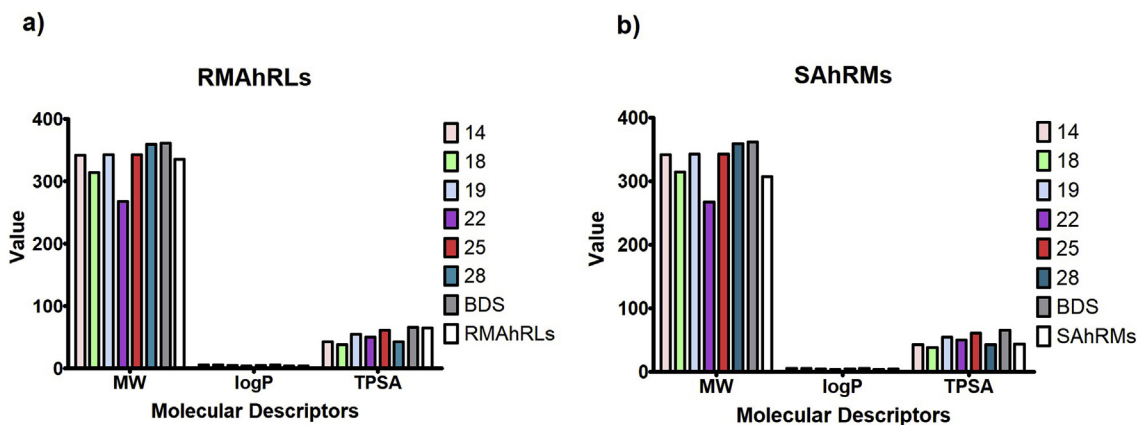
Fig. 4. Mean values of the molecular descriptors MW, logP and TPSA for compounds **14**, **18**, **19**, **22**, **25** and **28** suggested to classify rapidly metabolized AhR ligands (RMAhRLs) and Selective AhR modulators (SAhRMs) [63]. One-way ANOVA ($p < 0.05$) followed by Dunnett's post-test did not revealed significant differences between TAMs and RMAhRLs or SAhRMs, respectively.

Table 3
Prediction of ADME descriptors for the AhR-agonist TAMs and the drug bisacodyl.

ID	Caco-2 [nm/sec][a]	MDCK [nm/sec][b]	logBB[c]	logKp [nm/sec][d]	logKhsa [nm/sec][e]	Jm [µg cm^{-2} h^{-1}][f]	HOA[g]	#Metab[h]	Jorgensen's rule of three[i]
14	4561.02	2551.05	−0.07	−0.05	0.87	1.06	3	5	0
18	5349.67	3031.00	0.07	−0.07	0.93	0.83	3	4	0
19	2083.59	1093.83	−0.37	−0.88	0.80	0.08	3	3	1
22	2157.45	1872.28	−0.19	−1.07	0.47	1.11	3	5	0
25	4130.44	2291.76	−0.12	−0.08	0.70	1.12	3	4	0
28	5013.57	10000.00	0.26	−0.34	0.82	0.12	3	4	1
BSD	846.79	413.32	−0.87	−1.81	0.20	0.10	3	3	0

[a] Caco-2 (model for the gut-blood barrier): Predicted apparent Caco-2 cell permeability (non-active transport) [nm/sec]. Criteria: <25 poor permeability,>500 great permeability.

[b] MDCK (mimic for the blood-brain barrier): Predicted apparent MDCK cell permeability (non-active transport) [nm/sec]. Criteria: <25 poor permeability,>500 great permeability.

[c] logBB: Predicted brain/blood partition coefficient (model for orally delivery drugs). Recommended values: from −3.0 to 1.2.

[d] logKp: Predicted skin permeability. Recommended values: from −8.0 to −1.0.

[e] logKhsa: Predicted binding to human serum albumin. Recommended values: from −1.5 to 1.5.

[f] Jm: Predicted maximum transdermal transport rate obtain from: Kp × MW × S (µg cm^{-2}.h^{-1}).

[g] HOA (Human Oral Absorption): Predicted qualitative human oral absorption: low (1), medium (2), high (3).

[h] Metab: Number of likely metabolic reactions (listed in SI-3). Recommended values: 1−8.

[i] Jorgensen's rule of three (oral availability): Fewer (and preferably no) violations of the follow: logS > −5.7, Caco-2 > 22 nm/s, #PrimaryMetabolites< 7.

Gas chromatography-Mass spectrometry (GC-MS). TLC was performed on aluminium sheets, silica gel coated with fluorescent indicator F$_{254}$, Merck. TLC plates were visualized using irradiation with light at 254 nm or in an iodine chamber as appropriate. FCC was carried out when necessary using silica gel 60 (particle size 0.040−0.063 mm, Merck). The eluent mixture is specified for each purification.

4.1.2. Physical measurements

Melting points (Mp) were determined on a Leica VMHB system Kofler apparatus. The structure of the products prepared by different methods was checked by comparison of their NMR, IR and MS data and by the TLC behavior. ^{1}H and ^{13}C NMR spectra were acquired on a Bruker BioSpin GmbH spectrometer 400 MHz, at room temperature. Chemical shifts are reported in δ units, parts per million (ppm). Coupling constants (J) are measured in hertz (Hz). Splitting patterns are designed as follows: s, singlet; d, doublet; dd, doublet of doublets; dm, doublet of multiplets, ddd, doublet of doublets of doublets; m, multiplet; br, broad. Various 2D techniques and DEPT experiments were used to establish the structures and to assign the signals. For the assignments of the NMR signals, we use the convention presented in Fig. 5. GC-MS analyses were performed with an Agilent 6890 N instrument equipped with a 12 m × 0.20 mm dimethyl polysiloxane capillary column and an Agilent 5973 N MS detector-column temperature gradient 80−300 °C (method 160): 160 °C (1 min), 180 °C−260 °C (10 °C/min), 260 °C (4 min); (method 180): 180 °C (1 min), 180 °C−300 °C (10 °C/min), 300 °C (2 min), gradient 200−300 °C (method 200): 200 °C (1 min), 200 °C−300 °C (10 °C/min), 300 °C (4 min). Low-resolution mass spectra (LRMS) result from ionization by electronic impact. Infrared spectra were recorded over the 400-4000 cm^{-1} range with an Agilent Technologies Cary 630 FTIR/ATR/ZnSe spectrometer. High-resolution mass spectra (HRMS) analyses were acquired on a Thermo Scientific LTQ Orbitrap mass spectrometer. The HPLC analyses were carried out on a normal phase column Hypersil Si (length: 150 mm, diameter: 4.6 mm, stationary phase: 5 µm) and a reverse phase column Hypersil ODS C18 (length: 150 mm, diameter: 4.60 mm, stationary phase: 5 µm) using a Water 2998 Photodiode Array Detector (260−370 nm) and an isocratic system of elution. The retention time (R_t) is expressed in min in the decimal system. HPLC purity was determined on the Hypersil Si column, using *n*-heptane/ethyl acetate 7/3 with a flow rate of 0.8 mL per min and UV detection at $\lambda = 262−264$ nm, unless otherwise notified.
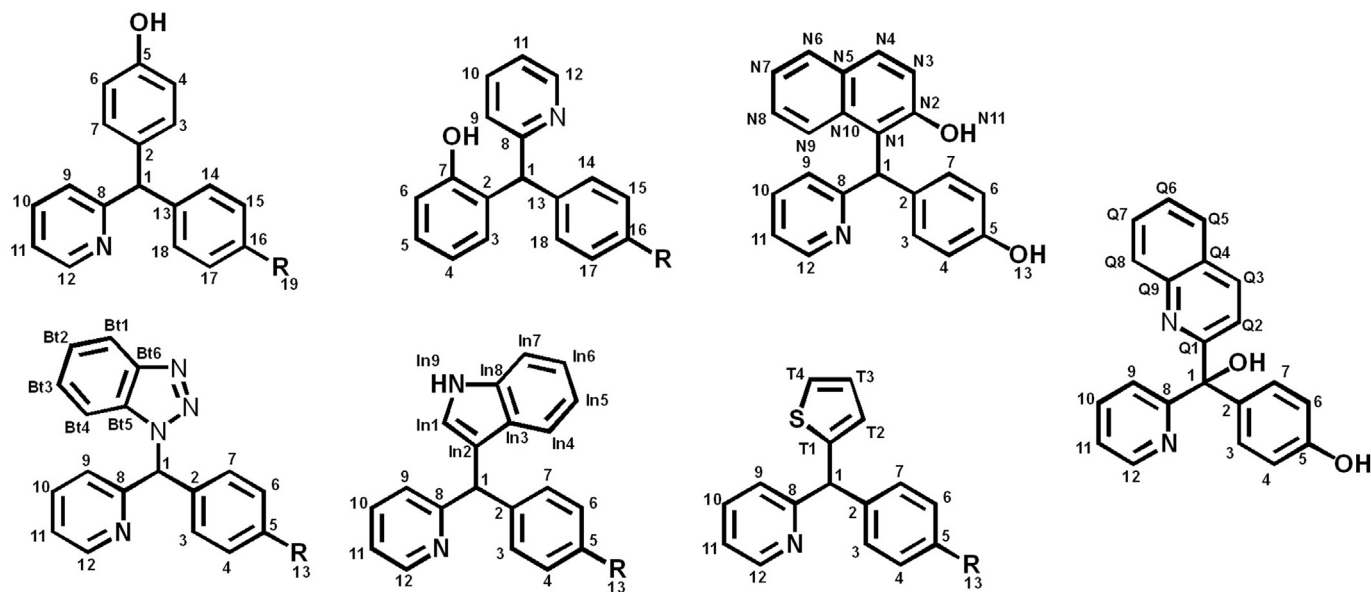
Fig. 5. Convention adopted to assign signals of ^{1}H and ^{13}C NMR spectra. Only the 32 TAM compounds evaluated *in vitro* are described herein. Intermediates and other TAMs obtained are detailed as SI-1.

4.1.3. General procedure for the preparation of p,p-triarylmethanes

Method A: To a solution of the corresponding carbinol (1 eq.) and phenol (1.2 eq.) in nitrobenzene (0.4 M) was added dropwise concentrated sulfuric acid (4 eq.) at 0 °C. The reaction progress was monitored by GC-MS and TLC (eluent DCM/MeOH 90/10). After 5 min at 80 °C the reaction was cooled to room temperature and neutralized with a saturated solution of NaHCO$_3$ (pH 7−8), then extracted with ethyl acetate three times. The combined organic phases were dried over anhydrous Na$_2$SO$_4$, filtered and concentrated. The crude residue was purified by FCC on silica gel (eluent gradient DCM, DCM/MeOH 98/2, DCM/MeOH 90/10) to afford the corresponding *p,p*-triarylmethane: Yields: **4a** (72%), **4b** (63%) and **4c** (39%).

Method B: To a solution of the corresponding carbinol (1 eq.) and phenol (1.2 eq.) in nitrobenzene (0.4 M) was added dropwise concentrated sulfuric acid (4 eq.) at 0 °C. The reaction progress was monitored by GC-MS and TLC (eluent CyHex/EtOAc 50/50). After stirring at 0 °C the reaction was cooled to room temperature and neutralized with a saturated solution of NaHCO$_3$ (formation of a gum which solubilizes once pH 7−8 is reached), then extracted with ethyl acetate four times. The combined organic phases were washed with brine, dried over anhydrous Na$_2$SO$_4$, filtered and concentrated. The crude residue was purified by FCC on silica gel (eluent gradient DCM, DCM/MeOH 98/2, DCM/MeOH 90/10) to afford the *p,p*-triarylmethane: Yields: **4d** (33%) and **4e** (47%).

4.1.3.1. 4-((4-(tert-butyl)phenyl) (pyridin-2-yl)methyl)phenol (4a).
Yield: 96 mg, 0.30 mmol, white solid, 72%. 13 mg (0.04 mmol, 9.6%) of 2-((4-(*tert*-butyl)phenyl) (pyridin-2-yl)methyl)phenol are also isolated, (Method A). Mp = 154−156 °C. **TLC** CyHex/EtOAc 50/50, Rf = 0.54, DCM/MeOH 90/10, Rf = 0.80; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 1.24 (s, 9H, H$_{20}$), 5.49 (s, 1H, H$_1$), 6.69 (d, 2H, J$_{ortho}$ = 8 Hz, H$_4$, H$_6$), 7.01 (d, 2H, J$_{ortho}$ = 8 Hz, H$_7$, H$_3$), 7.11 (d, 2H, J$_{ortho}$ = 8 Hz, H$_{14}$, H$_{18}$), 7.18−7.24 (m, 2H, H$_9$, H$_{11}$), 7.30 (d, 2H, J$_{ortho}$ = 8 Hz, H$_{15}$, H$_{17}$), 7.71 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 8 Hz, J$_{10-12}$ = 4 Hz, H$_{10}$), 8.52 (ddd, 1H, J$_{12-11}$ = 4.7 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.95 Hz, H$_{12}$), 9.27 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 31.12 (C$_{19}$), 34.03 (C$_{20}$), 57.06 (C$_1$), 114.96 (C$_6$, C$_4$), 121.38 (C$_{11}$),

123.37 (C$_9$), 124.87 (C$_{15}$, C$_{17}$), 128.60 (C$_{14}$, C$_{18}$), 129.88 (C$_3$, C$_7$), 133.31 (C$_2$), 136.56 (C$_{10}$), 140.59 (C$_{13}$), 148.22 (C$_{16}$), 149.02 (C$_{12}$), 155.58 (C$_5$), 163.15 (C$_8$); **GC-MS** method 180, R$_t$ = 7.82 min, *m/z* 317 [M$^{+\cdot}$] (100), 302 [M$^{+\cdot}$ - CH$_3$] (30), 286 [M$^{+\cdot}$ - 2 CH$_3$] (5), 260 [M$^{+\cdot}$ - C(CH$_3$)$_3$] (19), 239 [OHPhCH*t*-BuPh]$^+$(23), 224 [OHPhCH*t*-BuPh$^+$ - CH$_3$] (8); **IR** (ATR) (cm^{-1}) 3058, 3020 (νCsp2-H), 2957 (νCsp3-H), 1615, 1600, 1510 (νC = C), 1167 (νC-O), 810 (δCsp2-H *p*-disubst), 755(δCsp2-H *o*-disubst); **HPLC** purity: 96%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 264 nm, R$_t$ = 2.84 min).

4.1.3.2. 4-(pyridin-2-yl (p-tolyl)methyl)phenol (4b).
Yield: 73 mg, 0.26 mmol, beige solid, 63%. 18 mg (0.07 mmol, 15%) of 2-(pyridin-2-yl (*p*-tolyl)methyl)phenol are also isolated, (Method A). Mp = 156−158 °C. **TLC** CyHex/EtOAc 50/50, Rf = 0.44, DCM/MeOH 90/10, Rf = 0.53; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.24 (s, 3H, H$_{19}$), 5.50 (s, 1H, H$_1$), 6.69 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_4$, H$_6$), 6.99 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_7$, H$_3$), 7.03−7.11 (m, 4H, H$_{14}$, H$_{15}$,H$_{18}$, H$_{17}$), 7.16−7.24 (m, 2H, H$_9$, H$_{11}$), 7.70 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.80 Hz, H$_{10}$), 8.51 (ddd, 1H, J$_{12-11}$ = 5 Hz, J$_{12-10}$ = 2.8 Hz, J$_{12-9}$ = 0.8 Hz, H$_{12}$), 9.24 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.56 (C$_{19}$), 57.12 (C$_1$), 114.98 (C$_6$, C$_4$), 121.40 (C$_{11}$), 123.38 (C$_9$), 128.78 (C$_{14}$, C$_{15}$, C$_{17}$, C$_{18}$), 129.91 (C$_3$, C$_7$), 133.40 (C$_{16}$), 135.05 (C$_2$), 136.57 (C$_{10}$), 140.63 (C$_{13}$), 149.05 (C$_{12}$), 155.70 (C$_5$), 163.21 (C$_8$); **GC-MS** method 180, R$_t$ = 7.22 min, *m/z* 274 [M$^{+\cdot}$] (100), 259 [M$^{+\cdot}$ - CH$_3$] (12), 197 [OHPhCHPhCH$_3$]$^+$(34), 181 [OHPhCHPhCH$_3^+$ - OH] (24), 167 [PhCHPy]$^+$, (8), 78 [Py]$^+$(4); **IR** (ATR) (cm^{-1}) 3018, 3005 (νCsp2-H), 2915 (νCsp3-H), 1614, 1592, 1508 (νC = C), 1233 (νC-O), 754 (δCsp2-H *o*-disubst); **HPLC** purity: 98%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 264 nm, R$_t$ = 3.00 min).

4.1.3.3. 4-((4-bromophenyl) (pyridin-2-yl)methyl)phenol (4c).
Yield: 51 mg, 0.15 mmol, oil, 39%, (Method A). **TLC** DCM/MeOH 90/10, Rf = 0.60; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.55 (s, 1H, H$_1$), 6.69 (d, 2H, J$_{ortho}$ = 8.70 Hz, H$_4$, H$_6$), 7.00 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_7$, H$_3$), 7.14 (d, 2H, J$_{ortho}$ = 8.2 Hz, H$_{14}$, H$_{18}$), 7.20−7.28 (m, 2H, H$_9$, H$_{11}$), 7.46 (d, 2H, J$_{ortho}$ = 8.3 Hz, H$_{15}$, H$_{17}$), 7.73 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.8 Hz, H$_{10}$), 8.53 (dd, 1H, J$_{12-11}$ = 5.3 Hz, J$_{12-10}$ = 1.9 Hz, H$_{12}$), 9.32 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 56.56 (C$_1$), 115.15 (C$_6$, C$_4$), 119.32 (C$_{16}$), 121.68 (C$_{11}$), 123.58

(C$_9$), 129.94 (C$_3$, C$_7$), 131.02 (C$_{15}$, C$_{17}$), 131.23 (C$_{14}$, C$_{18}$), 132.79 (C$_2$), 136.83 (C$_{10}$), 143.16 (C$_{13}$), 149.21 (C$_{12}$), 159.91 (C$_7$), 162.44 (C$_8$); **GC-MS** method 180, R$_t$ = 8.81 min, *m/z* 340 [M$^{+\cdot}$] (100), 324 [M$^{+\cdot}$ - OH] (3), 259 [M$^{+\cdot}$ - Br] (52), 181 [OHPhCH$_2$Ph]$^+$(85), 167 [PhCH$_2$Py]$^+$(21), 78 [Py]$^+$ (15); **IR** (ATR) (cm^{-1}) 3055, 3018 (νCsp2-H), 2924 (νCsp3-H), 1593, 1511, 1486 (νC = C), 1168 (νC-O); **HPLC** purity: 95%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 263 nm, R$_t$ = 3.00 min).

4.1.3.4. 4-((4-chlorophenyl) (pyridin-2-yl)methyl)phenol (4d). Yield: 865 mg, 2.94 mmol, white solid, 33%, (Method B). Mp = 170−171 °C. **TLC** CyHex/EtOAc 50/50, Rf = 0.49, DCM/MeOH 90/10, Rf = 0.70; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.57 (s, 1H, H$_1$), 6.70 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 7 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_7$, H$_3$), 7.19−7.25 (d + m, 4H, J$_{ortho}$ = 8.5 Hz, H$_9$, H$_{11}$, H$_{14}$, H$_{18}$), 7.34 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_{15}$, H$_{17}$), 7.73 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 7.6 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.53 (dd, 1H, J$_{12-11}$ = 5.6 Hz, J$_{12-10}$ = 2,2 Hz, H$_{12}$), 9.31 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 56.48 (C$_1$), 115.12 (C$_4$, C$_6$), 121.64 (C$_{11}$), 123.55 (C$_9$), 128.07 (C$_{15}$, C$_{17}$), 129.91 (C$_3$, C$_7$), 130.77 (C$_{14}$, C$_{18}$), 130.80 (C$_{16}$), 132.85 (C$_2$), 136.80 (C$_{10}$), 142.70 (C$_{13}$), 149.19 (C$_{12}$), 155.90 (C$_5$), 162.50 (C$_8$); **GC-MS** method 180, R$_t$ = 8.04 min, *m/z* 295 [M$^{+\cdot}$] (100), 280 [M$^+$-OH] (2), 259 [M$^+$ − Cl] (25), 217 [OHPhCHPhCl]$^+$(36), 201 [OHPhCHPhCl$^+$-OH] (9), 181 [OHPhCHPhCl$^+$- Cl] (35), 167 [PhCHPy]$^+$(14), 78 [Py]$^+$ (7); **IR** (ATR) (cm^{-1}) 3054 (νO-H), ν 3019 (νCsp2-H), 2928 (νCsp3-H), 1615, 1592, 1510 (νC = C), 1235 (νC-O), 806 (δCsp2-H *p*-disubst), 756 (δCsp2-H *o*-disubst), 624(νC-Cl); **HPLC** purity: 95%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 263 nm, R$_t$ = 3.01 min).

4.1.3.5. 4-((4-fluorophenyl) (pyridin-2-yl)methyl)phenol (4e). Yield: 89 mg, 0.32 mmol, yellow oil, 47%, (Method B). **TLC** CyHex/EtOAc 50/50, Rf = 0.44, DCM/MeOH 90/10, Rf = 0.70; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.58 (s, 1H, H$_1$), 6.73 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_4$, H$_6$), 7.02 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_7$, H$_3$), 7.05−7.14 (m, 2H, H$_9$, H$_{11}$), 7.17−7.27 (m, 4H, H$_{14}$, H$_{15}$, H$_{17}$, H$_{18}$), 7.70 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 7.8 Hz, J$_{10-12}$ = 2 Hz, H$_{10}$), 8.53 (dd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.7 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 56.64 (C$_1$), 114.81 (d, 2C, J$_{C-F}$ = 21 Hz, C$_{15}$, C$_{17}$), 115.24 (C$_4$, C$_6$), 121.68 (C$_{11}$), 123.59 (C$_9$), 130.02 (C$_3$, C$_7$), 130.83 (d, 2C, J$_{C-F}$ = 8 Hz, C$_{14}$, C$_{18}$), 133.31 (C$_2$), 136.85 (C$_{10}$), 139.70 (d, 1C, J$_{C-F}$ = 3 Hz, C$_{13}$), 149.25 (C$_{12}$), 155.95 (C$_5$), 160.65 (d, 1C, J$_{C-F}$ = 240 Hz, C$_{16}$), 162.89 (C$_8$); **GC-MS** method 180, R$_t$ = 6.37 min, *m/z* 278 [M$^{+\cdot}$] (100), 261 [M$^{+\cdot}$-OH] (30), 201 [OHPhCHPhF]$^+$(42), 183 [OHPhCHPhCl$^+$-F] (23), 78 [Py]$^+$ (4); **IR** (ATR) (cm^{-1}) 3056, 3005 (νCsp2-H), 1592, 1506 (νC = C), 1221 (νC-O), 811 (δCsp2-H *p*-disubst), 757 (δCsp2-H *o*-disubst); **HPLC** purity: 100%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 264 nm, R$_t$ = 3.07 min).

4.1.4. General procedure for the preparation of triarylmethane acetates

To a solution of the corresponding compound (1 eq.) in acetic anhydride (110 eq.) was added an aqueous solution of sodium hydroxide 1 M (1.30 eq.) at 20 °C or 40 °C. The reaction progress was monitored by GC-MS and TLC (eluent CyHex/EtOAc 90/10). After stirring at room temperature, the reaction mixture was concentrated, and ethyl acetate was added. The solution was washed with water, an aqueous solution of NaHCO$_3$ then with brine and water. Then the organic phases were dried over anhydrous Na$_2$SO$_4$, filtered and concentrated. The expected compounds were isolated, and purity was checked by NMR. Yields: **5a** (90%), **5b** (96%), **5c** (77%), **5d** (76%), and **5e** (78%).

4.1.4.1. 4-((4-(tert-butyl)phenyl) (pyridin-2-yl)methyl)phenyl acetate (5a). The reaction was performed during 4 h at 20 °C then 15 h at 40 °C. Yield: 524 mg, 0.35 mmol, brown oil, 90%. **TLC** CyHex/EtOAc

50/50, Rf = 0.73, DCM/MeOH 90/10, Rf = 0.20; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 1.25 (s, 9H, H$_{19}$), 2.25 (s, 3H, H$_{21}$), 5.65 (s, 1H, H$_1$), 7.04 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 7.17 (d, 2H, J$_{ortho}$ = 8.3 Hz, H$_{14}$, H$_{18}$), 7.23−7.29 (m, 4H, H$_3$, H$_7$, H$_9$, H$_{11}$), 7.32 (d, 2H, J$_{ortho}$ = 8.4 Hz, H$_{15}$, H$_{17}$), 7.74 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 7.6 Hz, J$_{10-12}$ = 2 Hz, H$_{10}$), 8.51−8.56 (m, 1H, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.79 (C$_{21}$), 31.09 (C$_{19}$), 34.06 (C$_{20}$), 56.91 (C$_1$), 121.48 (C$_4$, C$_6$), 121.65 (C$_{11}$), 123.58 (C$_9$), 125.06 (C$_{15}$, C$_{17}$), 128.58 (C$_{14}$, C$_{18}$), 129.91 (C$_3$, C$_7$), 136.79 (C$_{10}$), 139.88 (C$_2$), 140.61 (C$_{13}$), 148.55 (C$_{16}$), 148.77 (C$_5$), 149.17 (C$_{12}$), 162.33 (C$_8$), 169.23 (C$_{22}$); **GC-MS** method 180, R$_t$ = 9.38 min, *m/z* 359 [M$^+$] (99), 344 [M$^+$ − CH$_3$] (11), 316 [M$^+$ − 2 CH$_3$] (100), 302 [M$^+$ − C(CH$_3$)$_3$] (33), 239 [OHPhCH*t*-BuPh]$^+$(35), 224 [OHPhCH*t*-BuPh$^+$ - CH$_3$](14), 209 [OHPhCH*t*-BuPh$^+$ - 2 CH$_3$] (25), 193 [OHPhCH*t*-BuPh$^+$ - C(CH$_3$)$_3$] (9), 167 [PhCHPy]$^+$(23); **IR** (ATR) (cm^{-1}) 3053 (νCsp2-H), 2960, 2905, 2868 (νCsp3-H), 1759 (νC = O), 1587, 1504, 1467 (νC = C), 1192 (νC-O), 749 (δCsp2-H *p*-disubst).

4.1.4.2. 4-(pyridin-2-yl (p-tolyl)methyl)phenyl acetate (5b). The reaction was performed during 2 h 20 min at 20 °C. Yield: 542 mg, 1.51 mmol, brown oil, 96%. **TLC** CyHex/EtOAc 50/50, Rf = 0.74, DCM/MeOH 90/10, Rf = 0.80; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.25 (s, 3H, H$_{20}$), 5.72 (s, 1H, H$_1$), 7.06 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 7.21 (d, 2H, J$_{ortho}$ = 8.4 Hz, H$_{14}$, H$_{18}$), 7.24−7.26 (m, 3H, H$_3$, H$_7$, H$_{11}$), 7.31 (dd, 1H, J$_{9-10}$ = 8 Hz, J$_{9-11}$ = 2 Hz, H$_9$), 7.51 (d, 2H, J$_{ortho}$ = 8.4 Hz, H$_{15}$, H$_{17}$), 7.76 (ddd, 1H, J$_{10-9}$ = J$_{10-11}$ = 7.6 Hz, J$_{10-12}$ = 1.80 Hz, H$_{10}$), 8.56 (ddd, 1H, J$_{12-11}$ = 7.6 Hz, J$_{12-10}$ = 2.7 Hz, J$_{12-9}$ = 0.9 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.83 (C$_{20}$), 56.40 (C$_1$), 119.64 (C$_{16}$), 121.68 (C$_4$, C$_6$), 121.90 (C$_{11}$), 123.76 (C$_9$), 129.94 (C$_3$, C$_7$), 131.19 (C$_{14}$, C$_{18}$), 131.24 (C$_{15}$, C$_{17}$), 137.02 (C$_{10}$), 140.08 (C$_2$), 142.39 (C$_{13}$), 148.97 (C$_5$), 149.35 (C$_{12}$), 161.66 (C$_8$), 169.24 (C$_{19}$); **GC-MS** method 180, *m/z* R$_t$ = 9.46 min, 382 [M$^+$] (61), 340 [M$^+$ − COCH$_3$] (100), 261 [M$^+$ − Cl−COCH$_3$] (38), 184 [OPhCH$_2$Ph]$^+$(46), 167 [PhCH$_2$Ph]$^+$(24); **IR** (ATR) (cm^{-1}) 3061, 3029, 3009 (νCsp2-H), 1754 (νC = O), 1585, 1504, 1487 (νC = C), 1202 (νC-O), 763 (δCsp2-H *p*-disubst); **HPLC** purity: 98%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 262 nm, R$_t$ = 2.79 min).

4.1.4.3. 4-((4-bromophenyl) (pyridin-2-yl)methyl)phenyl acetate (5c). The reaction was performed during 1 h at 20 °C. Yield: 344 mg, 0.90 mmol, brown oil, 77%. **TLC** CyHex/EtOAc 50/50, Rf = 0.82; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.25 (s, 3H, H$_{20}$), 5.74 (s, 1H, H$_1$), 7.07 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 7.24−7.29 (m, 5H, H$_3$, H$_7$, H$_{11}$, H$_{14}$, H$_{18}$), 7.31 (dd, 1H, J$_{9-10}$ = 7.8 Hz, J$_{9-11}$ = 1.20 Hz, H$_9$), 7.38 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_{15}$, H$_{17}$), 7.76 (dd, 1H, J$_{10-11}$ = 7.8 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.56 (ddd, 1H, J$_{12-11}$ = 7.5 Hz, J$_{12-10}$ = 2.7 Hz, J$_{12-9}$ = 0.9 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.83 (C$_{20}$), 56.36 (C$_1$), 121.68 (C$_4$, C$_6$), 121.90 (C$_{11}$), 123.75 (C$_9$), 128.27 (C$_{15}$, C$_{17}$), 129.94 (C$_3$, C$_7$), 130.85 (C$_{14}$, C$_{18}$), 131.11 (C$_{16}$), 137.70 (C$_{10}$), 140.16 (C$_2$), 141.95 (C$_{13}$), 148.97 (C$_5$), 149.35 (C$_{12}$), 161.73 (C$_8$), 169.24 (C$_{19}$); **GC-MS** method 180, R$_t$ = 8.67 min *m/z* 338 [M$^+$] (54), 294 [M$^+$ − COCH$_3$] (100), 259 [M$^+$ − Py] (17), 217 [OPhCH$_2$PhCl]$^+$ (28), 202 [PyCH$_2$PhCl]$^+$(7), 184 [OPhCH$_2$Ph]$^+$(18), 167 [PhCH$_2$Ph]$^+$(13); **IR** (ATR) (cm^{-1}) 3055, 3007 (νCsp2-H), 2917 (νCsp3-H), 1754 (νC = O), 1585, 1505, 1466 (νC = C), 1202 (νC-O), 817 (δCsp2-H *o*-disubst), 763 (δCsp2-H *p*-disubst); **HPLC** purity: 95%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ_{max} = 262 nm, R$_t$ = 2.78 min).

4.1.4.4. 4-((4-chlorophenyl) (pyridin-2-yl)methyl)phenyl acetate (5d). The reaction was performed during 1 h at 20 °C. Yield: 349 mg, 1.03 mmol, orange oil, 76%. **TLC** CyHex/EtOAc 50/50, Rf = 0.74, CyHex/EtOAc 50/50, Rf = 0.74; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.25 (s, 3H, H$_{20}$), 5.73 (s, 1H, H$_1$), 7.05 (d, 2H, J$_{4-6}$ = 8.6 Hz, H$_4$, H$_6$), 7.15 (dd, 2H, J$_{ortho}$ = J$_{H-F}$ = 8.9 Hz, H$_{15}$, H$_{17}$),

7.23−7.31 (m, 6H, H$_3$, H$_7$, H$_9$, H$_{11}$, H$_{14}$, H$_{18}$), 7.76 (ddd, 1H, J$_{10\text{-}9}$ = J$_{10\text{-}11}$ = 7.7 Hz, J$_{10\text{-}12}$ = 1.9 Hz, H$_{10}$), 8.55−8.57 (ddd, 1H, J$_{12\text{-}9}$ = 0.80 Hz, J$_{12\text{-}10}$ = 1.8 Hz, J$_{11\text{-}12}$ = 4.8 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.83 (C$_{20}$), 56.28 (C$_1$), 115.04 (d, 2C, J$_{C\text{-}F}$ = 20.90 Hz, C$_{15}$, C$_{17}$), 121.63 (C$_4$, C$_6$), 121.84 (C$_{11}$), 123.69 (C$_9$), 129.90 (C$_3$, C$_7$), 130.82 (d, 2C, J$_{C\text{-}F}$ = 7,9 Hz, C$_{14}$, C$_{18}$), 136.98 (C$_{10}$), 139.09 (d, 2C, J$_{C\text{-}F}$ = 3.20 Hz, C$_{13}$), 140.49 (C$_2$), 148.91 (C$_5$), 149.33 (C$_{12}$), 160.82 (d, 1C, J$_{C\text{-}F}$ = 239.60 Hz, C$_{16}$), 169.25 (C$_8$), 172.06 (C$_{19}$); **GC-MS** method 180, R$_t$ = 7.18 min, *m/z*, 321 [M$^+$] (6), 278 [M$^+$ − CH$_3$CO] (100),183) [PyCHPhO]$^+$(15); **IR** (ATR) (cm^{-1}) 3051, 3006 (νCsp2-H), 2927 (νCsp3-H), 1754 (νC = O), 1571, 1588, 1503 (νC = C), 1192 (νC-O), 1160 (νC-F), 819 (δCsp2-H *o*-disubst), 750 (δCsp2-H *p*-disubst); **HPLC** purity: 97%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ$_{max}$ = 262 nm, R$_t$ = 2.84 min).

4.1.4.5. 4-((4-fluorophenyl) (pyridin-2-yl)methyl)phenyl acetate (5e). The reaction was performed during 3 h at 40 °C. Yield: 470 mg, 1.46 mmol, brown oil, 78%. **TLC** CyHex/EtOAc 50/50, Rf = 0.60; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.25 (s, 3H, H$_{20}$), 5.73 (s, 1H, H$_1$), 7.05 (d, 2H, J$_{4\text{-}6}$ = 8.6 Hz, H$_4$, H$_6$), 7.15 (dd, 2H, J$_{ortho}$ = J$_{H\text{-}F}$ = 8.9 Hz, H$_{15}$, H$_{17}$), 7.23−7.31 (m, 6H, H$_3$, H$_7$, H$_9$, H$_{11}$, H$_{14}$, H$_{18}$), 7.76 (ddd, 1H, J$_{10\text{-}9}$ = J$_{10\text{-}11}$ = 7.7 Hz, J$_{10\text{-}12}$ = 1.9 Hz, H$_{10}$), 8.55−8.57 (ddd, 1H, J$_{12\text{-}9}$ = 0.80 Hz, J$_{12\text{-}10}$ = 1.8 Hz, J$_{11\text{-}12}$ = 4.8 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.83 (C$_{20}$), 56.28 (C$_1$), 115.04 (d, 2C, J$_{C\text{-}F}$ = 20.90 Hz, C$_{15}$, C$_{17}$), 121.63 (C$_4$, C$_6$), 121.84 (C$_{11}$), 123.69 (C$_9$), 129.90 (C$_3$, C$_7$), 130.82 (d, 2C, J$_{C\text{-}F}$ = 7,9 Hz, C$_{14}$, C$_{18}$), 136.98 (C$_{10}$), 139.09 (d, 2C, J$_{C\text{-}F}$ = 3.20 Hz, C$_{13}$), 140.49 (C$_2$), 148.91 (C$_5$), 149.33 (C$_{12}$), 160.82 (d, 1C, J$_{C\text{-}F}$ = 239.60 Hz, C$_{16}$), 169.25 (C$_8$), 172.06 (C$_{19}$); **GC-MS** method 180, R$_t$ = 7.18 min, *m/z*, 321 [M$^+$] (6), 278 [M$^+$ − CH$_3$CO] (100),183) [PyCHPhO]$^+$(15); **IR** (ATR) (cm^{-1}) 3051, 3006 (νCsp2-H), 2927 (νCsp3-H), 1754 (νC = O), 1571, 1588, 1503 (νC = C), 1192 (νC-O), 1160 (νC-F), 819 (δCsp2-H *o*-disubst), 750 (δCsp2-H *p*-disubst); **HPLC** purity: 97%, (Hypersyl Si, *n*-heptane/EtOAc 30/70, flow rate 0.80 mL/min, λ$_{max}$ = 262 nm, R$_t$ = 2.84 min).

4.1.5. General procedure for the preparation of triarylmethane acetate N-oxides

To a solution of the corresponding triarylmethane acetate (1 eq.) in anhydrous dichloromethane (0.22 M), was added in one portion *m*-chloroperbenzoic acid (3 eq.). The suspension was stirred at room temperature. The reaction medium became a clear solution and then pale yellow milky one. The reaction progress was monitored by GC-MS and TLC (eluent CyHex/EtOAc 60/40). At the end of the reaction, the reaction mixture was neutralized by a 40% aqueous solution of KOH (pH 7−8) then diluted with distilled water and extracted with dichloromethane four times. The combined organic phases were dried over anhydrous Na$_2$SO$_4$, filtered and concentrated. The crude product was purified by FCC on silica gel (eluent DCM/MeOH 97/3) to afford the corresponding *N*-oxides: Yields: **6a** (43%), **6b** (88%), **6c** (66%), **6d** (38%), and **6e** (80%).

4.1.5.1. 2-((4-acetoxyphenyl) (4-(tert-butyl)phenyl)methyl)pyridine 1-oxide (6a). The reaction was performed during 2 h at 20 °C. Yield: 45 mg, 0.12 mmol, light yellow oil, 43%. **TLC** DCM/MeOH 97/3, Rf = 0.20; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 1.27 (s, 9H, H$_{20}$), 2.26 (s, 3H, H$_{22}$), 6.13 (s, 1H, H$_1$), 6.98−7.02 (m, 3H, H$_{11}$, H$_{14}$, H$_{18}$), 7.07−7.12 (m, 4H, H$_3$, H$_4$, H$_6$, H$_7$), 7.29−7.38 (m, 4H, H$_9$, H$_{10}$, H$_{15}$, H$_{17}$), 8.27 (dd, 1H, J$_{12\text{-}11}$ = 5.4 Hz, J$_{12\text{-}9}$ = 0.8 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.84 (C$_{22}$), 31.12 (C$_{19}$), 34.17 (C$_{20}$), 48.71 (C$_1$), 121.87 (C$_4$, C$_6$), 124.63 (C$_9$,C$_{10}$), 125.42 (C$_{15}$,C$_{17}$), 126.51 (C$_{11}$), 128.66 (C$_{14}$, C$_{18}$), 129.86 (C$_3$, C$_7$), 137.43 (C$_2$), 138.11 (C$_{13}$), 139.20 (C$_{12}$), 149.13 (C$_8$, C$_{16}$), 152.55 (C$_5$), 169.19 (C$_{21}$); **IR** (ATR) (cm^{-1}) 3054 (νCsp2-H), 2962, 2954 (νCsp3-H), 1759 (νC = O), 1506, 1486, 1431 (νC = C), 1250 (νN-O), 1183 (νC-O), 852 (δCsp2-H *o*-disubst), 762 (δCsp2-H *p*-disubst); **HPLC** purity: 99%, (Hypersil ODS

C18, MeOH/H$_2$O 90/10, flow rate 0.80 mL/min, λ$_{max}$ = 264 nm, R$_t$ = 2.47 min).

4.1.5.2. 2-((4-acetoxyphenyl) (p-tolyl)methyl)pyridine 1-oxide (6b). The reaction was performed during 2 h at 20 °C. Yield: 92 mg, 0.28 mmol, white solid, 88%. Mp = 164−166 °C. **TLC** DCM/MeOH 97/3, Rf = 0.30; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.26 (s, 3H, H$_{21}$), 2.29 (s, 3H, H$_{19}$), 6.12 (s, 1H, H$_1$), 6.96−6.98 (m, 3H, H$_{11}$, H$_{14}$, H$_{18}$), 7.08 (s, 4H, H$_3$, H$_4$, H$_6$, H$_7$), 7.15 (d, 2H, J$_{ortho}$ = 7.8 Hz, H$_{15}$, H$_{17}$), 7.29 (ddd, 1H, J$_{10\text{-}9}$ = 7.7 Hz, J$_{10\text{-}11}$ = 7.8 Hz, J$_{10\text{-}12}$ = 1.30 Hz, H$_{10}$), 7.35 (ddd, 1H, J$_{9\text{-}10}$ = 7 Hz, J$_{9\text{-}11}$ = 6.4, J$_{9\text{-}12}$ = 2.2 Hz, H$_9$), 8.27 (dd, 1H, J$_{12\text{-}10}$ = 6.4 Hz, J$_{12\text{-}9}$ = 1.04 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.61 (C$_{21}$), 20.86 (C$_{19}$), 48.80 (C$_1$), 121.87 (C$_4$, C$_6$), 124.65 (C$_9$, C$_{10}$), 126.55 (C$_{11}$), 128.90 (C$_{14}$,C$_{18}$), 129.24 (C$_3$, C$_7$), 129.85 (C$_{15}$, C$_{17}$), 136.05 (C$_{16}$), 137.47 (C$_2$), 138.16 (C$_{13}$), 139.21 (C$_{12}$), 149.15 (C$_8$), 152.59 (C$_5$), 169.21 (C$_{20}$); **LRMS** (ESI, CV = 30) 356 [M+23]$^+$ (100), 357 [M + H+23]$^+$ (15), 689 [2 M + 23]$^+$ (12); **IR** (ATR) (cm^{-1}) 3071, 3049 (νCsp2-H), 2921 (νCsp3-H), 1756 (νC = O), 1607, 1501, 1488 (νC = C), 1250 (νN-O), 1202 (νC-O), 838 (δCsp2-H *o*-disubst), 769 (δCsp2-H *p*-disubst); **HPLC** purity: 97%, (Hypersil ODS C18, MeOH/H$_2$O 90/10, flow rate 0.80 mL/min, λ$_{max}$ = 263 nm, R$_t$ = 2.38 min).

4.1.5.3. 2-((4-acetoxyphenyl) (4-bromophenyl)methyl)pyridine 1-oxide (6c). The reaction was performed during 2 h 30 min at 20 °C. After purification by FCC, the product was solubilized in dichloromethane and then washed three times with NaHCO$_3$/Na$_2$CO$_3$ aqueous solution (1:1) to remove the *m*-chloroperbenzoic acid residue. Yield: 78 mg, 0.20 mmol, yellow oil, 66%. **TLC** DCM/MeOH 97/3, Rf = 0.36; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.27 (s, 3H, H$_{20}$), 6.14 (s, 1H, H$_1$), 7.10 (dd, 1H, J$_{9\text{-}10}$ = 7.8 Hz, J$_{9\text{-}11}$ = 2.1 Hz, H$_9$), 7.06 (d, 2H, J$_{ortho}$ = 8.4 Hz, H$_{14}$, H$_{18}$), 7.11 (s, 4H, H$_3$, H$_4$, H$_6$, H$_7$), 7.31 (ddd, 1H, J$_{10\text{-}9}$ = 7.8 Hz, J$_{10\text{-}11}$ = 7.7 Hz, J$_{10\text{-}12}$ = 1.3 Hz, H$_{10}$), 7.38 (ddd, 1H, J$_{11\text{-}10}$ = 7.7 Hz, J$_{11\text{-}12}$ = 7.4 Hz, J$_{11\text{-}9}$ = 2.1 Hz, H$_{11}$), 7.54 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_{14}$, H$_{18}$), 8.30 (ddd, 1H, J$_{12\text{-}11}$ = 7.4 Hz, J$_{12\text{-}10}$ = 2.8 Hz, J$_{12\text{-}9}$ = 1 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.85 (C$_{20}$), 48.74 (C$_1$), 120.07 (C$_{16}$), 122.02 (C$_4$, C$_6$), 124.81(C$_{11}$), 124.88 (C$_{10}$), 126.55 (C$_9$), 129.99 (C$_3$, C$_7$), 131.10 (C$_{14}$, C$_{18}$), 131.52 (C$_{15}$, C$_{17}$), 137.45 (C$_2$), 139.26 (C$_{12}$), 139.92 (C$_{13}$), 149.34 (C$_8$), 151.91 (C$_5$), 169.17 (C$_{19}$); **LRMS** (ESI, CV = 30) 420 [M+23]$^+$ (100), 689 [2 M + 23]$^+$ (12); **IR** (ATR) (cm^{-1}) 3083 (νCsp2-H), 1754 (νC = O), 1505, 1484, 1428 (νC = C), 1248 (νN-O), 1201 (νC-O), 850 (δCsp2-H *o*-disubst), 771 (δCsp2-H *p*-disubst); **HPLC** purity: 96%, (Hypersil ODS C18, MeOH/H$_2$O 90/1, flow rate 0,8 mL/min, λ$_{max}$ = 264 nm, R$_t$ = 2.42 min).

4.1.5.4. 2-((4-acetoxyphenyl) (4-chlorophenyl)methyl)pyridine 1-oxide (6d). The reaction was performed during 2 h 30 min at 20 °C. After purification by FCC, the product was solubilized in dichloromethane and then washed three times with NaHCO$_3$/Na$_2$CO$_3$ aqueous solution (1:1) to remove the *m*-chloroperbenzoic acid residue. Yield: 40 mg, 0.11 mmol, yellow oil, 38%. **TLC** DCM/MeOH 97/3, Rf = 0.38; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.27 (s, 3H, H$_{20}$), 6.15 (s, 1H, H$_1$), 6.97 (dd, 1H, J$_{9\text{-}10}$ = 7.8 Hz, J$_{9\text{-}11}$ = 2.1 Hz, H$_9$), 7.11 (d, 2H, J$_{ortho}$ = 8.4 Hz, H$_{14}$, H$_{18}$), 7.11 (s, 4H, H$_3$, H$_4$, H$_6$, H$_7$), 7.32 (ddd, 1H, J$_{10\text{-}9}$ = 7.8 Hz, J$_{10\text{-}11}$ = 7.7 Hz, J$_{10\text{-}12}$ = 1.3 Hz, H$_{10}$), 7.36−7.42 (m, 1H, H$_{11}$), 7.31 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_{14}$, H$_{18}$), 8.29 (ddd, 1H, J$_{12\text{-}11}$ = 7.3 Hz, J$_{12\text{-}10}$ = 2.8 Hz, J$_{12\text{-}9}$ = 1 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.84 (C$_{20}$), 48.67 (C$_1$), 122.02 (C$_4$, C$_6$), 124.81 (C$_{11}$), 124.87 (C$_{10}$), 126.55 (C$_9$), 128.60 (C$_{15}$, C$_{17}$), 129.99 (C$_3$, C$_7$), 130.74 (C$_{14}$, C$_{18}$), 131.54 (C$_{16}$), 137.52 (C$_2$), 139.27 (C$_{12}$), 139.49 (C$_{13}$), 149.34 (C$_8$), 151.98 (C$_5$), 169.17 (C$_{19}$); **IR** (ATR) (cm^{-1}) 3080, 3050 (νCsp2-H), 1755(νC = O), 1506, 1487, 1430 (νC = C), 1251 (νN-O), 1200 (νC-O), 857 (δCsp2-H *o*-disubst), 765 (δCsp2-H *p*-disubst); **HPLC** purity: 96%, (Hypersil ODS C18, MeOH/H$_2$O 90/10, flow rate 0.80 mL/min, λ$_{max}$ = 264 nm, R$_t$ = 2.38 min).

4.1.5.5. 2-((4-acetoxyphenyl) (4-fluorophenyl)methyl)pyridine 1-oxide (6e). The reaction was performed during 4 h 30 min at 20 °C. Yield: 220 mg, 0.65 mmol, yellow solid, 80%. Mp = 226-224 °C. **TLC** DCM/MeOH 97/3, Rf = 0.20; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.26 (s, 3H, H_{20}), 6.17 (s, 1H, H_1), 6.98 (dd, 1H, J_{9-11} = 2 Hz, J_{9-10} = 7.8 Hz, H_9), 7.10 (s, 4H, H_4, H_6, H_{15}, H_{17}), 7.12–7.20 (m, 4H, H_3, H_7, H_{14}, H_{18}), 7.31 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.2 Hz, H_{10}), 7.37 (ddd, 1H, J_{11-10} = J_{11-12} = 7.7 Hz, J_{11-9} = 2.1 Hz, H_{11}), 8.28 (dd, 1H, J_{12-11} = 6.4 Hz, J_{12-9} = 1.0 Hz, H_{12}); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.85 (C_{20}), 48.49 (C_1), 115.42 (d, 2C, J_{C-F} = 20.60 Hz, C_{15}, C_{17}), 121.98 (C_4, C_6), 124.79 (C_{10}, C_{11}), 126.53 (C_9), 129.90 (C_3, C_7), 130.85 (d, 2C, J_{C-F} = 8.30 Hz, C_{14}, C_{18}), 136.59 (d, 1C, J_{C-F} = 2.40 Hz, C_{13}), 137.86 (C_2), 139.26 (C_{12}), 149.28 (C_5), 152.29 (C_8), 161.08 (d, 1C, J_{C-F} = 241.90 Hz, C_{16}), 169.19 (C_{19}); **LRMS** (ESI, CV = 30) 360 (100) [M+23]$^+$, 361 360 (100) [M + H+23]$^+$; **IR** (ATR) (cm^{-1}) 3066 (vCsp2-H), v 2922 (vCsp3-H), 1753 (vC = O), 1603, 1504, 1427 (vC = C), 1275 (vN-O), 1193 (vC-O), 1160 (vC-F), 843 (δCsp2-H *o*-disubst), 765 (δCsp2-H *p*-disubst); **HPLC** purity: 97%, (Hypersil ODS C18, MeOH/H$_2$O 90/10, flow rate 0.80 mL/min, λ_{max} = 264 nm, R_t = 2.25 min).

4.1.6. General procedure for the preparation of o,p-triarylmethanes

To a solution of the corresponding carbinol (1 eq.) and phenol (1.2 eq.) in nitrobenzene (0.40 M) was added dropwise concentrated sulfuric acid (20 eq.) at 0 °C. The reaction progress was monitored by GC-MS and TLC (eluent DCM/MeOH 90/10). After 5 min at 80 °C the reaction was cooled to room temperature and neutralized with a saturated solution of NaHCO$_3$ (pH 7–8), then extracted with ethyl acetate three times. The combined organic phases were dried over anhydrous Na$_2$SO$_4$, filtered and concentrated. The crude residue was purified by FCC on silica gel (eluent gradient DCM, DCM/MeOH 95/5, DCM/MeOH 90/10) to afford the corresponding *o,p*-triarylmethane: Yields: **7a** (98%), **7b** (45%), **7c** (23%), **7d** (80%), and **7e** (64%).

4.1.6.1. 2-((4-(tert-butyl)phenyl) (pyridin-2-yl)methyl)phenol (7a). Yield: 134 mg, 0.42 mmol, beige solid, 98%, Mp = 202−204 °C. **TLC** DCM/MeOH 90/10, Rf = 0.16; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 1.25 (s, 9H, H_{19}), 5.54 (s, 1H, H_1), 6.71 (d, 1H, J_{ortho} = 8.4 Hz, H_6), 7.07 (dd, 1H, J_{4-6} = 2.3 Hz, J_{ortho} = 8.50 Hz, H_4), 7.13 (d, 2H, J_{ortho} = 8.2 Hz, H_{14}, H_{18}), 7.20−7.27 (m, 2H, H_9, H_{11}), 7.29−7.36 (m, 4H, H_{15}, H_{17}, H_5, H_3), 7.74 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 2 Hz, H_{10}), 8.54 (ddd, 1H, J_{12-11} = 4.7 Hz, J_{12-10} = 1,8 Hz, J_{12-9} = 0.9 Hz, H_{12}), 10.38 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 31.13 (C_{19}), 34.05 (C_{20}), 56.84 (C_1), 116.17 (C_6), 121.50 (C_{11}), 123.42 (C_9), 124.97 (C_{15}, C_{17}), 127.24 (C_3, C_5), 128.58 (C_{14}, C_{18}), 130.37 (C_2), 131.42 (C_4), 136.68 (C_{10}), 140.19 (C_{13}), 148.34 (C_{16}), 149.11 (C_{12}), 151.73 (C_7), 162.77 (C_8); **LRMS** (ESI, CV = 30) 318.18 (50) [M+H]$^+$; **IR** (ATR) (cm^{-1}) 3454 (vO-H), 3057 (vCsp2-H), v 2959, 2904 (vCsp3-H), 1489, 1592, 1473 (vC = C), 1168 (vC-O).

4.1.6.2. 2-(pyridin-2-yl (p-tolyl)methyl)phenol (7b). Yield: 53 mg, 0.21 mmol, beige solid, 45%, Mp = 210−212 °C. **TLC** DCM/MeOH 90/10, Rf = 0.20, DCM/MeOH 90/10, Rf = 0.10−0.20; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.26 (s, 3H, H_{19}), 5.54 (s, 1H, H_1), 6.70 (d, 1H, J_{ortho} = 8.4 Hz, H_6), 7.03 (dd, 1H, J_{4-3} = J_{4-5} = 8.5 Hz, J_{4-6} = 2.3 Hz, H_4), 7.05−7.13 (m, 5H, H_{14}, H_{15}, H_{18}, H_{17}, H_3), 7.19−7.25 (m, 2H, H_5, H_{11}), 7.29 (d, 1H, J_{9-10} = 2.3 Hz, H_9), 7.73 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.9 Hz, H_{10}), 8.53 (ddd, 1H, J_{12-11} = 7.5 Hz, J_{12-10} = 4.8 Hz, J_{12-9} = 0.9 Hz, H_{12}), 10.38 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.61 (C_{19}), 56.91 (C_1), 116.28 (C_6), 121.54 (C_9), 123.46 (C_5), 127.30 (C_3), 128.83 (C_2), 128.89 (C_{14}, C_{15}, C_{17}, C_{18}, C_{11}), 130.40 (C_{16}), 131.51 (C_4), 135.22 (C_{13}), 136.71 (C_{10}), 140.27 (C_7), 149.16 (C_{12}), 162.87 (C_8); **LRMS** (ESI, CV = 30) 298 (100) [M+23]$^+$; **IR** (ATR) (cm^{-1}) 3418 (vO-H), v 3052, 3007 (vCsp2-H), 2922 (vCsp3-H), 1590, 1509, 1474 (vC = C), 1166 (vC-O), 830 (δCsp2-H *p*-disubst).

4.1.6.3. 2-((4-bromophenyl) (pyridin-2-yl)methyl)phenol (7c). Yield: 31 mg, 0.09 mmol, orange oil, 23%. **TLC** DCM/MeOH 90/10, Rf = 0.10; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.61 (s, 1H, H_1), 6.73 (d, 1H, J_{ortho} = 8.4 Hz, H_6), 7.06 (ddd, 1H, J_{4-5} = J_{4-3} = 8,4 Hz, J_{4-6} = 2.4 Hz, H_4), 7.14−7.20 (m, 3H, H_{14}, H_{18}, H_5), 7.22−7.29 (m, 2H, H_9, H_{11}), 7.32 (d, 1H, J_{3-5} = 2.30 Hz, H_3), 7.49 (d, 2H, J_{ortho} = 8.5 Hz, H_{15}, H_{17}), 7.75 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.9 Hz, H_{10}), 8.54 (ddd, 1H, J_{12-11} = 4.8 Hz, J_{12-10} = 1.8 Hz, J_{12-9} = 0.9 Hz, H_{12}), 10.40 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 56.26 (C_1), 116.40 (C_6), 119.44 (C_{16}), 121.77 (C_{11}), 123.64 (C_9), 127.33 (C_3, C_5), 130.53 (C_2), 131 (C_{15}, C_{17}), 131.08 (C_{14}, C_{18}), 131.24 (C_4), 136.93 (C_{10}), 142.77 (C_{13}), 149.22 (C_{12}), 151.94 (C_7), 162.07 (C_8); **IR** (ATR) (cm^{-1}) 3403 (vO-H), v 3068, 2926 (vCsp2-H), 2922 (vCsp3-H), 1590, 1486 (vC = C), 1164 (vC-O).

4.1.6.4. 2-((4-chlorophenyl) (pyridin-2-yl)methyl)phenol (7d). Yield: 106 mg, 0.36 mmol, beige solid, 80%, Mp = 200−202 °C. **TLC** DCM/MeOH 90/10, Rf = 0.21; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.65 (s, 1H, H_1), 6.73 (d, 1H, J_{ortho} = 8.4 Hz, H_6), 7.06 (dd, 1H, J_{ortho} = 8.6 Hz, J_{4-6} = 2.5 Hz, H_4), 7.23 (d, 2H, J_{ortho} = 8.4 Hz, H_{14}, H_{18}), 7.26−7.32 (m, 4H, H_9, H_{11}, H_3, H_5), 7.37 (d, 2H, J_{ortho} = 8.5 Hz, H_{15}, H_{17}), 7.80 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.8 Hz, H_{10}), 8.57 (dd, 1H, J_{12-11} = 5.5 Hz, J_{12-10} = 1.8 Hz, H_{12}), 10.41 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 56 (C_1), 116.46 (C_6), 121.98 (C_{11}), 123.83 (C_9), 127.37 (C_3, C_5), 128.22 (C_{15}, C_{17}), 130.56 (C_2), 130.86 (C_{14}, C_{18}), 131.47 (C_4), 132.55 (C_{16}), 137.47 (C_{10}), 142.13 (C_{13}), 148.96 (C_{12}), 152 (C_7), 161.85 (C_8); **IR** (ATR) (cm^{-1}) 3422 (vO-H), 3064 (vCsp2-H), 2957 (vCsp3-H), 1591, 1488, 1450 (vC = C), 1163 (vC-O), 818 (δCsp2-H *p*-disubst), 623 (vC-Cl).

4.1.6.5. 2-((4-fluorophenyl) (pyridin-2-yl)methyl)phenol (7e). Yield: 78 mg, 0.28 mmol, light yellow solid, 64%, Mp = 222−224 °C. **TLC** DCM/MeOH 90/10, Rf = 0.39; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.62 (s, 1H, H_1), 6.71 (d, 1H, J_{ortho} = 8.5 Hz, H_6), 7.06 (dd, 1H, J_{3-4} = J_{4-5} = 8.5 Hz, J_{4-6} = 2.2 Hz, H_4), 7.09−7.16 (m, 2H, H_{14}, H_{18}), 7.22−7.28 (m, 4H, H_9, H_{11}, H_{15}, H_{17}), 7.31 (d, 2H, J_{ortho} = 2.1 Hz, H_3, H_5), 7.75 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.9 Hz, H_{10}), 8.54 (dd, 1H, J_{12-11} = 4.8 Hz, J_{12-10} = 1.9 Hz, H_{12}); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 56.18 (C_1), 114.83 (C_{15}, C_{17}), 115.04 (C_6), 121.70 (C_{11}), 123.57 (C_5, C_3), 127.30 (C_9), 130.52 (C_2), 130.83 (C_{14}, C_{18}), 131.43 (C_4), 136.88 (C_{10}), 139.47 (C_{13}), 149.27 (C_{12}), 157.33 (C_7), 159.54 (C_{16}), 162.46 (C_8); **IR** (ATR) (cm^{-1}) 3424 (vO-H), 3064 (vCsp2-H), 1595, 1506, 1432 (vC = C), 1156 (vC-O), 809 (δCsp2-H *p*-disubst), 750 (δCsp2-H *o*-disubst).

4.1.7. Preparation of benzotriazole-triarylmethanes

A mixture of benzotriazole (5 mmol, 1 eq.), (4-methoxyphenyl) (pyridin-2-yl)methanol **9** (4.6 mmol, 1 eq.) and *p*-toluenesulfonic acid monohydrate (13 mmol, 2.8 eq.) was stirred and refluxed overnight in perfluorooctane (20 mL, bp 104 °C) under argon. The perfluorocarbon fluid was removed on cooling. A solution of methanol saturated with KOH (20 mL) was added to the remaining solid and sonication was applied until solubilization was complete. Methanol was removed under vacuum and water (20 mL) was added. The mixture was extracted with ethyl acetate four times. The combined organic phases were dried over Na$_2$SO$_4$, filtered and concentrated. The crude product was purified by FCC on silica gel (eluent CyHex/EtOAc, 60/40) and afforded the desired compound **10** as a beige solid. The regioisomer 2-((4-methoxyphenyl) (pyridin-2-yl)methyl)-2,3-dihydro-1H-benzo [d] [1−3]triazole **10a** was also isolated and characterized (see SI-1).

4.1.7.1. 1-((4-methoxyphenyl) (pyridin-2-yl)methyl)-1H-benzo[d] [1–3] triazole (10). Yield: 730 mg (50%). Mp = 170–172 °C. **TLC** CyHex/EtOAc 60/40, R_f = 0.20; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 3.78 (s, 3H, CH$_3$), 7.00 (d, 2H, J$_{ortho}$ = 8.8 Hz, H$_4$, H$_6$), 7.32–7.38 (m, 3H, H$_3$, H$_7$, H$_9$), 7.39 (ddd, 1H, J$_{11-10}$ = 7.6 Hz, J$_{11-12}$ = 4.5 Hz, J$_{11-9}$ = 0.9 Hz, H$_{11}$), 7.42 (m, 1H, H$_{Bt3}$), 7.50 (ddd, 1H, J$_{Bt2-Bt1}$ = 8.3 Hz, J$_{Bt2-Bt3}$ = 6.9 Hz, J$_{Bt2-Bt4}$ = 1 Hz, H$_{Bt2}$), 7.61 (dt, 1H, J$_{Bt4-Bt3}$ = 8.4 Hz, H$_{Bt4}$), 7.69 (s, 1H, H$_1$), 7.88 (ddd, 1H, J$_{10-9}$ = 7.8 Hz, J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.8 Hz, H$_{10}$), 8.09 (dd, 1H, J$_{Bt1-Bt2}$ = J$_{Bt4-Bt3}$ = 8.3 Hz, J$_{Bt1-Bt3}$ = J$_{Bt4-Bt2}$ = 1.8 Hz, H$_{Bt1}$), 8.56 (ddd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.8 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 55.12 (C$_{13}$), 66.10 (C$_1$), 111.23 (C$_{Bt4}$), 114.02 (C$_6$, C$_4$), 119.20 (C$_{Bt1}$), 122.71 (C$_{11}$), 123.19 (C$_9$), 123.97 (C$_{Bt2}$), 127.30 (C$_{Bt3}$), 129.12 (C$_{Bt5}$), 130.15 (C$_7$ and C$_3$), 133.08 (C$_2$), 137.30 (C$_{10}$), 145.23 (C$_{Bt6}$), 149.37 (C$_{12}$), 157.44 (C$_5$), 159.14 (C$_8$); **IR** (ATR) (cm^{-1}) 3054, 3005 (νCsp2-H), 2933, 2905 (νCsp3-H), 2837 (νOMe), 1609, 1587, 1510, 1463 (νC = C), 1243 (νC - O); **GC-MS** method 200, R_t = 8.83 min m/z: 316 [M]$^+$(5), 198 [PyCHPhOCH$_3$]$^+$(100), 79 [PyH]$^+$(10); **HPLC**: purity: 95%, (Hypersil 250, 5 μm, A165, Isooctane/EtOAc 70/30, flow rate 1.20 mL/min, λ_{max} = 260 nm, R$_t$ = 12.3 min).

4.1.7.2. 4-((1H-benzo[d] [1–3]triazol-1-yl) (pyridin-2-yl)methyl) phenyl acetate (12). To a previously synthesized solution of 4-((1H-benzo [d] [1–3]triazol-1-yl) (pyridin-2-yl)methyl)phenol **11** (0.19 mmol, 1 eq.) (*see SI-1*) in acetic anhydride (40 mmol, 210 eq.) at 0 °C was slowly added a solution of NaOH 1 M (0.6 mmol, 3.1 eq.). The mixture was stirred for 24 h at room temperature, concentrated under vacuum and water was added (10 mL). The mixture was extracted with ethyl acetate three times then the organic phases were neutralized with saturated NaHCO$_3$ solution (pH ≈ 8), washed with brine and then dried over Na$_2$SO$_4$. The product was purified by FCC (eluent DCM/MeOH, 95/5) to afford **12** as a white solid.

Yield: 33.6 mg (52%). Mp = 172–174 °C. **TLC** DCM/MeOH 95/5, R_f 0.2; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.26 (s, 3H, CH$_3$), 7.17 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 7.31 (d, 1H, J$_{11-10}$ = 7.7 Hz, H$_9$), 7.36 (ddd, 1H, J$_{11-10}$ = 4.8, J$_{11-12}$ = 4.7 Hz, J$_{11-9}$ = 1.1 Hz, H$_{11}$), 7.39–7.42 (m, 1H, H$_{Bt2}$), 7.43 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_3$, H$_7$), 7.47–7.51 (m, 1H, H$_{Bt3}$), 7.64 (dt, 1H, J$_{Bt4-Bt3}$ = 8.4 Hz, H$_{Bt4}$), 7.75 (s, 1H, H$_1$), 7.86 (ddd, 1H, J$_{10-9}$ = 7.7 Hz, J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.8 Hz, H$_{10}$), 8.09 (dd, 1H, J$_{Bt1-Bt2}$ = J$_{Bt4-Bt3}$ = 8.3 Hz, J$_{Bt1-Bt3}$ = J$_{Bt4-Bt2}$ = 1.6 Hz, H$_{Bt1}$), 8.56 (ddd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.85 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.80 (CH$_3$), 65.85 (C$_1$), 111.10 (C$_{Bt4}$), 122.06 (C$_6$, C$_4$), 119.20 (C$_{Bt1}$), 122.84 (C$_9$), 123.37 (C$_{11}$), 124.12 (C$_{Bt2}$), 127.18 (C$_{Bt3}$), 130.02 (C$_7$ and C$_3$), 133.09 (C$_2$), 134.72 (C$_{Bt5}$), 137.44 (C$_{10}$), 145.21 (C$_{Bt6}$), 149.46 (C$_{12}$), 150.31 (C$_5$), 157.00 (C$_8$), 169.09 (C=O); **IR** (ATR) (cm^{-1}) 3054, 3005 (νCsp2-H), 2923, 2904 (νCsp3 - H), 1750 (νC-O), 1610, 1588, 1509 (νC = C), 1245 (νC - O), 835 (δCsp2-H p-disubst), 772 (δCsp2-H o-disubst); **GC-MS**: method 200, R_t = 9.92 min m/z: 344 [M]$^+$(12), 226 [M-Bt]$^+$ (25), 184 [PyCHNH-Ph]$^+$(100); **HRMS**: calcd. for C$_{20}$H$_{16}$N$_4$O$_2$ [M+Na]$^+$ (367.1165), found (367.1165).

4.1.8. Preparation of naphthol triarylmethanes
4.1.8.1. 1-((4-methoxyphenyl) (pyridin-2-yl)methyl)naphthalen-2-ol (14). 2-Naphthol **13** (2.70 mmol, 1.2 eq.) was mixed with (4-methoxyphenyl) (pyridin-2-yl)methanol (2.30 mmol 1.5 eq.) in dichloroethane (2.5 mL) and sulfamic acid (3.45 mmol, 1.5 eq.) under argon. The mixture was heated at 85 °C for 20 h. After reaction completion, the reaction mixture was cooled at room temperature, neutralized with a saturated solution of NaHCO$_3$ (pH ≈ 8) and extracted with dichloromethane three times. The organic layers were dried over MgSO$_4$, filtered and concentrated under vacuum. The product was purified by FCC on silica gel (DCM 100%) to afford **14** as a whitish solid.

Yield: 38.5 mg (49%). Mp = 158–160 °C. **TLC** CyHex/EtOAc 60/

40, R$_f$ 0.2; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) δ 3.67 (s, 3H, CH$_3$), 6.48 (s, 1H, H$_1$), 6.79 (d, 2H, J$_{ortho}$ = 8.7 Hz, H$_4$, H$_6$), 7.43 (d, 2H, J$_{ortho}$ = 8.7 Hz, H$_3$, H$_7$), 7.15 (d, 1H, J$_{N3-N4}$ = 8.8 Hz, H$_{N3}$), 7.25 (ddd, 1H, J$_{N7-N6}$ = 7.9 Hz, J$_{N7-N8}$ = 6.8 Hz, J$_{N7-N9}$ = 0.85 Hz, H$_{N7}$), 7.29 (ddd, 1H, J$_{11-10}$ = 7.5, J$_{11-12}$ = 4.9 Hz, J$_{11-9}$ = 1.1 Hz, H$_{11}$), 7.36 (ddd, 1H, J$_{N8-N9}$ = 8.5, J$_{N9-N8}$ = 7.9 Hz, J$_{N8-N6}$ = 1.3 Hz, H$_{N8}$), 7.51 (d, 1H, J$_{9-10}$ = 7.9 Hz, H$_9$), 7.74 (d, 1H, J$_{N3-N4}$ = 7.9 Hz, N$_{N4}$), 7.80 (d, 1H, J$_{N6-N7}$ = 7.9 Hz, H$_{N6}$), 7.78–7.83 (m, 1H, H$_{10}$), 8.18 (d, 1H, J$_{N9-N8}$ = 7.9 Hz, H$_{N9}$), 8.52 (ddd, 1H, J$_{12-11}$ = 4.9 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.7 Hz, H$_{12}$), 11.13 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 48.08 (C$_1$), 54.92 (CH$_3$), 113.39 (C$_6$, C$_4$), 119.91 (C$_{N3}$), 120.34 (C$_{N1}$), 121.89 (C$_{N9}$), 122.27 (C$_9$), 123.20 (C$_{N7}$), 124.07 (C$_{11}$), 126.16 (C$_{N8}$), 128.50 (C$_{N6}$), 128.57 (C$_{N5}$), 128.90 (C$_{N4}$), 129.24 (C$_7$ and C$_3$), 133.37 (C$_{N10}$), 133.69 (C$_2$), 137.59 (C$_{10}$), 148.10 (C$_{12}$), 153.63 (C$_{N2}$), 157.41 (C$_5$), 163.00 (C$_8$); **GC-MS**: method 180, R_t = 11.64 min m/z: 341 [M]$^+$(100), 324 [M − OH]$^+$(90), 393 [M − 47]$^+$(15); **IR** (ATR) (cm^{-1}) 3031 (νCsp2-H), 2952, 2929 (νCsp3-H), 2834 (νOMe), 1618, 1597, 1507 (νC = C), 1243 (νC-O); **HRMS**: calcd. for C$_{23}$H$_{19}$NO$_2$ [M+H]$^+$ (342.1489), found (342.1489).

4.1.8.2. 1-((4-hydroxyphenyl) (pyridin-2-yl)methyl)naphthalen-2-ol (15). To a solution of **14** (0.6 mmol, 1 eq.) in glacial acetic acid (2.5 mL) under argon, was added 0.6 mL of stabilized hydriodic acid (d = 1.701, 57%, 4.5 mmol, 7.5 eq.). The mixture was refluxed (T = 100 °C) for 5 h 30 then neutralized with a saturated solution of NaHCO$_3$ (pH ≈ 8). Then the mixture was extracted with ethyl acetate (3 × 20 mL), dried over Na$_2$SO$_4$ and filtrated. The filtrate was concentrated under vacuum and purified by FCC on silica gel (CyHex/EtOAc, 60/40) to provide **15** as a yellow solid.

Yield: 70.10 mg (35%). Mp = 218–220 °C. **TLC** CyHex/EtOAc 60/40, R$_f$ 0.45; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 3.67 (s, 3H, CH$_3$), 6.48 (s, 1H, H$_1$), 6.62 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_4$, H$_6$), 6.83 (d, 2H, J$_{ortho}$ = 8.5 Hz, H$_3$, H$_7$), 7.14 (d, 1H, J$_{N3-N4}$ = 7.1 Hz, H$_{N3}$), 7.23–7.31 (m, 2H, H$_{N7}$ and H$_{11}$), 7.37 m, 1H, H$_{N8}$), 7.53 (d, 1H, J$_{9-10}$ = 7.8 Hz, H$_9$), 7.73 (d, 1H, J$_{N3-N4}$ = 7.1 Hz, N$_{N4}$), 7.77–7.83 (m, 1H, H$_{N6}$ and H$_{10}$), 8.19 (d, 1H, J$_{N9-N8}$ = 8.6 Hz, H$_{N9}$), 8.50–8.54 (m, 1H, H$_{12}$), 9.19 (s, 1H, OH), 11.20 (s, 1H, OH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 48.12 (C$_1$), 114.77 (C$_6$, C$_4$), 119.99 (C$_{N3}$), 120.50 (C$_{N1}$), 121.85 (C$_{N9}$), 122.23 (C$_9$), 123.20 (C$_{N7}$), 124.09 (C$_{11}$), 126.12 (C$_{N8}$), 128.48 (C$_{N6}$), 128.56 (C$_{N5}$), 128.80 (C$_{N4}$), 129.14 (C$_7$, C$_3$), 131.92 (C$_{N10}$), 133.42 (C$_2$), 137.59 (C$_{10}$), 148.03 (C$_{12}$), 153.67 (C$_{N2}$), 155.44 (C$_5$), 163.21 (C$_8$); **GC-MS**: method 200, R_t = 9.12 min m/z: 327 [M]$^+$(90), 310 [M − OH]$^+$(100); **IR** (ATR) (cm^{-1}) 3339 (νO-H), 3077, 3016 (νCsp2-H), 1616, 1591, 1511 (νC = C), 1410 (δ C−O), 1227 (νC-O), 801 (δCsp2-H p-disubst), 740 (δCsp2-H o-disubst); **HRMS**: calcd. for C$_{22}$H$_{17}$NO$_2$ [M+H]$^+$ (328.1332), found (328.1332).

4.1.8.3. 4-((2-Acetoxynaphthalen-1-yl) (pyridin-2-yl)methyl)phenyl acetate (16). To a stirred solution of **15** (0.12 mmol, 1 eq.) in acetic anhydride (13 mmol, 110 eq.) at 0 °C, was slowly added a solution of NaOH 1 M (0.3 mmol, 2.6 eq.). The mixture was stirred for 24 h at room temperature, concentrated under vacuum and dissolved in 10 mL of water. The reaction mixture was extracted with ethyl acetate three times, neutralized with saturated NaHCO$_3$ solution (pH ≈ 8). The combined organic phases were washed with brine dried over Na$_2$SO$_4$, filtered and concentrated. FCC on silica gel (CyHex/EtOAc, 50/50) afforded **16** as a transparent oil.

Yield: 47.70 mg (97%). **TLC** CyHex/EtOAc 70/30, R$_f$ 0.60; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 1.84 (s, 3H, CH$_3$), 2.24 (s, 3H, CH$_3$), 3.67, 6.54 (s, 1H, H$_1$), 7.79 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 7.09 (d, 1H, J$_{N3-N4}$ = 7.9 Hz, H$_{N3}$), 7.22 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_3$, H$_7$), 7.20–7.24 (m, 1H, H$_{11}$), 7.30 (d, 1H, J$_{9-10}$ = 8.9 Hz, H$_9$), 7.43–7.51 (m, 2H, H$_{N7}$, H$_{N8}$), 7.72 (ddd, 1H, J$_{10-9}$ = 7.7, J$_{10-11}$ = 7.8 Hz, J$_{10-12}$ = 1.2 Hz, H$_{10}$), 7.92 (d, 1H, J$_{N6-N7}$ = 8.9 Hz, H$_{N6}$), 7.94–7.97 (m, 1H, N$_{N4}$), 8.16–8.21 (m, 1H, H$_9$), 8.47 (ddd, 1H, J$_{12-11}$ = 4.9 Hz, J$_{12-10}$ = 1.9 Hz, J$_{12-}$

$9 = 0.9$ Hz, H_{12}); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 25.66 and 26.05 (CH$_3$), 55.24 (C$_1$), 126.73 (C$_6$, C$_4$), 126.93 (C$_{N3}$),128.32 (C$_9$), 130.19 (C$_{N9}$), 130.56 (C$_{N7}$), 131.80 (C$_{N8}$), 133.76 (C$_{N5}$), 133.85 (C$_{N4}$ and C$_{N6}$), 129.14 (C$_7$, C$_3$), 137.01 (C$_{N1}$), 137.58 (C$_{N10}$), 141.94 (C$_{10}$), 143.90 (C$_2$), 152.00 (C$_{N2}$), 154.05 (C$_5$), 154.15 (C$_{12}$), 167.21 (C$_8$), 173.90 and 174.46 (C=O); **GC-MS**: method 180, R_t = 12.39 min m/z: 411 [M]$^+$(1), 369 [M − Ac]$^+$(3), 252 [PyCHPhNapth]$^+$(100); **IR (ATR) (cm^{-1})** 3059 (νCsp2-H). 2926 (νCsp3-H), 1749 (νC = O), 1572, 1503, 1469 (νC = C), 1180 (νC - O), 811 (δCsp2-H p-disubst). **HRMS:** calcd. for C$_{26}$H$_{21}$NO$_4$ [M+Na]$^+$ (434.1363), found (434.1363).

4.1.9. Preparation of indole triarylmethanes

4.1.9.1. 3-((4-methoxyphenyl) (pyridin-2-yl)methyl)-3a,7a-dihydro-1H-indole (18). Indole **17** (4.7 mmol, 2 eq.) was mixed with (4-methoxyphenyl) (pyridin-2-yl)methanol **9** (2.3 mmol, 1 eq.), dichloroethane (2.5 mL) and sulfamic acid (2.3 mmol, 1 eq.) under argon. The mixture was heated at 85 °C during 20 h. After reaction, the mixture was cooled to room temperature. A solution of methanol saturated with NaOH (20 mL) was added to the remaining solid and sonication was applied until complete solubilization. Methanol was removed under vacuum and water (20 mL) was added. The mixture was extracted with dichloromethane three times. The combined organic phases were dried over Na$_2$SO$_4$, filtered and concentrated. FCC on silica gel (CyHex/EtOAc, 50/50) afforded **18** as a brown solid.

Yield: 59 mg (81%). Mp = 136−138 °C. **TLC**: CyHex/EtOAc 50/50, R$_f$ 0.4; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 3.70 (s, 3H, CH$_3$), 5.69 (s, 1H, H$_1$), 6.84 (d, 2H, J$_{ortho}$ = 8.7 Hz, H$_4$, H$_6$), 6.82−6.88 (m, 2H, H$_{In1}$, H$_{In6}$), 7.0 (ddd, 1H, J$_{In7- In6}$ = 8.0 Hz, J$_{In7- In5}$ = 7.0 Hz, J $_{In7-}$ $_{In4}$ = 1.1 Hz, H$_{In7}$), 7.14 (d, 1H, J$_{11-10}$ = 7.7 Hz, H$_{11}$), 7.20 (ddd, 1H, J$_{9-10}$ = 7.8 Hz, J$_{9-11}$ = 4.9 Hz, J$_{9-12}$ = 1.1 Hz, H$_9$), 7.23 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_3$, H$_7$), 7.29 (dt, 1H, J$_{In4- In5}$ = 8.9 Hz, H$_{In4}$), 7.35 (dt, 1H, N$_{In5}$, J $_{In5-}$ $_{In4}$ = 8.9 Hz, H$_{In5}$), 7.68 (ddd, 1H, J$_{10-9}$ = 7.7 Hz, J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.8 Hz, H$_{10}$), 8.51 (ddd, 1H, J$_{12-11}$ = 4.9 Hz, J$_{12-10}$ = 1.9 Hz, J$_{12-9}$ = 0.8 Hz, H$_{12}$), 10.9 (s, 1H, NH); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 49.08 (C$_1$), 55.92 (CH$_3$),111.49 (C$_{In4}$), 114.56 (C$_{In2}$), 118.50 (C$_{11}$), 119.72 (C$_9$), 120.09 (C$_{In7}$), 120.36 (C$_6$, C$_4$), 122.56 (C$_{In5}$), 122.90 (C$_{In1}$), 122.94 (C$_{In6}$), 128.42 (C$_{In3}$), 130.56 (C$_7$, C$_3$), 135.40 (C$_2$), 136.80 (C$_{10}$), 139.81 (C$_{In8}$), 148.80 (C$_5$), 149.08 (C$_{12}$), 160.89 (C$_8$); **GC-MS:** method 160, R_t = 13.14 min m/z: 314 [M]$^+$(90), 299 [M − 15]$^+$(10), 236 [InCHPhOCH$_3$]$^+$(100); **IR (ATR) (cm^{-1})** 3412 (νN-H). 3052, 3009 (νCsp2-H), 2920 (νCsp3-H), 2857 (νOMe), 1592, 1507, 1456 (νC = C), 1246 (νC - O), 743 (δCsp2-H o-disubst); **HRMS:** calcd. for C$_{21}$H$_{18}$N$_2$O [M+H]$^+$ (315.1492), found (315.1492).

4.1.9.2. 4-((3a,7a-dihydro-1H-indol-3-yl) (pyridin-2-yl)methyl) phenyl acetate (19). To a stirred solution of **18a** (1.6 mmol, 1 eq.) in acetic anhydride (183 mmol, 115 eq.) at 0 °C, was slowly added a solution of NaOH 1 M (2.15 mmol, 1.3 eq.). The mixture was stirred for 3 h 30 at room temperature and concentrated under vacuum and dissolved in 10 mL of water. The reaction mixture was extracted with ethyl acetate three times and neutralized with saturated NaHCO$_3$ solution (pH ≈ 8). The combined organic phases were washed with brine, drying over Na$_2$SO$_4$, filtered and concentrated. FCC on silica gel (CyHex/EtOAc, 30/70) afforded **19** as a light brown solid.

Yield: 39 mg (68%). Mp = 152−154 °C. **TLC** CyHex/EtOAc 30/70, R$_f$ 0.60; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.23 (s, 3H, CH$_3$), 5.78 (s, 1H, H$_1$), 6.85−6.87 (m, 1H, H$_{In1}$), 6.88−6.89 (m, 1H, H$_{In6}$), 7.02 (d, 2H, J$_{ortho}$ = 8.7 Hz, H$_4$, H$_6$), 7.03 (dd, 1H, J$_{In7- In6}$ = 8.0 Hz, J$_{In7-In5}$ = 7.0 Hz, J $_{In7-}$ $_{In4}$ = 1.1 Hz, H$_{In7}$), 7.14 (d, 1H, J$_{11-10}$ = 7.9 Hz, H$_{11}$), 7.22 (ddd, 1H, J$_{9-10}$ = 7.5 Hz, J$_{9-11}$ = 4.8 Hz, J$_{9-12}$ = 1.1 Hz, H$_9$), 7.32 (m, 2H, H$_{In4}$, H$_{In5}$), 7.35 (d, 2H, J$_{ortho}$ = 8.7 Hz, H$_3$, H$_7$), 7.71 (ddd, 1H, J$_{10-9}$ = 8.6 Hz, J$_{10-11}$ = 7.6 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.51 (ddd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.8 Hz, H$_{12}$), 10.94 (s, 1H, NH); ^{13}C

NMR (DMSO-d_6, 100 mHz) (δ ppm) 20.80 (CH$_3$), 49.82 (C$_1$), 111.49 (C$_{In4}$), 116.56 (C$_{In2}$), 118.42 (C$_{11}$), 118.72 (C$_9$), 121.09 (C$_{In7}$), 121.36 (C$_6$, C$_4$), 121.56 (C$_{In5}$), 122.91 (C$_{In1}$), 123.94 (C$_{In6}$), 126.42 (C$_{In3}$), 129.56 (C$_7$, C$_3$), 136.40 (C$_2$), 136.70 (C$_{10}$), 140.81 (C$_{In8}$), 148.70 (C$_5$), 149.06 (C$_{12}$), 162.89 (C$_8$), 169.24 (C=O); **GC-MS:** method 180, R_t = 11.76 min m/z: 342 [M]$^+$(90), 300 [M − COCH$_3$]$^+$ (100), 264 [M-PyH]$^+$(15), 222 [PyCHIn]$^+$(100); **IR (ATR) (cm^{-1})** 3408 (νN-H). 3060 (νCsp2 - H), 2989 (νCsp3 - H), 1749 (νC = O), 1589, 1570, 1505 (νC = C), 1223 (νC - O), 743 (δCsp2-H o-disubst); **HRMS:** calcd. for C$_{22}$H$_{18}$N$_2$O$_2$ [M+Na]$^+$ (365.1260), found (365.1260).

4.1.10. Preparation of thiophen triarylmethanes

4.1.10.1. 2-((4-methoxyphenyl) (thiophen-2-yl)methyl)pyridine (21). A mixture of (4-methoxyphenyl) (pyridin-2-yl)methanol **9** (1.3 mmol, 1 eq.), thiophene **20** (13 mmol, 10 eq.) and methanesulfonic acid (0.7 mL, d = 1,48, 10 mmol, 8 eq.) in dichloroethane (6 mL) was placed in a vial. The vial was sealed, and the mixture was stirred and submitted to microwave irradiation for 2 h (300 W power, T = 80 °C). The mixture was neutralized with saturated NaHCO$_3$ solution (pH ≈ 8), extracted with dichloromethane (4 × 30 mL) and dried MgSO$_4$. Concentration under vacuum over SiO$_2$ (10 mL) and purification on FCC (CyHex/EtOAc, 70/30) afforded **21** as greenish oil.

Yield: 80 mg (25%). **TLC** CyHex/EtOAc 70/30, R$_f$ 0.4; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 3.71 (s, 3H, CH$_3$), 5.80 (s, 1H, H$_1$), 6.80 (dt, 1H, J$_{T2-T3}$ = 3.5 Hz, J$_{T2-T4}$ = 2.4 Hz, H$_{T2}$), 6.86 (d, 2H, J$_{ortho}$ = 8.8 Hz, H$_4$, H$_6$), 6.93 (dd, 1H, J$_{T3-T2}$ = 3.5 Hz, J$_{T3-T4}$ = 5.1 Hz, H$_{T3}$), 7.23−7.27 (m, 1H, H$_{T4}$), 7.25 (d, 2H, J$_{ortho}$ = 8.8 Hz, H$_3$, H$_7$), 7.34 (d, 1H, J$_{11-10}$ = 7.8 Hz, H$_{11}$), 7.38 (m, 1H, H$_9$), 7.74 (ddd, 1H, J$_{10-9}$ = 7.7 Hz, J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.51 (ddd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.8 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 52.53 (C$_1$), 54.99 (CH$_3$), 113.64 (C$_4$, C$_6$), 121.10 (C$_{11}$), 122.18 (C$_9$), 125.10 (C$_{T4}$), 125.10 (C$_{T2}$), 126.41 (C$_{T3}$), 129.47 (C$_3$, C$_7$), 135.09 (C$_2$), 136.93 (C$_{10}$), 146.68 (C$_{T1}$), 149.12 (C$_{12}$), 157.90 (C$_5$), 162.89 (C$_8$); **GC-MS:** method 160, R_t = 8.06 min m/z: 281 [M]$^+$(100), 266 [M − CH$_3$]$^+$(25), 203 [ThCHPhOCH$_3$]$^+$(100); **IR (ATR) (cm^{-1})** 3081, 3048 3005 (νCsp2-H), 2962, 2932 (νCsp3-H), 2837 (νOMe), 1604, 1584, 1509 (νC = C), 1242(νC-O), 808 (δCsp2-H p-disubst), 702 (δCsp2-H o-disubst); **HRMS:** calcd. for C$_{17}$H$_{15}$NOS [M+H]$^+$ (282.0947), found (282.0946).

4.1.10.2. 4-(Pyridin-2-yl(thiophen-2-yl)methyl)phenol (22). To a stirred solution of **21** (0.5 mmol, 1 eq.) in anhydrous dichloromethane (7 mL) under argon at 0 °C, was added dropwise a 1 M solution of BBr$_3$ in dichloromethane (2.4 mmol, 5 eq.). After 19 h at room temperature, MeOH (10 mL) was added. The solution was dried over Na$_2$SO$_4$, filtrated and concentrated under vacuum over SiO$_2$ (10 mL). The dry SiO$_2$ powder was loaded onto a silica gel column and eluted (DCM/MeOH, 95/5). Concentration under vacuum afforded **22** as a black oil.

Yield: 80 mg (60%). **TLC** DCM/MeOH 95/5, R$_f$ 0.3; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 5.72 (s, 1H, H$_1$), 6.67 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_4$, H$_6$), 6.80 (dt, 1H, J$_{T2-T3}$ = 3.4 Hz, J$_{T2-T4}$ = 2.1 Hz, H$_{T2}$), 6.93 (dd, 1H, J$_{T3-T2}$ = 3.5 Hz, J$_{T3-T4}$ = 5.1 Hz, H$_{T3}$), 7.11 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_3$, H$_7$), 7.24 (dd, 1H, J$_{T4-T3}$ = 7.5 Hz, J$_{T4-T2}$ = 2.1 Hz, H$_{T4}$), 7.32 (d, 1H, J$_{11-10}$ = 7.9 Hz, H$_{11}$), 7.36 (m, 1H, H$_9$), 7.73 (ddd, 1H, J$_{10-9}$ = 7.7 Hz, J$_{10-11}$ = 7.7 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.53 (ddd, 1H, J$_{12-11}$ = 4.9 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.9 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) (δ ppm) 50.50 (C$_1$), 114.68 (C$_4$, C$_6$), 120.10 (C$_{11}$), 122.20 (C$_9$), 125.24 (C$_{T4}$), 125.36 (C$_{T2}$), 126.40 (C$_{T3}$), 130.47 (C$_3$, C$_7$), 135.45 (C$_2$), 136.83 (C$_{10}$), 146.74 (C$_{T1}$), 149.10 (C$_{12}$), 157.76 (C$_5$), 162.40 (C$_8$); **GC-MS:** method 180, R_t = 6.5 min m/z: 267 [M]$^+$(100); **IR (ATR) (cm^{-1})** 3400 (νC-O), 3001 (νCsp2-H), 1590, 1511 (νC = C), 1249(νC-O), 816 (δCsp2-H p-disubst), 751(δCsp2-H o-disubst); **HRMS:** calcd. for C$_{16}$H$_{13}$NOS [M+H]$^+$ (268.0791), found (268.0790).

4.1.11. Preparation of quinoline triarylmethane

4.1.11.1. (4-methoxyphenyl) (pyridin-2-yl) (quinolin-2-yl)methanol (25). n-Butyllithium in hexane (1.6 M, 2.5 mmol, 1.3 eq.) was added dropwise to a stirred solution of 2-bromoquinoline **24** (2.2 mmol, 1.1 eq.) in anhydrous tetrahydrofuran (2 mL) at $-78\ ^{\circ}$C under argon. After 1 h 30, (4-methoxyphenyl) (pyridin-2-yl)methanone **23** (1.9 mmol, 1 eq.) in dry tetrahydrofuran (3 mL) was added dropwise. After 17 h at room temperature, water (10 mL) was added. The reaction mixture was extracted with ethyl acetate three times, dried over Na_2SO_4, filtered and concentrated. FCC (CyHex/EtOAc, 80/20) afforded **25** as a brown oil.

Yield: 53 mg (81%). **TLC** CyHex/EtOAc 80/20, R_f 0.3; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 3.70 (s, 3H, CH$_3$), 6.85 (d, 2H, J$_{ortho}$ = 8.9 Hz, H$_4$, H$_6$), 7.06 (s, 1H, OH), 7.24 (d, 2H, J$_{ortho}$ = 8.9 Hz, H$_3$, H$_7$), 7.29 (ddd, 1H, J$_{11-10}$ = 7.7 Hz, J$_{11-12}$ = 4.8 Hz, J$_{11-9}$ = 1.1 Hz, H$_{11}$), 7.59 (ddd, 1H, J$_{Q6-Q5}$ = 8 Hz, J$_{Q6-Q7}$ = 6.9 Hz, J$_{Q6-Q8}$ = 1.1 Hz, H$_{Q6}$), 7.72−7.77 (m, 3H, H$_9$, H$_{Q2}$, H$_{Q5}$), 7.82 (ddd, 1H, J$_{10-9}$ = 7.8 Hz, J$_{10-11}$ = 7.4 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 7.94−7.97 (m, 2H, H$_{Q7}$, H$_{Q8}$), 8.32 (d, 1H, J$_{Q3-Q2}$ = 8.0 Hz, H$_{Q3}$), 8.49 (ddd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.8 Hz, J$_{12-9}$ = 0.9 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 54.98 (CH$_3$), 90.63 (C$_1$), 113.01(C$_4$, C$_6$), 121.23 (C$_{11}$), 122.24 and 122.28 (C$_{Q2}$, C$_{Q5}$), 126.67 (C$_{Q8}$), 126.79 (C$_{Q4}$), 127.69 (C$_{Q7}$), 128.49 (C$_{11}$), 128.77 (C$_3$ and C$_7$), 129.70 (C$_9$), 136.15 (C$_{10}$), 136.74 (C$_{Q3}$), 137.86 (C$_2$), 145.16 (C$_{Q9}$), 147.49 (C$_{12}$), 158.15 (C$_8$), 163.81 (C$_{Q1}$), 163.99 (C$_5$); **GC-MS**: method 180, R_t = 10.71 min m/z: 342 [M]$^+$(75), 325 [M − OH]$^+$(75), 128 [C$_9$H$_6$N]$^+$(75); **IR** (ATR) (cm^{-1}) 3408 (νO-H). 3031, 3025 (νCsp2H), 2992 (νCsp3 −H), 2851 (νOMe), 1580, 1587, 1509 (νC = C), 1220(νC-O), 742 (δCsp2-H o-disubst); **HRMS:** calcd. for C$_{22}$H$_{18}$N$_2$O$_2$ [M+H]$^+$ (343.1368), found (343.1367).

4.1.12. Preparation of triarylmethanes bearing trifluoromethyl

4.1.12.1. (4-methoxyphenyl) (pyridin-2-yl) (4-(trifluoromethyl) phenyl)methanol (28). n-Butyllithium in hexane (1.6 M, 2.8 mmol, 1.6 eq.) was added dropwise to a stirred solution of 4-bromoanisole (2.5 mmol, 1.4 eq.) in anhydrous tetrahydrofuran (2 mL) at $-78\ ^{\circ}$C under argon. After 1½ h, pyridin-2-yl (4-(trifluoromethyl)phenyl) methanone **27** (1.7 mmol, 1 eq.) dissolved in dry tetrahydrofuran (1 mL) was added dropwise. After 17 h at room temperature, water (30 mL) was added. Extraction with ethyl acetate (4 × 30 mL), drying over Na_2SO_4, concentration under vacuum and column chromatography (CyHex/EtOAc, 70/30) afforded **28** as a white solid.

Yield: 39 mg (67%). Mp = 98−100 $^{\circ}$C. **TLC** CyHex/EtOAc 70/30, R_f 0.5; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 3.55 (s, 3H, CH$_3$), 6.45 (s, 1H, OH), 6.70 (d, 2H, J$_{ortho}$ = 9.0 Hz, H$_{16}$, H$_{18}$), 6.90 (d, 2H, J$_{ortho}$ = 9.0 Hz, H$_{15}$, H$_{19}$), 7.15 (ddd, 1H, J$_{11-10}$ = 7.6 Hz, J$_{11-12}$ = 4.8 Hz, J$_{11-9}$ = 1.1 Hz, H$_{11}$), 7.30 (d, 2H, J$_{ortho}$ = 8.2 Hz, H$_3$, H$_7$), 7.50 (d, 2H, J$_{ortho}$ = 8.2 Hz, H$_4$, H$_6$), 7.48 (d, 1H, J$_{9-10}$ = 8.1 Hz, H$_9$), 7.60 (ddd, 1H, J$_{10-9}$ = 7.9 Hz, J$_{10-11}$ = 7.5 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.31 (ddd, 1H, J$_{12-11}$ = 4.7 Hz, J$_{12-10}$ = 1. Hz, J$_{12-9}$ = 0.84 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 54.98 (CH$_3$), 80.15 (C$_1$), 112.94 (C$_{16}$, C$_{18}$), 121.30 (C$_9$), 122.04 (C$_{11}$), 124.21 and 124.25 (C$_4$, C$_6$), 126.60 (C$_3$,C$_7$), 128.94 (C$_{15}$, C$_{19}$), 129.10 (CF$_3$), 136.75 (C$_{10}$), 138.60 (C$_5$), 147.91 (C$_{12}$), 151.90 (C$_2$), 137.86 (C$_5$), 158.09 (C$_{17}$), 164.59 (C$_8$); **GC-MS:** method 160, R_t = 8.59 min m/z 359 [M$^+$](100), 281 [HOPhCOHPhCF$_3$]$^+$ (43), 252 [CH$_3$OPhCH$_2$PhCF$_3$]$^+$ (20), 173 [OCPhCF$_3$]$^+$ (65); **IR** (ATR) (cm^{-1}) 3423 (νOH), 3012 (νCsp2-H), 2934 (νCsp3-H), 2854 (νOMe), 1612, 1590, 1525 (νC = C), 1324 (νC-O), 1109 (νC-O-C), 1065 (νC-F), 831 (δCsp2-H *p*-disubst); **HRMS:** calcd. for C$_{20}$H$_{16}$F$_3$NO$_2$ [M+H]$^+$ (360.1206), found (360.1206).

4.1.12.2. 4-(hydroxy(pyridin-2-yl) (4-(trifluoromethyl)phenyl) methyl)phenyl acetate (30). To a stirred solution of 4-(pyridin-2-yl (4-(trifluoromethyl)phenyl)methyl)phenol **29** (0.4 mmol, 1 eq.) in acetic anhydride (44 mmol, 110 eq.) was slowly added a solution of NaOH 1 M (0.5 mmol, 1.3 eq.) at 0 $^{\circ}$C. The mixture was stirred for

4 h at room temperature then neutralized with a saturated solution of NaHCO$_3$ (pH $\approx$ 8). Acetone was added (10 mL) and the mixture was stirred for another 30 min (a white precipitate of sodium acetate formed). The mixture was extracted with ethyl acetate three times, dried with Na_2SO_4, filtrated and concentrated under vacuum. FCC on silica gel (CyHex/EtOAc, 50/50) afforded **30** as a yellow oil.

Yield: 130 mg (87%). **TLC**: CyHex/EtOAc 50/50, R_f 0.45; **^{1}H NMR (DMSO-d_6, 400 mHz)** (δ ppm) 2.30 (s, 3H, CH$_3$), 6.45 (s, 1H, C$_1$), 7.07 (d, 2H, J$_{ortho}$ = 8.6 Hz, H$_{16}$, H$_{18}$), 7.24−7.30 (m, 3H, H$_{11}$, H$_{15}$, H$_{19}$), 7.33 (d, 1H, J$_{9-10}$ = 7.9 Hz, H$_9$), 7.47 (d, 2H, J$_{ortho}$ = 8.4 Hz, H$_3$, H$_7$), 7.67 (d, 2H, J$_{ortho}$ = 8.2 Hz, H$_4$, H$_6$), 7.76 (ddd, 1H, J$_{10-9}$ = 8.6 Hz, J$_{10-11}$ = 7.6 Hz, J$_{10-12}$ = 1.9 Hz, H$_{10}$), 8.57 (ddd, 1H, J$_{12-11}$ = 4.8 Hz, J$_{12-10}$ = 1.9 Hz, J$_{12-9}$ = 0.9 Hz, H$_{12}$); **^{13}C NMR (DMSO-d_6, 100 mHz)** (δ ppm) 20.78 (CH$_3$), 56.65 (C$_1$), 121.72 (C$_{16}$, C$_{18}$), 121.95 (C$_{11}$), 122.93 (C$_9$), 123.83 (C$_4$, C$_6$), 129.73 (C$_3$, C$_7$), 129.79 (C$_{15}$, C$_{19}$), 129.96 (CF$_3$), 136.83 (C$_{10}$), 139.70 (C$_5$), 147.66 (C$_{14}$), 149.04 (C$_2$), 149.36 (C$_{12}$), 161.31 (C$_{17}$), 169.16 (C$_8$), 170.29 (C=O); **GC-MS**: method 180, R_t = 6.96 min m/z 371 [M$^+$] (50), 328 [M − Ac]$^+$ (100); **IR** (ATR) (cm^{-1}) 3051 (νCsp2 −H), 2931 (νCsp3-H), 1755 (ν C=O), 1618, 1588, 1505 (νC = C), 1323 (νC-O), 1108 (νC-O-C), 1066 (νC-F), 823 (δCsp2-H *p*-disubst); **HRMS:** calcd. for C$_{21}$H$_{16}$F$_3$O$_2$ Na [M+23]$^+$ (394.1025), found (394.1026).

4.2. AhR transcriptional activity

4.2.1. Materials

Unless otherwise specified, chemical reagents and biological products for *in vitro* assays were obtained from Gibco, Life Technologies (Thermo Fisher Scientific Inc), MilliporeSigma (Merck KGaA) or InvivoGen. Fungible material was provided by Falcon®, Corning ® or Eppendorf ®.

4.2.2. Cell culture

Cell-Line and luciferase assay system. AhR-Lucia™ Human liver carcinoma HepG2 (AhR-HepG2) reporter cells were obtained from InvivoGen engineered to detect endogenous AhR expression. This cell line is stably transfected with a pSELECT-zeo-Lucia plasmid that contains the resistance marker to the antibiotic Zeocin™. EF-1α/HTLV composite promoter is combined to the elongation factor 1 alpha core promoter and the 5' untranslated region of the Human T-cell Leukemia Virus. The secreted luciferase Lucia™ is expressed by a synthetic reporter gene codon optimized for prolonged mammalian cell expression. The promoter is coupled with the human *Cyp1a1* gene entire regulatory sequence, that contains six XREs. The secreted coelenterazine Lucia™ is a novel luciferase reporter technology that does not involve cells lysis to measure the bioluminescence.

Quality and sterility. Quality control of the reporter activity and guaranteed of mycoplasma-free contamination was provided by the cell line suppliers. Additional routine inspections were conducted as standard quality control procedures to avoid mycoplasma, fungi, yeast and/or viruses' contamination. All the experiments were steered with less than 20 passages after thawing as recommended. The cell culture facilities from the Central Service for Experimental Research (SCSIE) at the University of Valencia where the experiments were conducted have certified proficiency to maintain, subculturing and guarantee aseptic conditions.

Maintenance. AhR-HepG2 cells were handled and cultured according to supplier's information under strict sterility conditions in T-75 culture flasks under an aqueous saturated atmosphere with 5% CO$_2$ at 37 $^{\circ}$C. The growth medium was prepared as follow: to Minimum Essential Medium (MEM) containing non-essential amino acids (NEAA) was added 10% (v/v) of 56 $^{\circ}$C heat-inactivated fetal bovine serum (FBS) and a mixture of Penicillin-Streptomycin (100 U/mL-100 μg/mL). The antimicrobial formula

Normocin™ (0.1 mg/mL) and, after the third passage, the selective antibiotic Zeocin™ (0.2 mg/mL) were supplemented to prevent mycoplasma, bacterial and fungal contamination.

Subculturing. Once the cells reached 85% confluency in the culture flasks, they were rinsed twice with 10 mL of PBS and later detached through the incubation with 3−5 mL of 0.25% trypsin-EDTA during 6 min at 37 °C. After inactivation, the cell suspension was centrifuged at 1200 rpm during 5 min, the supernatant was removed, and the pellet resuspended in fresh medium. To dissociate the clumps during the passages, sterile 10 mL syringes with 18-gauge (18G) needles were used. This last step also guaranteed the accuracy of the cell counting performed by mixing 10 μL of cell suspension with 10 μL of 0.4% Trypan Blue Solution in a chamber slide read in cell counters on the Countess™ II instrument (Invitrogen™, Thermo Fisher Scientific).

4.2.3. Bioassays

Assay medium and seeding. The assay medium used was the growth medium without Normocin™ nor Zeocin™. A volume of 200 μL of cells/well was seeded into 96-wells microplates at a density of 2.0×10^5 cells/mL and incubated overnight prior to treatment.

Treatment. The 32 synthetized TAMs and the commercial TAM-drug bisacodyl were dissolved in DMSO (0.5% final maximum concentration/well). The cells were exposed during 24 h to at least 4 different concentration of the TAMs (0.1−10 μM), depending on their cytotoxicity and/or solubility. In the agonist assay, cells were exposed to 10 μL/well of the tested compounds while in the antagonist assay, cells were treated with 10 μL/well of the tested compounds plus 10 μL/well of the EC_{50} of FICZ.

Cell viability assay. Cell viability was assessed by the 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide (MTT) assay, which is a colorimetric method based on the reduction of the yellowish solution of the MTT tetrazolium salt to form the formazan precipitate, as an identification of the redox potential in metabolically actives cells [50]. After seeding, cells were treated with

coelenterazine substrate for the luciferase reaction was prepared according to suppliers by pouring the lyophilized powder protected from light in sterile water. Finally, 50 μL/well of the QUANTI-Luc solution was added and the light signal produced was immediately measured in a microplate reader (VICTORx3, PerkinElmer Inc., USA).

4.2.4. Activity results

Activity criteria. The capacity of the synthetized TAMs to activate (i.e. act as agonists) and to suppress (i.e. act as antagonists) AhR-mediated transcription was analyzed in terms of magnitude of the effects and based on the concentration at which such effects occurs. Therefore, the fold response induced (compared with vehicle control) was used to inform the magnitude of the effect while half effective/inhibitory concentrations (EC_{50} or IC_{50}) were estimated from dose-response curves of agonist or antagonist activity respectively. In the absence of a regulatory guidance for AhR, the threshold used to identify active from inactive compounds in both agonist and antagonist assays followed the OECD guidelines for ER transactivation, where the induced response is compared with a positive control (PC) [53]. Thus, the maximum response relative to the positive control (RPC_{max}) in this work represented the maximum fold response induced by each TAM (x) compared to the positive control FICZ.

In the AhR-agonist assay, the RPC_{max} was calculated as percentage of the maximum AhR induction of FICZ (in relative light units of the luciferase gene expression), that is expressed as *Fold response PC_{Max}(FICZ)* in Equation (1). Meanwhile, in the antagonist assay, the RPC_{max} was calculated as percentage of the effect induced by the EC_{50} of FICZ that is expressed as *Fold response EC_{50}(FICZ)* in Equation (2).

Compounds showing $RPC_{max} \geq 10\%$ were considered active in the agonist bioassay while compounds with $RPC_{max} \leq 70\%$ were considered active in the antagonist bioassay.

$$Agonist\ RPC_{max}\ (\%) = \frac{Fold\ response\ (x)}{Fold\ response\ PC_{Max}(FICZ)} \times 100 \tag{1}$$

$$Antagonistic\ RPC_{max}\ (\%) = \frac{Fold\ response\ (x)}{Fold\ response\ EC_{50}(FICZ)} \times 100 \tag{2}$$

different concentrations of the tested compounds and incubated during 24−72 h. Then, the medium was discarded and 100 μL/well of 0.5 mg/mL of MTT solution was added. Plates were incubated at 37 °C allowing the transformation, the supernatant was removed, and the formazan crystals were dissolved adding 100 μL/well of DMSO. Finally, the optical density was determined by reading the absorbance at 490 nm using a microplate reader (VICTORx3, PerkinElmer Inc., USA).

AhR-transactivation assay. The induction of AhR-mediated transcriptional activity was measured by transferring 20 μL of the supernatant of stimulated cells to white sterile and flat-bottom 96-wells microplates. The QUANTI-Luc™ assay reagent containing the

On the other hand, when possible, sigmoidal curves of x [log (concentration)] *vs. y* (Fold response) were designed constraining the Hill Slope value to 1.0 and using the *Top* and *Bottom* plateaus in the units of the *y*axis. The concentration of agonist required to provoke a response halfway between the baseline and maximum responses (EC_{50} or IC_{50}) was estimated from Equation (3).

$$y = Bottom + \left(\frac{Top - Bottom}{1 + 10^{(logEC_{50}-x)*Hill\ Slope}} \right) \tag{3}$$

Statistical analysis. All data informed represent means obtained from at least three independent experiments (n = 3) and sextuplicate in all cases. The precision of results was reported through

the standard error of the mean (SEM). All active compounds were re-tested at least once again (n = 4) and a greater number of concentrations was evaluated ($\geq$5). Statistical significance was determined with a one-way analysis of variance (ANOVA) following by Dunnett's post-test for comparison with controls or Bonferroni post-test to compare the studied compounds.

4.3. Computational Studies

4.3.1. Molecular docking simulations

Molecular docking analysis was performed with Autodock Vina [64] as implemented in YASARA [65]. The crystalized protein structure of HIF2α was used for docking analysis (PDB ID 3F1O). The sequence of the PAS-B domain of HIF2α shows the highest level of identity and similarity with AhR among all the PAS identified to date [66]. Hence, it is commonly used as template structure in molecular modeling of AhR ligand binding domain [67]. An additional evaluation of the aforementioned sequence identity and similarity is provided in Section 3, SI-3. The analyzed ligands were the strongest agonist identified **22**, the structurally related compound **21**, and the known AhR ligand/agonist compounds FICZ and TCCD. Besides, a preliminary off-targeting docking analysis of compound 22 with four nuclear receptors (ER, AR, PR and PXR) was performed.

All simulations were performed for the entire target structure making the protein ridig and the ligand compounds totally flexible. Protein Ligand Interaction Profiler server [68] was used to predict the interactions of the best protein/ligand complex for each ligand and molecular graphics and analyses were performed with UCSF Chimera [69].

4.3.2. Druglikeness and ADME profile

The physicochemical parameters and ADME descriptors were predicted using the QikProp v5.9 Panel from Schrödinger software v11.9. Conformational averages from OPLS-AA force field were used for calculations.

Author contribution

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Declaration of competing interest

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

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmech.2020.112777.

References

[1] C.A. Bradfield, E. Glover, A. Poland, Purification and N-terminal amino acid sequence of the Ah receptor from the C57BL/6J mouse, Mol. Pharmacol. 39 (1991) 13–19.

[2] K.W. Bock, Aryl hydrocarbon receptor (AHR): from selected human target genes and crosstalk with transcription factors to multiple AHR functions, Biochem. Pharmacol. 168 (2019) 65–70, https://doi.org/10.1016/j.bcp.2019.06.015.

[3] S.-H. Seok, W. Lee, L. Jiang, K. Molugu, A. Zheng, Y. Li, S. Park, C.A. Bradfield, Y. Xing, Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex, Proc. Natl. Acad. Sci. U. S. A 114 (2017) 5431–5436, https://doi.org/10.1073/pnas.1617035114.

[4] I.A. Murray, A.D. Patterson, G.H. Perdew, Aryl hydrocarbon receptor ligands in cancer: friend and foe, Nat. Rev. Canc. 14 (2014) 801–814, https://doi.org/10.1038/nrc3846.

[5] E.J. Wright, K.P. De Castro, A.D. Joshi, C.J. Elferink, Canonical and non-canonical aryl hydrocarbon receptor signaling pathways Toxicology, Curr. Opin. Toxicol. 2 (2017) 87–92, https://doi.org/10.1016/j.cotox.2017.01.001.

[6] L. Stejskalova, Z. Dvorak, P. Pavek, Endogenous and exogenous ligands of aryl hydrocarbon receptor: current state of art, Curr. Drug Metabol. 12 (2011) 198–212, https://doi.org/10.2174/138920011795016818.

[7] L.C. Quattrochi, R.H. Tukey, Nuclear uptake of the Ah (dioxin) receptor in response to omeprazole: transcriptional activation of the human CYP1A1 gene, Mol. Pharmacol. 43 (1993) 504–508.

[8] E.F. O'Donnell, 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, S.K. Kolluri, The anti-inflammatory drug leflunomide is an agonist of the aryl hydrocarbon receptor, PLoS One 5 (2010), https://doi.org/10.1371/journal.pone.0013128.

[9] H.P. Ciolino, P.J. Daschner, G.C. Yeh, Dietary flavonols quercetin and kaempferol are ligands of the aryl hydrocarbon receptor that affect CYP1A1 transcription differentially, Biochem. J. 340 (1999) 715–722, https://doi.org/10.1042/0264-6021:3400715.

[10] T.H. Scheuermann, D.R. Tomchick, M. Machius, Y. Guo, R.K. Bruick, K.H. Gardner, Artificial ligand binding within the HIF2α PAS-B domain of the HIF2 transcription factor, Proc. Natl. Acad. Sci. U. S. A 106 (2009) 450–455, https://doi.org/10.1073/pnas.0808092106.

[11] M.B. Kumar, P. Ramadoss, R.K. Reen, J.P. Vanden Heuvel, G.H. Perdew, The Q-rich subdomain of the human ah receptor transactivation domain is required for dioxin-mediated transcriptional activity, J. Biol. Chem. 276 (2001) 42302–42310, https://doi.org/10.1074/jbc.M104798200.

[12] F.J. Quintana, A.S. Basso, A.H. Iglesias, T. Korn, M.F. Farez, E. Bettelli, M. Caccamo, M. Oukka, H.L. Weiner, Control of Treg and TH17 cell differentiation by the aryl hydrocarbon receptor, Nature 453 (2008) 65–71, https://doi.org/10.1038/nature06880.

[13] A.A. Soshilov, M.S. Denison, Ligand promiscuity of aryl hydrocarbon receptor agonists and antagonists revealed by site-directed mutagenesis, Mol. Cell Biol. 34 (2014) 1707–1719, https://doi.org/10.1128/mcb.01183-13.

[14] Y. Xing, M. Nukaya, K.A. Satyshur, L. Jiang, V. Stanevich, E.N. Korkmaz, L. Burdette, G.D. Kennedy, Q. Cui, C.A. Bradfield, Identification of the Ah-receptor structural determinants for ligand preferences, Toxicol. Sci. 129 (2012) 86–97, https://doi.org/10.1093/toxsci/kfs194.

[15] D. Dolciami, M. Gargaro, B. Cerra, G. Scalisi, L. Bagnoli, G. Servillo, M.A. Della Fazia, P. Puccetti, F.J. Quintana, F. Fallarino, A. Macchiarulo, Binding mode and structure–activity relationships of ITE as an aryl hydrocarbon receptor (AhR) agonist, ChemMedChem 13 (2018) 270–279, https://doi.org/10.1002/cmdc.201700669.

[16] K.N. Chitrala, X. Yang, P. Nagarkatti, M. Nagarkatti, Comparative analysis of interactions between aryl hydrocarbon receptor ligand binding domain with its ligands: a computational study, BMC Struct. Biol. 18 (2018), https://doi.org/10.1186/s12900-018-0095-2.

[17] E. Goya-Jorge, T.Q. Doan, M.L. Scippo, M. Muller, R.M. Giner, S.J. Barigye, R. Gozalbes, Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective, SAR QSAR Environ. Res. 31 (2020) 209–226, https://doi.org/10.1080/1062936X.2019.1708460.

[18] J. Chen, C.A. Haller, F.E. Jernigan, S.K. Koerner, D.J. Wong, Y. Wang, J.E. Cheong, R. Kosaraju, J. Kwan, D.D. Park, B. Thomas, S. Bhasin, R.C. de la Rosa, A.M. Premji, L. Liu, E. Park, A.C. Moss, A. Emili, M. Bhasin, L. Sun, E.L. Chaikof, Modulation of lymphocyte-mediated tissue repair by rational design of heterocyclic aryl hydrocarbon receptor agonists, Sci. Adv. 6 (2020) 1–16, https://doi.org/10.1126/sciadv.aay8230.

[19] M. Mescher, T. Haarmann-Stemmann, Modulation of CYP1A1 metabolism: from adverse health effects to chemoprevention and therapeutic options, Pharmacol. Ther. 187 (2018) 71–87, https://doi.org/10.1016/j.pharmthera.2018.02.012.

[20] K.W. Bock, From TCDD-mediated toxicity to searches of physiologic AHR functions, Biochem. Pharmacol. 155 (2018) 419–424, https://doi.org/10.1016/j.bcp.2018.07.032.

[21] C. Esser, B.P. Lawrence, D.H. Sherr, G.H. Perdew, A. Puga, R. Barouki, X. Coumoul, Old receptor, new tricks—the ever-expanding universe of aryl hydrocarbon receptor functions. Report from the 4th AHR meeting, 29–31 August 2018 in Paris, France, Int. J. Mol. Sci. 19 (2018), https://doi.org/10.3390/ijms19113603.

[22] S. Zhang, P. Lei, X. Liu, X. Li, K. Walker, L. Kotha, C. Rowlands, S. Safe, The aryl

hydrocarbon receptor as a target for estrogen receptor-negative breast cancer chemotherapy, Endocr. Relat. Canc. 16 (2009) 835–844, https://doi.org/10.1677/erc-09-0054.

[23] K.W. Bock, Human AHR functions in vascular tissue: pro- and anti-inflammatory responses of AHR agonists in atherosclerosis, Biochem. Pharmacol. 159 (2019) 116–120, https://doi.org/10.1016/j.bcp.2018.11.021.

[24] N. Guerrina, H. Traboulsi, D.H. Eidelman, C.J. Baglole, The aryl hydrocarbon receptor and the maintenance of lung health, Int. J. Mol. Sci. 19 (2018), https://doi.org/10.3390/ijms19123882.

[25] C. Duval, E. Blanc, X. Coumoul, Aryl hydrocarbon receptor and liver fibrosis, Curr. Opin. Toxicol. 8 (2018) 8–13, https://doi.org/10.1016/j.cotox.2017.11.010.

[26] M. Puccetti, G. Paolicelli, V. Oikonomou, A. De Luca, G. Renga, M. Borghi, M. Pariano, C. Stincardini, L. Scaringi, S. Giovagnoli, M. Ricci, L. Romani, T. Zelante, Towards targeting the aryl hydrocarbon receptor in cystic fibrosis, Mediat. Inflamm. 2018 (2018), 1601486, https://doi.org/10.1155/2018/1601486, 7 pages.

[27] C. Dietrich, Antioxidant functions of the aryl hydrocarbon receptor, Stem Cell. Int. 2016 (2016), 7943495, https://doi.org/10.1155/2016/7943495, 10 pages.

[28] L. Juricek, X. Coumoul, The aryl hydrocarbon receptor and the nervous system, Int. J. Mol. Sci. 19 (2018) 2504, https://doi.org/10.3390/ijms19092504.

[29] T. Bradshaw, A. Westwell, The development of the antitumour benzothiazole prodrug, phortress, as a clinical candidate, Curr. Med. Chem. 11 (2005) 1009–1021, https://doi.org/10.2174/0929867043455530.

[30] S. Safe, Y. Cheng, U.H. Jin, The aryl hydrocarbon receptor (AhR) as a drug target for cancer chemotherapy, Curr. Opin. Toxicol. 1 (2017) 24–29, https://doi.org/10.1016/j.cotox.2017.01.012.

[31] B. Lamas, J.M. Natividad, H. Sokol, Aryl hydrocarbon receptor and intestinal immunity review-article, Mucosal Immunol. 11 (2018) 1024–1038, https://doi.org/10.1038/s41385-018-0019-2.

[32] J. Gao, K. Xu, H. Liu, G. Liu, M. Bai, C. Peng, T. Li, Y. Yin, Impact of the gut microbiota on intestinal immunity mediated by tryptophan metabolism, Front. Cell Infect. Microbiol. 8 (2018) 1–22, https://doi.org/10.3389/fcimb.2018.00013.

[33] C. Esser, A. Rannug, B. Stockinger, The aryl hydrocarbon receptor in immunity, Trends Immunol. 30 (2009) 447–454, https://doi.org/10.1016/j.it.2009.06.005.

[34] I. Marafini, D. Di Fusco, V. Dinallo, E. Franzè, C. Stolfi, G. Sica, G. Monteleone, I. Monteleone, NPD-0414-2 and NPD-0414-24, two chemical entities designed as aryl hydrocarbon receptor (AHR) ligands, inhibit gut inflammatory signals, Front. Pharmacol. 10 (2019) 1–9, https://doi.org/10.3389/fphar.2019.00380.

[35] P. Tarnow, T. Tralau, A. Luch, Chemical activation of estrogen and aryl hydrocarbon receptor signaling pathways and their interaction in toxicology and metabolism, Expet Opin. Drug Metabol. Toxicol. 15 (2019) 219–229, https://doi.org/10.1080/17425255.2019.1569627.

[36] E.F. O'Donnell, D.C. Koch, W.H. Bisson, H.S. Jang, S.K. Kolluri, The aryl hydrocarbon receptor mediates raloxifene-induced apoptosis in estrogen receptor-negative hepatoma and breast cancer cells, Cell Death Dis. 5 (2014) 1–12, https://doi.org/10.1038/cddis.2013.549.

[37] G. Guedes, Á. Amesty, R. Jiménez-Monzón, J. Marrero-Alonso, M. Díaz, L. Fernández-Pérez, A. Estévez-Braun, Synthesis of 4,4′-diaminotriphenylmethanes with potential selective estrogen receptor modulator (SERM)-like activity, ChemMedChem 10 (2015) 1403–1412, https://doi.org/10.1002/cmdc.201500148.

[38] Z. Dvořák, F. Kopp, C.M. Costello, J.S. Kemp, H. Li, A. Vrzalová, M. Štěpánková, I. Bartoňková, E. Jiskrová, K. Poulíková, B. Vyhlídalová, L.U. Nordstroem, C. V Karunaratne, H.S. Ranhotra, K.S. Mun, A.P. Naren, I.A. Murray, G.H. Perdew, J. Brtko, L. Toporova, A. Schön, W.G. Wallace, W.G. Walton, M.R. Redinbo, K. Sun, A. Beck, S. Kortagere, M.C. Neary, A. Chandran, S. Vishveshwara, M.M. Cavalluzzi, G. Lentini, J.Y. Cui, H. Gu, J.C. March, S. Chatterjee, A. Matson, D. Wright, K.L. Flannigan, S.A. Hirota, R.B. Sartor, S. Mani, Targeting the pregnane X receptor using microbial metabolite mimicry, EMBO Mol. Med. 12 (2020) 1–19, https://doi.org/10.15252/emmm.201911621.

[39] S. Mondal, G. Panda, Synthetic methodologies of achiral diarylmethanols, diaryl and triarylmethanes (TRAMs) and medicinal properties of diaryl and triarylmethanes-an overview, RSC Adv. 4 (2014) 28317–28358, https://doi.org/10.1039/c4ra01341g.

[40] J.L. Douglas, M.L. Panis, E. Ho, K.-Y. Lin, S.H. Krawczyk, D.M. Grant, R. Cai, S. Swaminathan, T. Cihlar, Inhibition of respiratory syncytial virus fusion by the small molecule VP-14637 via specific interactions with F protein, J. Virol. 77 (2003) 5054–5064, https://doi.org/10.1128/jvi.77.9.5054-5064.2003.

[41] G. Panda, Shagufta, A.K. Srivastava, S. Sinha, Synthesis and antitubercular activity of 2-hydroxy-aminoalkyl derivatives of diaryloxy methano phenanthrenes, Bioorg. Med. Chem. Lett 15 (2005) 5222–5225, https://doi.org/10.1016/j.bmcl.2005.08.045.

[42] S.K. Chauthe, S.B. Bharate, S. Sabde, D. Mitra, K.K. Bhutani, I.P. Singh, Biomimetic synthesis and anti-HIV activity of dimeric phloroglucinols, Bioorg. Med. Chem. 18 (2010) 2029–2036, https://doi.org/10.1016/j.bmc.2010.01.023.

[43] C. Ricco, F. Abdmouleh, C. Riccobono, L. Gueineche, F. Martin, E. Goya-Jorge, N. Lagarde, B. Liagre, M. Ben Ali, C. Ferroud, M. El Arbi, M.S.I. Veitía, Pegylated

[44] M.S.-I. Veitía, D. Siverio Mota, V. Lerari, M. Marín, R.M. Giner, L. Vicet Muro, Y.R. Guerra, F. Dumas, C. Ferroud, P.A.M. De Witte, A.D. Crawford, V.J. Arán, Y.M. Ponce, Fishing anti-inflammatories from known drugs: in silico repurposing, design, synthesis and biological evaluation of bisacodyl analogues, Curr. Top. Med. Chem. 17 (2017) 2866–2887, https://doi.org/10.2174/1568026617666170817161953.

[45] R.A. Al-Qawasmeh, Y. Lee, M.Y. Cao, X. Gu, A. Vassilakos, J.A. Wright, A. Young, Triaryl methane derivatives as antiproliferative agents, Bioorg. Med. Chem. Lett 14 (2004) 347–350, https://doi.org/10.1016/j.bmcl.2003.11.004.

[46] M. Seto, Y. Aramaki, H. Imoto, K. Aikawa, T. Oda, N. Kanzaki, Y. Iizawa, M. Baba, M. Shiraishi, Orally active CCR5 antagonists as anti-HIV-1 agents 2: synthesis and biological activities of anilide derivatives containing a pyridine N-oxide moiety, Chem. Pharm. Bull. 52 (2004) 818–829, https://doi.org/10.1248/cpb.52.818.

[47] F. Trécourt, G. Breton, V. Bonnet, F. Mongin, F. Marsais, G. Quéguiner, New syntheses of substituted pyridines via bromine-magnesium exchange, Tetrahedron 56 (2001) 1349–1360, https://doi.org/10.1016/S0040-4020(00)00027-2.

[48] M. Sylla-Iyarreta Veitía, C. Rampal, C. Ferroud, An efficient access to unsymmetrical triarylmethanes by regioselective Friedel-Crafts hydroxyalkylation, Trends Org. Chem. 20 (2019) 1–13.

[49] M. Görmen, M.S.I. Veitía, F. Trigui, M. El Arbi, C. Ferroud, Ferrocenyl analogues of bisacodyl: synthesis and antimicrobial activity, J. Organomet. Chem. 794 (2015) 274–281, https://doi.org/10.1016/j.jorganchem.2015.07.016.

[50] J.C. Stockert, R.W. Horobin, L.L. Colombo, A. Blázquez-Castro, Tetrazolium salts and formazan products in Cell Biology: viability assessment, fluorescence imaging, and labeling perspectives, Acta Histochem. 120 (2018) 159–167, https://doi.org/10.1016/j.acthis.2018.02.005.

[51] E. Goya-Jorge, L.E. Carpio, R.M. Giner, M.Sylla-I Veitía, Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic activities, Eur. J. Pharmaceut. Sci. 151 (2020) 105386, https://doi.org/10.1016/j.ejps.2020.105386.

[52] A. Mohammadi-Bardbori, M. Omidi, M.R. Arabnezhad, Impact of CH223191-induced mitochondrial dysfunction on its aryl hydrocarbon receptor agonistic and antagonistic activities, Chem. Res. Toxicol. 32 (2019) 691–697, https://doi.org/10.1021/acs.chemrestox.8b00371.

[53] OECD, Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation in Vitro Assays to Detect Estrogen Receptor Agonists and Antagonists, 2016, https://doi.org/10.1787/9789264265295-en.

[54] H. Kim, S. Reddy, R.F. Novak, 3-Methylcholanthrene and pyridine effects on CYP1A1 and CYP1A2 expression in rat renal tissue, Drug Metab. Dispos. 23 (1995) 818. LP – 824, http://dmd.aspetjournals.org/content/23/8/818.abstract.

[55] P. De Medina, R. Casper, J.F. Savouret, M. Poirot, Synthesis and biological properties of new stilbene derivatives of resveratrol as new selective aryl hydrocarbon modulators, J. Med. Chem. 48 (2005) 287–291, https://doi.org/10.1021/jm0498194.

[56] M.S. Denison, A. Pandini, S.R. Nagy, E.P. Baldwin, L. Bonati, Ligand binding and activation of the Ah receptor, Chem. Biol. Interact. 141 (2002) 3–24, https://doi.org/10.1016/S0009-2797(02)00063-7.

[57] M. Nambo, Z.T. Ariki, D. Canseco-Gonzalez, D.D. Beattie, C.M. Crudden, Arylative desulfonation of diarylmethyl phenyl sulfone with arenes catalyzed by scandium triflate, Org. Lett. 18 (2016) 2339–2342, https://doi.org/10.1021/acs.orglett.6b00744.

[58] R. Pohjanvirta, The AH Receptor in Biology and Toxicology, John Wiley and Sons, 2011, https://doi.org/10.1002/9781118140574.

[59] Á.C. Roman, J.M. Carvajal-Gonzalez, J.M. Merino, S. Mulero-Navarro, P.M. Fernández-Salguero, The aryl hydrocarbon receptor in the crossroad of signalling networks with therapeutic value, Pharmacol. Ther. 185 (2018) 50–63, https://doi.org/10.1016/j.pharmthera.2017.12.003.

[60] S. Safe, M. Wormke, Inhibitory aryl hydrocarbon Receptor–Estrogen receptor α cross-talk and mechanisms of action, Chem. Res. Toxicol. 16 (2003) 807–816, https://doi.org/10.1021/tx034036r.

[61] A.J. Lucas, J.L. Sproston, P. Barton, R.J. Riley, Estimating human ADME properties, pharmacokinetic parameters and likely clinical dose in drug discovery, Expet Opin. Drug Discov. 14 (2019) 1313–1327, https://doi.org/10.1080/17460441.2019.1660642.

[62] A. Leo, C. Hansch, D. Elkins, Partition coefficients and their uses, Chem. Rev. 71 (1971) 525–616, https://doi.org/10.1021/cr60274a001.

[63] D. Dolciami, M. Ballarotto, M. Gargaro, L.C. López-Cara, F. Fallarino, A. Macchiarulo, Targeting Aryl hydrocarbon receptor for next-generation immunotherapies: selective modulators (SAhRMs) versus rapidly metabolized ligands (RMAhRLs), Eur. J. Med. Chem. 185 (2020), https://doi.org/10.1016/j.ejmech.2019.111842.

[64] O. Trott, A. Olson, Software news and update. AutoDock Vina improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2009) 455–461, https://doi.org/10.1002/jcc.21334.

[65] E. Krieger, G. Vriend, YASARA View - molecular graphics for all devices - from smartphones to workstations, Bioinformatics 30 (2014) 2981–2982, https://doi.org/10.1093/bioinformatics/btu426.

[66] S. Giani Tagliabue, S.C. Faber, S. Motta, M.S. Denison, L. Bonati, Modeling the binding of diverse ligands within the Ah receptor ligand binding domain, Sci. Rep. 9 (2019) 1–14, https://doi.org/10.1038/s41598-019-47138-z.

[67] L. Bonati, D. Corrada, S. Giani Tagliabue, S. Motta, Molecular modeling of the AhR structure and interactions can shed light on ligand-dependent activation and transformation mechanisms, Curr. Opin. Toxicol. 1 (2017) 42–49, https://doi.org/10.1016/j.cotox.2017.01.011.

[68] S. Salentin, S. Schreiber, V.J. Haupt, M.F. Adasme, M. Schroeder, PLIP: fully automated protein-ligand interaction profiler, Nucleic Acids Res. 43 (2015) W443–W447, https://doi.org/10.1093/nar/gkv315.

[69] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.

SUPPLEMENTARY INFORMATION 1

Synthesis and characterization of intermediates and non-assayed compounds

Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes

Elizabeth Goya-Jorge [1,2], Celine Rampal [3], Nicolas Loones [3], Stephen J. Barigye [2], Laureano E. Carpio [2], Rafael Gozalbes [2], Clotilde Ferroud [3], Maité Sylla-Iyarreta Veitía[3*], Rosa M. Giner [1**]

[1] *Departament de Farmacologia, Facultat de Farmàcia, Universitat de València. Av. Vicente Andrés Estellés, s/n, 46100 Burjassot, Valencia, Spain.*

[2] *ProtoQSAR SL., CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia. Av. Benjamin Franklin 12, 46980 Paterna, Valencia, Spain.*

[3] *Equipe de Chimie Moléculaire du Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (EA 7528), Conservatoire National des Arts et Métiers (Cnam), 2 rue Conté, 75003, HESAM Université, Paris, France.*

Corresponding authors.

*E-mail (M.S.-I. Veitía): maite.sylla@lecnam.net

**E-mail (RM. Giner): rosa.m.giner@uv.es.

(M.S.-I. Veitía and RM. Giner equally contributed as the last authors)

Table of Contents

I. Synthesis of carbinols

I.1. Method 1. Reaction with *i*-PrMgCl

To a solution of 2-bromopyridine **1** (1 eq) in anhydrous THF (1 M) cooled to -78°C, was added a solution of 1M *i*-propylmagnesium chloride in 2-MeTHF (1 eq). After 2 h at room temperature, the corresponding aldehyde (1.03 eq) was slowly added in portions. The progress of the reaction was monitored by GC/MS. After 20 h of stirring at rt, the medium was hydrolyzed with distilled water (30 vol) and extracted with dichloromethane (or 2-MeTHF for the brominated derivative) (3 x 13 vol). The organic phases were dried over anhydrous Na_2SO_4, filtered and vacuum concentrated to give an oil that crystallized when cold. The precipitated product was filtered and rinsed with CyHex (4 x 3 vol) and after with CyHex/EtOAc 99.5/0.5 (2 x 3 vol), drained and finally vacuum dried. A second batch was obtained by purification of the filtrate using FCC on silica gel.

I.1.1. (4-bromophenyl)(pyridin-2-yl)methanol (3c)

CAS Number: 848605-31-2
(4-bromophenyl)(pyridin-2-yl)methanol
Chemical Formula: $C_{12}H_{10}BrNO$
Exact Mass: 262.99
Molecular Weight: 264.12

1.9 mL, 19.96 mmol   3.76 g, 20.32 mmol   2.19 g, Yield 42%
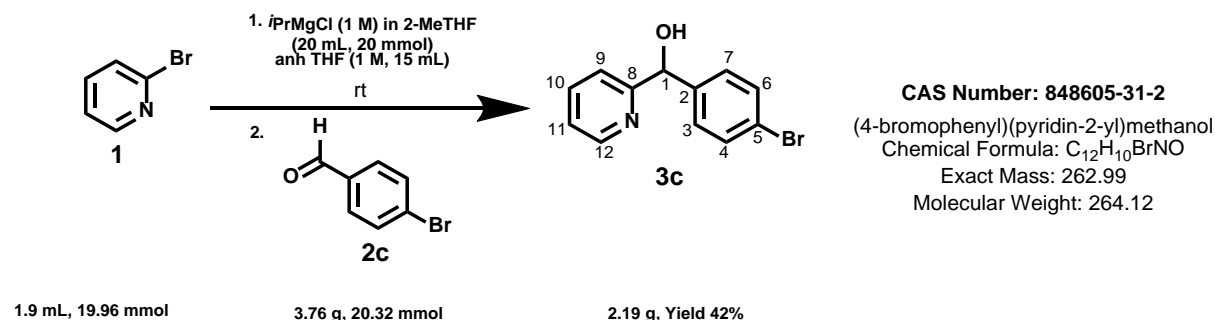
Yield: 2.19 g, 8.28 mmol, ocher solid, 42%. **TLC** CyHex/EtOAc 70/30, R_f = 0.16; **^{1}H NMR** (DMSO d6, δ in ppm), 5.71 (d, 1H, J_{1-OH} = 3.8 Hz, H$_1$), 6.22 (d, 1H, J_{1-OH} = 4.0 Hz, OH), 7.24 (m, 1H, H$_{11}$), 7.36 (d, 2H, J_{ortho} = 8.2 Hz, H$_4$, H$_6$), 7.49 (d, 2H, J_{ortho} = 8.2 Hz, H$_3$, H$_7$), 7.56 (d, 1H, J_{9-10} = 7.8 Hz, H$_9$), 7.77 (dd, 1H, J_{10-9} = J_{10-11} = 7.3 Hz, H$_{10}$), 8.45 (d, 1H, J_{12-11} = 3.9 Hz, H$_{12}$).

I.1.2. pyridin-2-yl(4-(trifluoromethyl)phenyl)methanol (26)

1.*i*PrMgCl (1 M) in 2-MeTHF
(10 mL, 10 mmol)
anh THF (1M, 15 mL)

rt

2.

CAS number 122377-18-8
pyridin-2-yl(4-(trifluoromethyl)phenyl)methanol
Chemical Formula: $C_{13}H_{10}F_3NO$
Exact Mass: 253.07
Molecular Weight: 253.22
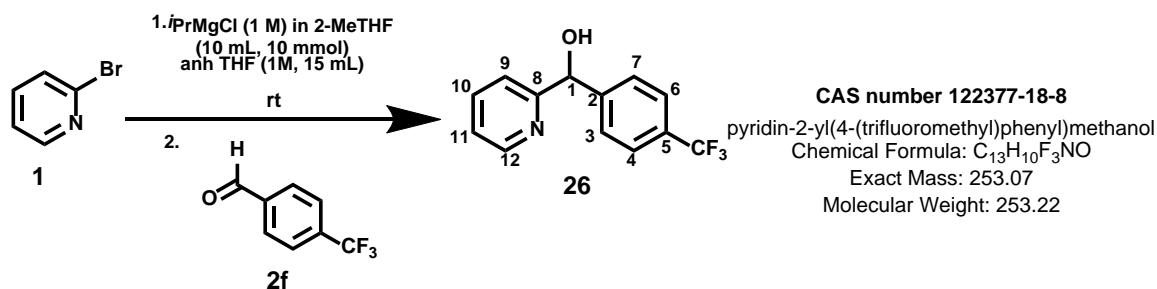
0.95 mL, 9.96 mmol   1.7 g, 9.76 mmol   1.2 g, Yield 48%

Yield: 1.2 g, 4.74 mmol, beige solid, 48%. **TLC** CyHex/EtOAc 50/50, R_f = 0.49, CyHex/EtOAc 70/30, R_f = 0.24; **^{1}H NMR** (DMSO d6, δ in ppm) 5.81 (d, 1H, J_{1-OH} = 4.4 Hz, H_1), 6.32 (d, 1H, J_{1-OH} = 4.5 Hz, OH), 6.47 (d, 2H, J_{ortho} = 8.5 Hz, H4, H6), 7.00 (d, 2H, J_{ortho} = 8.3 Hz, H3, H7), 7.19 (ddd, 1H, J_{9-10} = 7.4 Hz, J_{9-11} = 4.8 Hz, J_{9-12} = 1.2 Hz, H9), 7.52 (d, 1H, J_{11-10} = J_{11-12} = 7.9 Hz, H11), 7.75 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.8 Hz, H10), 8.46 (ddd, 1H, J_{12-11} = 4.8 Hz, J_{12-10} = 1.8 Hz, J_{12-11} = 0.9 Hz, H12); **GC-MS**, R_t = 9.94 min, method 80 (253/100) [M$^+$], (234/25) [PyCHOHPhCF2]$^+$, (236/15) [PyCHPhCF3]$^+$, (145/23) [PhCF3]$^+$, (108/62) [PyCHOH]$^+$, (79/99) [PyH]$^+$.

I.1.3. pyridin-2-yl(4-(trifluoromethyl)phenyl)methanone (27)

(4-trifluorophenyl)-pyridin-2-yl-methanol **26** was added to a solution of NaOH (14.6 mg, 0.366 mmol, 0.3 eq) in toluene (7.1 mL). The medium was brought to reflux (110 °C) for 24 h, then brought to room temperature. The mixture was concentrated, and the crude was purified by FCC on silica gel.

pyridin-2-yl(4-(trifluoromethyl)phenyl)methanone
Chemical Formula: $C_{13}H_8F_3NO$
Exact Mass: 251.06
Molecular Weight: 251.21

NaOH
toluene (7.1 mL)

110 ºC to rt

308 mg, 1.22 mmol, 1eq   14.6 mg, 0.366 mmol, 0.3eq   294 mg, 1.17 mmol Yield 96%
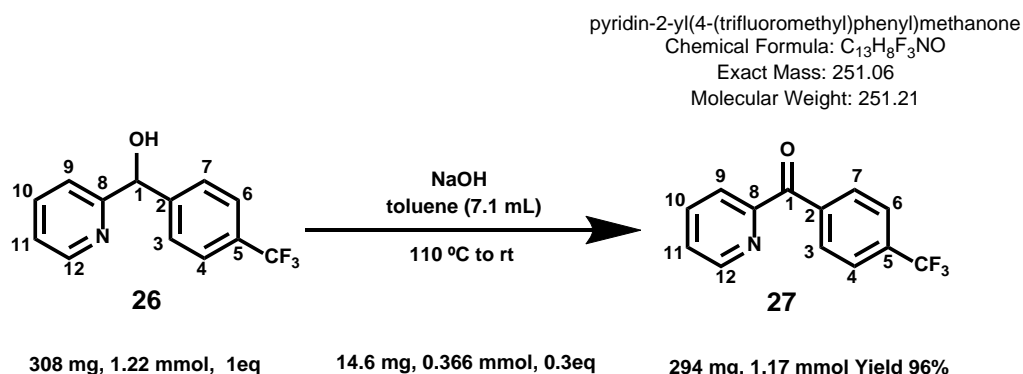
Yield: 294 mg, 1.17 mmol, orange solid, 96%. **TLC** CyHex/EtOAc 60/40, R_f = 0.70; **^{1}H NMR** (DMSO d6, δ en ppm) 7.72 (m, 1H, H11), 7.91 (dd, 2H, J_{ortho} = 8.1 Hz, J_{H-F} = 0.68 Hz, H4, H6), 8.10 (d, 1H, J_{9-11} = 1.28 Hz, H9), 8.11 (dd, 1H, J_{10-12} = 1.6 Hz, J_{10-9} = 2.6 Hz, H10), 8.14 (d, 2H, J_{ortho} = 8 Hz, H3, H7), 8.73 (ddd, 1H, J_{12-11} = 4.7 Hz, J_{12-10} = 1.6

Hz, J_{12-9} = 1.2 Hz, H_{12}). **GC-MS**, R_t = 3.77 min, method 140 (251/28) [M⁺], (173/65) [OCPhCF₃]⁺, (145/92) [PhCF₃]⁺, (145/23) [PhCF₃]⁺, (78/12) [PyH]⁺.

I.1.4. (4-nitrophenyl)(pyridin-2-yl)methanol (3g)

1.*i*PrMgCl (1 M) in 2-MeTHF
(10 mL, 10 mmol)
anh THF (1 M, 15 mL)
rt

2.

1

2g

3g

CAS number: 955-44-2
(4-nitrophenyl)(pyridin-2-yl)methanol
Chemical Formula: $C_{12}H_{10}N_2O_3$
Exact Mass: 230.07
Molecular Weight: 230.22
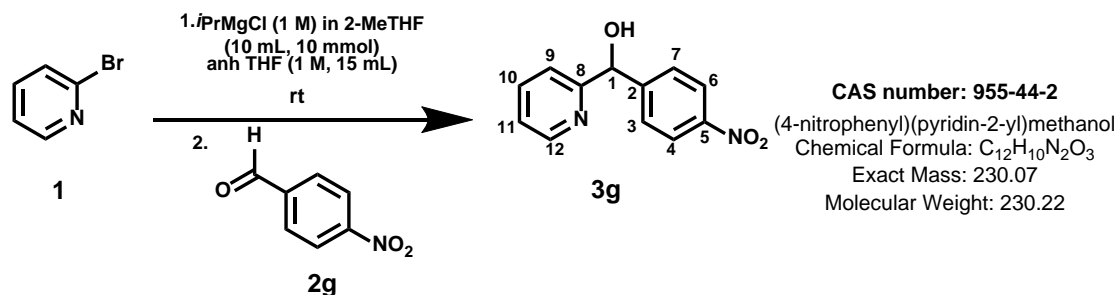
0.95 mL, 9.96 mmol   1.55 g, 10.25 mmol   884 mg, Yield 38%

Yield **3g**: 884 mg, 3.8 mmol, orange solid, 38%. **TLC** CyHex/EtOAc 50/50, R_f = 0.30; **¹H NMR** (DMSO d6, δ in ppm), 5.86 (d, 1H, J_{1-OH} = 4.3 Hz, H_1), 6.43 (d, 1H, J_{1-OH} = 4.4 Hz, OH), 7.25 (dd, 1H, J_{11-10} = 7.2 Hz, J_{11-12} = 5 Hz, H_{11}), 7.59 (d, 1H, J_{9-10} = 7.9 Hz, H_9), 7.69 (d, 2H, J_{ortho} = 8.5 Hz, H_3, H_7), 7.80 (dd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, H_{10}), 8.17 (d, 2H, J_{ortho} = 8.3 Hz, H_4, H_6), 8.46 (d, 1H, J_{12-11} = 4.7 Hz, H_{12}); **¹³C NMR** (DMSO d6, δ in ppm) 74.78 (C_1), 120.23 (C_{11}), 122.53 (C_9), 123.15 (C_4, C_6), 127.56 (C_3, C_7), 137.12 (C_{10}), 144.85 (C_2), 146.51 (C_5), 151.92 (C_{12}), 162.79 (C_8); **GC-MS**, R_t = 14.1 min, method 180 (230/93) [M⁺], (213/10) [PyCHPhNO2]⁺, (108/53) [PyCHOH]⁺, (79/100) [PyH]⁺.

Impurity **3g_(i)**: 298 mg of (4-nitrophenyl)methanol were isolated. **TLC** CyHex/EtOAc 50/50, R_f = 0.40; **¹H NMR** (DMSO, δ in ppm) 4.64 (d, 2H, J_{1-OH} = 5.6 Hz, H_1), 5.54 (t, 1H, J_{1-OH} = 5.7 Hz, OH), 7.59 (d, 2H, J_{4-8} = 8.16 Hz, H_4, H_8), 8.20 (d, 2H, J_{5-7} = 8.6 Hz, H_5, H_7).
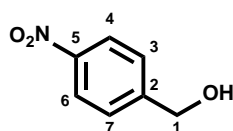
CAS number 619-73-8

(4-nitrophenyl)methanol
Chemical Formula: $C_7H_7NO_3$
Exact Mass: 153.04
Molecular Weight: 153.14

I.2.   Method 2. Reaction with *n*-BuLi

To a solution of 2-bromopyridine **1** (1 eq.) in anhydrous THF (1.3 M) cooled to -78°C, was added a solution of 1.6 M *n*-butyllithium in hexane (1.1 eq). After 1 h 30 min of stirring at -78°C, the corresponding aldehyde (1.03 eq.) is then slowly dispensed. During 17 h the mixture is stirring at room temperature (the progress of the reaction is monitored by GC/MS). Then, the medium is hydrolyzed using distilled water (8 vol.) and extracted with EtOAc (4 x 8 flight.). The organic phases are dried over anhydrous Na_2SO_4, filtered and concentrated in vacuum to obtain an oil that crystallizes by cooling. The precipitated product is then filtered and rinsed with CyHex (4 x 3 vol.), drained and vacuum dried. A second batch is obtained by purification of the filtrate by FCC on silica gel.

I.2.1. 4-(tert-butyl)phenyl)(pyridin-2-yl)methanol (3a)

2.16 mL, 22.68 mmol, 1 eq   3.9 mL, 23.31 mmol, 1.03 eq   3.57 g, 14.79 mmol, Yield 65%
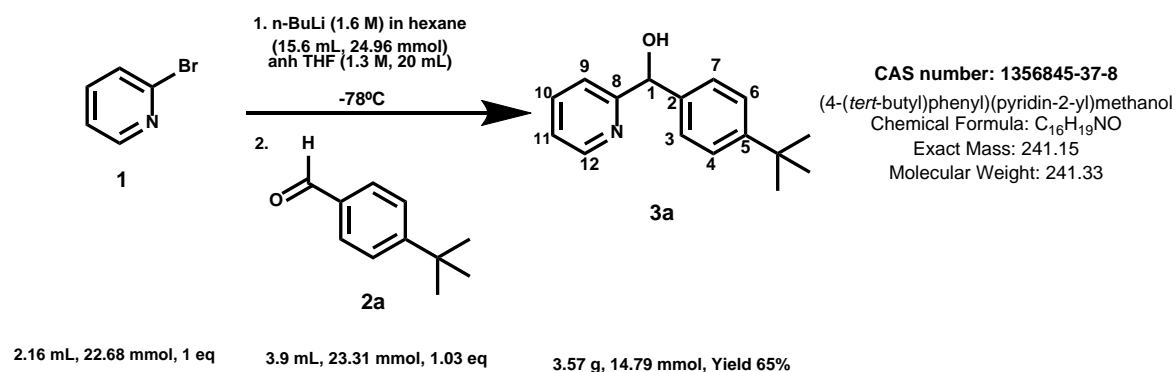
Yield: 3.57 g, 14.81 mmol, white solid, 65%. **TLC** CyHex/EtOAc 70/30, R_f = 0.26, CyHex/EtOAc 60/40, R_f = 0.33; **^{1}H NMR** (DMSO d6, δ in ppm) 1.24 (s, 9H, H_{14}), 5.67 (d, 1H, J_{1-OH} = 4 Hz, H1), 6.01 (d, 1H, J_{1-OH} = 4 Hz, OH), 7.22 (ddd, 1H, J_{11-9} = 1.24 Hz, J_{11-10} = 7.5 Hz, J_{11-12} = 4.8 Hz, H_{11}), 7.31 (s, 4H, H_6, H_7, H_4, H_3), 7.57 (ddd, 1H, J_{9-10} = 7.7 Hz, J_{9-11} = 1.2 Hz, J_{9-12} = 0.9 Hz, H_9), 7.77 (ddd, 1H, J_{10-9} = 7.7 Hz, J_{10-11} = 7.5 Hz, J_{10-12} = 1.8 Hz, H10), 8.44 (ddd, 1H, J_{12-11} = 4.8 Hz, J_{12-10} = 1.8 Hz, J_{12-9} = 0.9 Hz, H_{12}); **GC-MS**, Rt = 12.89 min, method 80 (241/100) [M$^+$], (226/20) [PyCHOHPhC (CH_2) 2]$^+$, (91/22) [PhCH3] $^+$, (108/37) [PyCHOH] $^+$, (184/6) [PyCHOHPh]$^+$.

I.2.2. pyridin-2-yl(p-tolyl)methanol (3b)
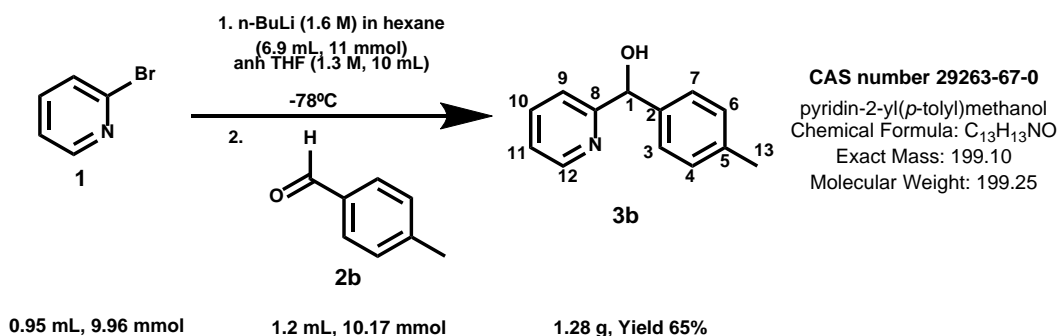
Yield: 1.28 g, 5.58 mmol, beige solid, 65%. **TLC** CyHex/EtOAc 70/30, R_f = 0.18; **^{1}H NMR** (DMSO d6, δ in ppm), 2.24 (s, 3H, H_{13}), 5.67 (d, 1H, J_{1-OH} = 4.4 Hz, H_1), 6.00 (d, 1H, J_{1-OH} = 4.3 Hz, OH), 7.08 (d, 2H, J_{ortho} = 7.9 Hz, H_4, H_6), 7.20 (ddd, 1H, J_{11-10} = 7.4 Hz, J_{11-12} = 4.8 Hz, J_{11-9} = 1.2 Hz, H_{11}), 7.27 (d, 2H, J_{ortho} = 8.0 Hz, H_7, H_3), 7.56 (d, 1H, J_{9-10} = 7.9 Hz, H_9), 7.75 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.8 Hz, H_{10}), 8.43 (ddd, 1H, J_{12-11} = 4.8 Hz, J_{12-10} = 1.8 Hz, J_{12-9} = 0.9 Hz, H_{12}); **GC-MS** R_t = 10.59 min, method 80 199/100 [M$^+$], (91/44) [PhCH$_3$]$^+$, (108/47) [PyCHOH]$^+$, (79/96) [PyH]$^+$.

I.2.3. (4-chlorophenyl)(pyridin-2-yl)methanol (3d)
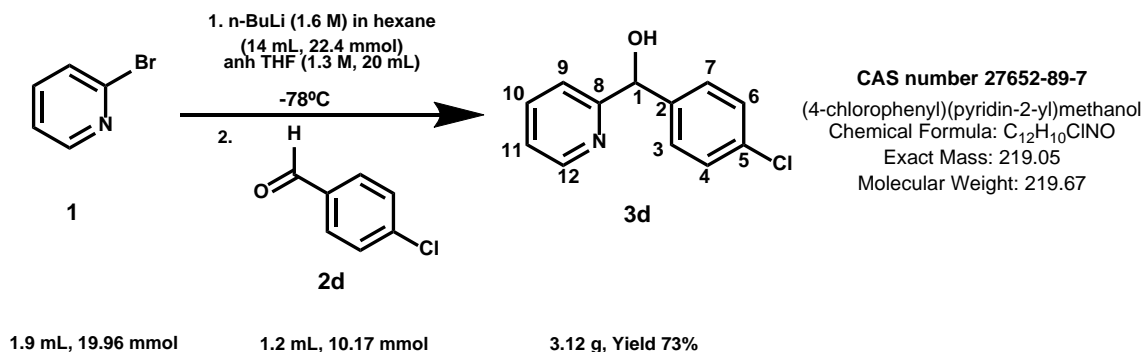
Yield: 3.12 g, 14.2 mmol, off-white solid, 73%. **TLC** CyHex/EtOAc 70/30, R_f = 0.10; **^{1}H NMR** (DMSO d6, δ in ppm) 5.71 (d, 1H, J_{1-OH} = 4.4 Hz, H_1), 6.19 (d, 1H, J_{1-OH} = 4.4 Hz, OH), 7.23 (ddd, 1H, J_{11-10} = 7.4 Hz, J_{11-12} = 4.8 Hz, J_{11-9} = 1.0 Hz, H_{11}), 7.35 (d, 2H, J_{ortho} = 8.6 Hz, H_4, H_6), 7.42 (d, 2H, J_{ortho} = 8.5 Hz, H_7, H_3), 7.56 (d, 1H, J_{9-10} = 7.8 Hz, H_9), 7.77 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 1.8 Hz, H_{10}), 8.45 (ddd, 1H, J_{12-11} = 4.8 Hz, J_{12-10} = 1.7 Hz, J_{12-9} = 0.9 Hz).

I.2.4. (4-fluorophenyl)(pyridin-2-yl)methanol (3e)
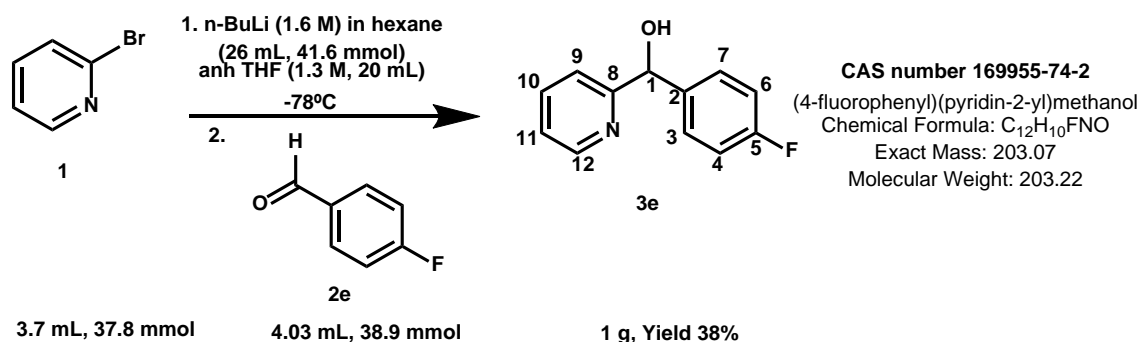
Yield: 1g, 4.93 mmol, red-orange solid, 38%. **TLC** CyHex/EtOAc 60/40, R_f = 0.30; **^{1}H NMR** (DMSO d6, δ in ppm) 5.72 (d, 1H, $J_{1\text{-OH}}$ = 4.2 Hz, H_1), 6.15 (d, 1H, $J_{1\text{-OH}}$ = 4.3 Hz, OH), 7.12 (dd, 2H, J_{ortho} = $J_{H\text{-}F}$ = 8.9 Hz, H_4, H_6), 7.23 (ddd, 1H, $J_{11\text{-}9}$ = 1.2 Hz, $J_{11\text{-}10}$ = 7.5 Hz, $J_{11\text{-}12}$ = 4.9 Hz, H_{11}), 7.43 (dd, 2H, J_{ortho} = 8.4 Hz, $J_{H\text{-}F}$ = 5.6 Hz, H_3, H_7), 7.56 (d, 1H, $J_{9\text{-}10}$ = 7.9 Hz, H_9), 7.78 (ddd, $J_{10\text{-}9}$ = $J_{10\text{-}11}$ = 7.7 Hz, $J_{10\text{-}12}$ = 1.8 Hz, 1H, H_{10}), 8.45 (ddd, 1H, $J_{12\text{-}9}$ = 0.88 Hz, $J_{12\text{-}10}$ = 1.8 Hz, $J_{12\text{-}11}$ = 4.8 Hz, H_{12}); **GC-MS**, R_t = 9.30 min, method 80 (203/71) [M$^+$], (186/30) [PyCHPhF], (108/36) [PyCHOH]$^+$, (79/100) [PyH]$^+$.

I.2.5. (4-methoxyphenyl)(pyridin-2-yl)methanol (9)

n-butyllithium in hexane (1.6 M, 100 mmol, 1.1 eq) was added dropwise to a stirred solution of 2-bromopyridine (87 mmol, 1 eq.) in anhydrous THF (90 mL) at -78°C under argon. After 1 h 30 min, 4-methoxybenzaldehyde **8** (93 mmol, 1.1 eq) was added dropwise. After 17 h at room temperature, water (30 mL) was added. Extraction with ethyl acetate (4 x 30 mL), drying over Na_2SO_4, concentration and crystallization in ethyl acetate afforded (4-methoxyphenyl)(pyridin-2-yl)methanol **9**.
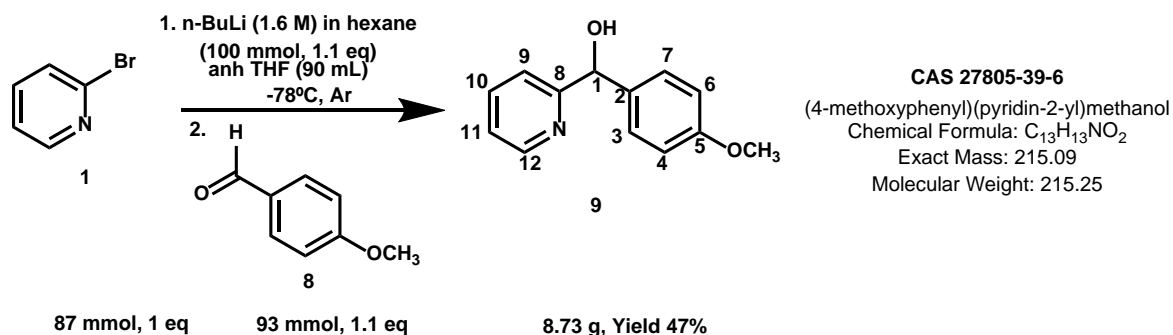
Yield: 8.73 g (47%). R_f = 0.2, SiO_2, (CyHex/EtOAc 50/50); **^{1}H NMR** ([400 MHz, DMSO-d6) δ 8.43 (ddd, 1H, $J_{12\text{-}11}$ = 4.8, $J_{12\text{-}10}$ = 1.8, $J_{12\text{-}9}$ = 0.9 Hz, H12), 7.76 (ddd, 1H, $J_{10\text{-}11}$ = 7.7 $J_{10\text{-}9}$ = 7.7, $J_{12\text{-}10}$ = 1.8 Hz, H10), 7.54 (ddd, 1H, $J_{9\text{-}10}$ = 7.9, $J_{9\text{-}11}$ = 1.2, $J_{9\text{-}12}$ = 0.9

Hz, H9), 7.28 (m, 2H, J_{3-4} or J_{7-6} = 8.4 Hz, H_3 and H_7), 7.21 (ddd, 1H, J_{11-10} = 7.5 J_{11-12} = 4.8, J_{11-9} = 1.2 Hz, H_{11}), 6.84 (d, 2H, J_{4-3} or J_{6-7} = 8.8 Hz, H_4 and H_6), 5.97 (d, 1H, J_{OH-1} = 4.2 Hz, OH), 5.65 (d, J_{1-OH} = 4.2 Hz, H_1), 3.70 (s, 3H, H_{13}); **^{13}C NMR** ([100 MHz, DMSO) δ 165.22 (C_8), 159.16 (C_5), 149.18 (C_{12}), 137.65 (C_{10}), 137.47(C_2), 128.54 (C_7 and C_3), 122.92 (C_9), 120.79 (C_{11}), 114.34 (C_6 and C_4), 76.13(C_1), 55.94(C_{13}); **GC-MS**, R_t = 11.48 min, method 100: (215/100) [M$^+$].

II.   Other triarylmethanes synthesized

II.1.   *p,p-triarylmethane*

2-((4-(tert-butyl)phenyl)(4-methoxyphenyl)methyl)pyridine (31)

2-((4-(*tert*-butyl)phenyl)(4-methoxyphenyl)methyl)pyridine
Chemical Formula: $C_{23}H_{25}NO$
Exact Mass: 331.19
Molecular Weight: 331.46

3a
101 mg, 0.42 mmol

H_2SO_4 **(88 µL, 4 eq), nitrobenzene (1 mL)**
80 °C, Ar

28
48 mg, 0.51 mmol

31
96 mg, Yield 72%
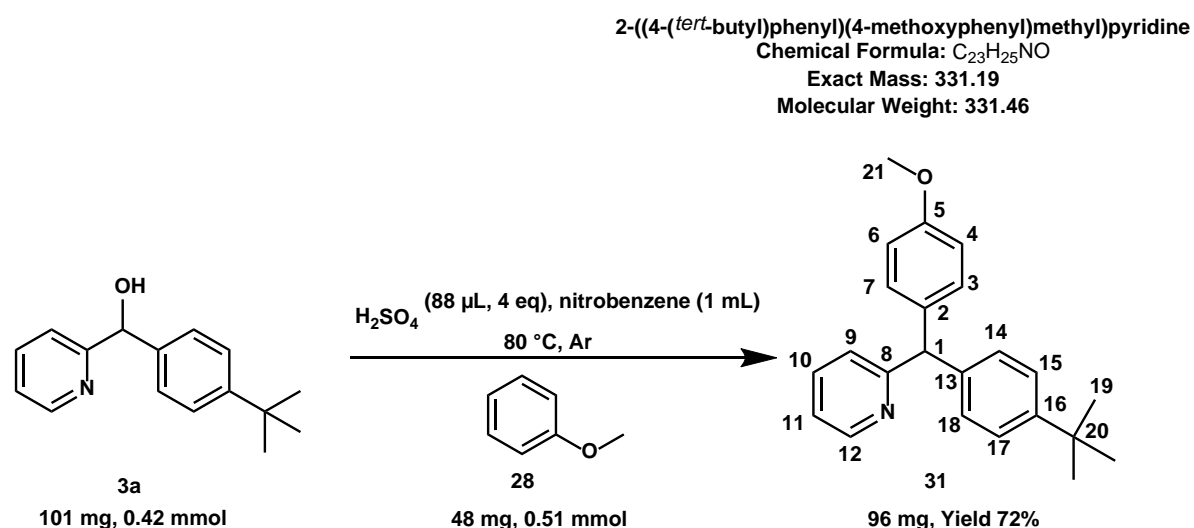
Yield: 98 mg, 0.71 mmol, yellow oil, 47%. **TLC** CyHex/EtOAc 70/30, R_f = 0.16, DCM/MeOH 90/10, R_f = 0.75; **^{1}H NMR** (DMSO d6, δ in ppm), 1.25 (s, 9H, H_{19}), 3.71 (s, 3H, H_{21}), 5.56 (s, 1H, H_1), 6.86 (d, 2H, J_{ortho} = 8.7 Hz, H_4, H_6), 7.09-7.17 (m, 4H, H_7, H_3, H_{14}, H_{18}) , 7.20-7.26 (m, 2H, H_9, H_{11}), 7.31 (d, 2H, J_{ortho} = 8.4 Hz, H_{15}, H_{17}), 7.73 (ddd, 1H, J_{10-9} = J_{10-11} = 7.7 Hz, J_{10-12} = 2 Hz, H_{10}), 8.53 (ddd, 1H, J_{12-11} = 5.6 Hz, J_{12-10} = 4 Hz, J_{12-9} = 1.76 Hz, H_{12}); **^{13}C NMR**(DMSO d6, δ in ppm), 31.16 (C_{19}), 34.09 (C_{20}), 55.0 (C_1), 56.91 (C_{21}), 113.64(C_6, C_4), 121.52 (C_{11}), 123.48 (C_9), 124.98 (C_{15}, C_{17}), 128.60 (C_{14}, C_{18}), 129.99 (C_3, C_7), 135.14 (C_2), 136.69 (C_{10}), 140.45 (C_{13}), 148.35 (C_{16}), 149.13 (C_{12}), 157.66 (C_8), 162.94 (C_5); **GC-MS**, R_t = 8.36 min, method 180: (331/100) [M$^{+\cdot}$], (316/34) [M$^+$-CH$_3$], (300/5) [M$^+$-OCH$_3$], (253/27) [OHPhCHt-BuPh]$^+$, (238/7) [OHPhCHt-BuPh$^+$-CH$_3$], (223/15) [OHPhCHt-BuPh$^+$-2CH$_3$], (208/7) [OHPhCHt-BuPh$^+$ - C(CH$_3$)$_3$], (167/10) [PhCHPy]$^+$.

II.2. benzotriazole-triarylmethanes

2-((4-methoxyphenyl)(pyridin-2-yl)methyl)-1H-benzo[d][1,2,3]triazole (10a)
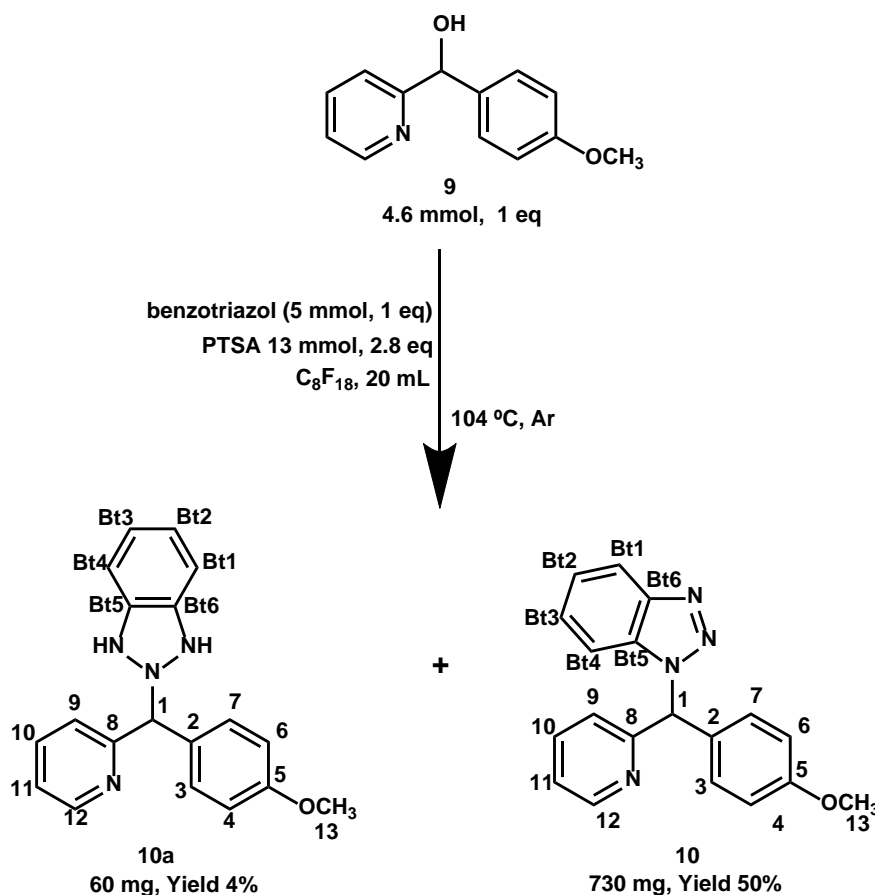
2-((4-methoxyphenyl)(pyridin-2-yl)methyl)-2,3-dihydro-1H^-
benzo[d][1,2,3]triazole
Chemical Formula: $C_{19}H_{18}N_4O$
Exact Mass: 318.15
Molecular Weight: 318.38

Yield: 60 mg, green solid, 4%. **TLC** CyHex/EtOAc 60/40, R_f = 0.4; **¹H NMR** (400 MHz DMSO d6, δ in ppm), 3.78 (s, 3H, H_{13}), 6.96 (d, 2H, J_{4-3} or J_{6-7} = 8.8 Hz, H_4 and H_6), 7.28-7.33 (m, 3H $_{(H11+H7+H3)}$), 7.34-7.40 (m, 2H $_{(H16+H11)}$), 7.46 (ddd, 1H, J_{15-14} = 8.3, J_{15-16} = 6.9, J_{15-17} = 1 Hz, H_{15}), 7.57 (dt, 1H, J_{14-15} = 8.36, J_{14-16} = 1 Hz, H_{14}), 7.65 (s, 1H, H_1), 7.65 (s, 1H, H_1), 7.84 (ddd, 1H, J_{10-9} = 7.8, J_{10-11} = 7.8, J_{10-12} = 1.8 Hz, H_{10}), 7.84 (td, 1H, $J_{10-11+10-9}$ = 7.70, J_{10-12} = 1.8 Hz, H_{10}), 8.06 (dt, 1H, J_{17-16} = 8.3, J_{17-15} = 0.8 Hz, H_{17}), 8.53 (ddd, 1H, J_{12-11} = 4.8, J_{12-10} = 1.8, J_{12-9} = 0.8 Hz, H_{12}); **¹³C NMR** (100 MHz, DMSO) δ, 55.12 (C_{13}), 66.10 (C_1), 111.23 (C_{14}), 114.02 (C_6 and C_4), 119.20 (C_{17}), 122.71 (C_{11}), 123.19 (C_9), 123.97 (C_{16}), 127.30 (C_{15}), 129.12 (C_{19}), 130.15 (C_7 and C_3), 133.08 (C_2), 137.30 (C_{10}), 145.23 (C_{18}), 149.37 (C_{12}), 157.44 (C_5), 159.14 (C_8); **GC-MS**, R_t = 8.83 min, method 200: (316/100) [M⁺].

4-((1H-benzo[d][1,2,3]triazol-1-yl)(pyridin-2-yl)methyl)phenol (11)

To a cold (0°C) stirred solution of 1-((4-methoxyphenyl)(pyridin-2-yl)methyl)-1H-benzo[d][1,2,3]triazole (1.1 mmol, 1 eq.) in anhydrous dichloromethane (6 mL) under argon, was added dropwise a 1 M solution of BBr_3 in dichloromethane (5.2 mmol, 4.7 eq). After 6 h at room temperature, MeOH (50 mL) was added. The mixture was then concentrated under vacuum over SiO_2 (30 mL), purified on flash column chromatography (100% EtOAc) and concentrated under vacuum afforded 4-((1H-benzo[d][1,2,3]triazol-1-yl)(pyridin-2-yl)methyl)phenol **11**.

4-((1*H*-benzo[*d*][1,2,3]triazol-1-yl)(pyridin-2-yl)methyl)phenol
Chemical Formula: $C_{18}H_{14}N_4O$
Exact Mass: 302.1168
Molecular Weight: 302.3370

10
1.1 mmol, 1eq

BBr_3, DCM (5.2 mmol, 4.7 eq)
0 °C to rt, Ar

11
166 g, Yield 50%
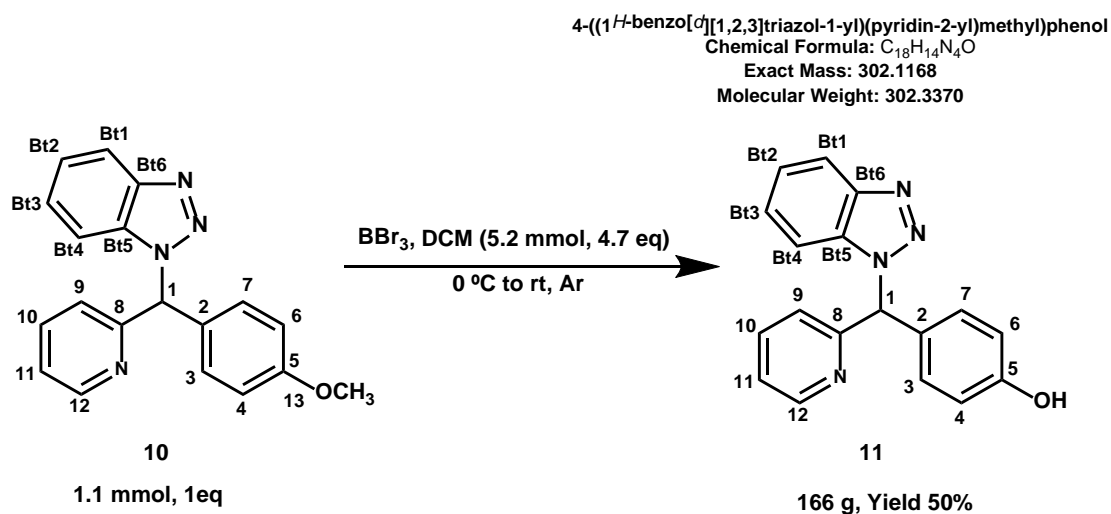
Yield: 166g (50%). R_f = 0.7, SiO_2, (EtOAc 100%); **^{1}H NMR** (400 MHz, DMSO d6, δ in ppm), 6.77 (d, 2H, J_{ortho} = 8.7 Hz, H4, H6), 7.18 (d, 2H, J_{ortho} = 8.7 Hz , H3, H7), 7.30 (d, 1H, $J_{9\text{-}10}$= 8.2 Hz , H9), 7.33-7.39 (m, 2H, H11, HBt3), 7.43-7.45 (m, 1H, HBt2), 7.51-7.55 (m, 1H, HBt4), 7.58 (s, 1H, H1), 7.83 (ddd, 1H, $J_{10\text{-}9}$ = 7.8 Hz, $J_{10\text{-}11}$ = 7.7 Hz, $J_{10\text{-}12}$ = 1.8 Hz, H10), 8.09 (dd, 1H, $J_{Bt1\text{-}Bt2}$ = $J_{Bt4\text{-}Bt3}$ = 8.3 Hz, $J_{Bt1\text{-}Bt3}$ = $J_{Bt4\text{-}Bt2}$ = 1.8 Hz, HBt1), 8.56 (ddd, 1H, $J_{12\text{-}11}$ = 4.8 Hz, $J_{12\text{-}10}$ = 1.8 Hz, $J_{12\text{-}9}$ = 0.8 Hz, H12), 9.62 (s, 1H, HOH); **^{13}C NMR** (100 MHz, DMSO d6, δ in ppm), 55.12 (C13), 66.10 (C1), 111.23 (C14), 114.02 (C6 and C4), 119.20 (C17), 122.71 (C11), 123.19 (C9), 123.97 (C16), 127.30 (C15), 129.12 (C19), 130.15 (C7 and C3), 137.30 (C10), 133.08 (C2), 149.37 (C12), 145.23 (C18), 159.14 (C8), 157.44 (C5); **GC-MS**, R_t = 8.83 min**,** method 200: (316/100) [M$^+$].

II.3. Indole triarylmethanes

4-((1H-indol-3-yl)(pyridin-2-yl)methyl)phenol (18a)

To stirred solution of **18** (1.4 mmol, 1 eq.) in anhydrous dichloromethane (20 mL) under argon at 0°C, was added dropwise a 1 M solution of BBr_3 in dichloromethane (6.6 mmol, 4.7 eq.). After 19 h at room temperature, MeOH (30 mL) was added. The solution was dried over Na_2SO_4, filtrated, and concentrated under vacuum over SiO_2 (10 mL). FCC (CyHex/AcOEt, 60/40) afforded **18a** as a yellow oil.

4-((1*H*-indol-3-yl)(pyridin-2-yl)methyl)phenol
Chemical Formula: $C_{20}H_{16}N_2O$
Exact Mass: 300,13
Molecular Weight: 300,36

1.4 mmol, 1eq
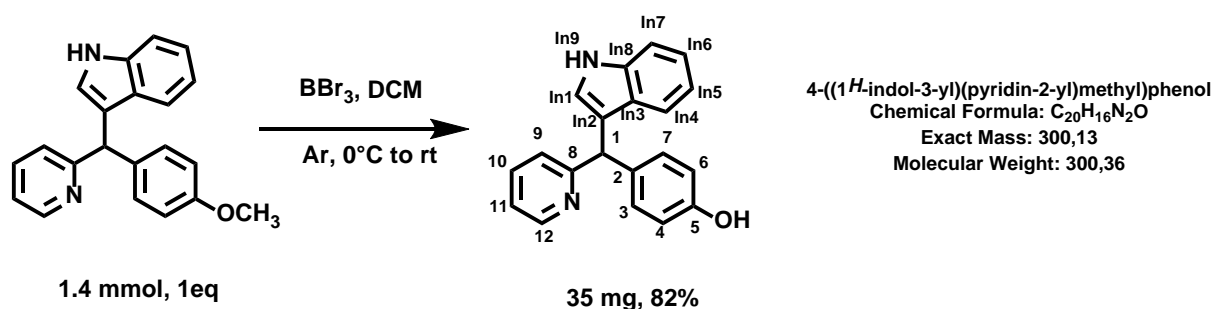
35 mg, 82%

Yield: 35 mg (81%). **TLC**: CyHex/EtOAc 50/50, R_f 0.4; **^{1}H NMR (DMSO d$_6$, 400 mHz)** (δ ppm) 6.10 (s, 1H, H_1), 6.83 (d, 2H, J_{ortho} = 8.7 Hz, H_4, H_6), 6.88- 6.90 (m, 2H, H_{In1}, H_{In6}), 7.0-7.2 (m, 2H, H_{In7} and H_{11}), 7.21 (d, 2H, J_{ortho} = 8.6 Hz, H_3, H_7), 7.41 (ddd, 1H, J_{9-10} = 7.8 Hz, J_{9-11} = 4.9 Hz, J_{9-12} = 0.7 Hz, H_9), 7.77 (dt, 1H, $J_{In4- In5}$ = 8.9 Hz, H_{In4}), 7.90 (dt, 1H, N_{In5}, $J_{In5- In4}$ = 8.9 Hz, H_{In5}), 8.51 (ddd, 1H, J_{10-9} = 7.8 Hz, J_{10-11} = 7.8 Hz, J_{10-12} = 2.1 Hz, H_{10}), 8.80 (ddd, 1H, J_{12-11} = 4.9 Hz, J_{12-10} = 2.1 Hz, J_{12-9} = 0.7 Hz, H_{12}), 11.2 (s, 1H, NH).

II.4. CF$_3$ triarylmethanes

4-(pyridin-2-yl(4-(trifluoromethyl)phenyl)methyl)phenol (29)

To a solution of (4-methoxyphenyl)(pyridin-2-yl)(4-(trifluoromethyl)phenyl)methanol (199.8 mg, 0.56 mmol, 1 eq.) in glacial acetic acid (2.5 mL) under argon was added 1.2 mL of stabilized hydriodic acid d = 1.701, 57%, 1.2 mmol, 2 eq). The mixture was refluxed (T = 120°C) for 24 h and then neutralized with a saturated solution of NaHCO$_3$ (pH ≈ 8). Few grains of $Na_2S_2O_3$ and water was added after until the reddish-yellow colour disappeared. The mixture was extracted with ethyl acetate (3 x 20 mL), dried

with Na_2SO_4 and filtrated. The filtrated was concentrated under vacuum and purified by column chromatography on silica gel (CyHex/EtOAc, 60/40) to provide 4-(pyridin-2-yl(4-(trifluoromethyl)phenyl)methyl)phenol **29**.

4-(pyridin-2-yl(4-(trifluoromethyl)phenyl)methyl)phenol
Chemical Formula: $C_{19}H_{14}F_3NO$
Exact Mass: 329.10
Molecular Weight: 329.32

28

0.56 mmol, 1 eq

HI 57% (1.2 mmol, 2 eq)
120 ºC, Ar

29

150.3 mg, Yield 81 %
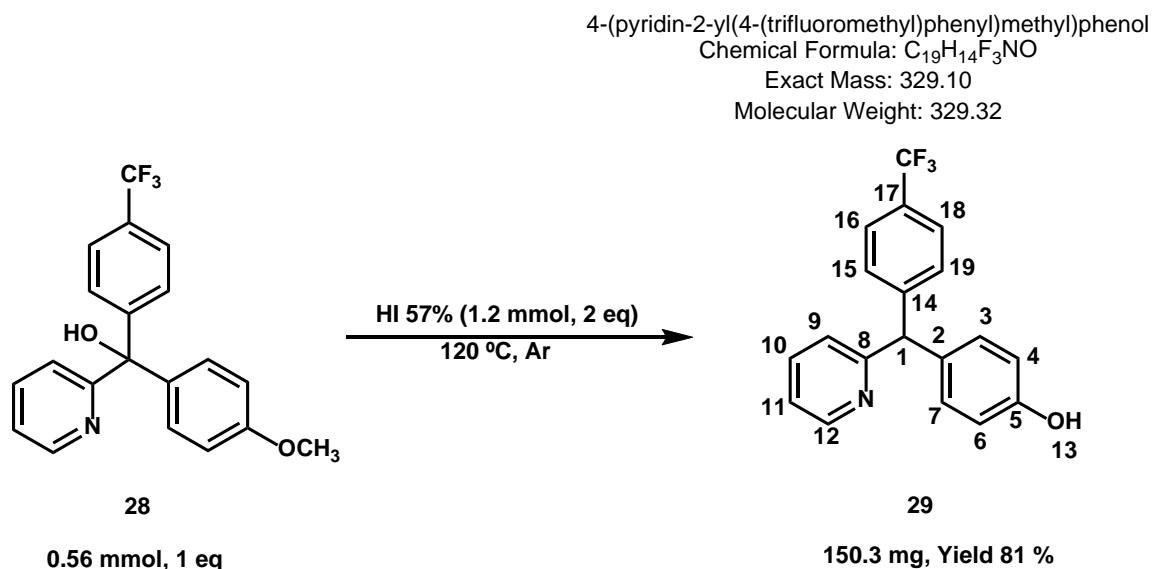
Yield: 150.3 mg, orange solid, 81%. **TLC** CyHex/EtOAc 60/40, R_f = 0.40; **^{1}H NMR** (DMSO d_6, δ ppm), 6.70 (d, 2H, $J_{ortho(16-15)}$ = 8.50 Hz, H_{16}, H_{18}), 7.02 (d, 2H, $J_{ortho(15-16)}$ = 8.50 Hz, H_{15}, H_{19}), 7.25 (m, 2H, H_9, H_{11}), 7.40 (d, 2H, $J_{ortho(4-3)}$ = 8.10 Hz, H_4, H_6), 7.64 (d, 2H, $J_{ortho(3-4)}$ = 8.20 Hz, H_3, H_7), 7.74 (dd, 1H, J_{10-11} = 7.64 Hz, J_{10-9} = 1.72 Hz, H_{10}), 8.53 (d, 1H, J_{12-10} = 3.92 Hz, H_{12}), 9.34 (s, 1H, OH); **GC-MS**, R_t = 6.32 min, method 180: (328/100) [M$^+$], (251/12) [PyCOPhCF$_3$]$^+$, (183/10) (139/2) [HOCPhOCH$_3$]$^+$, (79/3) [PyH]$^+$.

SUPPLEMENTARY INFORMATION 2

Dose-Response Curves

Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes

Elizabeth Goya-Jorge [1,2], Celine Rampal [3], Nicolas Loones [3], Stephen J. Barigye [2], Laureano E. Carpio [2], Rafael Gozalbes [2], Clotilde Ferroud [3], Maité Sylla-Iyarreta Veitía[3*], Rosa M. Giner [1**]

[1] *Departament de Farmacologia, Facultat de Farmàcia, Universitat de València. Av. Vicente Andrés Estellés, s/n, 46100 Burjassot, Valencia, Spain.*

[2] *ProtoQSAR SL., CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia. Av. Benjamin Franklin 12, 46980 Paterna, Valencia, Spain.*

[3] *Equipe de Chimie Moléculaire du Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (EA 7528), Conservatoire National des Arts et Métiers (Cnam), 2 rue Conté, 75003, HESAM Université, Paris, France.*

Corresponding authors.

E-mail (M.S.-I. Veitía): maite.sylla@lecnam.net

**E-mail* (RM. Giner): rosa.m.giner@uv.es.

(M.S.-I. Veitía and RM. Giner equally contributed as the last authors)

Section 1. Dose-response curves of the standard controls for the agonist (FICZ) and antagonist (CH223191) reporter gene bioassays.

FICZ

$EC_{50} = 9.06 \mu M$
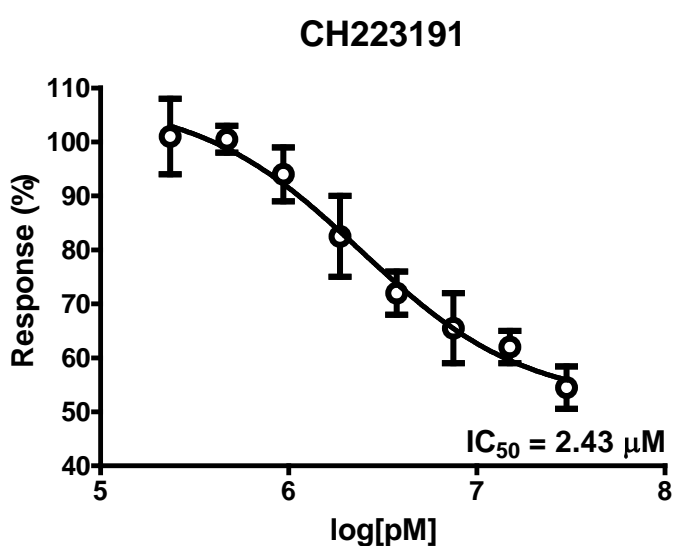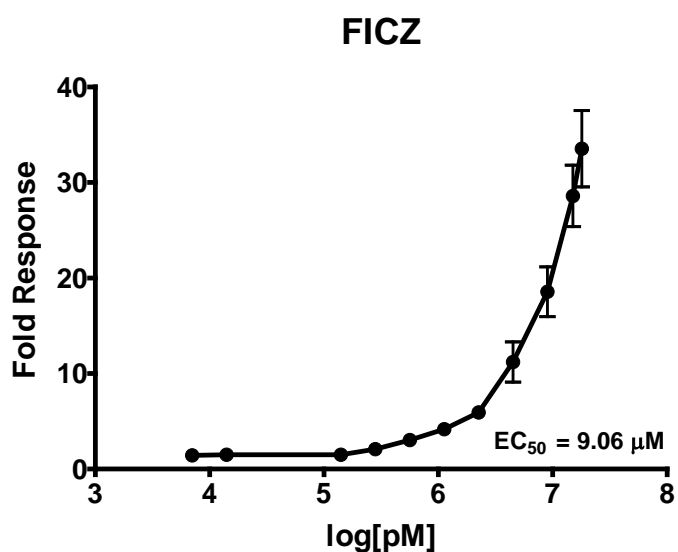
Figure 1. Dose-response curves of the **a)** AhR agonist [6-formylindolo[3,2-b]carbazole (FICZ)] and **b)** AhR antagonist [2-methyl-2H-pyrazole-3-carboxylic acid (CH223191)] controls. The dosage is represented as logarithm of the concentration expressed in pM (1 x 10^{12} M), while responses are:
a) Fold response ± SEM of FICZ
b) % response FICZ [EC_{50}] ± SEM [1]

CH223191

$IC_{50} = 2.43 \mu M$

[1]   E. Goya-Jorge, F. Abdmouleh, L.E. Carpio, R.M. Giner, M. Sylla-Iyarreta Veitía, Discovery of 2-aryl and 2-pyridinylbenzothiazoles endowed with antimicrobial and aryl hydrocarbon receptor agonistic activities, Eur. J. Pharm. Sci. 151 (2020) 105386. https://doi.org/10.1016/j.ejps.2020.105386.

Section 2. Dose-response curves of the agonist triarylmethane compounds discovered in this study.
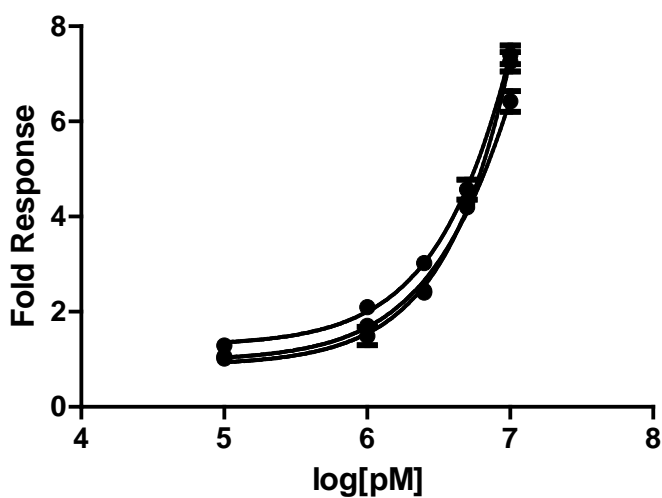
Compound **14**
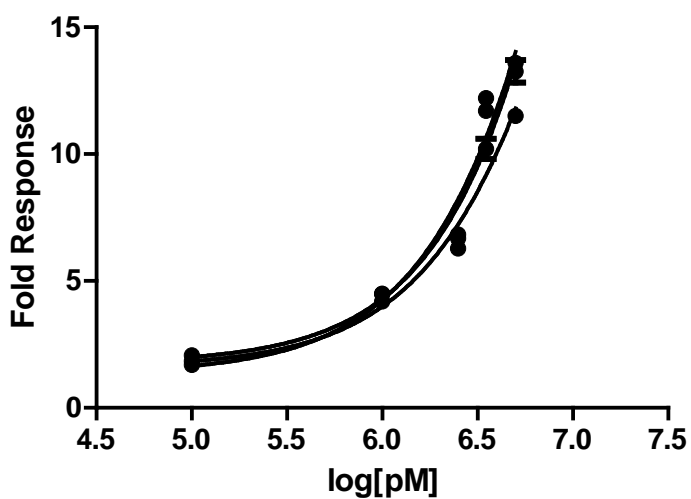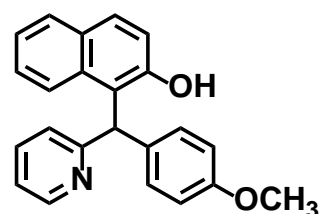
Compound **18**
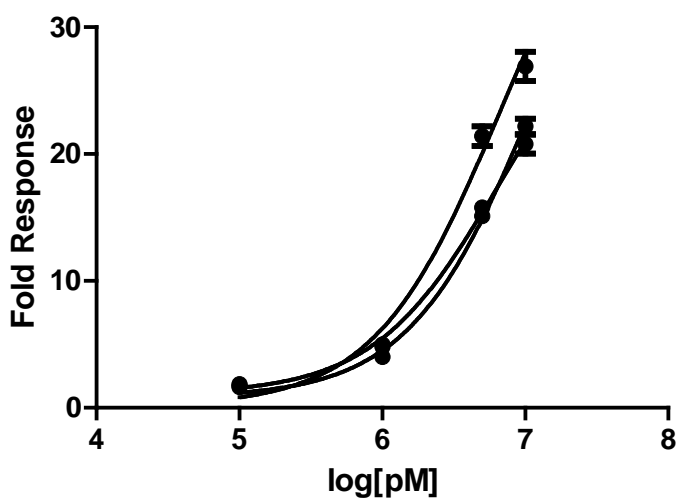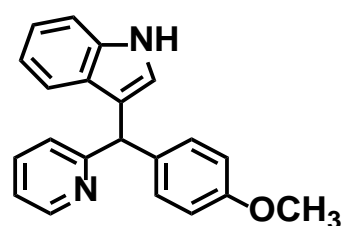
Compound **22**
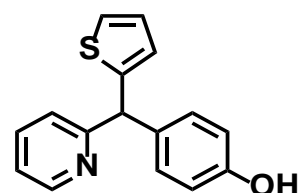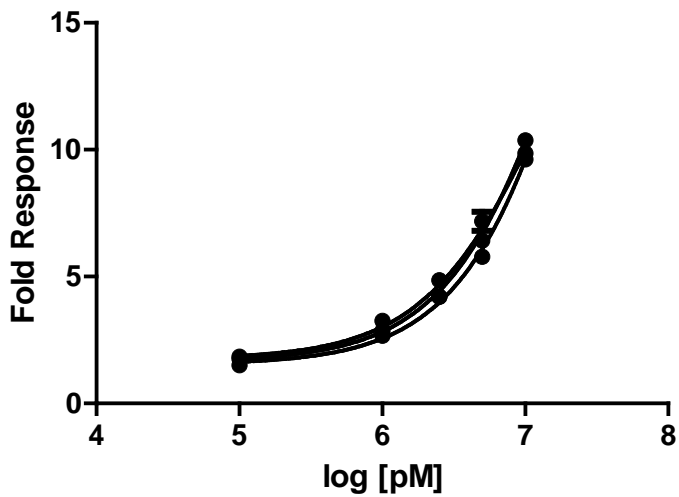
Compound **25**
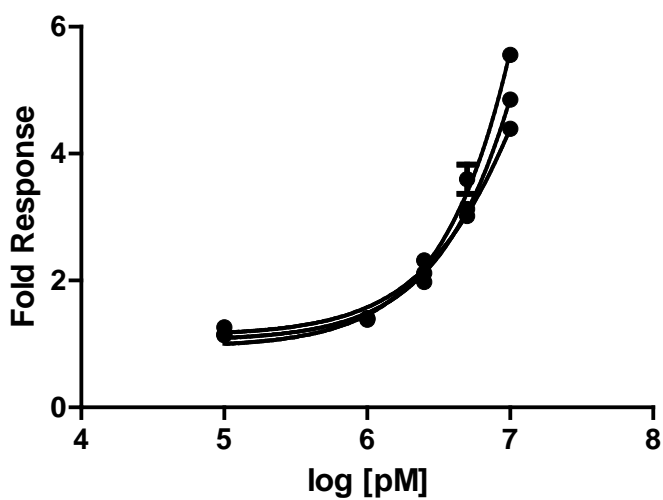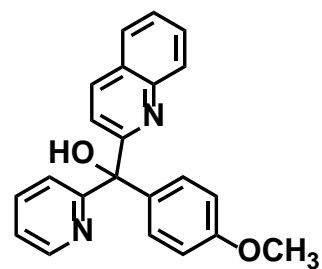
Compound **28**
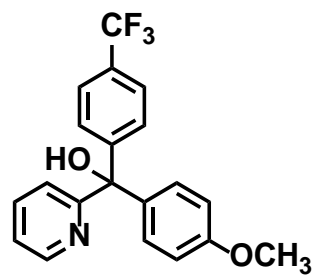
SUPPLEMENTARY INFORMATION 3

Computational Studies

Targeting the Aryl Hydrocarbon Receptor with a novel set of Triarylmethanes

Elizabeth Goya-Jorge [1,2], Celine Rampal [3], Nicolas Loones [3], Stephen J. Barigye [2], Laureano E. Carpio [2], Rafael Gozalbes [2], Clotilde Ferroud [3], Maité Sylla-Iyarreta Veitía[3*], Rosa M. Giner [1**]

[1] *Departament de Farmacologia, Facultat de Farmàcia, Universitat de València. Av. Vicente Andrés Estellés, s/n, 46100 Burjassot, Valencia, Spain.*

[2] *ProtoQSAR SL., CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia. Av. Benjamin Franklin 12, 46980 Paterna, Valencia, Spain.*

[3] *Equipe de Chimie Moléculaire du Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (EA 7528), Conservatoire National des Arts et Métiers (Cnam), 2 rue Conté, 75003, HESAM Université, Paris, France.*

Corresponding authors.

*E-mail (M.S.-I. Veitía): maite.sylla@lecnam.net

**E-mail (RM. Giner): rosa.m.giner@uv.es

(M.S.-I. Veitía and RM. Giner equally contributed as the last authors)

Table of Contents

Section 1. Binding pockets residues for AhR agonists

The binding pockets residues in the molecular docking analysis for the three studied ligands are represented in Figure 1 and listed in Table 1.
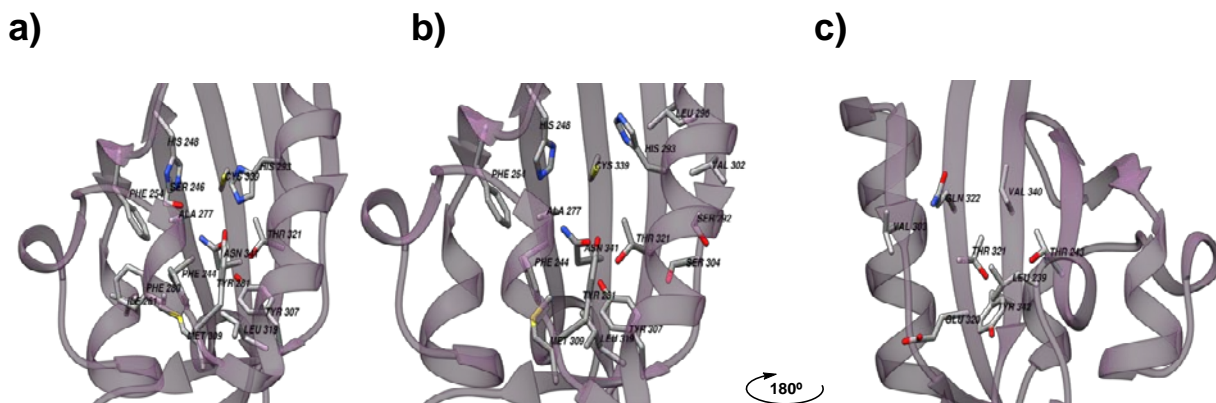
Figure 1. Binding pocket residues identified in the molecular docking for **a)** Compound **22**, **b)** FICZ and **c)** TCDD.

Table 1. Listed residues of the binding pockets of **22**, FICZ and TCDD, respectively.

Ligand	Binding pocket residues [a]
Compound **22**	**PHE 244**, SER 246, **HIS 248**, **PHE 254**, ILE 261, **ALA 277**, PHE 280, **TYR 281**, **HIS 293**, **TYR 307**, **MET 309**, **LEU 319**, **THR 321**, **CYS 339**, **ASN 341**
FICZ	**PHE 244**, **HIS 248**, **PHE 254**, **ALA 277**, **TYR 281**, SER 292, **HIS 293**, LEU 296, VAL 302, SER 304, **TYR 307**, **MET 309**, **LEU 319**, **THR 321**, **CYS 339**, **ASN 341**
TCDD	LEU 239, THR 243, VAL 303, GLU 320, **THR 321**, GLN 322, VAL 340, TYR 342

[a] *The matching residues are in bold.*

Section 2. Binding and Interactions of compounds 22 *vs.* 21

Table 2. Binding energies of compounds **22** and **21** in the molecular docking analysis and predicted interactions for the two best poses.

Compound **21**

Pose 1	Pose 2
Binding Energy: 9.351 Kcal/mol	Binding Energy: 9.239 Kcal/mol

Pose 1

Hydrophobic Interactions ····

Index	Residue	AA	Distance	Ligand Atom	Protein Atom
1	296A	LEU	3.65	1751	972
2	296A	LEU	3.67	1746	972
3	302A	VAL	3.75	1753	1055
4	321A	THR	3.56	1763	1363
5	337A	ILE	3.91	1744	1528

Hydrogen Bonds ──

Index	Residue	AA	Distance H-A	Distance D-A	Donor Angle	Protein donor?	Sidechain	Donor Atom	Acceptor Atom
1	293A	HIS	2.70	3.38	125.03	✔	✔	919 [Npl]	1738 [O3]

π-Cation Interactions ····

Index	Residue	AA	Distance	Offset	Protein charged?	Ligand Group	Ligand Atoms
1	293A	HIS	5.67	1.38	✔		Aromatic	1750, 1751, 1753, 1755, 1757

Pose 2

Hydrophobic Interactions ····

Index	Residue	AA	Distance	Ligand Atom	Protein Atom
1	244A	PHE	3.40	1763	144
2	254A	PHE	3.65	1751	309
3	277A	ALA	3.52	1744	666
4	277A	ALA	3.66	1751	666
5	280A	PHE	3.62	1753	712
6	281A	TYR	3.76	1753	740

Hydrogen Bonds ──

Index	Residue	AA	Distance H-A	Distance D-A	Donor Angle	Protein donor?	Sidechain	Donor Atom	Acceptor Atom
1	248A	HIS	3.75	4.08	102.59	✘	✔	1767 [Nar]	211 [N2]
2	341A	ASN	3.46	3.86	106.67	✔	✔	1591 [Nam]	1767 [Nar]

π-Stacking ···· ····

Index	Residue	AA	Distance	Angle	Offset	Type	Ligand Atoms
1	254A	PHE	4.75	83.70	1.22	T	1761, 1763, 1765, 1767, 1758, 1759
2	281A	TYR	5.40	89.65	1.57	T	1739, 1740, 1742, 1744, 1746, 1748

Compound **22**

Pose 1	Pose 2
Binding Energy: 9.958 Kcal/mol	Binding Energy: 9.378 Kcal/mol

Pose 1

Hydrophobic Interactions ····

Index	Residue	AA	Distance	Ligand Atom	Protein Atom
1	244A	PHE	3.39	1749	144
2	254A	PHE	3.44	1755	309
3	254A	PHE	3.30	1747	309
4	277A	ALA	3.51	1740	666
5	277A	ALA	3.88	1757	666
6	280A	PHE	3.35	1759	712
7	281A	TYR	3.58	1759	740

π-Cation Interactions ····

Index	Residue	AA	Distance	Offset	Protein charged?	Ligand Group	Ligand Atoms
1	248A	HIS	4.41	1.50	✔		Aromatic	1735, 1736, 1738, 1740, 1742, 1744

Pose 2

Hydrophobic Interactions ····

Index	Residue	AA	Distance	Ligand Atom	Protein Atom
1	296A	LEU	3.75	1747	972
2	296A	LEU	3.72	1755	972
3	302A	VAL	3.74	1749	1055
4	337A	ILE	3.91	1757	1528

Hydrogen Bonds ──

Index	Residue	AA	Distance H-A	Distance D-A	Donor Angle	Protein donor?	Sidechain	Donor Atom	Acceptor Atom
1	281A	TYR	3.05	3.41	103.58	✔	✔	745 [O3]	1763 [N2]
2	293A	HIS	3.12	3.98	146.01	✔	✔	919 [Npl]	1763 [N2]

π-Cation Interactions ····

Index	Residue	AA	Distance	Offset	Protein charged?	Ligand Group	Ligand Atoms
1	293A	HIS	5.76	1.57	✔		Aromatic	1746, 1747, 1749, 1751, 1753

Section 3. Sequential alignment PAS B domain

The sequential alignment of the PAS B domain of human AhR (UniProt ID: P35869) and HIF2-α (PDB ID: 3F1O) is shown in Figure 2. Comparison plots of the polarity and the hydrophobicity of both receptors are shown in Figure 3 a) and b), respectively.

```
sp|P35869|AHR_HUMAN   GKLKYLHGQKKKGKDGSILPPQLALFAIATPLQPPSILEIRTKNFIFRTKHKLDFTPIGC   300
3F1O_1|Chain          --------------------------------GEFKGLDSK--TFLSEHSMDMKFTYC   24
                                                       .:   :  :*   *  ::*.:*:.    *

sp|P35869|AHR_HUMAN   DAKGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRLLTKNNRW   360
3F1O_1|Chain          DDRITELIGYHPEELLGR-SAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGY   83
                      *:    ::*:*   **   **  * *.*:* **     ::**  .:  ..*:    :*:*:*:.. :

sp|P35869|AHR_HUMAN   TWVQSNARLLYKNGR--PDYIIVTQRPLTDBEGTEHLRKRNTKLPFMFTTGEAVLYEATN   418
3F1O_1|Chain          VWLETQGTVIYNPRNLQPQCIMCVNYVLSEIBKN--------------------------   117
                      .*:::*:. ::*:   .  *: *: .:  *::  *  .
```

Figure 2. Sequential alignment of AhR-PAS B domain and HIF2α made by ClustalΩ [1]. Residues conserved in both sequences are marked in red boxes and with "*****". Conserved substitutions are marked with "**:**" and "**.**" is used for indicating semi-conserved substitutions.

a)

b)
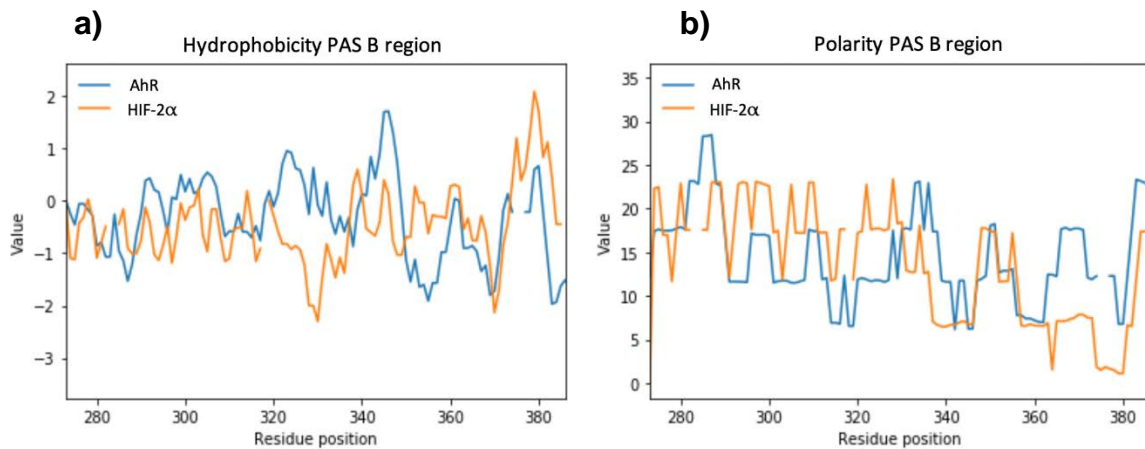
Figure 3. Representation of a) Hydrophobicity [2] and b) Polarity [3] of the aligned PAS B regions of AhR and HIF2α.

Section 4. Off-targeting study of compound 22

The off-targeting potential of compound **22** was studied by molecular docking in four nuclear receptors: Progesterone Receptor (PR), Estrogen Receptor (ER), Pregnane X Receptor (PXR) and Androgen Receptor (AR).

To study the off-targeting potential of compound **22**, four different docking analysis were performed for this compound with different receptors and their specific ligands. The receptors studied were PR, ER, PXR and AR. As shown in Figures 5-8, the binding region of compound **22** in each of the four receptors is different from that of each specific ligand. Referring to binding energy, this parameter was calculated as the mean of the best poses of each ligand with the receptors. Figures 5,6,8 show that three of the four systems have a similar binding energy (PR, ER and AR, respectively) and only one of them (PXR, Figure 7) possesses a significant difference of this parameter.

These results showed favorable binding energies for compound **22** with the four different receptors. In this sense, it is possible that this compound has an off-target potential. Nonetheless additional studies will be necessarily for more conclusive inferences.
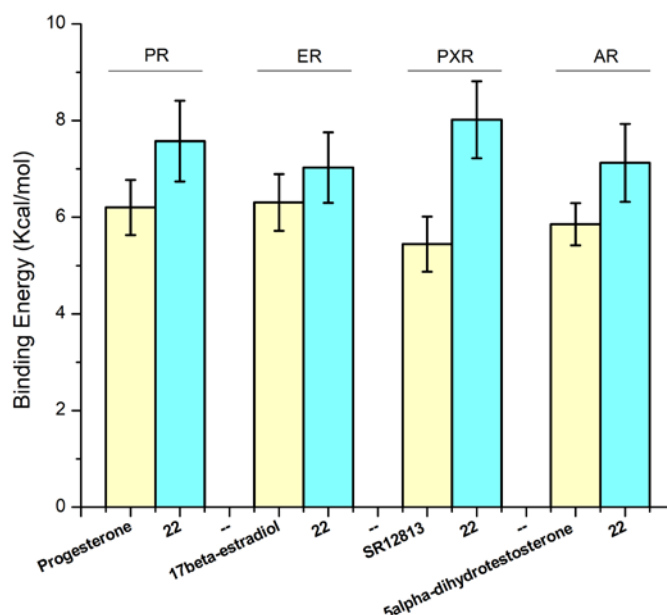
Figure 4. Summary of the binding energies (Kcal/mol ± SD) of compound **22** and the specific ligands with Progesterone Receptor (PR), Estrogen Receptor (ER) and Pregnane X Receptor (PXR) obtained by molecular docking analysis.
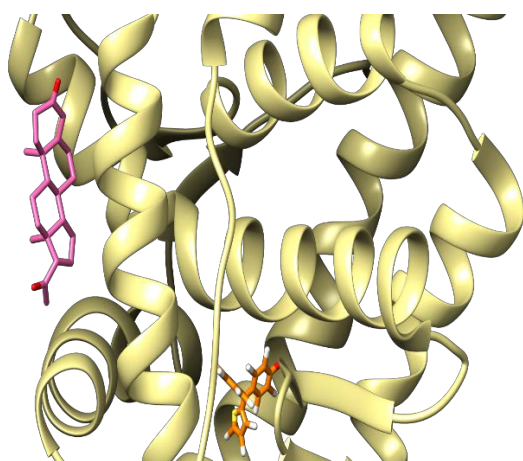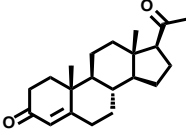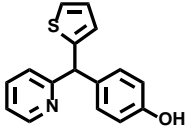
Progesterone Receptor	
Progesterone	**22**
Binding Energy (Kcal/mol ± SD)	
6.20 ± 0.57	7.57 ± 0.83

Figure 5. Molecular representations of the best pose for compound **22** (colored in orange) and the specific ligand progesterone (colored in pink) with Progesterone Receptor (colored in yellow) after molecular docking experiments. Molecular binding energies (Kcal/mol) are expressed as the mean of the best poses ± SD.
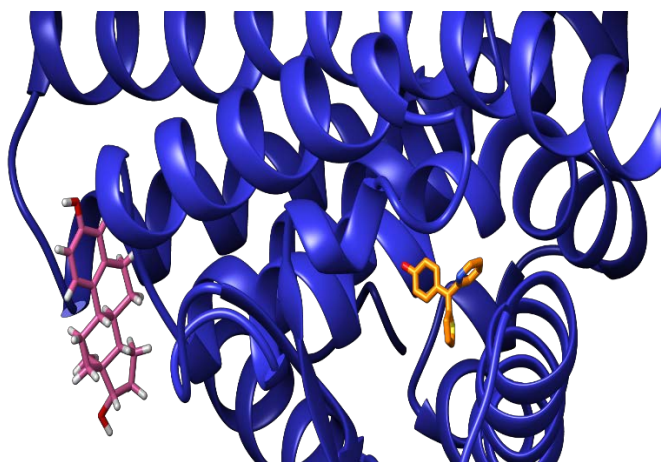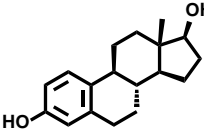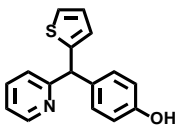
Estrogen Receptor	
17beta-estradiol	**22**
Binding Energy (Kcal/mol)	
6.31 ± 0.59	7.03 ± 0.73

Figure 6. Molecular representations of the best pose for compound **22** (colored in orange) and the specific ligand 17beta-estradiol (colored in pink) with Estrogen Receptor (colored in blue) after molecular docking experiments. Molecular binding energies (Kcal/mol) are expressed as the mean of the best poses ± SD.
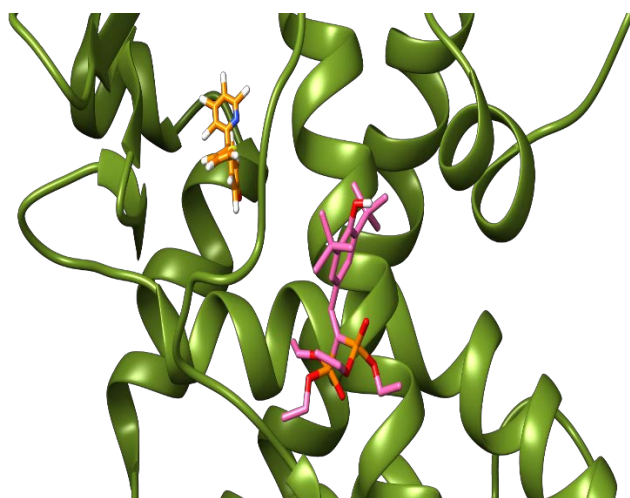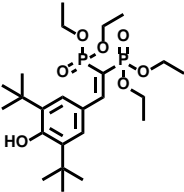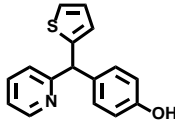
Pregnane X Receptor	
SR12813	**22**
Binding Energy (Kcal/mol)	
5.44 ± 0.57	8.02 ± 0.80

Figure 7. Molecular representations of the best pose for compound **22** (colored in orange) and the specific ligand SR12813 (colored in pink) with Pregnane X Receptor (colored in green) after molecular docking experiments. Molecular binding energies (Kcal/mol) are expressed as the mean of the best poses ± SD.
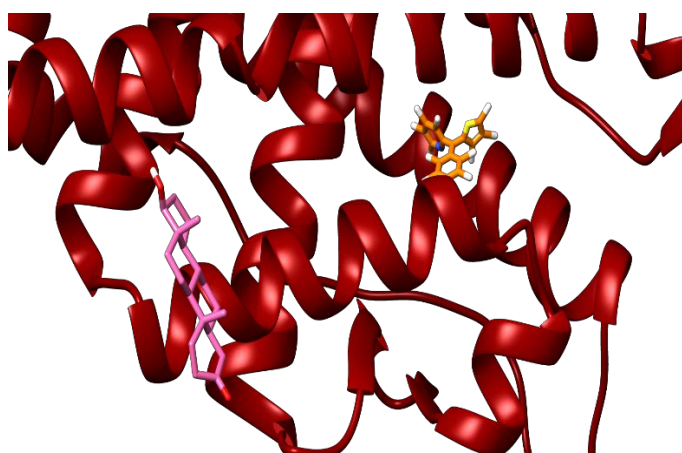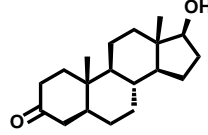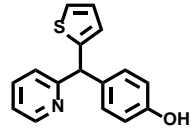
Androgen Receptor	
5alpha-dihydrotestosterone	**22**
Binding Energy (Kcal/mol)	
5.85 ± 0.44	7.13 ± 0.81

Figure 8. Molecular representations of the best pose for compound **22** (colored in orange) and the specific ligand 5alpha-dihydrotestosterone (colored in pink) with Androgen Receptor (colored in red) after molecular docking experiments. Molecular binding energies (Kcal/mol) are expressed as the mean of the best poses ± SD.

Section 5. Additional information on the druglikeness and ADME profile.

Details on #rtvFG definition: Number of reactive functional groups. The presence of all the following reactive functional groups contribute to the *#rtvFG* descriptor: acyl halide, hetero-halogen bond, NAS substrate, alkyl halide, halogen alpha to W-group, heteroatom in 3-ring, activated cyclopropane, aluminum present (toxic), silicon present (toxic), hetero-hetero single bond, azo, diazo, or azide, acceptor carbonyl or derivative, anhydride or analog, unhindered ester, sulfonate or relative, phosphonate or relative, acetal or analog, carbonate, thiol—oxidation possible, carbonyl in 3-ring.

Details on #metab definition: Number of likely metabolic reactions. The following metabolic reactions contribute to the *#metab* descriptor: aromatic OH oxidation, enol oxidation, benzylic-like H → alcohol, allylic H → alcohol, secondary alcohol → ketone, primary alcohol → acid, tertiary alcohol E1 or SN1, amine dealkylation, ether dealkylation, pyridine C2 hydroxylation, aniline NH → NOH or NCOR, low IP—easily oxidized, alpha hydroxylation of cyclic ether, sulfoxide → sulfone, alpha hydroxylation of carbonyl, alpha or beta dehydrogenation at carbonyl, thiol SH → SSR, SR, para hydroxylation of aryl, aryl sulfide → S=O, reduction of aryl nitro to amine, oxidative deamination of primary amine.

Table 3. Additional properties and descriptors

ID	#stars [a]	#rotor [b]	CNS [c]	FOSA [d]	FISA [e]	PISA [f]	WPS [g]	logPoct [h]	logPw [i]	Glob [j]	IP(eV) [k]	EA(eV) [l]
14	**	5	0	95.74	35.52	467.78	0	15.36	7.41	0.85	8.70	0.43
18	**	4	1	99.39	28.22	450.83	0	14.47	6.68	0.86	8.14	-0.08
19	0	4	0	108.41	71.40	445.55	0	16.62	8.77	0.83	8.16	0.04
22	*	4	0	9.39	69.81	385.13	39.63	12.41	6.35	0.87	9.07	0.34
25	*	5	0	93.29	40.06	482.44	0	16.59	8.56	0.83	8.87	0.78
28	0	5	1	93.85	31.19	360.73	120.54	15.79	6.71	0.83	9.11	0.73

[a] #stars: Number of property or descriptor values that fall outside the 95% range of similar values for known drugs. Outlying descriptors and predicted properties are denoted with asterisks (*). A large number of stars suggests that a molecule is less drug-like than molecules with few stars. Properties and descriptors included in the determination of #stars: MW, dipole, IP, EA, SASA, FOSA, FISA, PISA, WPSA, PSA, volume, #rotor, donorHB, accptHB, glob, Polrz, logPC16, logPoct, logPw, logPo/w, logS, LogKhsa, logBB, #metabol. Recommended values: 0-2.
[b] #rotor: Number of non-trivial (not CX3), non-hindered (not alkene, amide, small ring) rotatable bonds. Recommended values: 0 – 15
[c] CNS: Predicted central nervous system activity on a −2 (inactive) to +2 (active) scale. Values: −2 (inactive), +2 (active)
[d] FOSA: Hydrophobic component of the SASA (saturated carbon and attached hydrogen). Recommended values: 0 – 750
[e] FISA: Hydrophilic component of the SASA (SASA on N, O, H on heteroatoms, carbonyl C). Recommended values: 7 – 330
[f] PISA: π (carbon and attached hydrogen) component of the SASA. Recommended values: 0 – 450
[g] WPSA: Weakly polar component of the SASA (halogens, P, and S). Recommended values: 0 – 175
[h] logPoct: Predicted octanol/gas partition coefficient. Recommended values: 8 – 35
[i] logPw: Predicted water/gas partition coefficient. Recommended values: 4.0 – 45.0
[j] Glob: Globularity descriptor, (equation) where r is the radius of a sphere with a volume equal to the molecular volume. Globularity is 1.0 for a spherical molecule. Recommended values: 0.75 – 0.95.
[l] IP(eV): PM3 calculated ionization potential (negative of HOMO energy). Recommended values: 7.9 – 10.5
[m] EA(eV): PM3 calculated electron affinity (negative of LUMO energy). Recommended values: −0.9 – 1.7

Chapter 5

CHAPTER 5. Predictive modeling of AhR antagonism

- **Publication Title**: "Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective". (https://doi.org/10.1080/1062936X.2019.1708460).
- **Authors**: <u>Elizabeth Goya-Jorge</u>, Thi Que Doan, Marie Louise Scippo, Marc Muller, Rosa M. Giner, Stephen J. Barigye, Rafael Gozalbes.

Summary

This fifth Chapter presents an *in silico*/*in vitro* study of chemical compounds as antagonists of AhR-mediated transcription. The studied compounds corresponded to organohalogenated (chlorinated, brominated, fluorinated), polyaromatic hydrocarbon, flavonoid, quinone, imidazole, bisphenol, and pyrethroid chemical classes. Many of these profiled compounds are reported as potential endocrine disruptors or as persistent organic pollutants and have been banned due to their toxic spectrum through different mechanisms. Computational models (QSARs and Toxicophore mapping) were built to elucidate the structural and physicochemical factors influencing the AhR antagonistic activity of chemical compounds.

Highlights

- Results from *in vitro* AhR-CALUX bioassays were modeled *in silico* using QSAR methods and toxicophoric mapping.
- Stringent validation demonstrated the robustness and predictive power of the built computational models.
- The presence of aromatic rings and electron-acceptor moieties were found to be important chemical features for AhR antagonism.
- The hydrophobic constraints seem to influence the AhR antagonistic capacity of toxic chemicals.

Taylor & Francis
Taylor & Francis Group

Check for updates
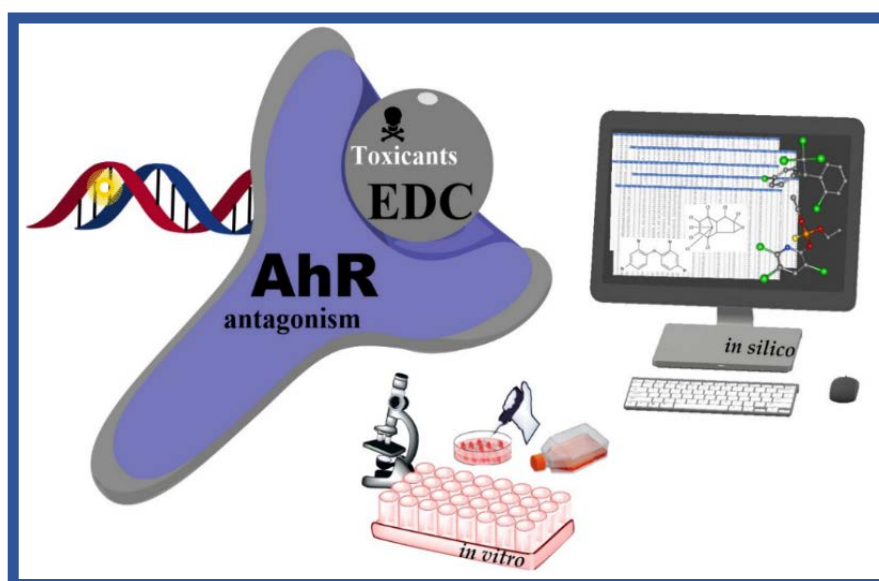
Elucidating the aryl hydrocarbon receptor antagonism from a chemical-structural perspective

E. Goya-Jorge [ID][a,b], T.Q. Doan[c], M.L. Scippo[c], M. Muller[d], R.M. Giner[b], S.J. Barigye[a] and R. Gozalbes[a,e]

[a]CEEI (Centro Europeo de Empresas Innovadoras), ProtoQSAR SL, Parque Tecnológico de Valencia, Valencia, Spain; [b]Departament de Farmacologia, Facultat de Farmàcia, Universitat de València, Valencia, Spain; [c]Laboratory of Food Analysis, FARAH-Veterinary Public Health, ULiège, Liège, Belgium; [d]Laboratory for Organogenesis and Regeneration, GIGA-Research, ULiège, Liège, Belgium; [e]R&D Department, MolDrug AI Systems SL, Valencia, Spain

ABSTRACT

The aryl hydrocarbon receptor (AhR) plays an important role in several biological processes such as reproduction, immunity and homoeostasis. However, little is known on the chemical-structural and physicochemical features that influence the activity of AhR antagonistic modulators. In the present report, in vitro AhR antagonistic activity evaluations, based on a chemical-activated luciferase gene expression (AhR-CALUX) bioassay, and an extensive literature review were performed with the aim of constructing a structurally diverse database of contaminants and potentially toxic chemicals. Subsequently, QSAR models based on Linear Discriminant Analysis and Logistic Regression, as well as two toxicophoric hypotheses were proposed to model the AhR antagonistic activity of the built dataset. The QSAR models were rigorously validated yielding satisfactory performance for all classification parameters. Likewise, the toxicophoric hypotheses were validated using a diverse set of 350 decoys, demonstrating adequate robustness and predictive power. Chemical interpretations of both the QSAR and toxicophoric models suggested that hydrophobic constraints, the presence of aromatic rings and electron-acceptor moieties are critical for the AhR antagonism. Therefore, it is hoped that the deductions obtained in the present study will contribute to elucidate further on the structural and physicochemical factors influencing the AhR antagonistic activity of chemical compounds.

Introduction

The aryl hydrocarbon receptor (AhR) is an evolutionary conserved transcription factor member of the basic helix–loop–helix (bHLH) family of receptors [1]. The AhR acts as a cytoplasmatic chemical sensor mediating intracellular and cellular signals [2] with its main transcriptional regulatory function being the up-regulation of cytochrome P450 family 1 (CYP1) of metabolizing enzymes [3]. The AhR Ligand Binding Domain (LBD) has been suggested to be one of its two PER-ARNT-SIM (PAS) domains, the PAS-B [4].

CONTACT R. Gozalbes ✉ rgozalbes@protoqsar.com
📄 Supplementary material for this article can be accessed at: https://doi.org/10.1080/1062936X.2019.1708460.

However, comprehensive understanding of the structural and functional profile of the AhR has been limited by the unavailability of an experimentally determined AhR structure co-crystallized with the corresponding functional domains. Consequently, studies aimed at elucidating the interaction modes involved in the AhR signal transduction pathway have relied on homologous systems [5].

AhR owes its first discovery to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD), which is one of the most toxic synthetic compounds known to date and whose effects in biological systems are attributed to the AhR binding and activation [3]. Along with TCDD, several dioxin and dioxin-like compounds have been studied for their AhR agonistic activity and associated toxicological effects [6,7]. However, the ability of AhR to interact with a structurally diverse spectrum of xenobiotic and endogenous ligands suggests that it is highly promiscuous [8].

The induction of the AhR/CYP1A1 axis, as a consequence of AhR agonism, has been reported to enhance oestrogenic detoxification [9], regulate the oxidative balance and to propagate the metabolism of proinflammatory and tumour-promoting metabolites, among other effects [10]. Moreover, AhR plays an important role in the physiological functions of reproductive organs, immune system, liver and vascular development, cardiac function, cell growth, differentiation, homoeostasis and circadian rhythms [2,11–13]. On the other hand, several toxicants and high concern substances have been reported as antagonists of AhR transcriptional activity [7,14]. However, in contrast to AhR agonism, little attention has been paid to the analysis of the biochemical consequences of AhR antagonism and even less on the subsequent toxicological implications [15–17].

Computational models, coupled with in vitro assays, have in the recent times gained increasing utility as alternative tools for providing insight on the pharmacological and/or toxicological effects of chemical compounds [18]. Indeed when proper interpretations are feasible, these techniques jointly provide solid understanding on the structural and functional characteristics relevant for studied bioactivity profiles [19], in addition to their inherent ethical, economic and predictive advantages [20].

It should be noted that most of the AhR computational models reported in the literature have been based on the binding affinity as the endpoint, which is an important drawback since it does not discriminate between agonists and antagonists [21,22]. In addition, as an experimentally determined AhR structure including the respective functional domains is unavailable, in silico initiatives have relied on homology models of the PAS-B domain to provide insights on the possible ligand binding modes [5]. The differences observed in these studies between the agonistic and antagonistic interaction modes suggested greater distortions in the structure of the LBD in case of the latter [5,23]. Nonetheless, little is still known on the chemical structural and functional features that favour the binding of AhR antagonists [24].

Therefore, further analyses of the agonist and/or antagonist responses following AhR binding are needed to gain better understanding of role of the AhR in both toxicological and pharmacological contexts [25]. Certainly, informative experimental bioassays will be necessary to obtain greater insight on the AhR-ligand interaction modes, and particularly to discriminate between agonistic and antagonistic modulations. One of the most popularized assays is the in vitro AhR Chemically Activated LUciferase eXpression (CALUX) bioassay, which has been reported to be useful in the detection of the AhR antagonistic and agonistic effects of dioxin and dioxin-like chemical compounds [26,27]. This assay is

based on the analysis of the induced luciferase response in a recombinant cell line driven by several CYP1A1 dioxin response elements (DREs) as a direct reflection of AhR-mediated transcriptional activity.

The goal of the present manuscript is to analyse the factors, from a chemical perspective, that influence the AhR antagonism of chemical compounds using a combination of in vitro and in silico methods. Firstly, a CALUX reporter gene assay for the AhR antagonistic activity for a set of selected xenobiotics with known toxicity profiles was performed. Then, based on the results obtained from the in vitro assays and an extensive literature review, QSAR and toxicophoric models were built to examine the chemical structural and physicochemical features that modulate AhR antagonism. It should be noted that while QSAR methods have been employed to model and predict the AhR binding capacity [28,29], this is the first time that in vitro, QSAR and toxicophoric approaches are collectively employed to analyse AhR antagonism.

Materials and methods

In vitro evaluation

Chemicals and reagents

A set of 68 chemical compounds was tested in vitro for their AhR antagonistic activity due to their suspected or suggested toxic effects. Twenty-nine of them correspond to Persistent Organic Pollutants (POPs) prevalent in Scandinavian human blood [30] and the remaining 39 corresponded to a chemical library belonging to the Laboratory of Food Analysis of the University of Liège. The tested compounds were mostly obtained from Sigma-Aldrich (Missouri, USA), for details see Supporting Information SI-1. All the chemicals were dissolved in dimethyl sulphoxide (DMSO) (Acros Organics, Molinons, France) as the stock solution of 20 mg/mL and kept in −20°C. The standard solution of 2,3,7,8-tetra-chlorodibenzodioxin (TCDD) (purity > 98%) in DMSO was supplied from Wellington laboratories (Guelph, Canada). The reagent 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetra-zolium bromide (MTT) was obtained from Sigma Aldrich (Missouri, USA).

AhR CALUX assay

The AhR-CALUX bioassay was developed based on a stably transfected dioxin response (DR) rat hepatoma (H4IIE) cell line, consistent with previous reports in the literature [6,31]. These cells were obtained from BioDetection System (BDS) (Amsterdam, The Netherlands) transfected with four native dioxin response elements from the upstream region of the mouse *cyp1a1* gene leading the Mouse Mammary Tumour Virus (MMTV) promoter and controlling expression of the luciferase gene were stably integrated into the cell's genome. The AhR transactivation activity of the compounds was reported as the expression of the inserted luciferase and measured by light production.

The cells were maintained in *α*-MEM (ThermoFisher, Massachusetts, USA) with 10% v/v foetal bovine serum (Greiner, Kremsmünster, Austria) and 50 IU/mL penicillin and 50 µg/mL streptomycin (Sigma Aldrich, USA) and were incubated at 37°C in a water saturated atmosphere injected with 5% CO_2.

The AhR antagonistic tests for dioxin responsive chemical-activated luciferase gene expression (AhR-CALUX) bioassays were performed as indicated by the provider BDS.

Briefly, after reaching about 90% of confluence in the culture flask, the cells were trypsinized and seeded homogenously in white clear-bottomed 96 well microplates (Greiner, Kremsmünster, Austria). The cells were then incubated for 24 h and afterwards treated with the test compounds for another 24 h. The experiment was terminated by cell lysis using lysis solution containing Triton X100 (Sigma Aldrich, Missouri, USA). After adding luciferin (Promega, Wisconsin, United States) and ATP (Roche Diagnostics, Rotkreuz, Switzerland) to the cell lysate, the plates were read by a luminometer (ORION II, Berthold Detection System, Pforzheim, Germany).

For antagonistic tests, the cells were co-exposed to the test compounds and 15 pM TCDD corresponding to TCDD EC_{50} in DR-H4IIE. All the experiments were repeated at least three times independently. The maximum concentration tested to determine the AhR antagonistic activity was 40 μM with some exceptions (see SI-1), and the final concentration of DMSO in the culture medium was 0.4%. To ensure the adequacy of the test method and provide a basis for comparisons, a reference curve of the positive control TCDD, as well as concurrent negative and solvent controls were added on each plate.

Cell viability assay

An analysis of the MTT cell viability as well as a visual inspection of the cell morphology and attachment were performed to detect the cytotoxic compounds. After exposure to the test compounds, 25 μL of MTT solution 5 mg/mL was added into each well. The plates were then incubated for 4 h at 37°C to allow for the formation of the purple metabolite formazan from the tetrazolium dye MTT via the activity of the mitochondrial succinate dehydrogenase. Later, the formazan crystals were dissolved during 2 h by adding 100 μL of isopropanol. The MTT formazan absorbance was read at 550/630 nm using a microplate spectrophotometer (ELX800[TM] BioTek Inc., Winooski, USA).

Results for both reporter gene and MTT assays were presented as relative responses, either as the percentage of the cell response to the tested compound compared to the cell response to TCDD EC_{50} on the same plate for CALUX assays, or to solvent control DMSO for MTT test. A compound was considered an AhR antagonist when it was able to reduce the activity of the TCDD EC_{50} from 100% to at least 70%, while reductions in cell population greater than 15% were considered cytotoxic [32].

QSAR models

Based on both the experimental in vitro results obtained herein and those compiled from an extensive review of the literature, binary classifiers of AhR antagonism were built using statistical modelling methods. The molecular characterization was conducted by means of simple and interpretable chemical structural descriptors.

Structures, descriptors and activity

A dataset of 116 chemicals was built to develop the QSAR classification models. This set comprised of 68 compounds tested in the present report for their AhR antagonistic activity and 48 compounds retrieved from literature. From this dataset, compounds with undetermined or inconsistent activity were excluded from the analysis and therefore minimizing potential error sources (for details see SI-1). A binary scale of activity values was considered in the sense that the compounds observed or reported in the literature to

induce the AhR antagonistic effect, based on the AhR-CALUX method, were labelled as active while those that did not induce any effect were considered as inactive.

The calculated molecular descriptors were implemented in-house, based on definitions obtained from DRAGON [33] and PADEL [34] software. Consistent with the parsimony principle, only the simplest characterizations of the molecular structures were considered. A total of 1929 DRAGON descriptors were calculated corresponding to: constitutional descriptors, information indices, functional group counts, atom-centred fragments, 2D binary fingerprints and 2D frequency fingerprints. From this set, constant variables or those with pair-correlations greater than 0.9 were removed, retaining only 171 descriptors. As for the PADEL software, the Estate and MACCS fingerprints were computed yielding 245 descriptors from which 15 constants were removed. Globally, a total of 401 molecular descriptors were retained and subsequently used as input variables for the chemoinformatic modelling.

Variables and QSAR methods

The selection of molecular descriptors for modelling the AhR antagonistic activity was carried out using the information gain filter as provided by IMMAN software [35]. The information gain defined as the reduction in variable entropy (or uncertainty) given that another variable (response variable in this context) is known. This measure allows variables to be ranked based on their capacity to discriminate cases that belong to different classes and consequently filter out informative variables.

A Principal Components Analysis (PCA) was performed on the selected set of variables and 10 principal components (PCs), explaining more than 90% of the total variance, were obtained. Subsequently these PCs were used as input variables for modelling the AhR antagonistic activity. The PCs were rotated using the orthogonal rotation scheme varimax to maximize the sum of variances of the squared loadings. For the PCA the MATLAB software [36] was employed.

Linear Discriminant Analysis (LDA) and Logistic Regression (LReg) methods were employed for the classification model building. The metrics considered to evaluate the performance of the built classification models were the accuracy, precision, sensitivity and specificity (see below) where: TA = true active; FA = false active; TI = true inactive and FI = false inactive.

$$\text{Accuracy} \; = \; \frac{(\text{TA} + \text{TI})}{(\text{TA} + \text{FI} + \text{FA} + \text{TI})} \times 100\%$$

$$\text{Precision} \; = \; \frac{\text{TA}}{(\text{TA} + \text{FA})} \times 100\%$$

$$\text{Sensitivity} \; = \; \frac{\text{TA}}{(\text{TA} + \text{FI})} \times 100\%$$

$$\text{Specificity} \; = \; \frac{\text{TI}}{(\text{TI} + \text{FA})} \times 100\%$$

Validation methods and applicability domain

Internal and external validation experiments were designed to assess the robustness and predictiveness of the built classification models. To this end, a cluster analysis was performed to rationally divide the dataset in training and test sets, based on the following procedure: first, with hierarchical clustering (based on the square-Euclidean distance as the similarity criterion) an optimal number of clusters (*k*) was established using the graph of the amalgamation schedule. Then, k-means method was employed to stratify the structures into clusters according to their similarity. An appropriate distribution of active and inactive compounds was considered during the clustering. With this procedure, two subsets of compounds were generated with 75% of the data comprising the training set and the 25% test set. Internal leave-group-out cross validation (25% of the training set was held out for validation for each fold) was performed to check for possible overfitting during the model building procedure. Ten different combinations of training and test sets were created using cluster analysis and a 10-fold external validation was performed to evaluate the average predictive capacity of the models. This approach has been recommended as the most stringent validation protocol for QSAR models [19]. For the cluster analysis, the STATISTICA [37] software was employed.

The applicability domain (AD) of the built QSAR models was established by using the Euclidean distances and Tanimoto coefficients as the structural similarity measures and the k-nearest neighbour algorithm (k-NN) as the feature space approximator. For this AD analysis the AmbitDiscovery software [38] was used.

Toxicophoric modelling

A diverse set of active AhR antagonistic compounds was selected from the dataset built in the present report for mapping the toxicophoric hypothesis. The selection criterion was based on the hierarchical and k-means clustering previously described. Additionally, a Tanimoto matrix score provided by the PubChem [39] platform was considered. For this study, the PHASE module of Schrödinger [40] software was used.

The three-dimensional structures of the selected set of active AhR antagonistic compounds were generated and employed to build toxicophoric models following a ligand-dependent approach exclusively. The mapped space was unrestricted by considering all volumes and shapes, but with locations and orientations in 3D-spaces allowing the enantiomeric discrimination by absolute coordinated distances. All the available features included in PHASE were analysed at distances of 1.8 Å and with a tolerance of 2.0 Å to match the hypothesis. These features were represented in the toxicophoric models as (A) acceptor, (D) donor, (H) hydrophobic, (N) negative ionic, (P) positive ionic, and (R) aromatic ring. The best alignment was searched considering 50 conformations for each compound. The generated hypotheses were accepted if they fitted at least 50% of the training set. Three to seven features were considered as optimum for hypothesis acceptability.

The goodness of fit of the toxicophoric hypotheses was evaluated using the Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC), which is a generalization of the ROC metric to deal with the early recognition problem. The difference criterion between toxicophoric models was set to 0.5, while the Phase Hypo Score was employed as the scoring function. The mathematical definitions for all

these metrics are detailed elsewhere and the performance is measured from 0–1 values where 1 represents perfect fitness [41].

Finally, an external set of decoy structures was generated using as seeds the diverse set of active AhR antagonistic compounds employed in the toxicophore hypothesis building. For decoy structure generation, the Directory of Useful Decoys (DUD•E) online platform was employed. DUD•E constitutes the largest and most comprehensive public dataset for benchmarking virtual screening programs and was employed here to validate the predictive capacity of the toxicophoric hypotheses. DUD•E is designed to provide sets of structures with similar physicochemical properties but dissimilar topology to the active compounds being, therefore, unlikely binders [42].

Results and discussion

In vitro results

Out of the 68 test compounds evaluated with the AhR-CALUX assay, 24 of these induced AhR antagonistic transactivity, suppressing the cell response to the spike-in control 15 pM TCDD from 100% to at least 70%. Three of these, two insecticide isomers (*α*-endosulfan and *β*-endosulfan) and the fungicide thiram, reduced the cell viability by more than 15%. Moreover, the former two showed cytotoxic effects in more than 60% of the cell population at a concentration of 40 μM while the latter showed a similar effect even at 20 μM. These three chemicals were therefore excluded from the analysis of AhR antagonism (see SI-1). The remaining 41 chemicals did not show any AhR antagonistic potential and were thus considered as inactive.

QSAR modelling

Database of structures and AhR antagonism

A dataset of 116 chemical compounds with defined AhR antagonistic activity profiles based on the DR-H4IIE cell line was constructed. This database contained organohalogenated (chlorinated, brominated, fluorinated), polyaromatic hydrocarbon, flavonoid, quinone, imidazole, bisphenol and pyrethroid chemical classes. Although a few antibiotics and phytochemicals were included, most of the modelled structures are or have been used as industrial chemicals, pesticides, fungicides, algaecides, insecticides, rodenticides and herbicides. Many of these profiled compounds are reported as potential endocrine disruptors or have been already banned due to their toxic spectrum through different mechanisms.

During the construction of the dataset, interest was placed on the homogeneity of the results obtained from the experimental protocols, in order to obtain logical and interpretable conclusions. Therefore, compounds with ambiguous and/or unclear experimental determinations were discarded, especially for those results extracted from literature. Such stringent selectiveness has been strongly supported by several reports in the literature [19].

The hexachlorocyclohexane (HCH) conformational isomers (α-, β-, γ-, δ-HCH) were excluded since they possessed contrasting activity profiles (the descriptors used in the present study are insensitive to stereochemical differences). For the same reason, two

polycyclic aromatic hydrocarbons were excluded. The hexabromocyclododecane (HBCD) in vitro activity was the only one considered, while the rest of the isomers (α-, β-, γ-HBCD) reported in the literature were discarded. Finally, one structural outlier (chlormequat chloride) was removed during the modelling due to disconnected fragments in the molecular structure.

Moreover, BDE-49 (i.e. 2,2′,4,5′-tetrabromodiphenyl ether) [43] and nine polychlorinated biphenyls (PCBs) [7] were excluded as their activity profiles were not clearly defined in the literature (see SI-1). Three cytotoxic compounds in the MTT assay were not considered (see above).

A standard protocol was followed to guarantee the homogeneity of the final dataset presented in this study that included 93 compounds, 40 active AhR antagonists and 53 inactive. Therefore, this dataset may be considered of acceptable quality and is, to the best of our knowledge, the largest for AhR antagonistic activity reported so far.

Selection and transformation of molecular descriptors

The calculated molecular descriptors were all based on simple and interpretable features of the chemical compounds. Due to the large number of generated descriptors, dimensionality reduction was deemed necessary.

From the calculated 401 molecular descriptors, a set of 96 descriptors was retrieved using the aforementioned information gain filter, and subsequently used to generate orthogonal descriptors using the PCA method. Ten PCs were determined to be the optimum number of PCs based on the eigenvalues of the generated covariance matrix and these described 90% of the total variance in the dataset. The matrix of the PCs coefficients obtained for the set of molecular descriptors and the scores for all the PCs are provided as supplementary information (SI-2).

QSAR models obtained

For the QSAR modelling of the AhR antagonistic activity, simple and interpretable classification algorithms were preferred. Hence, five models were selected following commonly established criteria [19], three corresponding to LDA and two corresponding to LReg methods. The training metrics obtained for these models are shown in Table 1 and additional parameters are provided as SI-3. One of the significant advantages offered by these statistical models, compared to most machine learning algorithms such as random forest, support vector machine or artificial neural networks, is the straightforward

Table 1. Evaluation metrics of the QSAR models.

Model[a]	PCs	Accuracy		Precision		Sensitivity		Specificity	
		Train[b]	Ext.V[c]	Train[b]	Ext.V[c]	Train[b]	Ext.V[c]	Train[b]	Ext.V[c]
LDA	10	88.51	81.48	87.02	73.49	88.01	83.56	89.21	79.95
LDA	9	83.02	81.94	87.02	73.58	88.12	83.33	78.11	79.62
LDA	6	78.01	74.07	76.04	65.33	63.23	77.58	89.30	68.95
LReg	7	88.50	73.61	88.51	64.89	88.20	74.78	89.22	72.14
LReg	6	93.51	78.24	87.01	70.32	88.14	74.89	99.03	79.00

[a] QSAR models based on linear discriminant analysis (LDA) and logistic regression (LReg).
[b] Value of the classification metrics obtained during the training of the models (Additional parameters in SI-3).
[c] Average of the classification metrics made during the 10-fold external validation procedure.

interpretation and consequently applicability in the rationalization of the influence of the different chemical structural features on the modelled property [44].

Moreover, both the internal and external validation procedures yielded adequate statistical parameters, demonstrating the robustness and predictive capacity of the five QSAR models obtained herein. Indeed, all the selected models showed good external validation performance, yielding percentages higher than 60% for accuracy, precision, sensitivity and specificity (Table 1).

These models are the first QSARs for AhR antagonistic activity of which we are aware. Previous QSAR models reported in the literature focused on AhR agonism or non-specific AhR-ligand binding [28,45]. Moreover, the models presented herein were built over a diverse dataset of chemicals and were observed to be consistent with the OECD validation principles [46]. Therefore, these models can reliably predict the AhR antagonistic potential of untested chemicals in their AD.

Toxicophoric mapping

Toxicophoric (or Pharmacophoric) models are defined as an ensemble of steric and electronic features that are necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response [47]. Toxicophoric modelling aims to identify a set of features or restrictions in the molecules and/or their receptors that determine their capacity to interact in any mode and consequently cause toxic effects. These models may be built following structure-based or ligand-based approaches depending on whether the information of the target receptor is employed or not [48,49].

Herein, as a fully crystalized structure of the AhR protein (containing the PAS-B ligand binding domain) is not yet available [25], a ligand-based toxicophoric modelling was performed. The cluster analysis performed during the QSAR modelling was employed to select the most dissimilar compounds for the toxicophoric mapping.

Seven most dissimilar structures were selected (Figure 1) and used to explore toxicophoric hypothesis of AhR antagonism. These chemical compounds are considered to be an environmental and health threat and they showed strong antagonistic activity in the in vitro AhR-CALUX assay.

A total of 8 toxicophoric hypotheses were obtained and their performance assessed based on the internal validation enrichment metrics. Consequently, two toxicophoric models were selected as they showed adequate performance and contributed dissimilar structural information.

Figure 2(a,b) show the spatial representations of the selected Toxicophoric-AhR antagonistic hypotheses for the first: Hydrophobic_Hydrophobic_Hydrophobic_ Aromatic Ring (HHHR) and the second: Acceptor_Hydrophobic_Hydrophobic_ Aromatic Ring (AHHR) models, respectively. Both arrangements included a common segment of two hydrophobic constraints and an aromatic ring as critical features. On the other hand, the latter model included an electro-acceptor group while the former possessed an additional hydrophobic constraint.

An external validation of the two selected AhR-antagonistic toxicophoric models was conducted as recommended [42], using a diverse dataset of 350 decoy structures generated from the seven active molecules employed in the hypotheses exploration.
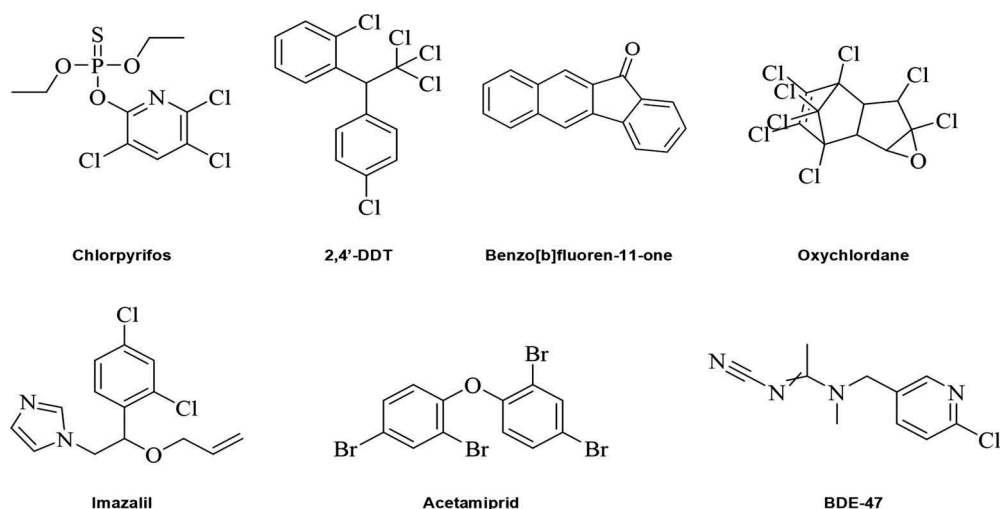
Figure 1. Active structures selected for toxicophoric modelling (For details see SI-1).
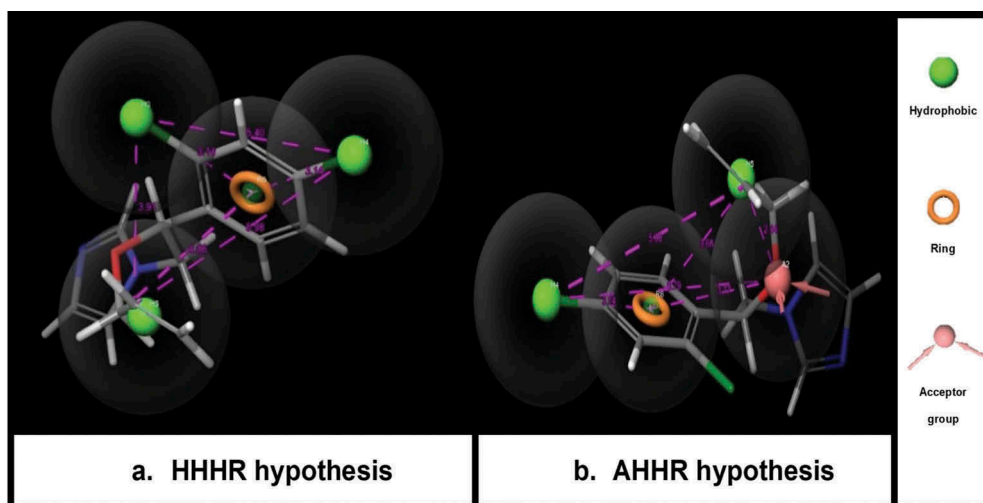
Figure 2. Representation of the toxicophoric models designed in Schrödinger software. Hydrophobic (green-ball), ring (orange ring), acceptor group (pink ball with signalling arrows). a. HHHR: Hydrophobic_Hydrophobic_Hydrophobic_Aromatic Ring. b. AHHR: Acceptor_Hydrophobic_Hydrophobic_Aromatic Ring.

The training and the corresponding external validation results of the HHHR and AHHR toxicophoric hypotheses are shown in Table 2. Overall, the built hypotheses showed good performance as all validation metrics were greater than 0.5 and thus demonstrating their robustness and predictive capacity.

Several pharmacophoric models considering nuclear receptors ERα and ERβ as drug targets [50,51] have been reported in the literature. However, limited distinctions between agonist and antagonistic ligands have been suggested [52]. Furthermore, the few toxicophoric considerations available are generally restricted to drug-safety contexts

Table 2. Results of the training and the external validation for HHHR and AHHR toxicophoric hypotheses.

	Training performance		External validation	
Hypothesis	PHASE[c]	BEDROC[d]	PHASE[c]	BEDROC[d]
HHHR[a]	0.93	0.68	0.93	0.98
AHHR[b]	0.82	0.69	0.82	0.92

[a] HHHR: Hydrophobic hydrophobic hydrophobic aromatic ring.
[b] AHHR: Acceptor hydrophobic hydrophobic aromatic ring.
[c] Scoring function phase hypo score [41].
[d] Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC) [41].

[48,49]. Nevertheless, some successful and experimentally validated studies on enzymatic-mediated endocrine disruptive mechanisms have been reported based on the inhibition of 17β-hydroxysteroid dehydrogenases [53].

The first pharmacophoric/toxicophoric estimations based on the blocking capacity of chemicals over AhR are herein proposed. Moreover, considering the good predictive power offered by these computational models, they could be a useful virtual screening tool for chemical compounds with potential AhR antagonistic activity.

Structural and physicochemical interpretations
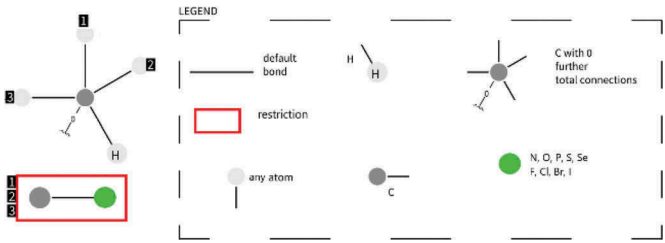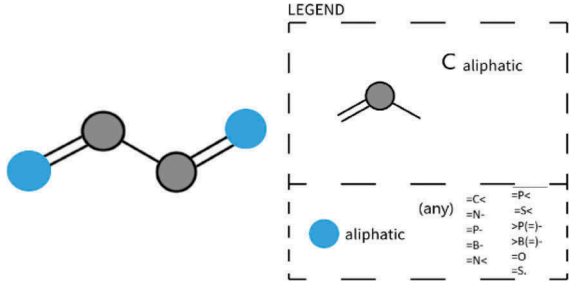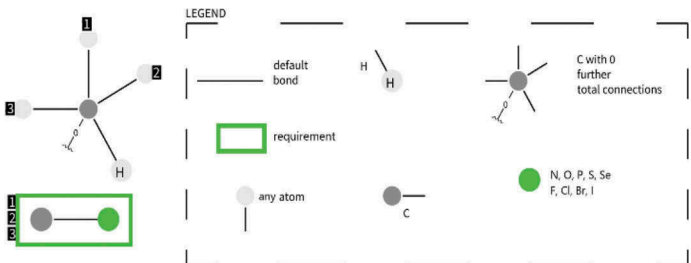
As an essential receptor whose cellular implications are still being discovered, the disruption of the AhR functions could lead to toxic effects that are yet to be fully understood [3,8,11,12]. Hence, the mapping of chemical structural features critical for the AhR agonistic/antagonistic potential could enhance the identification of specific modulators with possible toxic profiles. In this sense, based on the developed QSAR and toxicophore-based models, analyses were performed to further understand how different chemical-structural features influence the AhR antagonistic activity.

The interpretation of computational models is highly recommended by the OECD guidelines to enhance the utility of the QSAR predictive results [46]. Here, only simple 0-2D descriptors were calculated and used for the model building, therefore the elucidation of their contribution to the AhR antagonistic activity in chemical structural terms is straightforward. On the other hand, the toxicophore-based methods provide information on the steric and electrostatic features deemed critical for a given toxicity (or activity) profile. Hence both QSAR and toxicophore models may serve as complementary approaches providing a more complete outlook of the structural and physicochemical factors that influence the AhR antagonistic activity.

The molecular descriptors with relevant contribution were determined based on the PCA factor loadings (SI-2). It has been previously reported that coefficients greater than 0.7 are considered remarkably influential to the generated PCA components [54]. Therefore, the same cut-off was considered to identify the most influential descriptors on the modelled AhR antagonism. Definitions and SMARTS [55] representations of the top five variables obtained are provided in Table 3.

From the toxicophoric models, hydrophobic interactions were identified as an important constraint in active AhR antagonistic molecules. Moreover, the presence of an aromatic ring and an electron acceptor group was suggested as relevant for AhR

Table 3. Molecular descriptors most relevant for the AhR antagonistic activity of the analysed chemical compounds.

ID Code	Molecular descriptors*
H-046	
	H attached to C (sp3) no X attached to next C
nCconj	
	Sum of the number of non-aromatic conjugated C(sp2) belonging to any conjugated system, excluding aromatic rings. Including the following atom types: = C<, = N-, = P-, = B-, = N<, = P<, = S<, >P(=)-, >B (=)-, = O, = S
H-052	
	H attached to C(sp3) with 1X attached to next C
nBM	Multiple bonds
nCIR	Number of circuits

* Brief definitions/representations of the descriptors using the SMARTs [55] formats of the codified features are provided. The range of values obtained was: H-046 [0–10]; nCconj [0–6]; H-052 [0–12]; nBM [0–26]; nCIR [0–28].

antagonism. These features will now be analysed in detail and the possible complementarity with the QSAR model variables also explored.

Hydrophobic constraints

The hydrophobic features comprising the toxicophoric model and the QSAR model descriptors codifying information on hydrophobicity (i.e. hydrocarbon chains without attached heteroatoms (H-046) and closed hydrophobic circuits (nCIR) of saturated cyclic

hydrocarbons) suggest that non-polar structural segments probably play an important role in AhR-ligand interactions. Indeed, early studies on TCDD where AhR was simply considered as an intracellular, soluble and unknown receptor protein, already recognized its hydrophobic character [56].

Some other reports in the literature have also highlighted the importance of hydrophobicity in AhR-LBD interactions e.g. a strong correlation has been found between the suppression of AhR on TCDD-induced activity and the hydrophobicity of curcumin derivatives [17]. It has also been suggested that the affinity of stilbene derivatives of resveratrol to AhR is enhanced by substituents with high hydrophobicity, among other factors [57]. Moreover, an analysis of the AhR ligand binding pocket revealed that it is comprised predominantly of hydrophobic residues with some polar segments located near the medial positions of the ligands [58]. Therefore, structural fragments with strong hydrophobic character should probably be prioritized in the design of ligands with possible AhR antagonistic activity.

Ring feature

Similarly, both the toxicophoric hypotheses and the QSAR models included ring features [circuit number (nCIR), in the case of the latter] as critical for AhR antagonism. Bearing in mind that the name AhR is attributed to the polycyclic aromatic hydrocarbons (PAHs), which were in fact the very first ligands found to interact with this receptor, it is unsurprising that ring features are suggested to play a critical role in AhR-mediated interactions probably through non-covalent interactions (i.e. between electron-rich and electron-deficient aromatic moieties) [59]. These interactions, also named aromatic donor–acceptor interactions, have been analysed for several decades in the AhR context [60,61]. Firstly, early studies hypothesized that π-π-interactions with phenyl and tyrosyl groups of the AhR protein occurred in the TCDD-mode of action [56]. Also, these interactions have been evoked to rationalize the effect of PBDEs and PCBs upon contact with dissolved organic matter in the environment [28], and have been suggested to play an important role in endocrine disruptive mechanisms e.g. in ER disruption [62].

Nonetheless, it is important to highlight that there ligands without the neutral, hydrophobic and non-polar molecular characteristics of PAHs that have also been shown to have effects on AhR, demonstrating that these interactions depend on the contribution of multiple factors [2,63].

Acceptor-group presence

The charge transfer between molecules and biological systems commonly influences the potential toxicity of the former [64]. Herein, one of the proposed toxicophoric hypothesis suggested the presence of an acceptor group as a relevant structural feature for AhR antagonistic activity. Likewise, the QSAR models comprise various key molecular descriptors that could be related to this toxicophoric feature. For example, heteroatoms attached to sp3 carbons (H-046 and H-052), non-aromatic conjugated systems (nCconj) or multiple bonds (nBM) in general could all contribute to electron deficient fragments in the ligand's structure whose interactions may potentially influence the AhR antagonistic effects.

Altogether, the results obtained herein suggest that the presence of electron acceptor groups, connected ring systems (preferably aromatic or delocalized moieties), and most importantly hydrophobic groups seems to be critical for the potential AhR antagonistic effects.

Conclusions

Structural and physicochemical determinants of the AhR antagonistic capacity were assessed using a combination of in vitro (AhR-CALUX bioassay) and in silico (QSAR and toxicophoric mapping) methods. In the former, several chemical compounds were evaluated, and their antagonistic activity profiles determined. In the latter, QSAR algorithms based on LDA and LReg were obtained and validated. These models were shown to possess adequate robustness and predictive power, based on the quality of the obtained statistical validation parameters. Moreover, toxicophoric models were derived and the structural features potentially favouring the AhR antagonistic potential analysed. Interpretations of the QSAR and toxicophore models revealed that electron acceptor groups, aromatic or delocalized ring systems, as well hydrophobic moieties probably favour the AhR antagonistic activity, demonstrating the complementarity of the two approaches notwithstanding their dissimilar conceptual basis. The models proposed herein could be useful in the prediction of the AhR antagonistic capacity of chemical compounds with toxicological and pharmacological applications.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

E. Goya-Jorge   http://orcid.org/0000-0003-0515-989X

References

[1] D.W. Nebert, *Aryl hydrocarbon receptor (AHR): "Pioneer member" of the basic-helix/loop/helix per-Arnt-sim (bHLH/PAS) family of "sensors" of foreign and endogenous signals*, Prog. Lipid Res. 67 (2017), pp. 38–57. doi:10.1016/j.plipres.2017.06.001.

[2] K. Kawajiri and Y. Fuji-Kuriyama, *The aryl hydrocarbon receptor: A multifunctional chemical sensor for host defense and homeostatic maintenance*, Exp. Anim. 66 (2016), pp. 75–89. doi:10.1538/expanim.16-0092.

[3] K.W. Bock, *From TCDD-mediated toxicity to searches of physiologic AHR functions*, Biochem. Pharmacol. 155 (2018), pp. 419–424. doi:10.1016/j.bcp.2018.07.032.

[4] R. Pohjanvirta, *The AH Receptor in Biology and Toxicology*, John Wiley and Sons, Hoboken, NJ, 2011.

[5] L. Bonati, D. Corrada, S. Giani Tagliabue, and S. Motta, *Molecular modeling of the AhR structure and interactions can shed light on ligand-dependent activation and transformation mechanisms*, Curr. Opin. Toxicol. 2 (2017), pp. 42–49. doi:10.1016/j.cotox.2017.01.011.

[6] T. Hamers, J.H. Kamstra, P.H. Cenijn, K. Pencikova, L. Palkova, P. Simeckova, J. Vondracek, P. L. Andersson, M. Stenberg, and M. Machala, *In vitro toxicity profiling of ultrapure non-dioxin-*

like polychlorinated biphenyl congeners and their relative toxic contribution to PCB mixtures in humans, Toxicol. Sci. 121 (2011), pp. 88–100. doi:10.1093/toxsci/kfr043.

[7] P. Brenerová, T. Hamers, J.H. Kamstra, J. Vondráček, S. Strapáčová, P.L. Andersson, and M. Machala, *Pure non-dioxin-like PCB congeners suppress induction of AhR-dependent endpoints in rat liver cells*, Environ. Sci. Pollut. Res. 23 (2016), pp. 2099–2107. doi:10.1007/s11356-015-4819-6.

[8] L. Stejskalova, Z. Dvorak, and P. Pavek, *Endogenous and exogenous ligands of aryl hydrocarbon receptor: Current state of art*, Curr. Drug Metab. 12 (2011), pp. 198–212. doi:10.2174/138920011795016818.

[9] G.A. Reed, K.S. Peterson, H.J. Smith, J.C. Gray, D.K. Sullivan, M.S. Mayo, J.A. Crowell, and A. Hurwitz, *A phase I study of indole-3-carbinol in women: Tolerability and effects*, Cancer Epidemiol. Biomarkers Prev. 14 (2005), pp. 1953–1960. doi:10.1158/1055-9965.EPI-05-0121.

[10] O. Hankinson, *The role of AHR-inducible cytochrome P450s in metabolism of polyunsaturated fatty acids*, Drug Metab. Rev. 48 (2016), pp. 342–350. doi:10.1080/03602532.2016.1197240.

[11] A. Puga, Y. Xia, and C. Elferink, *Role of the aryl hydrocarbon receptor in cell cycle regulation*, Chem. Biol. Interact. 141 (2002), pp. 117–130. doi:10.1016/S0009-2797(02)00069-8.

[12] C. Esser, A. Rannug, and B. Stockinger, *The aryl hydrocarbon receptor in immunity*, Trends Immunol. 30 (2009), pp. 447–454. doi:10.1016/j.it.2009.06.005.

[13] C. Duval, E. Blanc, and X. Coumoul, *Aryl hydrocarbon receptor and liver fibrosis*, Curr. Opin. Toxicol. 8 (2018), pp. 8–13. doi:10.1016/j.cotox.2017.11.010.

[14] K. Pěnčíková, L. Svržková, S. Strapáčová, J. Neča, I. Bartoňková, Z. Dvořák, M. Hýžďalová, J. Pivnička, L. Pálková, H.J. Lehmler, X. Li, J. Vondráček, and M. Machala, *In vitro profiling of toxic effects of prominent environmental lower-chlorinated PCB congeners linked with endocrine disruption and tumor promotion*, Environ. Pollut. 237 (2018), pp. 473–486. doi:10.1016/j.envpol.2018.02.067.

[15] S. Safe, S.O. Lee, and U.H. Jin, *Role of the aryl hydrocarbon receptor in carcinogenesis and potential as a drug target*, Toxicol. Sci. 135 (2013), pp. 1–16. doi:10.1093/toxsci/kft128.

[16] Z. Xue, D. Li, W. Yu, Q. Zhang, X. Hou, Y. He, and X. Kou, *Mechanisms and therapeutic prospects of polyphenols as modulators of the aryl hydrocarbon receptor*, Food Funct. 8 (2017), pp. 1414–1437. doi:10.1039/C6FO01810F.

[17] R. Nakai, S. Fukuda, M. Kawase, Y. Yamashita, and H. Ashida, *Curcumin and its derivatives inhibit 2,3,7,8,-tetrachloro-dibenzo-p-dioxin-induced expression of drug metabolizing enzymes through aryl hydrocarbon receptor-mediated pathway*, Biosci. Biotechnol. Biochem. 82 (2017), pp. 616–628. doi:10.1080/09168451.2017.1386086.

[18] H. Kandárová and S. Letaáiová, *Alternative methods in toxicology: Pre-validated and validated methods*, Interdiscip. Toxicol. 4 (2011), pp. 107–113. doi:10.2478/v10102-011-0018-6.

[19] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.l. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, and A. Tropsha, *QSAR modeling: Where have you been? Where are you going to?* J. Med. Chem. 57 (2014), pp. 4977–5010. doi:10.1021/jm4004285.

[20] L.G. Valerio, C. Yang, K.B. Arvidson, and N.L. Kruhlak, *A structural feature-based computational approach for toxicology predictions*, Expert Opin. Drug Metab. Toxicol. 6 (2010), pp. 505–518. doi:10.1517/17425250903499286.

[21] D. Szöllösi, Á. Erdei, G. Gyimesi, C. Magyar, and T. Hegedüs, *Access path to the ligand binding pocket may play a role in xenobiotics selection by AhR*, PLoS ONE 11 (2016), pp. 1–22. doi:10.1371/journal.pone.0146066.

[22] A.D. Şahin and M.T. Saçan, *Understanding the toxic potencies of xenobiotics inducing TCDD/TCDF-like effects*, SAR QSAR Environ. Res. 29 (2018), pp. 117–131. doi:10.1080/1062936X.2017.1414075.

[23] M.K. Gadhwal, S. Patil, P. D'mello, and A. Joshi, *Homology modeling of aryl hydrocarbon receptor and docking of agonists and antagonists*, Int. J. Pharm. Pharm. Sci. 5 (2013), pp. 76–81.

[24] D. Dolciami, M. Gargaro, B. Cerra, G. Scalisi, L. Bagnoli, G. Servillo, M.A.D. Fazia, P. Puccetti, F. J. Quintana, F. Fallarino, and A. Macchiarulo, *Binding mode and structure–activity relationships of ITE as an aryl hydrocarbon receptor (AhR) agonist*, ChemMedChem 13 (2018), pp. 270–279. doi:10.1002/cmdc.201700669.

[25] S.-H. Seok, W. Lee, L. Jiang, K. Molugu, A. Zheng, Y. Li, S. Park, C.A. Bradfield, and Y. Xing, *Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex*, Proc. Natl. Acad. Sci. 114 (2017), pp. 5431–5436. doi:10.1073/pnas.1617035114.

[26] G. Suzuki, M. Nakamura, C. Michinaka, N.M. Tue, H. Handa, and H. Takigami, *Dioxin-like activity of brominated dioxins as individual compounds or mixtures in in vitro reporter gene assays with rat and mouse hepatoma cell lines*, Toxicol. Vitro 44 (2017), pp. 134–141. doi:10.1016/j. tiv.2017.06.025.

[27] S. Sciuto, M. Prearo, R. Desiato, C. Bulfon, E.A.V. Burioli, G. Esposito, C. Guglielmetti, L. Dell'atti, G. Ru, D. Volpatti, P.L. Acutis, and F. Martucci, *Dioxin-like compounds in lake fish species: Evaluation by DR-CALUX bioassay*, J. Food Prot. 81 (2018), pp. 842–847. doi:10.4315/0362-028X.JFP-17-476.

[28] A.L.J. Nuerla, X.L. Qiao, J. Li, D.M. Zhao, X.H. Yang, Q. Xie, and J.W. Chen, *Effects of substituent position on the interactions between PBDEs/PCBs and DOM*, Chin. Sci. Bull. 58 (2013), pp. 884–889. doi:10.1007/s11434-012-5464-9.

[29] R. Panda, A.S.S. Cleave, and P. Suresh, *In silico predictive studies of mAHR congener binding using homology modelling and molecular docking*, Toxicol. Ind. Health 30 (2014), pp. 765–776. doi:10.1177/0748233712463774.

[30] H.F. Berntsen, V. Berg, C. Thomsen, E. Ropstad, and K.E. Zimmer, *The design of an environmentally relevant mixture of persistent organic pollutants for use in in vivo and in vitro studies*, J. Toxicol. Environ. Health Part A 24 (2017), pp. 1002–1016. doi:10.1080/15287394.2017.1354439.

[31] A.J. Murk, J. Legler, M.S. Denison, J.P. Giesy, C. Van de Guchte, and A. Brouwer, *Chemical-activated luciferase gene expression (CALUX): A novel in vitro bioassay for Ah receptor active compounds in sediments and pore water*, Fundam. Appl. Toxicol. 33 (1996), pp. 149–160. doi:10.1006/faat.1996.0152.

[32] OECD, *Test No. 455: Performance-Based Test Guideline for Stably Transfected Transactivation in Vitro Assays to Detect Estrogen Receptor Agonists and Antagonists*, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris, 2016.

[33] R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics*, WILEY-VCH, Weinheim, Germany, 2009.

[34] C.W. Yap, *PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints*, J. Comput. Chem. 32 (2011), pp. 1466–1474. doi:10.1002/jcc.v32.7.

[35] R.W.P. Urias, S.J. Barigye, Y. Marrero-Ponce, C.R. García-Jacas, J.R. Valdes-Martiní, and F. Perez-Gimenez, *IMMAN: Free software for information theory-based chemometric analysis*, Mol. Divers. 19 (2015), pp. 305–319. doi:10.1007/s11030-014-9565-z.

[36] *MATLAB MathWorks Inc R2017b 9.3.0.713579*, MathWorks Inc, USA, 2017. software. Available at https://www.mathworks.com/.

[37] *STATISTICA StatSoft Inc* v8.0. StatSoft Inc, USA, 2007. software. Available at http://www.statsoft.com.

[38] J. Jaworska and N. Nikolova-Jeliazkova, *How can structural similarity analysis help in category formation?* SAR QSAR Environ. Res. 18 (2007), pp. 195–207. doi:10.1080/10629360701306050.

[39] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S.H. Bryant, *PubChem substance and compound databases*, Nucleic Acids Res. 44 (2016), pp. D1202–D1213. doi:10.1093/nar/gkv951.

[40] S.L. Dixon, A.M. Smondyrev, and S.N. Rao, *PHASE: A novel approach to pharmacophore modeling and 3D database searching*, Chem. Biol. Drug Des. 67 (2006), pp. 370–372. doi:10.1111/j.1747-0285.2006.00384.x.

[41] J.F. Truchon and C.I. Bayly, *Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem*, J. Chem. Inf. Model. 47 (2007), pp. 488–508. doi:10.1021/ci600426e.

[42] M.M. Mysinger, M. Carchia, J.J. Irwin, and B.K. Shoichet, *Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking*, J. Med. Chem. 55 (2012), pp. 6582–6594. doi:10.1021/jm300687e.

[43] T. Hamers, J.H. Kamstra, E. Sonneveld, A.J. Murk, M.H.A. Kester, P.L. Andersson, J. Legler, and A. Brouwer, *In vitro profiling of the endocrine-disrupting potency of brominated flame retardants*, Toxicol. Sci. 92 (2006), pp. 157–173. doi:10.1093/toxsci/kfj187.

[44] C. Sammut and G. Webb (eds.), *Encyclopedia of Machine Learning and Data Mining*, 2nd ed., Vol. 32, Springer, New York, NY, 2018.

[45] M. Ghorbanzadeh, K.I. Van Ede, M. Larsson, M.B.M. Van Duursen, L. Poellinger, S. Lücke-Johansson, M. Machala, K. Pěnčíková, J. Vondráček, M. van den Berg, M.S. Denison, T. Ringsted, and P.L. Andersson, *In vitro and in silico derived relative effect potencies of ah-receptor-mediated effects by PCDD/Fs and PCBs in rat, mouse, and guinea pig CALUX cell lines*, Chem. Res. Toxicol. 27 (2014), pp. 1120–1132. doi:10.1021/tx5001255.

[46] OECD, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, 2014. OECD Series on Testing and Assessment OECD, No. 69, OECD Publishing, Paris.

[47] O.F. Güner and J.P. Bowen, *Setting the record straight: The origin of the pharmacophore concept*, J. Chem. Inf. Model. 54 (2014), pp. 1269–1283. doi:10.1021/ci5000533.

[48] D.P. Williams, *Toxicophores: Investigations in drug safety*, Toxicology 226 (2006), pp. 1–11. doi:10.1016/j.tox.2006.05.101.

[49] P.K. Singh, A. Negi, P.K. Gupta, M. Chauhan, and R. Kumar, *Toxicophore exploration as a screening technology for drug design and discovery: Techniques, scope and limitations*, Arch. Toxicol. 90 (2016), pp. 1785–1802. doi:10.1007/s00204-015-1587-5.

[50] L. Chen, D. Wu, H.P. Bian, G.L. Kuang, J. Jiang, W.H. Li, G.X. Liu, S. Zou, J. Huang, and Y. Tang, *Selective ligands of estrogen receptor β discovered using pharmacophore mapping and structure-based virtual screening*, Acta Pharmacol. Sin. 35 (2014), pp. 1333–1341. doi:10.1038/aps.2014.69.

[51] S.S. Ashtekar, N.M. Bhatia, and M.S. Bhatia, *Development of leads targeting ER-α in breast cancer: An in silico exploration from natural domain*, Steroids 131 (2018), pp. 14–22. doi:10.1016/j.steroids.2017.12.016.

[52] N. Lagarde, S. Delahaye, A. Jérémie, N. Ben Nasr, H. Guillemain, C. Empereur-Mot, V. Laville, T. Labib, M. Réau, F. Langenfeld, J.F. Zagury, and M. Montes, *Discriminating agonist from antagonist ligands of the nuclear receptors using different chemoinformatics approaches*, Mol. Inf. 36 (2017), pp. 1–16. doi:10.1002/minf.201700020.

[53] R.T. Engeli, S.R. Rohrer, A. Vuorinen, S. Herdlinger, T. Kaserer, S. Leugger, D. Schuster, and A. Odermatt, *Interference of paraben compounds with estrogen metabolism by inhibition of 17β-hydroxysteroid dehydrogenases*, Int. J. Mol. Sci. 18 (2017). doi:10.3390/ijms18092007.

[54] I.T. Jolliffe, *Principal Component Analysis*, 2nd ed., Vol. 98, Springer-Verlag, New York, NY, 2002.

[55] K. Schomburg, H. Ehrlich, K. Stierand, and M. Rarey, *From structure diagrams to visual chemical patterns*, J. Chem. Inf. Model. 50 (2010), pp. 1529–1535. doi:10.1021/ci100209a.

[56] L. Poellinger and D. Gullberg, *Characterization of the hydrophobic properties of the receptor for 2,3,7,8-tetrachlorodibenzo-p-dioxin*, Mol. Pharmacol. 27 (1985), pp. 271–276.

[57] T. Tripathi and A.K. Saxena, *2D- QSAR studies on new stilbene derivatives of resveratrol as a new selective aryl hydrocarbon receptor*, Med. Chem. Res. 17 (2008), pp. 212–218. doi:10.1007/s00044-007-9055-2.

[58] Y. Xing, M. Nukaya, K.A. Satyshur, L. Jiang, V. Stanevich, E.N. Korkmaz, L. Burdette, G. D. Kennedy, Q. Cui, and C.A. Bradfield, *Identification of the Ah-receptor structural determinants for ligand preferences*, Toxicol. Sci. 129 (2012), pp. 86–97. doi:10.1093/toxsci/kfs194.

[59] K. Bekki, H. Takigami, G. Suzuki, N. Tang, and K. Hayakawa, *Evaluation of toxic activities of polycyclic aromatic hydrocarbon derivatives using in vitro bioassays*, J. Health Sci. 55 (2009), pp. 601–610. doi:10.1248/jhs.55.601.

[60] C. Gu, X. Ju, X. Jiang, K. Yu, S. Yang, and C. Sun, *Improved 3D-QSAR analyzes for the predictive toxicology of polybrominated diphenyl ethers with CoMFA/CoMSIA and DFT*, Ecotoxicol. Environ. Saf. 73 (2010), pp. 1470–1479. doi:10.1016/j.ecoenv.2009.11.003.

[61] C.R. Martinez and B.L. Iverson, *Rethinking the term "pi-stacking,"* Chem. Sci. 3 (2012), pp. 2191–2201.

[62] R.C. Kolanczyk, J.S. Denny, B.R. Sheedy, P.K. Schmieder, and M.A. Tapper, *Estrogenic activity of multicyclic aromatic hydrocarbons in rainbow trout (*Oncorhynchus mykiss*) in vitro assays*, Aquat. Toxicol. (2019), pp. 43–51. doi:10.1016/j.aquatox.2018.11.023.

[63] C. Yun, J.A. Weiner, D.S. Chun, J. Yun, R.W. Cook, M.S. Schallmo, A.S. Kannan, S.M. Mitchell, R. D. Freshman, C. Park, W.K. Hsu, and E.L. Hsu, *Mechanistic insight into the effects of Aryl hydrocarbon receptor activation on osteogenic differentiation*, Bone Rep. 6 (2017), pp. 51–59. doi:10.1016/j.bonr.2017.02.003.

[64] C.A. McQueen (ed.), *Comprehensive Toxicology*, 3rd ed., Vol. 35, Elsevier, Oxford, UK, 2018.

Conclusions

THESIS CONCLUSIONS

During the research work of this Ph.D. thesis, more than one hundred structurally diverse chemical compounds from natural and synthetic origins were studied as AhR modulators. Except for some phytochemicals, most of the compounds evaluated *in vitro* are for the first time herein reported. The *in silico* methods built and based in QSARs, Pharmacophore Modeling and Molecular Docking simulations are novel and with validated predictive capacity. Therefore, the general objective proposed in this Ph.D. thesis was successfully accomplished. The conclusions that corroborate the fulfillment of the specific objectives proposed for each Chapter are:

CHAPTER 1. The literature on AhR mediated effects and existing theoretical and experimental methods to investigate the receptor was explored, highlighting that:

☑ The transcription factor AhR regulates multiple signaling networks.

☑ AhR is a key receptor in physiological conditions but it also mediates several adverse health effects.

☑ AhR is a promiscuous receptor with the capacity to accommodate diverse endogenous and exogenous ligands.

☑ The luciferase reporter gene assays are important *in vitro* techniques widely used to explore the effects of chemical compounds on AhR expression.

☑ Some *in silico* studies on AhR and its ligands have been reported, although, important gaps remain.

CHAPTER 2. Several QSAR models were developed and experimentally validated to predict the AhR agonist activity of chemical compounds.

☑ Five classification models for AhR agonism were built using a large and balanced dataset

☑ A 10-fold external validation demonstrated the robustness and predictiveness of built QSARs.

☑ An *in vitro* experimental validation corroborated the predictivity of the ensemble classifier.

☑ The benzothiazoles were the most prominent group among the AhR agonists.

☑ The built QSARs may be useful in guiding the screening of AhR agonists.

CHAPTER 3. Sixteen 2-aryl and 2-pyridinylbenzothiazoles were evaluated *in vitro* as AhR modulators and as antimicrobial agents.

☑ All the benzothiazoles showed noticeable antimicrobial effects against Gram-positive and Gram-negative pathogens and against the yeast *C. albicans.*

☑ Six benzothiazoles exhibited significant AhR agonist effects in a cell-based reporter gene assay.

☑ A structure-activity relationship analysis revealed relevant contributions of the substituents to the studied biological effects.

☑ One compound displayed promising *in vitro* biocide effects and AhR agonist activity, as well as an adequate ADMET profile and binding similarities with FICZ.

CHAPTER 4. A novel set of triarylmethane compounds was evaluated *in vitro* as AhR ligands.

☑ Series of novel triarylmethane compounds were straightforwardly synthesized.

☑ 32 triarylmethanes were screened in a cell-based bioassay as AhR modulators and 8 of them exhibited AhR agonist activity.

☑ The introduction of heteroaromatic or naphthol substituents was favorable for AhR agonist activity.

☑ Adequate ADME profiles and druglikeness were predicted *in silico* for the active AhR agonist compounds.

☑ Docking studies revealed AhR binding resemblances for the most potent TAM identified and the endogenous ligand used as positive control (FICZ).

CHAPTER 5. The AhR antagonism was studied in a set of potentially harmful substances using computational methods.

☑ Results from *in vitro* AhR-CALUX bioassays were modeled and validated using QSAR methods and Toxicophore mapping.

☑ The presence of aromatic rings and electron-acceptor moieties were demonstrated to be important chemical features for the AhR antagonism.

☑ The hydrophobic constraints seem to influence the AhR antagonistic capacity of toxic chemicals.

Annexes

ANNEXE 1. Review article: QSAR modeling of endocrine disruption

- **Publication Title:** PROTECTion against Endocrine Disruption through Quantitative Structure-Activity Relationship modelling [*Revisión de los modelos computacionales que relacionan la estructura química con la disrupción del sistema endocrino*].

- **Authors**: Elizabeth Goya-Jorge, Jesús V. de Julián-Ortiz, Rafael Gozalbes.

Revisión de los modelos computacionales que relacionan la estructura química con la disrupción del sistema endocrino

Goya-Jorge, E. [1,2], de Julián-Ortiz, J.V. [3], Gozalbes, R. [1]*

[1] ProtoQSAR SL., CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia. Av. Benjamin Franklin 12, 46980 Paterna, Valencia, España.
[2] Departamento de Farmacología, Facultad de Farmacia, Universidad de Valencia, Av. Vicente Andrés Estellés, s/n, 46100 Burjassot, Valencia, España.
[3] Unidad de Diseño de Fármacos y Topología Molecular, Departamento de Química-Física. Facultad de Farmacia, Universidad de Valencia, Av. Vicente Andrés Estellés, s/n, 46100 Burjassot, Valencia, España.

Resumen: El término interruptor endocrino define una amplia y diversa clase de sustancias de origen natural o antropogénico con la capacidad de interferir con alguna función del sistema endocrino y provocar efectos adversos en un organismo o su descendencia. La disrupción endocrina se asocia con cáncer, obesidad, diabetes, disfunción reproductiva e inmunológica. Constituye una forma específica de toxicidad cuyas regulaciones y legislación actualmente carecen de consenso. Los métodos computacionales, y particularmente los estudios quimioinformáticos como las relaciones cuantitativas estructura-actividad (QSAR), son herramientas valiosas de investigación que han ocupado gradualmente un importante espacio en los estudios toxicológicos. Esta revisión propone un análisis del más reciente estado del arte relativo a la modelación QSAR de la disrupción endocrina. Los casos de estudio seleccionados se centran en tres mecanismos importantes que representan la biosíntesis, el transporte y la interacción con los receptores hormonales mediados por la capacidad inhibitoria de la enzima aromatasa, y los efectos sobre la proteína transportadora de transtiretina y el receptor de andrógenos, respectivamente. Estas herramientas predictivas pueden ayudar a priorizar sustancias como posibles alteradores endocrinos y, por lo tanto, son contribuciones importantes que garantizan el ahorro de tiempo, materiales y recursos humanos.

Palabras clave: QSAR, modelos computacionales, interruptor endocrino, TTR, aromatasa, hormona.

Abstract: *Protection against Endocrine Disruption through Quantitative Structure-Activity Relationship modelling*

The endocrine disruptors are defined as a broad and diverse class of substances of natural or anthropogenic origin with the ability to interfere with some function of the endocrine system and, in doing so, cause adverse effects on an organism or its descendants. Endocrine disruption, associated with pathologies such as cancer, obesity, diabetes, and reproductive and immunological dysfunction, constitutes a specific form of toxicity whose regulation and legislation currently lack consensus. Computational methods, and within them chemoinformatic studies such as the prediction of quantitative structure-activity relationships (QSAR), are valuable research tools that have gradually occupied an important space in toxicological studies. This review proposes an analysis of the most recent state of the art related to QSAR modelling in the context of endocrine disruption. For this, case studies reported on three important hormonal mechanisms were selected, which represent synthesis, transport, and interaction with receptors. The summarized QSARs modelled the inhibitory capacity of the aromatase enzyme and the effects on the transthyretin transporter protein and the androgen receptor. These predictive tools can assist in prioritizing substances as potential endocrine disruptors and are therefore important contributions that guarantee the saving of time, material, and human resources.

Keywords: QSAR, endocrine disruptor, TTR, aromatase, androgen receptor, hormone.

Introducción

Desde mediados del siglo XX, la capacidad de ciertos compuestos químicos de irrumpir en el organismo y causar modificaciones en las funciones de las hormonas ha preocupado a la comunidad científica internacional (Gassner et al., 1958). Por su terminología en inglés "*endocrine disruptors*" o "*endocrine disrupting chemicals*", se ha extendido y popularizado su nombre indistintamente como "disruptores endocrinos" (anglicismo), interruptor o alterador endocrino. En mamíferos, aves, reptiles y peces, los efectos de estas sustancias se han detectado en alteraciones del comportamiento sexual, la masculinización de hembras y la feminización de machos, que incluye la reducción del tamaño de los órganos sexuales, así como la pérdida de su funcionalidad. Otras consecuencias son el desarrollo de malformaciones y procesos cancerosos, niveles séricos anómalos de las principales hormonas, problemas de fertilidad, perturbaciones en la respuesta inmune, disminución de la capacidad de aprendizaje y de la memoria a corto y largo plazo. Niveles elevados de compuestos interruptores endocrinos en fluidos como la leche materna han revelado el riesgo de exposición temprana a estos tóxicos (Schug et al., 2016; Yilmaz et al., 2020).

La publicación del libro de Rachel Carson *"Silent Spring"* en 1962 supuso para el público en general una llamada de atención ante las consecuencias a largo plazo de la contaminación medioambiental, y en particular sobre los riesgos potenciales en la reproducción de productos químicos sintéticos como el dicloro difenil tricloroetano (DDT). Sin embargo, la primera tentativa global de establecer una definición formal de los compuestos que provocan efectos adversos al interactuar con el sistema endocrino data de 1991 (Bern et al., 1991).

A nivel europeo, la primera iniciativa en relación con los alteradores endocrinos fue la aprobación en el Parlamento, en octubre de 1998 de la "Resolución sobre las sustancias químicas que provocan perturbaciones endocrinas" propuesta por la Comisión de Medio Ambiente, Salud Pública y Protección del Consumidor. Comenzó a raíz de esta resolución a aplicarse en la Unión Europea la aún vigente "Estrategia Comunitaria para los Alteradores Endocrinos" con diferentes proyectos de investigación asociados (Karjalainen et al., 2017).

En una de las primeras resoluciones adoptadas por el Parlamento Europeo en el año 2000, se enfatizó la necesidad de desarrollar sistemas de control de riesgo rápidos y efectivos, y de actuar inmediatamente en la identificación de sustancias con potencial toxicidad endocrina. El Parlamento y el Consejo de Ministros se harían eco de este imperativo al proyectar la disrupción endocrina como parte del Reglamento (CE) N.º 1907/2006 (REACH, acrónimo de Registro, Evaluación, Autorización y Restricción de sustancias y mezclas químicas), así como la creación de la Agencia Europea de Sustancias y Mezclas Químicas (ECHA, del inglés "*European Chemicals Agency*"). De conformidad con REACH, los alteradores endocrinos se identifican como sustancias extremadamente preocupantes, junto con las consideradas carcinógenas, mutagénicas y tóxicas para la reproducción (sustancias CMR), con objetivo de reducir su uso y sustituirlas progresivamente por otras alternativas más seguras.

Junto a la creciente llamada de atención en torno a la disrupción endocrina en la década de los 90, la Organización Mundial de la Salud (OMS) y la Organización para la Cooperación y el Desarrollo Económico (OECD) publicaron numerosas directivas en pro de reducir la experimentación animal en la investigación científica. Se intensificaría así el ya notable interés en la aplicación de métodos alternativos al uso de animales, y entre ellos los estudios

*e-mail: rgozalbes@protosar.com

computacionales o modelación "QSAR" (del inglés *Quantitative Structure-Activity Relationships*).

Actualmente, el desarrollo de métodos QSAR continúa siendo una temática de creciente impacto en el campo toxicológico y farmacológico (Gozalbes et al., 2015; Lin et al., 2020). La aplicación de esta metodología para la predicción de la toxicidad endocrina causada por compuestos químicos a través de la enzima aromatasa, la proteína transportadora transtiretina y el receptor de andrógenos será objeto de revisión en este artículo.

El estudio que se presenta fue desarrollado en el contexto del proyecto europeo H2020- PROTECTED: "*Protection against endocrine disruptors; detection, mixtures and health effects, risk assessment and communication*" (http://protected.eu.com/) que tiene como objetivo integrar diversos métodos de evaluación de la toxicidad endocrina de compuestos y mezclas químicas, así como promover la conciencia social de los riesgos que supone la disrupción endocrina.

Panorama conceptual y teórico

2.1. El sistema endocrino y su disrupción

2.1.1. Las hormonas

Las hormonas son mensajeros químicos que se encargan de regular y coordinar las funciones metabólicas principales, el crecimiento y el desarrollo de diferentes poblaciones de células en un organismo. No existe ningún órgano o sistema en el cuerpo humano fuera del control de una o varias rutas del sistema hormonal, más conocido como sistema endocrino (Tata, 2017). De acuerdo con su estructura química, las hormonas se clasifican en tres grupos: péptidos/proteínas, esteroides y

aminas. Algunas de las principales hormonas humanas y las glándulas endocrinas donde son secretadas se muestran en la Figura 1.

La complejidad del sistema endocrino queda evidenciada en la diversidad estructural y funcional de las moléculas hormonales representada en la Figura 1 (Leblebicioglu et al., 2013). Existen hormonas de naturaleza puramente hidrofóbica como es el caso de las esteroidales que derivan del lípido colesterol y entre las cuales se encuentran las asociadas con la reproducción como el estradiol, la progesterona y la testosterona. Sin embargo, hormonas peptídicas/proteicas como la insulina secretada en el páncreas o la calcitonina en las glándulas tiroideas, están formadas por múltiples cadenas de aminoácidos que le atribuyen propiedades hidrofílicas. Las hormonas de tipo aminas contienen representantes hidrofílicos como la serotonina o la dopamina y otros lipofílicos como la tiroxina (Soto et al., 2009).

2.1.2. La actividad hormonal

La secreción o síntesis de hormonas es un proceso que implica múltiples tejidos debido a las variadas funciones que tienen estas moléculas en el organismo. En el caso de las hormonas esteroideas, los mecanismos enzimáticos involucrados se agrupan como: enzimas específicas del citocromo P450 (CYPs), hidroxiesteroide-deshidrogenasas (HSDs) y esteroide reductasas (Miller WL, 1988). Un ejemplo característico es la enzima P450 aromatasa (CYP19A1), la única capaz de catalizar en vertebrados la reacción de aromatización que transforma andrógenos en estrógenos (Sanderson, J.T., 2006). Otra enzima importante es la hemoproteína transmembrana glicosilada tiroperoxidasa (TPO), que desempeña un papel crucial en la síntesis de hormonas tiroideas (Yen, 2001).

Glándula endocrina	Hormona
Pituitaria	LH, FSH, TSH, GH
Hipotálamo	GHR, TRH, GRH
Páncreas	Insulina, Glucagón Somatostatina
Tiroides	Calcitonina

Glándula Endocrina	Hormona
Tracto Gastrointestinal	Serotonina
Médula Adrenal	Epinefrina Norepinefrina
Hipotálamo	Dopamina
Tiroides	Tiroxina (T4) Triiyodotironina (T3)

Glándula Endocrina	Hormona
Corteza Adrenal	Aldosterona, Cortisol
Ovarios	Estradiol, Progesterona Testosterona
Testículos	Testosterona, Estradiol

Figura 1: Diagrama de las clases químicas de hormonas y algunos ejemplos de las glándulas endocrinas que las secretan.

El proceso de señalización endocrina involucra a un gran número de receptores y rutas celulares. Un grupo importante es la superfamilia de receptores nucleares (NR), reguladores de la transcripción y activados por ligandos endógenos y exógenos. Estos receptores suelen compartir patrones de organización estructural que incluyen un dominio N-terminal variable, uno central de unión al ADN y uno C-terminal al cual se unen ligandos como las hormonas, pero también prostaglandinas, ciertas vitaminas y ácidos grasos (McEwan, 2016). Están relacionados con diversas funciones fisiológicas en los metazoos que van desde el crecimiento, desarrollo y reproducción hasta el metabolismo y la homeostasia. Algunos de los receptores nucleares más conocidos son el receptor de estrógenos (ESR), el receptor de andrógenos (AR), y los receptores de hormonas tiroideas (TR). Son estos receptores también la diana de múltiples xenobióticos que, mimetizando o antagonizando el efecto de las hormonas, interrumpen la homeostasia endocrina causando diversas consecuencias tóxicas (Balaguer et al., 2019). El ESR tiene como ligando endógeno la hormona 17β-estradiol (E2). Existen dos subtipos (ESRα y ESRβ), ambos con un papel importante en la fisiología humana. Sin embargo, sus efectos pueden llegar a ser antagónicos en muchos casos. Con el descubrimiento de ESRβ se reveló que las funciones estrogénicas no se limitaban a la reproducción femenina, sino que tenían relación con muchos otros órganos y sistemas en ambos sexos (McEwan, 2016; Tata, 2017). Por otro lado, la activación del AR está ligada a las hormonas androgénicas testosterona y 5α-dihidrotestosterona (5α-DHT). Fisiológicamente el AR es el responsable de la diferenciación sexual masculina, desde el útero, hasta los cambios que acontecen en la pubertad. En el hombre adulto sus efectos han sido relacionados con la masa y la fuerza muscular (Leblebicioglu et al., 2013). Por último, los receptores tiroideos se pueden encontrar expresados a partir de dos genes independientes (α y β), y cada gen tiene a su vez 2 isoformas del receptor. Intervienen entre otras funciones en el crecimiento, el desarrollo y la regulación del metabolismo (Little, 2016).

2.1.3. *La disrupción endocrina*

La acción sobre receptores nucleares, como se ha mencionado anteriormente, está entre los mecanismos más comunes por los cuales sustancias de todo tipo pueden alterar el efecto de las hormonas (Balaguer et al., 2019). Sin embargo, muchas otras rutas de interrupción del sistema endocrino han sido identificadas a consecuencia de la acción de entidades químicas (WHO/UNEP, 2013). Los más importantes se representan en la Figura 2.
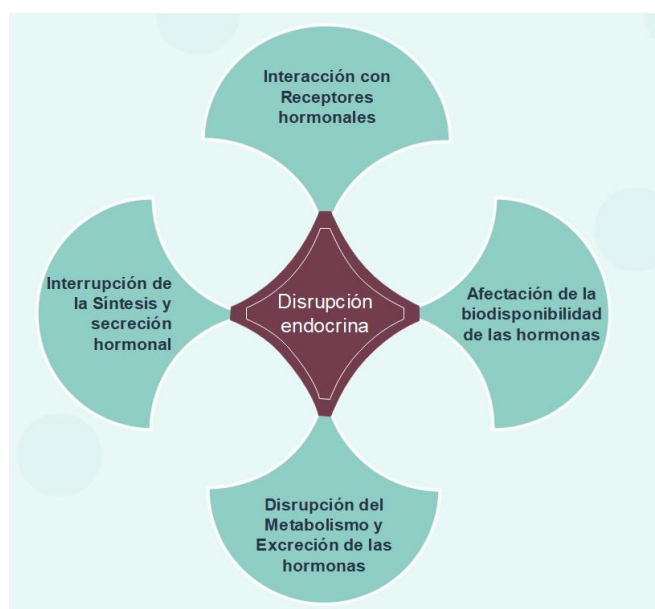
Figura 2: Principales mecanismos de disrupción del sistema endocrino causado por agentes químicos.

La toxicidad sobre el sistema endocrino puede ser causada por la alteración de la síntesis y secreción de hormonas, afectando por ejemplo la aromatasa P450 o inhibiendo la TPO. Diferentes compuestos químicos son capaces de causar también modificaciones en la biodisponibilidad y en la unión a proteínas trasportadoras como las globulinas de enlace a hormonas sexuales (SHBG), las de corticosteroides (CBG) y la transtiretina (TTR) (Cheek et al., 1999; Hong et al., 2015). Además, han sido descritos otros mecanismos de disrupción del metabolismo y la excreción de hormonas.

Por otro lado, los elementos que regulan la activación o desactivación no fisiológica de rutas hormonales dependen, entre otros factores, de la capacidad de respuesta de las distintas células y tejidos, lo que a su vez está condicionado por la afinidad hacia el receptor, el número de receptores disponibles, la concentración hormonal y la presencia de cofactores que puedan favorecer o reprimir sus efectos. Los perfiles de expresión de los receptores suelen variar también durante el desarrollo, y una hormona es capaz de cumplir diferentes roles en un momento u otro del crecimiento. Los múltiples elementos involucrados complican por tanto el estudio de las acciones celulares de las hormonas, así como el perfil toxicológico de los compuestos que interfieren con los mecanismos endocrinos (Vandenberg et al., 2013).

Durante décadas de estudio e investigación, diversos compuestos han demostrado tener capacidad para alterar el sistema endocrino ocasionando numerosas consecuencias fisiológicas. Sin embargo, no existe una definición universalmente aceptada por las entidades científicas y regulatorias para los llamados "interruptores endocrinos". La principal limitación consiste en la dificultad de integrar en un único concepto la diversidad estructural, los mecanismos de acción y los escenarios de exposición asociados con la toxicidad endocrina (Schug et al., 2016).

Así por ejemplo, la Comisión Europea asume la definición establecida por la OMS en el informe emitido por el Programa Internacional de Seguridad Química de 2002 según la cual un interruptor endocrino es *"una sustancia o mezcla exógena que altera la(s) función(es) del sistema endocrino, causando consecuentemente efectos adversos para la salud en un organismo, su progenie o (sub)poblaciones"* (WHO, 2002). Sin embargo, la *Endocrine Society* de los Estados Unidos, acordó en 2012 definir los alteradores endocrinos como aquellos "compuestos químicos exógenos o mezclas que pueden interferir con cualquier aspecto de la acción hormonal" (Thomas Zoeller et al., 2012). La inclusión pues, de los efectos adversos a la hora de considerar un compuesto como interruptor endocrino ha abierto una brecha entre ambos conceptos, lo que se traduce en significativas diferencias en el número de sustancias a considerar según una u otra definición.

La falta de coherencia en este sentido no ha impedido que los términos como disrupción, interruptores o alteración endocrina se empleen alegando indistintamente cualquiera de los conceptos. Los interruptores endocrinos se han asumido globalmente como compuestos cuya determinación permite establecer parámetros o criterios de toxicidad relativos a las alteraciones que, por múltiples mecanismos, pueden desencadenar diversos agentes químicos sobre las rutas hormonales. Se obvia, por tanto, en muchos casos la definición establecida por la OMS, y se reportan ensayos, mediciones, determinaciones o modelos de disrupción endocrina en los cuales no se demuestra que el efecto sobre el organismo, su progenie o (sub)poblaciones sea adverso, sino que pueda ser considerado, simplemente, un efecto potencial. Existe, claro está, una necesidad de aunar esfuerzos para sistematizar la definición de los compuestos que deberían considerarse interruptores endocrinos, de manera que su estudio pueda ser abordado en distintos contextos sin generar contradicción.

2.2. *Generalidades de la modelación QSAR*

La información toxicológica es cada vez más específica y amplia debido a los esfuerzos internacionales de proveer datos científicos para regular las sustancias químicas y garantizar la salud de las personas y del medioambiente. Sin embargo, persisten limitaciones y lagunas de

información. En el caso de la toxicidad endocrina, aun cuando algunos mecanismos han sido dilucidados, existen otros en los cuales el volumen de información es escaso o que las mediciones experimentales son objeto de controversia (Delgado et al., 2012).

Los métodos alternativos a la experimentación animal son particularmente importantes en el campo toxicológico donde han ganado una marcada aceptación. Dentro de las múltiples ventajas que ofrecen destacan el ahorro de tiempo y recursos. Además, la capacidad predictiva de los modelos *in vitro* e *in silico* se ha ido perfeccionando durante décadas, llegando a sustituir hoy día a la experimentación animal en la industria cosmética europea y en gran parte de los estudios de toxicidad ocular y cutánea de sustancias (de Ávila et al., 2019; Tourneix et al., 2020).

2.2.1. Conceptualización

En el caso de los métodos computacionales, destacan los estudios basados en la relación existente entre la estructura química de los compuestos y la actividad/propiedad objeto de interés. Difundidos como QSAR/QSPR (de sus siglas en inglés: *Quantitative Structure-Activity Relationships/Quantitative Structure-Property Relationships*), esta metodología de modelación quimioinformática fue introducida hace más de medio siglo en la práctica farmacéutica y toxicológica, con utilidad además en la agroquímica y la química medioambiental. Su aplicabilidad se refleja en los múltiples trabajos científicos publicados que, como se representa en la Figura 3, superan los ochocientos anuales en la última década.
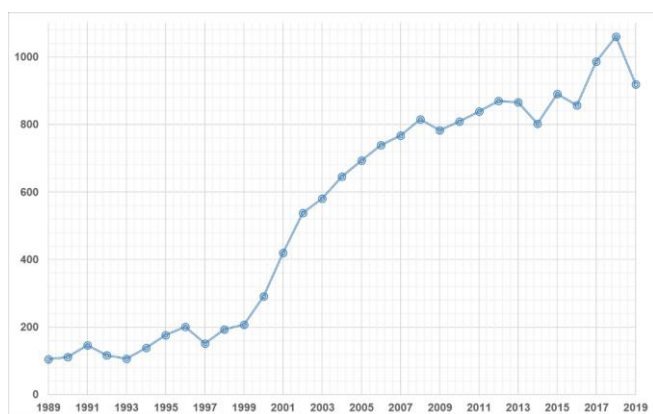
Figura 3: Línea temporal que representa el conteo de publicaciones científicas asociadas con el término MeSH: quantitative structure-activity relationship en la base de datos PubMed

Los modelos QSAR asumen que las características y propiedades de las moléculas, tanto estéricas, geométricas como electrónicas, determinan finalmente su comportamiento en diversos ambientes biológicos, así como las propiedades fisicoquímicas atribuibles a las mismas. Es de esta forma que los tres elementos fundamentales de un estudio QSAR/QSPR son: el concepto de estructura molecular, la definición de descriptores moleculares y la determinación estadística de la relación existente entre la una y la otra.

El concepto de estructura molecular fue un importante aporte de la ciencia de finales del siglo XIX, mientras que la Teoría de Crum-Brown sería el primer sistema de representación gráfica de compuestos químicos, preludio de los que hoy se conocen como descriptores moleculares. Crum-Brown and Fraser introducirían en esa época la llamada "primera formulación de las relaciones cuantitativas estructura-actividad" en la que propusieron que la actividad biológica (Φ) es una función que depende de la constitución química (C), de modo que $\Phi = f(C)$. Por tanto, variaciones en una (ΔC) se reflejan en la otra ($\Delta \Phi$) (Brown and Fraser, 1868).

Transformar la estructura molecular en parámetros que matemáticamente pudieran ser relacionados con las mediciones experimentales de efectos o con las propiedades fisicoquímicas de los compuestos se convirtió por tanto en elemento crucial. Esta parametrización cuantitativa o descripción de las moléculas no es más que "el resultado final de un procedimiento matemático y lógico que trasforma la información química codificada en la representación simbólica, en un número útil o el resultado de algún experimento estandarizado" (Todeschini and Consonni, 2009).

Derivados del análisis de un gran número de teorías, se han definido miles de descriptores moleculares. El contenido de información en sus definiciones ha avanzado también junto al surgimiento de nuevas herramientas estadísticas para el análisis de datos. En general es posible identificar dos clases de descriptores, los basados en medidas experimentales (ej. polarizabilidad, log P, refractividad molar), y los derivados de la representación simbólica de las moléculas, también conocidos como descriptores teóricos. A su vez, los descriptores teóricos se clasifican en descriptores de 0-5D dependiendo de la información estructural que contienen. En los descriptores 0D no se considera la estructura molecular, únicamente la fórmula química de los compuestos (ej. peso atómico, número de átomos). Los descriptores 1D son listas de fragmentos estructurales de las moléculas, pero sin considerar la estructura molecular en su conjunto (ej. conteo de grupos funcionales o de sustituyentes de interés). Los descriptores 2D (o topológicos), ponderan los átomos con propiedades como la electronegatividad o el volumen y consideran la conectividad, así como las interacciones intra/intermoleculares. Los descriptores 3D (o geométricos) permiten la consideración de elementos estereoquímicas de los átomos, mientras que los 4D incluyen además información de campos de interacción como es el caso de las relaciones con los sitios activos de los receptores. Por último, la descripción 5D estima también elementos estereodinámicos en la parametrización de la estructura química de las moléculas (Sahoo et al., 2016).

2.2.2. Sistemática en la modelación QSAR

El desarrollo de modelos QSAR se encuentra en crecimiento y actualización continuos (Cherkasov et al., 2014). Sin embargo, siete pasos comunes rigen su sistemática (Figura 4).

El primer paso en un estudio QSAR, como se muestra en la secuencia de la Figura 4, es la formulación del problema. Es importante desde el comienzo del estudio determinar el objeto de análisis y el nivel de información requerido. Además, se agrupará el mayor volumen posible de información relevante para la construcción de bases de datos amplias o específicas de acuerdo con el interés del trabajo. Seguidamente, se procederá a la parametrización cuantitativa de las estructuras químicas contenidas en la base de datos, empleando para ello la clase de descriptores moleculares que se estime oportuna. La selección de variables permite escoger aquéllas estadísticamente relevantes a través de técnicas como la de pasos sucesivos (SW). Es posible también emplear técnicas de agrupamiento como el análisis de componentes principales (PCA) para generar nuevos descriptores que contengan mayor volumen de información.

La medición de la propiedad de interés suele requerir estrategias de "pretratamiento" de los datos, necesario para garantizar la homogeneidad de la variable respuesta a modelar, dado fundamentalmente el hecho de que a menudo el origen de dichos datos es muy diverso (diferentes fuentes de información, técnicas de medición experimental, etc.). La selección del tipo de modelo QSAR es un paso crucial, y dependerá tanto de la variable respuesta como del conjunto de datos en su totalidad. Según la propiedad o actividad analizada, se identifican globalmente dos tipos de modelos, los de regresión, cuando la variable respuesta es cuantitativa, y los de clasificación, cuando se modelan categorías o clases (Dudek et al., 2006). En ambos tipos de QSARs, la dependencia estadística con los datos estructurales puede o no ser lineal.

Los métodos estadísticos lineales más comunes para relacionar la estructura con la actividad de los compuestos químicos son el análisis discriminante lineal (LDA) para los datos categóricos y la regresión lineal múltiple (MLR) para los datos cuantitativos. También es muy

Sistemática de un estudio QSAR

Pasos Generales

Figura 4: Representación de los siete pasos generales de la modelación QSAR.

empleado el método de mínimos cuadrados parciales (PLS) basado en la factorización de las variables independientes (descriptores). En los modelos PLS se proyectan los descriptores y la variable respuesta en un nuevo espacio y, aunque su aplicación más común es en el análisis de regresión, se ha combinado con el análisis de discriminante (PLS-DA) para modelar categorías binarias.

Las funciones genéticas de aproximación (GFA) son métodos también muy empleados en la construcción de modelos QSAR/QSPR. Inspirados en teorías evolutivas y adaptativas, los GFAs son capaces de generar múltiples modelos de forma eficiente analizando grandes volúmenes de datos. Se han combinado también con los modelos PLS (G/PLS) aprovechando las ventajas de uno y otro (Rogers, 1999). De igual forma, otras funciones que permiten la transformación de los datos como es la basada en el método de Kernel se han acoplado con PLS (KPLS) en el desarrollo de modelos QSAR. En el caso de la clasificación, la técnica de reconocimiento de patrones k-vecinos más cercanos (k-NN) ha demostrado una marcada utilidad en la modelación quimioinformática.

Los métodos predictivos no lineales más comunes y que se emplean para modelar todo tipo de variables en el análisis QSAR, son los algoritmos de aprendizaje automático (*"machine learning"*). Dentro de ellos son muy conocidos los árboles de decisión (DT), que en función de la variable de destino serán árboles de clasificación o de regresión. Además, por su elevada capacidad predictiva algoritmos como las máquinas de soporte vectorial (SVM) y las redes neuronales artificiales (ANN) son frecuentemente utilizados tanto para problemas de regresión y como de clasificación.

Vale la pena mencionar además dos tipos de metodologías que, por su amplia aplicabilidad en el descubrimiento de fármacos, se encuentran entre las más renombradas en el campo de modelación QSAR. Éstas son el análisis comparativo de campo molecular (CoMFA) y el análisis comparativo de índices de similitud molecular (CoMSIA). Los métodos CoMFA y CoMSIA son análisis 3D-QSAR con la particularidad del tipo de descriptores moleculares que se emplean para definir las estructuras químicas. CoMFA emplea los campos potenciales de Lennard-Jones y Coulomb para calcular las energías

estéricas y electrostáticas de interacción (Cramer et al., 1988), mientras que CoMSIA calcula las energías de interacción entre los átomos introduciendo una función gaussiana. Por lo general, ambos métodos producen un elevado número de parámetros numéricos que caracterizan las moléculas y que suelen ser agrupados y luego modelados junto a la variable respuesta empleando PLS.

Durante la modelación, es común la división del conjunto de datos en series más pequeñas que son empleadas para el entrenamiento y para la evaluación de la capacidad predictiva de los modelos QSAR. El diseño estadístico de estos subconjuntos de entrenamiento y de predicción puede ser aleatorio o empleando técnicas multivariantes de reconocimiento de patrones para garantizar la distribución racional del conjunto de datos. En cualquier caso, resulta imprescindible aplicar procedimientos de validación para garantizar que los modelos son capaces de generar predicciones confiables, como se detalla a continuación. Por último, la modelación QSAR suele concluirse con la interpretación mecanicista y que generalmente está basada en el análisis de los descriptores moleculares empleados. La mayor aplicabilidad de un modelo predictivo basado en QSAR será en el cribado virtual de compuestos químicos que no han sido estudiados anteriormente.

2.2.3. Principios de validación definidos por la OECD

El proceso de validación de un modelo QSAR demanda la obtención de pruebas estadísticas que demuestren que el método desarrollado es lo suficientemente fiable y reproducible en la generación de los resultados previstos. El espacio químico en el cual un modelo QSAR es capaz de generar predicciones fiables se conoce como dominio de aplicación (Cherkasov et al., 2014; Tropsha et al., 2003). Con el objetivo de estandarizar el desarrollo de los modelos QSAR, la OECD agrupa el proceso de validación en una serie de principios regulatorios (OECD, 2014), representados en la Figura 5 a continuación.

El primer principio de la regulación que establece la OECD, se refiere a definir adecuadamente las variables a modelar (propiedades fisicoquímicas, efectos biológicos o medioambientales). De esta forma, se pretende asegurar la transparencia en la determinación de los parámetros a modelar, ya que pueden haberse obtenido a través de

diferentes protocolos experimentales, o bajo distintas condiciones. La falta de homogeneidad en los datos puede acarrear consecuencias en el desarrollo de los modelos y errores en las predicciones, de ahí la importancia que tiene definir el punto de medición. En el segundo principio se establece que los algoritmos empleados no sean ambiguos, tanto aquellos que definen los descriptores moleculares como los que se emplean en la construcción de las series y los propios modelos estadísticos de QSAR. El establecimiento del dominio de aplicabilidad es un elemento fundamental, que como tercer principio supone la declaración de un espacio de confiabilidad en el cual los modelos desarrollados son capaces de generar predicciones acertadas con una alta probabilidad. La validación propiamente dicha es el cuarto principio, referido a que todo modelo QSAR debe presentar una adecuada bondad de ajuste, robustez y predictividad, sustentado por criterios estadísticos que demuestren el desempeño del modelo (validación interna), y su capacidad para predecir (validación externa). Por último, aun cuando es recomendable, el quinto principio está sujeto a las particularidades de cada tipo de modelo QSAR. En este paso final de control y retroalimentación se sugiere, cuando sea posible, interpretación mecanicista o teórica, que suele implicar el análisis de los descriptores más significativos empleados en la modelación (Tropsha, 2010).

Es así como el modelado QSAR, como importante método alternativo a la experimentación animal es aprobado, bajo determinadas directrices, por entidades científicas y regulatorias. Su aplicación ha sido propuesta para la identificación de múltiples actividades toxicológicas que incluye a la disrupción endocrina. A continuación, se expone una revisión de los principales modelos QSAR disponibles para tres importantes mecanismos de disrupción endocrina.

La disrupción endocrina a través de métodos QSAR: selección de casos de estudio

El desarrollo de modelos QSAR depende de la información disponible, ya que sus inferencias se basan en la extrapolación de los datos existentes para predecir la actividad de compuestos no testeados. Esta es la razón por la cual el mayor número de estudios QSAR en el contexto de la disrupción endocrina se basa en el mecanismo más abordado, la acción de compuestos químicos como moduladores de los receptores estrogénicos.

En esta revisión se han querido compendiar, debido a su importancia, métodos QSAR recientes de otros mecanismos de interrupción del sistema endocrino. Como se encuentra esquematizado en la Figura 6, los tres ejemplos seleccionados representan las principales rutas
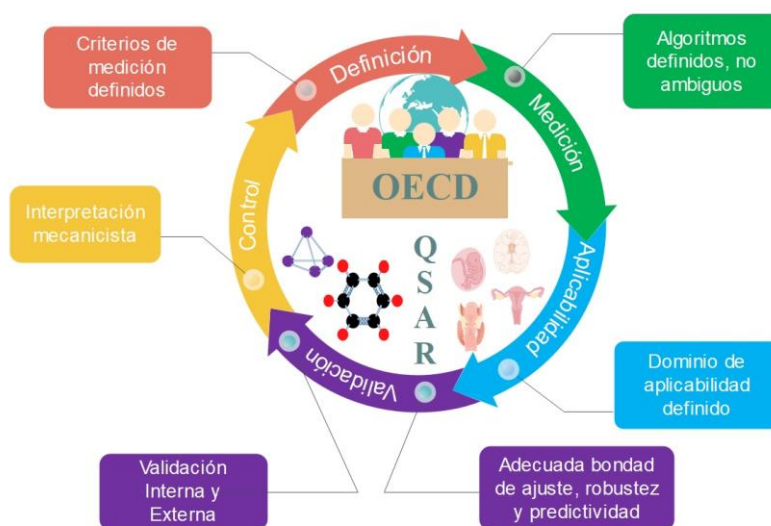
Figura 5: Principios establecidos por la OECD para la validación de los modelos QSAR.
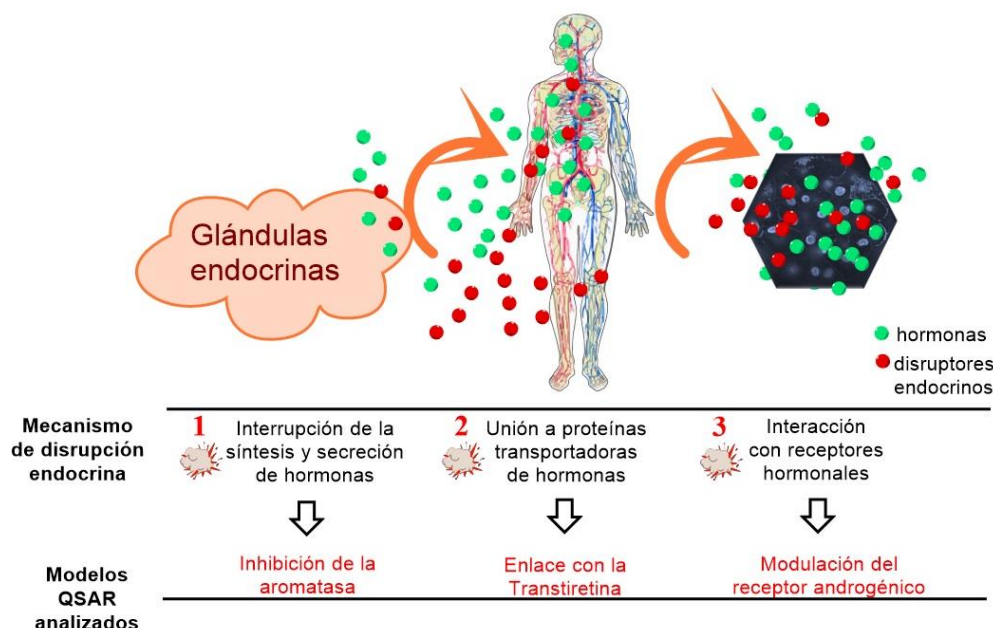
Figura 6: Mecanismos de disrupción endocrina seleccionados para el análisis de los modelos QSAR reportados en la literatura.

potenciales de toxicidad endocrina y en el contexto de hormonas estrogénicas, tiroideas y androgénicas, indistintamente.

En el primer mecanismo objeto de estudio, relacionado con la interrupción de la biosíntesis de estrógenos, se analizan los modelos QSAR desarrollados para predecir la inhibición de la enzima citocromo P450 aromatasa. Compitiendo con las hormonas tiroideas, la unión de compuestos químicos a la proteína transportadora transtiretina, o la interrupción de este mecanismo de transporte hormonal es la segunda ruta potencial de disrupción endocrina abordada. Por último, se analizan los más recientes métodos QSAR enfocados en la modulación, y especialmente el antagonismo sobre el receptor de andrógenos.

3.1 Estudios QSAR basados en los efectos sobre la enzima aromatasa

La interrupción de las rutas biosintéticas de hormonas es un mecanismo potencial de disrupción endocrina. Sin embargo, la información químico-biológica disponible hoy día procede en su mayoría de estudios farmacológicos (Mori et al., 2018). La principal aplicación terapéutica de la inhibición de la enzima P450 aromatasa ha sido en enfermedades neoplásicas asociadas a la producción de estrógenos como es el caso del cáncer de mama (Adhikari et al., 2017b).

Diversos estudios QSAR han pretendido dilucidar la relación entre los rasgos químico-estructurales y los efectos inhibidores de los compuestos sobre la CYP19A1. Los más significativos publicados en la última década se encuentran resumidos en la Tabla 1 a continuación.

De la modelación QSAR reportada entre 2010-2020 en la literatura basada en la inhibición de la enzima aromatasa y representada en los 37 ejemplos de la Tabla 1, destacan los siguientes elementos:

1. Los modelos lineales, particularmente la regresión lineal múltiple o el método de mínimos cuadrados parciales, son las metodologías estadísticas preferidas en el estudio QSAR de la inhibición de la enzima aromatasa, probablemente por su simplicidad e interpretabilidad.

2. Se ha apostado por la alta capacidad predictiva de los métodos de aprendizaje automático como redes neuronales o máquinas de soporte vectorial tanto en problemas de regresión como en problemas de clasificación.

3. Aun cuando la mayoría de los modelos han llevado a cabo la validación interna, muchos no fueron validados con el uso de series externas o con metodologías de 5 o 10-*folds* como es recomendado para garantizar la confiabilidad y capacidad predictiva.

4. En la mayoría de los modelos disponibles, los descriptores moleculares empleados son conocidos y fácilmente interpretables.

5. Metodologías innovadoras en la parametrización de las estructuras han sido propuestas como el análisis multivariante de imágenes empleando la transformada discreta bidimensional de Fourier o descriptores derivados de hipótesis toxicofóricas.

6. Aun cuando la complejidad y el volumen de los datos ha ido en aumento, el uso de bases de datos de congéneres con limitado dominio de aplicabilidad sigue siendo habitual. Especies químicas como flavonoides, cumarinas, benzimidazoles, bis-sulfonamidas, naftoquinonas y arilsulfonamidas han sido modeladas con sus respectivos derivados estructurales cercanos en series que no sobrepasan los 60 compuestos químicos.

7. Hasta el momento, la base de datos más diversa reportada en la inhibición de la enzima aromatasa cuenta con 973 compuestos, e incluye esteroides y estructuras no esteroidales. Modelos de regresión y de clasificación han sido diseñados a partir de esta base de datos y estrictamente validados para la predicción de nuevas entidades con actividad desconocida.

Una de las principales aplicaciones de los modelos computacionales, y en particular de los estudios QSAR, es el cribado virtual de grandes bases de datos de compuestos con la finalidad de encontrar los llamados "compuestos líderes" de la actividad modelada. Esta utilidad, particularmente ventajosa en el contexto farmacológico, ha sido aprovechada por varios estudios de predicción de los efectos de compuestos químicos sobre la enzima aromatasa.

Dawood *et al*. cribaron 1 720 constituyentes fitoquímicos presentes en 29 plantas medicinales y otros productos naturales usados en la medicina tradicional egipcia. La disrupción endocrina y/o el potencial terapéutico de estos compuestos naturales mediada por la inhibición de la enzima aromatasa, fue predicha de esta forma empleando modelos QSAR (Dawood et al., 2018). En el trabajo de Xie *et al*., el cribado virtual de la "NCI2000 *database*" lo desarrollaron con los modelos farmacofóricos previamente y la actividad inhibitoria de los seis compuestos identificados fue predicha a través de los modelos 3D-QSAR diseñados (Xie et al., 2014). Barigye *et al*. cribaron 1 593 compuestos y realizaron la validación experimental *in vitro* de 10 sustancias predichas como inhibidores activos de la aromatasa según el consenso mayoritario de los modelos que propusieron. Esta última constituye la más rigurosa validación del poder predictivo de los modelos QSAR, conocida como validación experimental prospectiva (Barigye et al., 2018).

Finalmente, vale la pena señalar, que no solo la inhibición de la enzima aromatasa ha sido considerada variable respuesta en el estudio de la disrupción endocrina por este mecanismo. La habilidad de actuar como ligandos de la enzima, y que no implica información acerca de sus efectos agonistas/antagonistas ha sido también recomendada en el análisis toxicológico. Como ejemplo representativo, Du H *et al*. modelaron la capacidad de enlace a la aromatasa de 1 552 compuestos diversos obteniendo una elevada robustez y predictividad en los modelos QSAR resultantes (Du et al., 2017). Estas estimaciones son importantes en las fases preliminares de priorización de compuestos, así como en el análisis de los riesgos potenciales. Sin embargo, modelos que logren predecir el efecto consecuente de la interacción con la enzima aportan una información mucho más relevante desde el punto de vista biológico.

3.2 Estudios QSAR basados en los efectos sobre la transtiretina

Varios modelos QSAR han pretendido dilucidar la relación entre xenobióticos, fundamentalmente contaminantes ambientales, y la proteína transtiretina, encargada del transporte de la hormona tiroidea. Los modelos de regresión en su mayoría están basados en porcentajes de inhibición de la unión transtiretina-tiroxina, o valores de IC_{50} para los compuestos objeto de análisis. Mientras tanto los estudios categóricos suelen basarse en resultados de ensayos de potencia competitiva con la hormona, asignando dos o más clases a estos efectos (ej. activo/inactivo o bajo/moderado/alto) (Kar et al., 2017). En la Tabla 2 se encuentran resumidos los principales modelos QSAR de disrupción del mecanismo de transporte de la hormona tiroxina.

Algunos elementos destacables en la modelación QSAR de la disrupción de la TTR son:

1. La mayor parte de los compuestos cuya evaluación experimental se encuentra disponible para la modelación son sustancias que, por otros mecanismos, han sido sugeridas como potenciales interruptores endocrinos, ej. poli y perfluorocarbonos (PFCs) y los polibromodifenil éteres (PBDEs).

2. Descriptores interpretables han sido empleados en todos los trabajos con la finalidad de dilucidar su relación con la disrupción de la transtiretina.

3. El escaso volumen de información disponible para este mecanismo conlleva a que la regresión lineal, especialmente el método de PLS y la clasificación basada en k-vecinos más cercanos (técnicas además muy interpretables), sean las metodologías estadísticas más comúnmente empleadas.

Tabla 1. *Modelos QSAR de compuestos inhibidores de la aromatasa.*

n.º	Base de Datos [a]	Descriptores [b]	Método QSAR	Validación [c]	Ref.
1	116 compuestos diversos	Algunos de los siguientes: Jurs_FNSA_3+0.063, NCOSV, Hbondacceptor-2, 4.134-AlogP, Jurs_PNSA_3+34.086, AlogP-4.273, Jurs_FNSA_1-0.414, Chiralcenters.	3 modelos GFA (regresión)	R^2=0.68-0.71, Q^2_{int}=0.64-0.69, Q^2_{ext}=0.62-0.67	(Roy and Roy, 2010)
			G/PLS (regresión)	R^2=0.69, Q^2_{int}=0.63, Q^2_{ext}=0.63	
2	57 flavonoides	Parámetros energéticos y electrónicos del enlace ligando-receptor	2 modelos SW-MLR (regresión)	R^2=0.82-0.87, Q^2_{int}=0.63-0.79, Q^2_{ext}=0.81-0.83	(Narayana et al., 2012)
3	280 esteroides	MW, RBN, nCIC, Q_m, nHDon, nHAcc, TPSA, HOMO, LUMO, LogP, Energy, HOMO-LUMO gap, dipole moment	DT (clasificación)	Acc=71.67 % Sn=0.76 Sp=0.67	(Nantasenamat et al., 2013a)
4	690 no-esteroides			Acc=76.79 % Sn=0.58 Sp=0.82	
5	47 triazoles (análogos del letrozol)	-Número de anillos -ALogP -HOMO-LUMO	MLR (regresión)	R^2=0.81, Q^2_{int}=0.76, Q^2_{ext}=0.65	(Nantasenamat et al., 2013b)
6			ANN (regresión)	R^2=0.82, Q^2_{int}=0.76, Q^2_{ext}=0.65	
7			SVM (regresión)	R^2=0.88, Q^2_{int}=0.83, Q^2_{ext}=0.66	
8	12 benzimidazoles	E_ADJ_equ, SC_1, CIC, Jurs_TASA, Shadow_XZ	GFA-PLS (regresión)	R^2=0.99, Q^2_{int}=0.96	(Dai et al., 2014)
9	44 flavonoides	Algunos o todos los descriptores siguientes: -RBN -nCIC -nHDon -nHAcc -ALogP -Energy -Dipole moment	MLR (regresión)	n=31 R^2=0.99, Q^2_{int}=0.99, Q^2_{ext}=0.99	(Nantasenamat et al., 2014)
10			ANN (regresión)	n=30 R^2=0.99, Q^2_{int}=0.98, Q^2_{ext}=0.99	
11			SVM (regresión)	n=33 R^2=0.99, Q^2_{int}=0.98, Q^2_{ext}=0.98	
12			DT (clasificación)	n=26 Acc=84.62% Sn=0.85	
13	973 compuestos diversos	Descriptores basados en SMILES	Método Monte Carlo (regresión)	R^2=0.63–0.71 Q^2_{int}=0.62–0.70 Q^2_{ext}=0.61-0.66	(Worachartcheewan et al., 2014a)
14	34 cumarinas disustituidas	F10[N-O], MAXDP BELe1, Inflammat-50, B10[C-O], H-047, Psychotic-80.	MLR (regresión)	R^2=0.96 Q^2_{int}=0.92 Q^2_{ext}=0.73	(Worachartcheewan et al., 2014b)
15	66 esteroides	descriptores CoMFA	PLS c=10 (regresión)	R^2=0.99 Q^2_{int}=0.66 Q^2_{ext}=0.64	(Xie et al., 2014)
16		descriptores CoMSIA	PLS c=17 (regresión)	R^2=0.99 Q^2_{int}=0.60 Q^2_{ext}=0.84	
17	45 flavonoides	Descriptores CoMFA	PLS c=6 (regresión)	R^2=0.92, Q^2_{int}=0.83, Q^2_{ext}=0.71	(Awasthi et al., 2015)
18	180 esteroides	C-025, ESpm13r, ESpm14u, ESpm10r, ESpm12x, ESpm15d, Spm10x, H-050, nBM, MATS6p, MATS6e, GATS6m, GATS6p, piPC07, GGI1	Efficient linear method (ELM) (clasificación)	Acc=80.83 ± 0.71% Sn=0.88±0.02 Sp=0.75±0.03 MCC=0.63±0.01	(Shoombuatong et al., 2015)
19	474 diversos (no esteroides)			Acc=80.76± %0.33 Sn=0.83±0.01 Sp=0.73±0.01 MCC=0.54±0.01	
20	84 diversos (no esteroides)	CoMFA (modelo farmacofórico)	PLS c=7 (regresión)	R^2=0.99 Q^2_{int}=0.74 Q^2_{ext}=0.63	(Xie et al., 2015)
21		CoMSIA (modelo farmacofórico)	PLS c=6 (regresión)	R^2=0.93 Q^2_{int}=0.71 Q^2_{ext}=0.67	
22	76 compuestos: -diarilalquilimidazoles -diarilalquiltriazoles	nTB, BEHv2, ATS3m, ATS7m, nCp, G(N…N), Hyd.E, G(O…O), SRW07	MLR (regresión)	R^2=0.83 R^2_{pred}=0.70 Q^2_{ext}=0.91	(Ghodsi and Hemmateenejad, 2016)
23	67 no esteroides (análogos del letrozol)	SaaN, JGI7, MATS8c, CLogP, CMR	SW-MLR (regresión)	R^2=0.90 Q^2_{int}=0.86 Q^2_{ext}=0.85	(Adhikari et al., 2017a)
24		*"fingerprints"* moleculares (hologramas)	PLS c=6 (regresión)	R^2=0.87 Q^2_{int}=0.76 Q^2_{ext}=0.61	
25		Fragmentos Subestructurales	Model bayesiano (clasificación)	Acc=88.0% Sen=0.89 Spe=0.88	
26		descriptores CoMFA	PLS c=7 (regresión)	R^2=0.94 Q^2_{int}=0.54 Q^2_{ext}=0.62	
27	11 compuestos (1,4-naftoquinonas)	Mor04m, H8m, Mor08e, G1v, SIC2	MLR	R^2=1.00 Q^2_{int}=0.98	(Prachayasittikul et al., 2017)

28	973 compuestos diversos	Descriptores basados en el análisis multivariado de imágenes (MIA) empleando la transformada discreta bidimensional de Fourier (2D-DFT)	ANN (clasificación)	Acc=88.74% Sn=0.86 Sp=0.91	(Barigye et al., 2018)
29			LDA (clasificación)	Acc=84.23% Sn=0.72 Sp=0.95	
30			SVM (clasificación)	Acc=85.14% Sn=0.76 Sp=0.93	
31			BT (clasificación)	Acc=84.68% Sn=0.81 Sp=0.88	
32			kNN (clasificación)	Acc=85.59% Sn=0.82 Sp=0.88	
33	16 esteroides	Descriptores derivados de hipótesis farmacofóricas (PHASE)	PLS c=3 (regresión)	R^2=1.00 Q^2_{int}=0.49 Q^2_{ext}=0.21	(Dawood et al., 2018)
34	25 no esteroides		PLS c=5 (regresión)	R^2=1.00 Q^2_{int}=0.34 Q^2_{ext}=0.09	
35	17 bis-sulfonamidas	GATS6v, Mor03m	MLR (regresión)	R^2=0.80 Q^2_{int}=0.68	(Leechaisit et al., 2019)
36	30 arilsulfonamidas	aLogP, HBA, HBD, RB, HAC, RC, PSA, E-state, MR, Polar	MLR (regresión)	R^2=0.83 Q^2_{int}=0.76	(Fantacuzzi et al., 2020)

[a] Número de compuestos usados en la modelación (incluyendo series de entrenamiento y de validación)
[b] Descriptores que se incluyeron en los modelos finales. Para más detalles, se recomienda consultar las referencias de cada modelo o el *Handbook* (Todeschini and Consonni, 2000)
[c] En modelos de regresión: el coeficiente de correlación (R^2), el coeficiente de validación interna (generalmente resultantes de la validación cruzada) (Q^2_{int}) y el de validación externa (Q^2_{ext}). En modelos de clasificación, solamente se muestran los resultados de la validación con la serie de predicción a través de la exactitud (Acc), sensibilidad (Sn), especificidad (Sp) y el coeficiente de correlación de Matthews (MCC).

Tabla 2. *Modelos QSAR de compuestos moduladores de la transtiretina.*

n.º	Base de Datos [a]	Descriptores [b]	Método QSAR	Validación [c]	Ref.
1	24 poli y perfluorocarbonos (PFCs)	Diversos: -propiedades moleculares -índices topológicos y de carga -tiempos de retención (HPLC)	PLS c=2	R^2=0.61 Q^2_{int}=0.41	(Weiss et al., 2009)
2	17 Polibromodifenil éteres (PBDEs)	Qpmax, MATS6v	MLR	R^2=0.96 Q^2_{int}=0.93 Q^2_{ext}=0.90	(Papa et al., 2010)
3	29 PBDEs	DISPe, nArOH	k-NN k=3	n=20 Acc_{clase1} = 87.5% Acc_{clase2} = 100% Acc_{clase3} = 100% Sn=1.00 Sp=0.87	(Kovarich et al., 2011)
			k-NN k=3	n=29 Acc_{clase1} = 83.3% Acc_{clase2} = 88.9% Acc_{clase3} = 100% Sn=0.94 Sp=0.83	
4	28 PBDEs	Descriptores CoMSIA	PLS c=4	R^2=0.98, Q^2_{int}=0.75, Q^2_{ext}=0.93	(Yang et al., 2011)
5	19 PFCs	AMW, HATS6m	k-NN k=1	Acc=100% Sn= Sp=1.00	(Kovarich et al., 2012)
		nH, HATS6m		Acc=100% Sn= Sp=1.00	
		nH, F06[C-O]		Acc=90% Sn=1.00, Sp=0.75	
		T(F..F), HATS6m		Acc=100% Sn= Sp=1.00	
6	53 halogenados (PFCs & PBDEs)	nArOH, (F03[Br-Br]), HATS6m	k-NN k=5	n=53 Acc=89% Sn=1.00, Sp=0.86	(Papa et al., 2013)
7		R5u, F07[C-O], nArOH	MLR	n=32 R^2=0.89, Q^2_{int}=0.81	
8	47 compuestos -36 fenólicos -11 no ionizables	Diversos: Ej: log D, MW, nX, E_{HOMO}, E_{LUMO}, qH^+, qC^-, qO^-	PLS c=2	R^2=0.86, Q^2_{int}=0.84, Q^2_{ext}=0.93	(Yang et al., 2013)
9	178 compuestos diversos	14 descriptores 2D -Propiedades fisicoquímicas ej. hidrofobicidad, carga parcial. -conteo de fragmentos: ej. hidroxilos, átomos de halógeno	PLS c=5	Acc=79% Sn=0.94 Sp=0.69	(Zhang et al., 2015)
10			kNN k=4	Acc=82% Sn=0.78 Sp=0.85	
11			SVM	Acc=77% Sn=0.90 Sp=0.73	
12	15 PFCs	IC3, $\sum \beta'_S$	GFA-MLR	R^2=0.86, Q^2_{int}=0.64, Q^2_{ext}=0.73	(Kar et al., 2017)

13	24 PFCs	Me, nCsp2, H-050	LDA	Acc=100% Sn=1.00 Sp=1.00 MCC=0.87	(Kar et al., 2017)
14	29 compuestos -24 PFCs -6 ácidos grasos	HATS6m, qO⁻, δ_A^-	PLS	R^2=0.93, Q^2_{int}=0.67, Q^2=0.65	(Yang et al., 2017)

[a] Número de compuestos usados en la modelación (incluyendo series de entrenamiento y de validación)

[b] Descriptores que se incluyeron en los modelos finales. Para más detalles, se recomienda consultar las referencias de cada modelo o el *Handbook* (Todeschini and Consonni, 2000)

[c] En modelos de regresión: el coeficiente de correlación (R^2), el coeficiente de validación interna (generalmente resultantes de la validación cruzada) (Q^2_{int}) y el de validación externa (Q^2_{ext}). En modelos de clasificación, solamente se muestran los resultados de la validación con la serie de predicción a través de la exactitud (Acc), sensibilidad (Sn) y especificidad (Sp).

El estudio de clasificación de Zhang *et al.* es posiblemente el más completo hasta el momento para este mecanismo debido al número de estructuras y los diversos métodos aplicados. Los autores desarrollaron además un cribado virtual y seleccionaron un grupo de 23 compuestos para llevar a cabo una validación experimental *in vitro* de los efectos que denominaron "alteradores de la hormona tiroidea" (mediados por la TTR). En el cribado se analizaron 485 contaminantes orgánicos presentes en el polvo que fueron extraídos de la literatura y 433 potenciales metabolitos -determinados *in silico*- de los mismos. Cuatro nuevos compuestos químicos interruptores fueron identificados combinando las metodologías *in silico*/*in vitro*. En este trabajo se sugirió además la importancia de la activación metabólica para el enlace a la transtiretina (Zhang et al., 2015).

Por otro lado, los estudios de Yang *et al.* han dilucidado la importancia de las formas químicas y la ionización de éstas en las interacciones con TTR. El primer trabajo estuvo enfocado en compuestos fenólicos mientras que el siguiente abordó diversos PCFs. En ambos estudios se concluyó que las diferencias entre formas neutras y aniónicas de los compuestos influyen en la capacidad de enlace con la proteína transportadora TTR (Yang et al., 2017, 2013).

3.3 Estudios QSAR basados en los efectos sobre el receptor de andrógenos

Los andrógenos, entre otras muchas funciones, participan en el desarrollo y el mantenimiento de la próstata. La expresión del receptor androgénico es por tanto una importante diana terapéutica en patologías y malignidades prostáticas, en las cuales se han enfocado diversos estudios de descubrimiento de fármacos. Además, dada su notable importancia fisiológica, la inhibición del AR deriva en reacciones adversas que conllevan a su inclusión como un potencial mecanismo de disrupción endocrina. Sin embargo, la información actual disponible ha derivado en su mayoría de estudios farmacológicos enfocados en el tratamiento del cáncer de próstata a través del efecto anti androgénico de compuestos químicos (Heinlein and Chang, 2004). Los más recientes estudios QSAR enfocados en la inhibición del receptor de estrógenos se encuentran resumidos en la Tabla 3.

La predicción del antagonismo sobre el receptor androgénico ha sido estimada empleando métodos computacionales para grandes bases de datos de compuestos cuya información biológica deriva de ensayos *in vitro* principalmente (Vinggaard et al., 2008). En la última década el estudio de Li *et al.* destaca debido al amplio espacio químico modelado que lo sitúa como una importante herramienta predictiva del antagonismo sobre el receptor androgénico tanto en el panorama farmacológico como toxicológico. Los autores de este trabajo fusionaron varias estrategias de cribado virtual y demostraron la efectividad de esta combinación en comparación con metodologías clásicas (H. Li et al., 2013).

Otros trabajos recientes han aportado modelos de predicción que, aun con un menor número de compuestos, permiten la estimación del efecto inhibitorio en % relativos o en IC$_{50}$. Son especialmente importantes desde la perspectiva de la disrupción endocrina, la predicción de derivados de contaminantes (ej. PCBs, PBDEs) o de compuestos presentes en altas dosis en la dieta humana como los flavonoides.

Una combinación de las diferentes interacciones ligando-receptor a través de metodologías de dinámica molecular fue propuesta por Wu Y *et al.* en la clasificación de compuestos activos/inactivos sobre el receptor androgénico. La combinación de estas simulaciones con estudios QSAR resultó en una significativa mejora de la capacidad predictiva de los modelos respecto a metodologías QSAR convencionales (Wu et al., 2016). Sin embargo, este estudio fue desarrollado exclusivamente con flavonoides, por tanto, se necesitan pruebas adicionales para asumir generalizaciones o corroborar la aplicabilidad que tiene esta propuesta en un espacio químico más amplio.

A propósito de esta necesidad, en un trabajo de Trisciuzzi *et al.* se sugiere un innovador y práctico protocolo para el desarrollo de modelos *in silico* altamente predictivos como clasificadores de la actividad androgénica. Estos autores emplearon el acoplamiento molecular (conocido como *docking*), que consiste en la predicción de la energía y de los modos de enlace entre ligandos y proteínas . La exigente validación externa de los modelos y la aplicación de éstos en el cribado de más de 55 mil compuestos químicos relevantes (debido al grado de exposición humana a los mismos), corroboró su capacidad predictiva. Además, la integración de diferentes métodos computacionales se definió como una importante pauta para la interpretación mecanicista de los resultados (Trisciuzzi et al., 2017).

Por otro lado, desde hace varias décadas, diferentes métodos QSAR han sido empleados en la predicción de la capacidad de enlace al AR, la cual resulta útil fundamentalmente en análisis preliminares de riesgo (Loughney and Schwender, 1992; Waller et al., 1996).

Conclusiones

Miles de productos químicos y mezclas que hoy día se emplean en alimentos, suplementos nutricionales, fármacos, cosméticos, envases, productos agrícolas, etc. representan un riesgo potencial como interruptores endocrinos. La evaluación del riesgo de disrupción endocrina es altamente compleja, y las tentativas de aunar esfuerzos para el análisis y la legislación de estos tóxicos son aún insuficientes. Además, la información experimental disponible es limitada y en algunos casos contradictoria, incrementando las dificultades en el establecimiento de pautas apropiadas.

Considerando el número significativo de compuestos químicos con potencial efecto tóxico sobre el sistema endocrino, resulta necesario y ventajoso el uso de métodos alternativos que garanticen el análisis rápido, certero y menos costoso de sustancias. Los modelos computacionales no sólo permiten a los investigadores reducir el número de experiencias en el laboratorio, sino que además asisten y extraen la información relevante de los resultados experimentales.

El empleo de metodologías predictivas, particularmente modelos quimioinformáticos es recomendado en todas las normativas vigentes para el control toxicológico de compuestos químicos. Desde hace algunos años, diversos estudios QSAR han sido descritos en la predicción de la disrupción endocrina.

Los modelos QSAR empleados en la predicción de la actividad inhibitoria sobre la enzima aromatasa han sido revisados (Adhikari et al., 2017c; Shoombuatong et al., 2018). De igual forma, varias revisiones han considerado de forma general los mecanismos de disrupción endocrina y la importancia de las metodologías QSAR para

Tabla 3. *Modelos QSAR de compuestos antagonistas del receptor androgénico.*

n.º	Base de Datos [a]	Descriptores [b]	Método QSAR	Validación [c]	Ref.
1			DT	Acc=77.60% Sn=0.74 Sp=0.71	
2	1220 compuestos diversos	De 5 a 30 descriptores diversos INDUCTIVE MOE & DRAGON	ANN	Acc=80.90% Sn=1.00 Sp=0.79	(H. Li et al., 2013)
3			agregación de *Bootstrap* (*bagging*)	Acc=80.90% Sn=0.92 Sp=0.71	
4			kNN	Acc=70.80% Sn=0.81 Sp=0.71	
5	23 PCBs	Descriptores CoMSIA	PLS c=6	R^2=0.95, Q^2_{int}=0.67	(X. Li et al., 2013)
6		análisis de campo molecular (MFA) CH3/334, H+/467, CH3/543, H+/600, CH3/193, CH3/187, CH3/606		R^2=0.84, Q^2_{int}=0.79	
7	60 compuestos diversos	análisis de superficie del receptor (RSA) TOT/1380, TOT/1104, TOT/2615, TOT/1618, TOT/1974, TOT/2150, TOT/2344	PLS c=8	R^2=0.91, Q^2_{int}=0.86	(Suresh, 2013)
8		descriptors 2D CH3/77, CH3/156, CH3/127, CH3/141, H+/171		R^2=0.86, Q^2_{int}=0.74	
9	53 flavonoides	Descriptores CoMSIA	PLS c=3	R^2=0.95 Q^2_{int}=0.60	(Wu et al., 2016)
10	20 bisphenols	nROH, $V_{s,max}$, E_{LUMO}	MLR	R^2=0.92 Q^2_{int}=0.82 Q^2_{ext}=0.85	(Yang et al., 2016)
11	36 derivados 7α sustituidos de DHT	IC5, GATS5e, DISPp, HATS3u	GA-MLR	R^2=0.76 Q^2_{int}=0.66	(Wang et al., 2017)
11		Descriptores CoMFA	PLS c=1	R^2=0.69 Q^2_{int}=0.38	
12	21 PBDEs	Descriptores CoMSIA	PLS c=1	R^2=0.61 Q^2_{int}=0.23	(Wu et al., 2017)

[a] Número de compuestos usados en la modelación (incluyendo series de entrenamiento y de validación)

[b] Descriptores que se incluyeron en los modelos finales. Para más detalles, se recomienda consultar las referencias de cada modelo o el *Handbook* (Todeschini and Consonni, 2000)

[c] En modelos de regresión: el coeficiente de correlación (R^2), el coeficiente de validación interna (generalmente resultantes de la validación cruzada) (Q^2_{int}) y el de validación externa (Q^2_{ext}). En modelos de clasificación, solamente se muestran los resultados de la validación con la serie de predicción a través de la exactitud (Acc), sensibilidad (Sn) y especificidad (Sp).

el análisis de riesgo y la predicción de la toxicidad (Vuorinen et al., 2015). Además, la disrupción del receptor de estrógenos, a través de múltiples variables respuesta y bioensayos, se ha modelado en diversas ocasiones (Porta et al., 2016) y sus resultados, junto a otros receptores nucleares y diversas técnicas computacionales se han revisado también (Ruiz et al., 2017). Sin embargo, esta es la primera vez, que, a nuestro conocimiento, se revisan los modelos QSAR disponibles en el estudio de la disrupción de la TTR. También es el primer compendio que analiza, en forma de casos de estudio, exclusivamente la modelación QSAR de la disrupción endocrina mediada por los tres mecanismos de: inhibición de la aromatasa, modulación de la transtiretina y del receptor androgénico.

Muchos de los modelos disponibles para estos tres mecanismos se encuentran rigurosamente validados y sus predicciones pueden contribuir en gran medida al estudio de la disrupción endocrina. Especialmente importante es el uso de las herramientas QSAR en la selección de los compuestos a probar de forma experimental. Además, las predicciones combinadas de estos mecanismos y otros como los estrogénicos pueden proveer de una visión general que garantice la prohibición de sustancias altamente preocupantes.

Agradecimientos

Bibliografía

1. *Adhikari, N., Amin, S.A., Jha, T., Gayen, S.*, 2017a. Integrating regression and classification-based QSARs with molecular docking analyses to explore the structure-antiaromatase activity relationships of letrozole-based analogs. Can. J. Chem. 95, 1285–1295. https://doi.org/10.1139/cjc-2017-0419

2. *Adhikari, N., Amin, S.A., Saha, A., Jha, T.*, 2017b. Combating breast cancer with non-steroidal aromatase inhibitors (NSAIs): Understanding the chemico-biological interactions through comparative SAR/QSAR study. Eur. J. Med. Chem. 137, 365–438. https://doi.org/10.1016/j.ejmech.2017.05.041

3. *Adhikari, N., Amin, S.A., Saha, A., Jha, T.*, 2017c. Combating breast cancer with non-steroidal aromatase inhibitors (NSAIs): Understanding the chemico-biological interactions through comparative SAR/QSAR study. Eur. J. Med. Chem. 137, 365–438. https://doi.org/10.1016/j.ejmech.2017.05.041

4. *Awasthi, M., Singh, S., Pandey, V.P., Dwivedi, U.N.,* 2015. Molecular docking and 3D-QSAR-based virtual screening of flavonoids as potential aromatase inhibitors against estrogen-dependent breast cancer. J. Biomol. Struct. Dyn. 33, 804–819. https://doi.org/10.1080/07391102.2014.912152

5. *Balaguer, P., Delfosse, V., Bourguet, W.,* 2019. Mechanisms of endocrine disruption through nuclear receptors and related pathways. Curr. Opin. Endocr. Metab. Res. 7, 1–8. https://doi.org/10.1016/j.coemr.2019.04.008

6. *Barigye, S.J., Freitas, M.P., Ausina, P., Zancan, P., Sola-Penna, M., Castillo-Garit, J.A.,* 2018. Discrete Fourier Transform-Based Multivariate Image Analysis: Application to Modeling of Aromatase Inhibitory Activity. ACS Comb. Sci. 20, 75–81. https://doi.org/10.1021/acscombsci.7b00155

7. *Bern, H., Blair, P., Brasseur, S., Colborn, T., Cunha, G., Davis, W., Al, E.,* 1991. Statement from the Work Session on Chemically Induced Alterations in Sexual Development: the Wildlife/human

Connectionitle, in: The Wingspread Conference. Racine, Wisconsin.

8. *Brown, A.C., Fraser, T.R.,* 1868. On the Connection between Chemical Constitution and Physiological Action; with special reference to the Physiological Action of the Salts of the Ammonium Bases derived from Strychnia, Brucia, Thebaia, Codeia, Morphia, and Nicotia. J. Anat. Physiol. 2, 224–242.

9. *Cheek, A.O., Kow, K., Chen, J., McLachlan, J.A.*, 1999. Potential mechanisms of thyroid disruption in humans: Interaction of organochlorine compounds with thyroid receptor, transthyretin, and thyroid-binding globulin. Environ. Health Perspect. 107, 273–278. https://doi.org/10.1289/ehp.99107273

10. *Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, I.I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'Min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A.,* 2014. QSAR modeling: Where have you been? Where are you going to? J. Med. Chem. 57, 4977–5010. https://doi.org/10.1021/jm4004285

11. *Cramer, R.D., David E, P., Bunce, J.D.,* 1988. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc 110, 5959–5967. https://doi.org/10.1021/ja00226a005

12. *Dai, Y., Xiao, Q., Wang, S., Wei, X., Zhang, Z., Ma, H., Zheng, M., Hou, M., Zhang, T.,* 2014. Syntheses and QSAR Studies of Benzylimidazole Derivatives and Benzylcarbazole as Potential Aromatase Inhibitors. Asian J. Chem. 26, 70–73. https://doi.org/10.1093/jae/ejp002

13. *Dawood, H.M., Ibrahim, R.S., Shawky, E., Hammoda, H.M., Metwally, A.M.,* 2018. Integrated in silico-in vitro strategy for screening of some traditional Egyptian plants for human aromatase inhibitors. J. Ethnopharmacol. 224, 359–372. https://doi.org/10.1016/j.jep.2018.06.009

14. *de Ávila, R.I., Lindstedt, M., Valadares, M.C.,* 2019. The 21st Century movement within the area of skin sensitization assessment: From the animal context towards current human-relevant in vitro solutions. Regul. Toxicol. Pharmacol. 108, 104445. https://doi.org/https://doi.org/10.1016/j.yrtph.2019.104445

15. *Delgado, L.F., Charles, P., Glucina, K., Morlay, C.,* 2012. QSAR-like models: A potential tool for the selection of PhACs and EDCs for monitoring purposes in drinking water treatment systems - A review. Water Res. 46, 6196–6209. https://doi.org/10.1016/j.watres.2012.08.016

16. *Du, H., Cai, Y., Yang, H., Zhang, H., Xue, Y., Liu, G., Tang, Y., Li, W.,* 2017. In Silico Prediction of Chemicals Binding to Aromatase with Machine Learning Methods. Chem. Res. Toxicol. 30, 1209–1218. https://doi.org/10.1021/acs.chemrestox.7b00037

17. *Dudek, A., Arodz, T., Galvez, J.*, 2006. Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. Comb. Chem. High Throughput Screen. 9, 213–228. https://doi.org/10.2174/138620706776055539

18. *Fantacuzzi, M., De Filippis, B., Gallorini, M., Ammazzalorso, A., Giampietro, L., Maccallini, C., Aturki, Z., Donati, E., Ibrahim, R.S., Shawky, E., Cataldi, A., Amoroso, R.,* 2020. Synthesis, biological evaluation, and docking study of indole aryl sulfonamides as aromatase inhibitors. Eur. J. Med. Chem. 185, 111815. https://doi.org/10.1016/j.ejmech.2019.111815

19. *Gassner, F.X., Reifenstein, E.C.J., Algeo, J.W., Mattox, W.E.*, 1958. Effects of hormones on growth, fattening, and meat production potential of livestock. Recent Prog. Horm. Res. 14, 183–187.

20. *Ghodsi, R., Hemmateenejad, B.,* 2016. QSAR study of

diarylalkylimidazole and diarylalkyltriazole aromatase inhibitors. Med. Chem. Res. 25, 834–842. https://doi.org/10.1007/s00044-016-1530-1

21. *Gozalbes, R., de Julián-Ortiz, J.V., Fito-López, C.*, 2015. Métodos computacionales en toxicología predictiva: Aplicación a la reducción de ensayos con animales en el contexto de la legislación comunitaria REACH. Rev. Toxicol. 31, 157–167.

22. *Heinlein, C.A., Chang, C.,* 2004. Androgen Receptor in Prostate Cancer. Endocr. Rev. 25, 276–308. https://doi.org/10.1210/er.2002-0032

23. *Hong, H., Branham, W.S., Ng, H.W., Moland, C.L., Dial, S.L., Fang, H., Perkins, R., Sheehan, D., Tong, W.*, 2015. Human sex hormone-binding globulin binding affinities of 125 structurally diverse chemicals and comparison with their binding to androgen receptor, estrogen receptor, and α-Fetoprotein. Toxicol. Sci. 143, 333–348. https://doi.org/10.1093/toxsci/kfu231

24. *Kar, S., Sepúlveda, M.S., Roy, K., Leszczynski, J.*, 2017. Endocrine-disrupting activity of per- and polyfluoroalkyl substances: Exploring combined approaches of ligand and structure based modeling. Chemosphere 184, 514–523. https://doi.org/10.1016/j.chemosphere.2017.06.024

25. *Karjalainen, T., Hoeveler, A., Draghia-Akli, R.*, 2017. European Union research in support of environment and health: Building scientific evidence base for policy. Environ. Int. 103, 51–60. https://doi.org/10.1016/j.envint.2017.03.014

26. *Kovarich, S., Papa, E., Gramatica, P.,* 2011. QSAR classification models for the prediction of endocrine disrupting activity of brominated flame retardants. J. Hazard. Mater. 190, 106–112. https://doi.org/10.1016/j.jhazmat.2011.03.008

27. *Kovarich, S., Papa, E., Li, J., Gramatica, P.*, 2012. QSAR classification models for the screening of the endocrine-disrupting activity of perfluorinated compounds. SAR QSAR Environ. Res. 23, 207–220. https://doi.org/10.1080/1062936X.2012.657235

28. *Leblebicioglu, B., Connors, J., Mariotti, A.,* 2013. Principles of endocrinology. Periodontol. 2000 61, 54–68. https://doi.org/10.1111/j.1600-0757.2011.00440.x

29. *Leechaisit, R., Pingaew, R., Prachayasittikul, Veda, Worachartcheewan, A., Prachayasittikul, S., Ruchirawat, S., Prachayasittikul, Virapong,* 2019. Synthesis, molecular docking, and QSAR study of bis-sulfonamide derivatives as potential aromatase inhibitors. Bioorganic Med. Chem. 27, 115040. https://doi.org/10.1016/j.bmc.2019.08.001

30. *Li, H., Ren, X., Leblanc, E., Frewin, K., Rennie, P.S., Cherkasov, A.*, 2013. Identification of novel androgen receptor antagonists using structure- and ligand-based methods. J. Chem. Inf. Model. 53, 123–130. https://doi.org/10.1021/ci300514v

31. *Li, X., Ye, L., Wang, X., Shi, W., Liu, H., Qian, X., Zhu, Y., Yu, H.,* 2013. In silico investigations of anti-androgen activity of polychlorinated biphenyls. Chemosphere 92, 795–802. https://doi.org/10.1016/j.chemosphere.2013.04.022

32. *Lin, Xiaoqian, Li, X., Lin, Xubo,* 2020. A review on applications of computational methods in drug screening and design. Molecules 25, 1–18. https://doi.org/10.3390/molecules25061375

33. *Little, A.G.,* 2016. A review of the peripheral levels of regulation by thyroid hormone. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 186, 677–688. https://doi.org/10.1007/s00360-016-0984-2

34. *Loughney, D.A., Schwender, C.F.,* 1992. A comparison of progestin and androgen receptor binding using the CoMFA technique. J. Comput. Aided. Mol. Des. 6, 569–581. https://doi.org/10.1007/BF00126215

35. *McEwan, I.J.*, 2016. The Nuclear Receptor Superfamily at Thirty, in: The Nuclear Receptor Superfamily. Springer Science+Business Media, New York, pp. 3–13. https://doi.org/10.1007/978-1-60327-575-0

36. *Miller WL,* 1988. Molecular biology of steroid hormone synthesis. Endocr. Rev. 9, 295–318.

37. *Mori, T., Ito, F., Koshiba, A., Kataoka, H., Tanaka, Y., Okimura, H., Khan, K.N., Kitawaki, J.,* 2018. Aromatase as a target for treating endometriosis. J. Obstet. Gynaecol. Res. 44, 1673–1681. https://doi.org/10.1111/jog.13743

38. *Nantasenamat, C., Li, H., Mandi, P., Worachartcheewan, A., Monnor, T., Isarankura-Na-Ayudhya, C., Prachayasittikul, V.,* 2013a. Exploring the chemical space of aromatase inhibitors. Mol. Divers. 17, 661–677. https://doi.org/10.1007/s11030-013-9462-x

39. *Nantasenamat, C., Worachartcheewan, A., Mandi, P., Monnor, T., Isarankura-Na-Ayudhya, C., Prachayasittikul, V.,* 2014. QSAR modeling of aromatase inhibition by flavonoids using machine learning approaches. Chem. Pap. 68, 697–713. https://doi.org/10.2478/s11696-013-0498-2

40. *Nantasenamat, C., Worachartcheewan, A., Prachayasittikul, S., Isarankura-Na-Ayudhya, C., Prachayasittikul, V.,* 2013b. QSAR modeling of aromatase inhibitory activity of 1-substituted 1,2,3-triazole analogs of letrozole. Eur. J. Med. Chem. 69, 99–114. https://doi.org/10.1016/j.ejmech.2013.08.015

41. *Narayana, B.L., Pran Kishore, D., Balakumar, C., Rao, K.V., Kaur, R., Rao, A.R., Murthy, J.N., Ravikumar, M.,* 2012. Molecular Modeling Evaluation of Non-Steroidal Aromatase Inhibitors. Chem. Biol. Drug Des. 79, 674–682. https://doi.org/10.1111/j.1747-0285.2011.01277.x

42. OECD, 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment. OECD Publishing, Paris. https://doi.org/10.1787/9789264085442-en

43. *Papa, E., Kovarich, S., Gramatica, P.,* 2013. QSAR prediction of the competitive interaction of emerging halogenated pollutants with human transthyretin£. SAR QSAR Environ. Res. 24, 333–349. https://doi.org/10.1080/1062936X.2013.773374

44. *Papa, E., Kovarich, S., Gramatica, P.,* 2010. QSAR Modeling and Prediction of the Endocrine-Disrupting Potencies of Brominated Flame Retardants. Chem. Res. Toxicol. 23, 946–954. https://doi.org/10.1021/tx1000392

45. *Porta, N., Roncaglioni, A., Marzo, M., Benfenati, E.*, 2016. QSAR Methods to Screen Endocrine Disruptors. Nucl. Recept. Res. 3. https://doi.org/10.11131/2016/101203

46. *Prachayasittikul, V., Pingaew, R., Worachartcheewan, A., Sitthimonchai, S., Nantasenamat, C., Prachayasittikul, S., Ruchirawat, S., Prachayasittikul, V.,* 2017. Aromatase inhibitory activity of 1,4-naphthoquinone derivatives and QSAR study. EXCLI J. 16, 714–726. https://doi.org/10.17179/excli2017-309

47. *Rogers, D.*, 1999. Genetic Function Approximation: Evolutionary Construction of Novel, Interpretable, Nonlinear Models of Experimental Data, in: Truhlar, D.G., Howe, W.J., Hopfinger, A.J., Blaney, J., Dammkoehler, R.A. (Eds.), Rational Drug Design. Springer New York, New York, NY, pp. 163–189. https://doi.org/10.1007/978-1-4612-1480-9_13

48. *Roy, P.P., Roy, K.,* 2010. Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. J. Mol. Model. 16, 1597–1616. https://doi.org/10.1007/s00894-010-0667-y

49. *Ruiz, P., Sack, A., Wampole, M., Bobst, S., Vracko, M.,* 2017. Integration of in silico methods and computational systems biology to explore endocrine-disrupting chemical binding with nuclear hormone receptors. Chemosphere 178, 99–109. https://doi.org/10.1016/j.chemosphere.2017.03.026

50. *Sahoo, S., Adhikari, C., Kuanar, M., Mishra, B.K.*, 2016. A Short Review of the Generation of Molecular Descriptors and Their Applications in Quantitative Structure Property/Activity Relationships. Curr. Comput. Aided. Drug Des. 12, 181–205. https://doi.org/10.2174/1573409912666160525112114

51. *Schug, T.T., Johnson, A.F., Birnbaum, L.S., Colborn, T., Guillette, L.J., Crews, D.P., Collins, T., Soto, A.M., Vom Saal, F.S., McLachlan, J.A., Sonnenschein, C., Heindel, J.J.,* 2016. Minireview: Endocrine disruptors: Past lessons and future directions. Mol. Endocrinol. 30, 833–847. https://doi.org/10.1210/me.2016-1096

52. *Shoombuatong, W., Prachayasittikul, V., Prachayasittiku, V., Nantasenamat, C.,* 2015. Prediction of aromatase inhibitory activity using the efficient linear method (ELM). EXCLI J. 14, 452–464. https://doi.org/10.17179/excli2015-140

53. *Shoombuatong, W., Schaduangrat, N., Nantasenamat, C.,* 2018. Towards understanding aromatase inhibitory activity via QSAR modeling. EXCLI J. 17, 688–708. https://doi.org/10.17179/excli2018-1417

54. *Soto, A.M., Rubin, B.S., Sonnenschein, C.,* 2009. Interpreting endocrine disruption from an integrative biology perspective. Mol. Cell. Endocrinol. 304, 3–7. https://doi.org/10.1016/j.mce.2009.02.020

55. *Suresh, N.,* 2013. Human Androgen Receptor Inhibitors: Computational 3D QSAR Studies to Design Lead Compounds for Treatment of Prostate Cancer. Turkish J. Biochem. 38, 262–279. https://doi.org/10.5505/tjb.2013.38258

56. *Tata, J.R.,* 2017. How Hormones, as Ancient Signalling Molecules, Regulate Diverse Biological Processes Through Evolution BT - Hormones in Ageing and Longevity, in: Rattan, S., Sharma, R. (Eds.), . Springer International Publishing, Cham, pp. 3–20. https://doi.org/10.1007/978-3-319-63001-4_1

57. *Thomas Zoeller, R., Brown, T.R., Doan, L.L., Gore, A.C., Skakkebaek, N.E., Soto, A.M., Woodruff, T.J., Vom Saal, F.S.,* 2012. Endocrine-disrupting chemicals and public health protection: A statement of principles from the Endocrine Society. Endocrinology 153, 4097–4110. https://doi.org/10.1210/en.2012-1422

58. *Todeschini, R., Consonni, V.,* 2009. Molecular Descriptors for Chemoinformatics. WILEY-VCH.

59. *Todeschini, R., Consonni, V.*, 2000. Handbook of Molecular Descriptors, 1st ed. WILEY-VCH, Weinheim. https://doi.org/10.1002/9783527613106

60. *Tourneix, F., Alépée, N., Detroyer, A., Eilstein, J., Ez-Zoubir, M., Teissier, S.M., Noçairi, H., Piroird, C., Basketter, D., Del Bufalo, A.,* 2020. Skin sensitisation testing in practice: Applying a stacking meta model to cosmetic ingredients. Toxicol. Vitr. 66, 104831. https://doi.org/https://doi.org/10.1016/j.tiv.2020.104831

61. *Trisciuzzi, D., Alberga, D., Mansouri, K., Judson, R., Novellino, E., Mangiatordi, G.F., Nicolotti, O.*, 2017. Predictive Structure-Based Toxicology Approaches to Assess the Androgenic Potential of Chemicals. J. Chem. Inf. Model. 57, 2874–2884. https://doi.org/10.1021/acs.jcim.7b00420

62. *Tropsha, A.,* 2010. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol. Inform. 29, 476–488. https://doi.org/10.1002/minf.201000061

63. *Tropsha, A., Gramatica, P., Gombar, V.K.,* 2003. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR Comb. Sci. 22, 69–77. https://doi.org/10.1002/qsar.200390007

64. *Vandenberg, L.N., Colborn, T., Hayes, T.B., Heindel, J.J., Jacobs, D.R., Lee, D.H., Myers, J.P., Shioda, T., Soto, A.M., vom Saal, F.S., Welshons, W. V., Zoeller, R.T.,* 2013. Regulatory decisions on endocrine disrupting chemicals should be based on the principles of endocrinology. Reprod. Toxicol. 38, 1–15. https://doi.org/10.1016/j.reprotox.2013.02.002

65. *Vinggaard, A.M., Niemela, J., Wedebye, E.B., Jensen, G.E.,* 2008. Screening of 397 chemicals and development of a quantitative structure--activity relationship model for androgen receptor antagonism. Chem. Res. Toxicol. 21, 813–823. https://doi.org/10.1021/tx7002382

66. *Vuorinen, A., Odermatt, A., Schuster, D.*, 2015. Reprint of "in silico methods in the discovery of endocrine disrupting chemicals." J. Steroid Biochem. Mol. Biol. 153, 93–101. https://doi.org/10.1016/j.jsbmb.2015.08.015

67. *Waller, C.L., Juma, B.W., Earl Gray, L., Kelce, W.R.,* 1996. Three-dimensional quantitative structure-activity relationships for androgen receptor ligands. Toxicol. Appl. Pharmacol. 137, 219–227. https://doi.org/10.1006/taap.1996.0075

68. *Wang, Y., Han, R., Zhang, H., Liu, Hongli, Li, J., Liu, Huanxiang, Gramatica, P.,* 2017. Combined Ligand/Structure-Based Virtual Screening and Molecular Dynamics Simulations of Steroidal Androgen Receptor Antagonists. Biomed Res. Int. 2017. https://doi.org/10.1155/2017/3572394

69. *Weiss, J.M., Andersson, P.L., Lamoree, M.H., Leonards, P.E.G., Van Leeuwen, S.P.J., Hamers, T.*, 2009. Competitive binding of poly- and perfluorinated compounds to the thyroid hormone transport protein transthyretin. Toxicol. Sci. 109, 206–216. https://doi.org/10.1093/toxsci/kfp055

70. WHO/UNEP, 2013. State of the science of endocrine disrupting chemicals - 2012 An assessment of the state of the science of endocrine disruptors prepared by a group of experts for the United Nations Environment Programme (UNEP) and WHO, World Health Organization.

71. WHO, 2002. Global assessment of the state-of-the-science of endocrine disruptors., WHO publication no. WHO/PCS/EDC/02.2.

72. *Worachartcheewan, A., Mandi, P., Prachayasittikul, V., Toropova, A.P., Toropov, A.A., Nantasenamat, C*., 2014a. Large-scale QSAR study of aromatase inhibitors using SMILES-based descriptors. Chemom. Intell. Lab. Syst. 138, 120–126. https://doi.org/10.1016/j.chemolab.2014.07.017

73. *Worachartcheewan, A., Suvannang, N., Prachayasittikul, S., Prachayasittikul, V., Nantasenamat, C.*, 2014b. Probing the origins of aromatase inhibitory activity of disubstituted coumarins via qsar and molecular docking. EXCLI J. 13, 1259–1274. https://doi.org/10.17877/DE290R-7420

74. *Wu, Y., Doering, J.A., Ma, Z., Tang, S., Liu, H., Zhang, X., Wang, X., Yu, H.,* 2016. Identification of androgen receptor antagonists: In vitro investigation and classification methodology for flavonoid. Chemosphere 158, 72–79. https://doi.org/10.1016/j.chemosphere.2016.05.059

75. *Wu, Y., Shi, W., Xia, P., Zhang, X., Yu, H.,* 2017. Qualitative and quantitative simulation of androgen receptor antagonists: A case study of polybrominated diphenyl ethers. Sci. Total Environ. 603–604, 495–501. https://doi.org/10.1016/j.scitotenv.2017.06.106

76. *Xie, H., Qiu, K., Xie, X.,* 2015. Pharmacophore modeling, virtual screening, and 3D-QSAR studies on a series of non-steroidal aromatase inhibitors. Med. Chem. Res. 24, 1901–1915. https://doi.org/10.1007/s00044-014-1257-9

77. *Xie, H., Qiu, K., Xie, X.,* 2014. 3D QSAR studies, pharmacophore modeling and virtual screening on a series of steroidal aromatase

inhibitors. Int. J. Mol. Sci. 15, 20927–20947. https://doi.org/10.3390/ijms151120927

78. *Yang, W., Shen, S., Mu, L., Yu, H.,* 2011. Structure-activity relationship study on the binding of PBDEs with thyroxine transport proteins. Environ. Toxicol. Chem. 30, 2431–2439. https://doi.org/10.1002/etc.645

79. *Yang, X., Liu, H., Yang, Q., Liu, J., Chen, J., Shi, L.,* 2016. Predicting anti-androgenic activity of bisphenols using molecular docking and quantitative structure-activity relationships. Chemosphere 163, 373–381. https://doi.org/10.1016/j.chemosphere.2016.08.062

80. *Yang, X., Lyakurwa, F., Xie, H., Chen, J., Li, X., Qiao, X., Cai, X.*, 2017. Different binding mechanisms of neutral and anionic poly-/perfluorinated chemicals to human transthyretin revealed by In silico models. Chemosphere 182, 574–583. https://doi.org/10.1016/j.chemosphere.2017.05.016

81. *Yang, X., Xie, H., Chen, J., Li, X.,* 2013. Anionic phenolic compounds bind stronger with transthyretin than their neutral forms: Nonnegligible mechanisms in virtual screening of endocrine disrupting chemicals. Chem. Res. Toxicol. 26, 1340–1347. https://doi.org/10.1021/tx4001557

82. *Yen, P.M.*, 2001. Physiological and Molecular Basis of Thyroid Hormone Action. Physiol. Rev. 81, 1097–1142. https://doi.org/10.1152/physrev.2001.81.3.1097

83. *Yilmaz, B., Terekeci, H., Sandal, S., Kelestimur, F.,* 2020. Endocrine disrupting chemicals: exposure, effects on human health, mechanism of action, models for testing and strategies for prevention. Rev. Endocr. Metab. Disord. 21, 127–147. https://doi.org/10.1007/s11154-019-09521-z

84. *Zhang, J., Kamstra, J.H., Ghorbanzadeh, M., Weiss, J.M., Hamers, T., Andersson, P.L.,* 2015. In Silico Approach To Identify Potential Thyroid Hormone Disruptors among Currently Known Dust Contaminants and Their Metabolites. Environ. Sci. Technol. 49, 10099–10107. https://doi.org/10.1021/acs.est.5b01742

ANNEXE 2. Review article: Dietary flavonoids as AhR modulators

- **Publication Title:** Plant occurring flavonoids as ligands of the aryl hydrocarbon receptor.
- **Authors**: <u>Elizabeth Goya-Jorge</u>, Elisa Jorge Rodríguez, Maité Sylla-Iyarreta Veitía, Rosa M. Giner

(manuscript in preparation)

Abstract

The aryl hydrocarbon receptor (AhR) is a transcription factor deeply implicated in health and disease. Hence, targeting AhR has become an emerging pharmacological target in cancer, immunology, inflammatory diseases, and aging. Flavonoids are considered the largest group of natural dietary ligands of AhR. However, structural and cellular context dependencies are acknowledged on the agonist and/or antagonist effects caused by flavonoids on AhR. Herein is presented an update and summary of the state of the art (last 5 years) in an extensive list of plant occurring flavonoids that have been studied as AhR modulators. With the aim of highlight the structural features causing AhR agonism and/or antagonism, a structure-activity relationship analysis was addressed on the main subclasses of flavonoids (*i.e.* flavones, isoflavones, flavonols, and flavanones) reported in the literature. We hope that this review contributes to elucidate further on the structural determinants of flavonoids as AhR ligands, and therefore their potentially harmful and beneficial effects through this pathway.

Section of Plant flavonoids: the main class of natural AhR ligands in human diet

1. Flavonoids: chemistry and bioactivity profiles

1.1. Plant polyphenols

The (poly)phenols compounds comprise a large and heterogeneous group of natural products. In the context of plant derivatives, it is suggested to consider exclusively the term "plant polyphenols" to define those secondary natural metabolites from the shikimate derived phenylpropanoid and/or the polyketide pathway(s). Therefore, strictly featuring the presence of more than one phenolic ring and without pondered simple nitrogen-based functions [1,2].

Plant polyphenols are secondary metabolites widely distributed throughout the plant kingdom in both edible and non-edible species. Moreover, they are implicated in diverse ecological and physiological roles including pigmentation, protection of plants from ultraviolet radiation and plant-defense against excess light, low temperatures, pathogen infection, herbivores, and deficiency of nutrients. The importance of polyphenols is virtually associated with all classes of environmental stresses that ultimately lead to increased production of free radicals and oxidative species in plants. Furthermore, as a large reservoir of chemical diversity, plant polyphenols are involved in a varied spectrum of gene and enzymatic regulation mechanisms [3–5].

Several thousand phenolic compounds synthesized by higher plants are known to date, and this number is continually increasing. Based on their basic chemical skeleton, the phenolic and polyphenolic compounds found in plants can be grouped into the following five main classes: flavonoids, phenolic acids, tannins, lignans and stilbenes [6]. The largest group identified are the flavonoids with thousand structures characterized [7]. In multiple vegetables, flavonoids are responsible for the color and flavor [8]. The aglycone and glycoside (i.e., glucosides, rhamnoglucosides, and rutinosides) forms of flavonoids are found in the human daily diet like fruits, vegetables, grains, seeds, bark, herbs, roots, flowers, stems, and spices. Moreover, the bioactivity profiles of these compounds have made them the object of research for decades in pharmacological and toxicological fields [9,10].

1.2. Chemistry of flavonoids

Structurally, flavonoids are characterized by their basic skeleton arranged in the form C6-C3-C6, which corresponds to two benzene rings linked by a unit of three carbon atoms, which may or may not give rise to a third ring. Despite the various criteria regarding their classification, three general classes are recognized according to the linkage position of the aromatic ring with respect to the benzopyrane (chromane) moiety: flavonoids (2-phenyl-1,4-benzopyrones), isoflavonoids (3-phenyl-1,4-benzopyrones), and neoflavonoids (4-phenyl-1,2-benzopyrones) [4]. Moreover, they can be subclassified into subgroups according to the degree of oxidation of the pyranic ring, which can be opened and recycled in the furan cycle, and linking chain unsaturation: 2-phenylchromones (flavones, flavaonols, flavanones, and dihydroflavonols); 2-phenylchromanes (flavans, flavan-3-ols, flavan-3,4-diols), anthocyanins (2-phenyl-benzopyrilium) and chalcones, dihydrochalcones and aurones [9–12]. Figure 1 shows some representative examples of these classes.
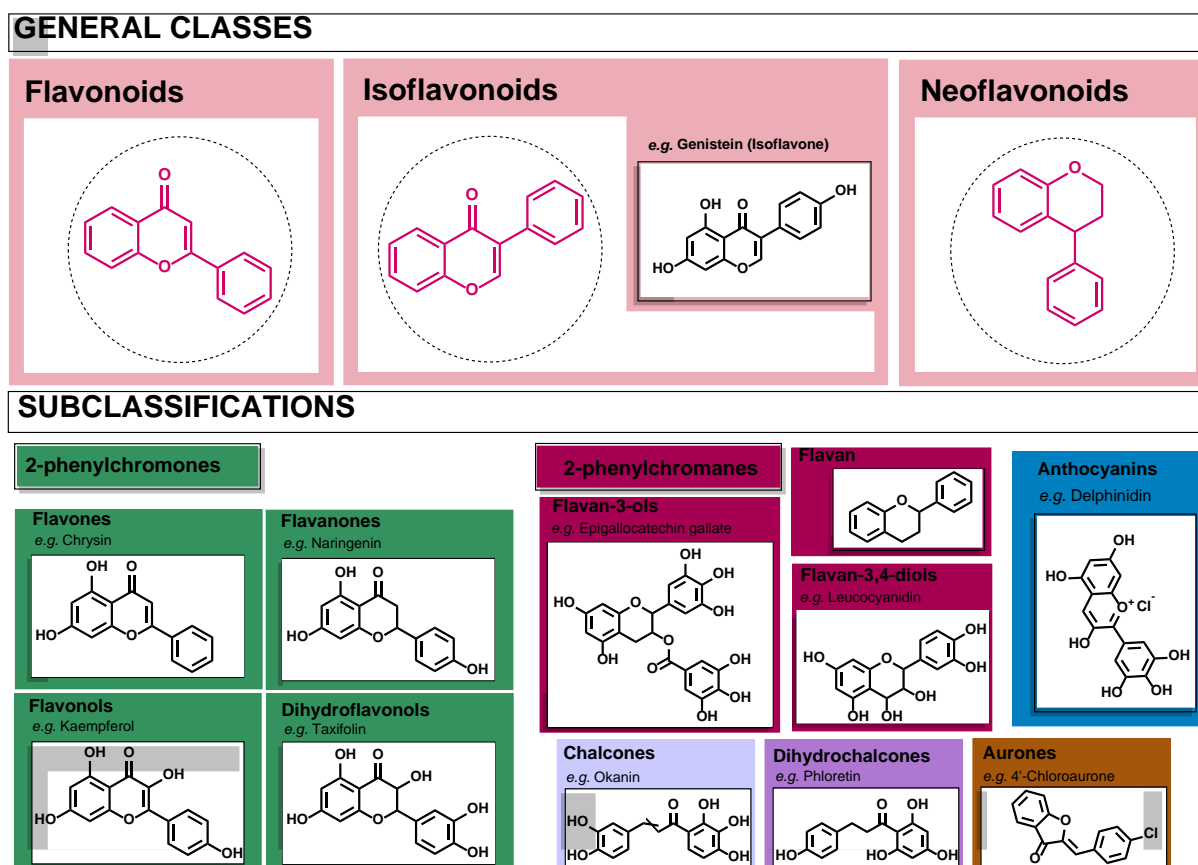
Figure 1. General classes and subclasses of flavonoids and some representative examples.

Regarding their chemical reactivity, flavonoids occur as hydroxylated, methylated, and prenylated aglycones or glycosylated derivatives, either O-glycosides formed via linkage of the sugar unit to the hydroxyls or C-glycosides directly to the carbon atom of the flavonoid skeleton [9,13]. In the cytosol (pH 7.4), most flavonoids form a mixture of phenolate anions and neutral phenols. Their proportion depends on the pKa of each phenolic group in the structure. Since most flavonoids are weak hydrophobic acids, depending on their lipophilicity they have potential to cross cellular and mitochondrial membranes and act as protonophores. Sometimes even minimal structural modifications alter the solubility, reactivity, and stability of flavonoids that ultimately lead to their bioavailability differences and the diversity of biological effects across the groups [14,15].

1.3. Occurrence & Bioactivity of plant flavonoids

Flavonoids are a large group of polyphenolic compounds omnipresent in plants that are commonly found in foods and beverages. Furthermore, the crucial roles of phenolic molecules and particularly flavonoids as protective dietary constituents have become an increasingly important area of human nutrition research. Countless literature reports indicate a link between flavonoid consumption and health benefits (or risks) due to its antioxidant, antiproliferative, estrogenic or anti-estrogenic properties [16,17]. Some observational and metabolic studies suggest that flavonoids may lower the risk of some diet-related chronic degenerative diseases. Other clinical and pre-clinical studies have indicated important adverse effects such as liver toxicity as well as the interference of flavonoid-rich diets with common medications [18–21]. Therefore, it is important to accurately assess flavonoids intake to ensure safety and take advantage only of their beneficial effects [22,23].

The current uses of plant flavonoids cover cosmetic, nutraceutical and pharmaceutical industries [1,10,24]. Their prophylactic efficacy against neoplastic [25–27], cardiovascular [28–30], bone [31], and neurological disorders [32], are among the most widespread. Some other pharmacological studies have revealed anti-inflammatory [33,34] and antibacterial [5] potential of flavonoids as well as their applications in the prevention and treatment of obesity [35] and diabetes [36,37]. Besides, antioxidant effects have been attributed to all flavonoid subclasses [38]. Thus, most of the anticancer properties attributed to plant flavonoids and natural extracts from traditional medicine are explained through flavonoids' ability to scavenge free radicals and other reactive oxygen species (ROS), modulate ROS-scavenging enzymes, chelate metal ions and prevent

oxidative stress-related conditions [12,39]. However, the mechanisms mediating the diverse bioactivity profiles of flavonoids are still pending of further investigations, since in many cases they are contradictory [9,40].

2. Recent Studies of Flavonoids as AhR modulators

Several well-known flavonoids have been studied as AhR modulators. Representative compounds and bioassays reported in the last five years (the scope of this review) are summarized in Table 1.

Table 1. Representative examples of plant occurring flavonoids studied as AhR ligands through different bioassays.

Nº	Flavonoid	Occurrence[a]		Tested [µM] or dose [mg/kg]	Bioassay[b]	Model details[c]	Ref[d]
		Plants	Food				
1	Acacetin	*Turnera diffusa*	ginkgo nuts orange mint	10 µM, 50 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1)	41
				10 µM	RT-qPCR	YAMC (CYP1A1)	
				10 µM	RT-qPCR	YAMC (CYP1B1, UGT1A1)	
2	Apigenin	*Malus domestica*	apple, apricots, turnip	50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
				50 µM 100 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1)	41
				10 µM, 50 µM	RT-qPCR	YAMC (CYP1A1, CYP1B1, UGT1A1)	
				10 µM, 50 µM 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
				25 µM	LucRGA	H1L6.1c2 (CYP1A1)	44
				30 µM	WB	3T3-L1 (AhR)	45
3	Chrysin	*Passiflora caerulea*	mushrooms, carrots	25 µM	LucRGA	H1L6.1c1 (CYP1A1)	44
				100 µM	LucRGA	HCT116 (CYP1A1)	46
				30 µM	WB	3T3-L1 (AhR)	45
4	Genkwanin	*Daphne genkwa*		10 µM,50 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1) & YAMC (CYP1A1, CYP1B1, UGT1A1)	41
5	Icaritin	*Herba Epimedii*		10 µM,20 µM, 30 µM, 40 µM,50 µM	RT-PCR	LNCaP & CWR22Rv1 (CYP1A1)	47
6	Luteolin	*Reseda luteola*	pistachio, peppers, artichokes	10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
				30 µM	WB	3T3-L1 (AhR)	45
7	Nobiletin	*Citrus reticulata*		30 µM	WB	3T3-L1 (AhR)	45
8	Scutellarein	*Scutellaria lateriflora*		50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
9	Tangeretin	*Citrus maxima*		30 µM	WB	3T3-L1 (AhR)	45
				50 mg/kg	WB, PCR, IF, IHC	Hepatic tissue (rats)	48
10	Tricetin	*Stachys scardica*		10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				50 µM 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	

11	Robinetin	*Robinia pseudoacacia*		10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
12	Baicalin	*Scutellaria baicalensis*		5 µM, 10 µM, 20 µM	IF, WB, RT-PCR	DLN (CYPA1, AhR)	49
				1 µM	LucRGA	H1L6.1c2 (CYP1A1)	44
				30 mg/L	RT-PCR, WB, IHC	Myocardial tissue (mice)	50
				30 µM	WB	3T3-L1 (AhR)	45
13	Taxifolin	*Larix sibirica*	chestnut, corn, cucumber	10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
14				50 µM 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
15	Dihydromyricetin	*Ampelopsis cantoniensis*		From 0.1 µM to 100 µM	LucRGA	HepG2 (CYP1A1, CYP1A2, CYP2B1, CYP2B2)	51
16	Silymarin	*Silybum marianum*		25 µM	LucRGA	H1L6.1c2 (CYP1A1)	44
17	Fisetin	*Senegalia greggii*	persimmons, strawberry, cucumber	10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
				25 µM	LucRGA	H1L6.1c2 (CYP1A1)	44
				30 µM	WB	3T3-L1 (AhR)	45
18	Galangin	*Allium fistulosum*	onions, wine	30 µM	WB	3T3-L1 (AhR)	45
19	Gossypetin	*Hibiscus sabdariffa*		50 µM 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
20	Isorhamnetin	*Ginkgo biloba*		30 µM	WB	3T3-L1 (AhR)	45
21	Kaempferol	*Allium cepa*	apricots, honey, kale	50 µM 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
				30 µM	WB	3T3-L1 (AhR)	45
22	Morin	*Prunus dulcis*	apple, almond	50 µM 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
23	Myricetin	*Solanum pimpinellifolium*	tomato, artichokes, honey	10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
24	Quercetin	*Persea americana*	arugula, avocados, beans	10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
				25 µM	qPCR	PBMEC/C1-2	44
				25 µM	LucRGA	H1L6.1c2 (CYP1A1)	
				30 µM	WB	3T3-L1 (AhR)	45
25	Guaijaverin	*Psidium guajava*	onions, strawberries, guava	50 µM 100 µM	LucRGA	HepG2-AhR Lucia (CYP1A1)	42
26	Rutin	*Ruta graveolens*	grapes, asparagus, raspberry	50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
				25 µM	LucRGA	H1L6.1c2 (CYP1A1)	44
27	6-Prenylnaringenin	*Humulus lupulus*		10 µM	LucRGA	HepG2 (CYP1A1)	52
				5 µM	LucRGA	MCF-7 (CYP1A1)	
28	8-Prenylnaringenin	*Humulus lupulus*		10 µM	LucRGA	HepG2 (CYP1A1)	52
				5 µM	LucRGA	MCF-7 (CYP1A1)	
29	Hesperetin	*Allium fistulosum*		50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
				0.1 µM, 1 µM, 10 µM	LucRGA, WB, RT-PCR	MCF-7 (CYP1A1, CYP1A2, CYP1B1)	53
				1 µM	LucRGA	H1L6.1c2 (CYP1A1)	44

30	Naringenin	*Vitis vinifera*	oregano, artichokes, grapefruit	10 µM, 50 µM, 100 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1)	
				10 µM,50 µM	RT-qPCR	YAMC (CYP1A1)	41
				10 µM,50 µM	RT-qPCR	YAMC (CYP1B1, UGT1A1)	
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (CYP1A1)	43
				10 µM, 50 µM, 100 µM	RT-qPCR, ChIP, WB	Caco2 (UGT1A1)	
				30 µM	WB	3T3-L1 (AhR)	45
31	Naringin	*Citrus sinensis*	orange, pomelo	50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
				200 mg/kg	RT-PCR	liver tissue (rats)	54
32	Hesperidin	*Mentha × piperita*	orange, grapefruit, lemon	25 µM	qPCR	PBMEC/C1-2	44
				25 µM,	qPCR	PBMEC/C1-2	44
				50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
33	Sakuranetin	*Pyricularia oryzae*		12.5 µM 25 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
34	Alpinetin	*Alpinia katsumadai*		3 µM, 10 µM, 30 µM	Co-IP, WB	Caco2 (CYPA1, AhR)	55
				30 µM	EROD	EL-4	56
35	Epigallocatechin gallate	*Matricaria chamomilla*	kiwi, avocado, cranberries,	50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
				30 µM	WB	3T3-L1 (AhR)	45
36	Cardamonin	*Alpinia conchigera*	orange	3 µM, 10 µM, 30 µM	qPCR, WB	THP-1 (CYP1A1)	57
37	Biochanin A	*Trifolium pratense*	peanuts, chickpea, soy	10 µM,50 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1)	
				10 µM,50 µM	RT-qPCR	YAMC (CYP1A1)	41
				10 µM	RT-qPCR	YAMC (CYP1B1, UGT1A1)	
				25 µM	LucRGA	H1L6.1c2 (CYP1A1)	44
				From 1 µM to 20 µM	RT-qPCR	MCF-7 (CYP1A1, CYP1B1)	58
				10 µM	LucRGA	HC-04	
				From 0.1 µM to 1 mM	LucRGA	HepG2 (AZ-AhR)	59
38	Daidzein	*Maackia amurensis*	tomato, soy	From 1 µM to 20 µM	RT-qPCR	MCF-7 (CYP1A1, CYP1B1)	58
				10 µM	LucRGA	HC-04	
				5 µM	WB	Hepa-1c1c7 (CYP1A1)	60
				From 0.1 µM to 1 mM	LucRGA	HepG2 (AZ-AhR)	59
39	Formononetin	*Arachis hypogaea*	soy, peanut	From 1 µM to 20 µM	RT-qPCR	MCF-7 (CYP1A1, CYP1B1)	58
				10 µM	LucRGA	HC-04	
				From 0.1 µM to 1 mM	LucRGA	HepG2 (AZ-AhR)	59
40	Genistein	*Genista tinctoria*	chickpeas, soy	50 µM 100 µM	LucRGA	HepG2 AhR-Lucia (CYP1A1)	42
				10 µM, 50 µM 100 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1)	41
				10 µM	RT-qPCR	YAMC (CYP1A1, CYP1B1, UGT1A1)	
				From 1 µM to 20 µM	RT-qPCR	MCF-7 (CYP1A1, CYP1B1)	58
				10 µM	LucRGA	HC-04	
				5 µM	WB	Hepa-1c1c7 (CYP1A1)	60
				0.1 µM-1 mM	LucRGA	HepG2 (AZ-AhR)	59
41	Prunetin	*Prunus emarginata*	plum, cherry	10 µM, 50 µM	RT-qPCR	Caco2 (CYP1A1, CYP1B1, UGT1A1)	41
				10 µM	RT-qPCR	YAMC (CYP1A1, CYP1B1, UGT1A1)	

[a] **References**: [61–63]. [b] **Bioassays**: Chromatin immunoprecipitation (ChIP), Co-immunoprecipitation (Co-IP), EROD, ethoxyresorufin-O-deethylase (EROD), Immunofluorescence (IF), Immunohistochemical (IHC), Luciferase reporter gene assay

(LucRGA), reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR), western blotting (WB). [c] **Cell lines**: draining lymph node (DLN), human breast adenocarcinoma (MCF-7), human colon cancer (HCT116), human epithelial colorectal adenocarcinoma (Caco2), human hepatocyte (HC-04), human hepatoma (HepG2), human monocytic leukemia (THP-1), human prostate carcinoma (LNCaP, CWR22Rv1), mouse adipocytes (3T3-L1), mouse colonic epithelium (YAMC), mouse hepatoma (H1L6.1c2, Hepa-1c1c7), mouse lymphoblast (EL-4), porcine brain microvascular endothelium (PBMEC/C1-2). [d] References are exclusively linked to the tested concentrations and bioassays showed.

As presented in Table 1, in recent years accurate and informative techniques such as RT-qPCR, LucRGA, ChIP, Co-IP, EROD, IF, IHC, and WB have been used to assess the modulatory capacity of flavonoids on AhR. Meanwhile, the maximum tested concentrations have been around 100 μM, except for some isoflavones. In Figure 2, the flavonoids included in Table 1 and some others detailed later were quantified, revealing that the main subclasses of flavonoids recently studied as AhR agonists and/or antagonist have been: flavones, isoflavones, flavanones, and flavonols.
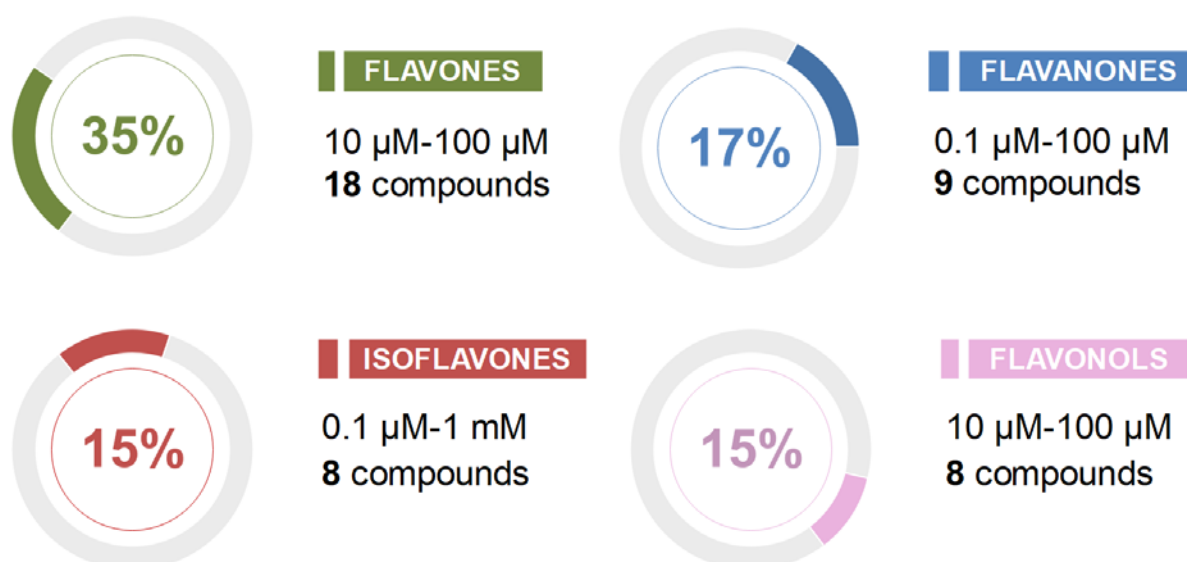
Figure 2. Main subclasses of flavonoids reported in the literature as AhR modulators. Estimated percentage from all the compounds analyzed in this article (*i.e.* 52 flavonoids), the ranges of tested concentrations, and the number of compounds, are shown.
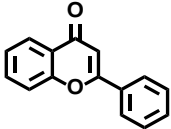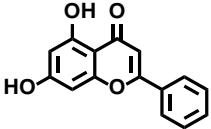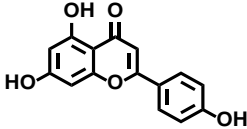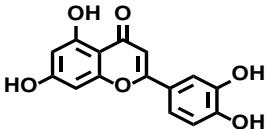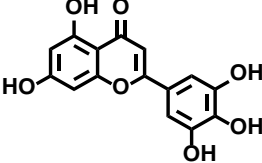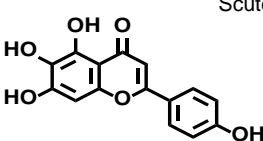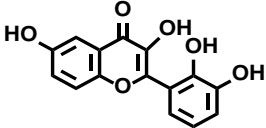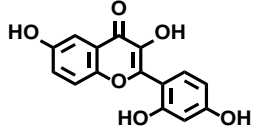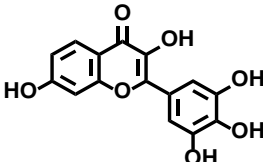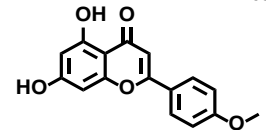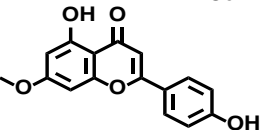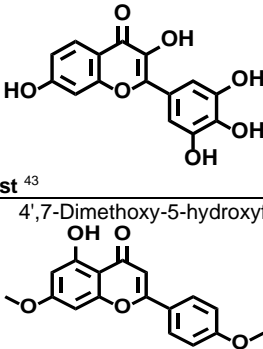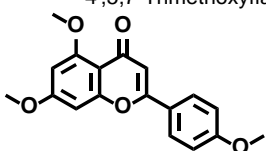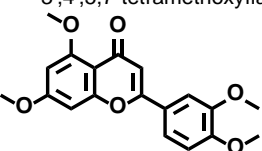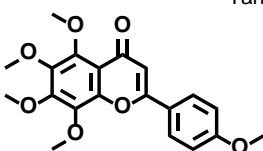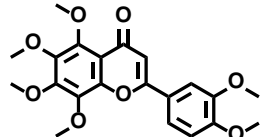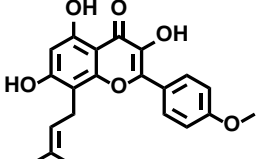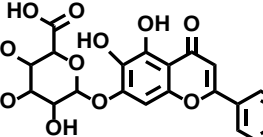
In the following sections, a structure-activity relationship analysis is addressed regarding the main subclasses of flavonoids suggested as AhR modulators.

3. SAR analysis of flavonoids as AhR ligands

3.1. Flavones
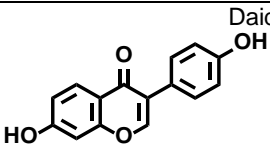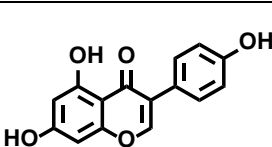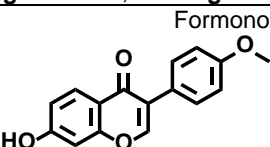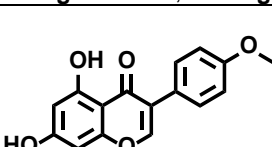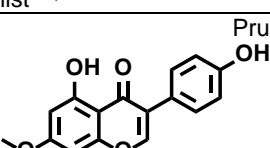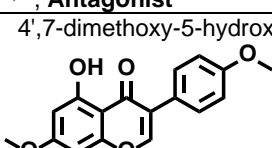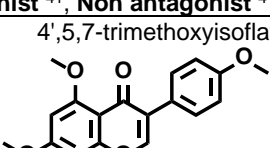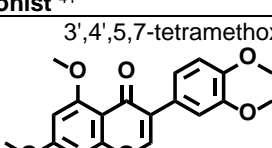
In Table 2 are shown the flavones that have been recently evaluated as AhR modulators and their suggested activity outcome.

Table 2. Structures & AhR activity reported for flavone class of compounds

Flavones

Flavone	Chrysin	Apigenin
Agonist [44]	Agonist [44,46] Non-antagonist [45]	Agonist [41–44] Non-antagonist [45]
Luteolin	Tricetin	Scutellarein
Agonist [43], Antagonist [45]	Agonist [43]	Agonist [42]
3,6,2',3'-Tetrahydroxyflavone	3,6,2',4'-Tetrahydroxyflavone	Robinetin
Agonist [43]	Agonist [43]	Agonist [43]
Acacetin	Genkwanin	4',7-Dimethoxy-5-hydroxyflavone
Antagonist [41]	Non-antagonist [41]	Non-antagonist [41]
4',5,7-Trimethoxyflavone	3',4',5,7-tetramethoxyflavone	Tangeretin
Non-antagonist [41]	Non-antagonist [41]	Antagonist [48], Non-antagonist [45]
Nobiletin	Icaritin	Baicalin
Non-antagonist [45]	Agonist [47]	Agonist [44,49], Antagonist [50] Non-antagonist [45]
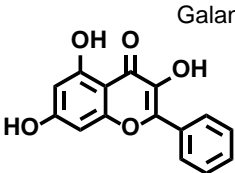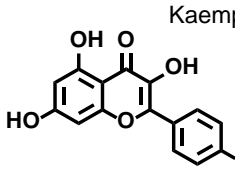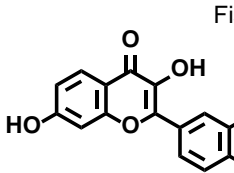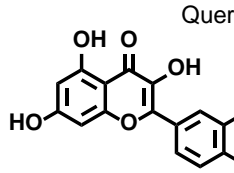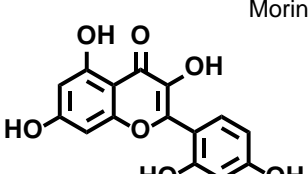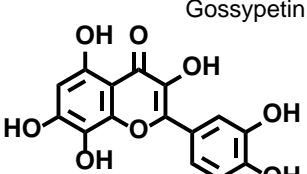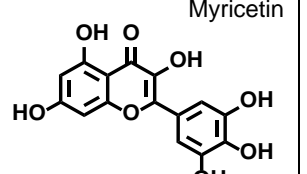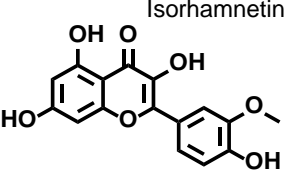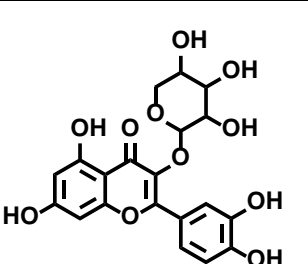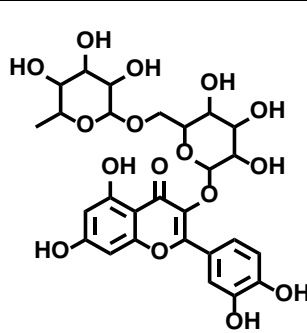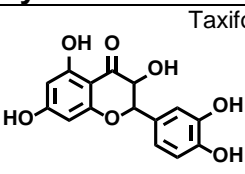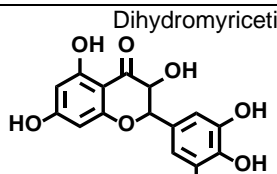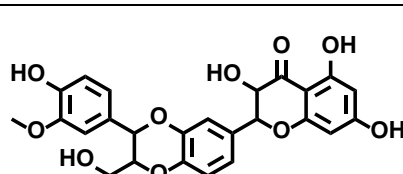
3.2. Isoflavones

Table 3. Structures & AhR activity reported for isoflavone class of compounds

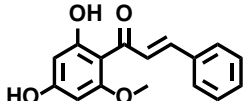| Isoflavones	
Daidzein **Antagonist** [58,60], **Non-Agonist** [59]	Genistein **Agonist** [59], **Antagonist** [58,60], **Non-agonist** [41,42]
Formononetin Agonist [58,59]	Biochanin A **Agonist** [44,58,59], **Antagonist** [41]
Prunetin **Agonist** [41], **Non antagonist** [41]	4',7-dimethoxy-5-hydroxy isoflavone **Non-antagonist** [41]
4',5,7-trimethoxyisoflavone **Non-antagonist** [41]	3',4',5,7-tetramethoxyisoflavone **Non-antagonist** [41]
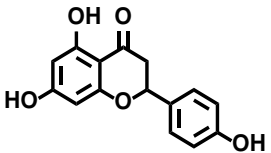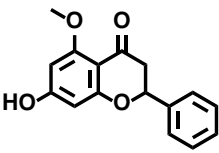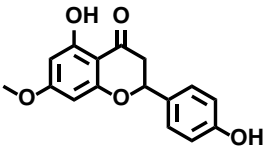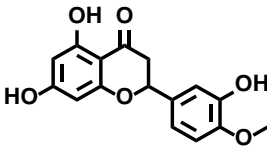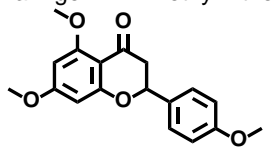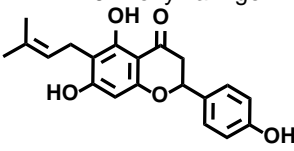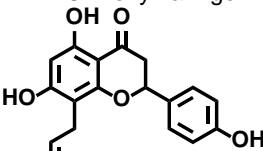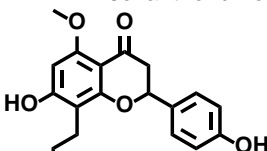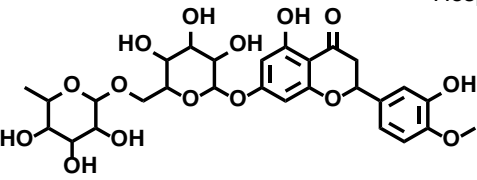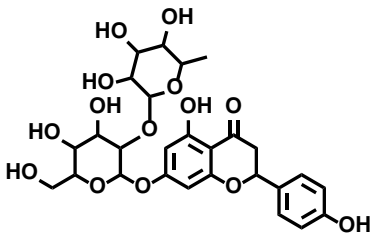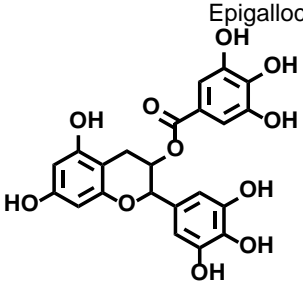
3.3. Flavonols & Dihydroflavonols

Table 4. Structures & AhR activity reported for flavonol and dihydroflavonol classes of compounds

Flavonols

Galangin	Kaempferol	Fisetin	Quercetin
Non-antagonist [45]	**Agonist** [43] **Non-antagonist** [45]	**Agonist** [43,44] **Non-antagonist** [45]	**Agonist** [43,44] **Non-antagonist** [45]
Morin	Gossypetin	Myricetin	Isorhamnetin
Agonist [43]	**Agonist** [43]	**Agonist** [43]	**Non-antagonist** [45]

Guaijaverin	Rutin
Agonist [42]	**Agonist** [44] **Non-agonist** [42]

Dihydroflavonols

Taxifolin	Dihydromyricetin	Silymarin
Agonist [43]	**Non-agonist** [51]	**Agonist** [44]

3.4. Flavanones

Table 5. Structures & AhR activity reported for flavanone class of compounds

Flavanones

Naringenin	Alpinetin	4',5-dihydroxy-7-methoxy flavanone	Hesperetin
Agonist [41], **Non-Agonist** [43], **Non-antagonist** [45]	**Agonist** [55,56]	**Antagonist** [41] **Non-Agonist** [42]	**Agonist** [42,44], **Antagonist** [53]
Naringenin Trimethyl Ether	6-Prenylnaringenin	8-Prenylnaringenin	Isoxanthohumol
Non-antagonist [41]	**Agonist** [52]	**Agonist** [52]	**Non-agonist** [52]

Hesperidin	Naringin
Agonist [42,44]	**Agonist** [44,54], **Non-agonist** [42]

Flavan-3-ol

Dihydrochalcone

Epigallocatechin gallate	Cardamonin
Antagonist [45], **Non-Agonist** [42]	**Agonist** [57]

References

1. Quideau, S., Deffieux, D., Douat-Casassus, C. & Pouységu, L. Plant Polyphenols: Chemical Properties, Biological Activities, and Synthesis. *Angew. Chemie Int. Ed.* **50**, 586–621 (2011).

2. Lattanzio, V., Kroon, P. A., Quideau, S. & Treutter, D. Plant Phenolics – Secondary Metabolites with Diverse Functions. in *Recent Advances in Polyphenol Research* 1–35 (John Wiley & Sons, Ltd, 2009). doi:10.1002/9781444302400.ch1

3. Agati, G. & Tattini, M. Multiple functional roles of flavonoids in photoprotection. *New Phytol.* **186**, 786–793 (2010).

4. Lattanzio, V. Phenolic Compounds: Introduction. in *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes* (eds. Ramawat, K. G. & Mérillon, J.-M.) 1543–1580 (Springer Berlin Heidelberg, 2013). doi:10.1007/978-3-642-22144-6_57

5. Abbas, M. *et al.* Natural polyphenols: An overview. *Int. J. Food Prop.* **20**, 1689–1699 (2017).

6. Bodoira, R. & Maestri, D. Phenolic Compounds from Nuts: Extraction, Chemical Profiles, and Bioactivity. *J. Agric. Food Chem.* **68**, 927–942 (2020).

7. Corradini, E. *et al.* Flavonoids: chemical properties and analytical methodologies of identification and quantitation in foods and plants. *Nat. Prod. Res.* **25**, 469–495 (2011).

8. Scarano, A., Chieppa, M. & Santino, A. Looking at flavonoid biodiversity in horticultural crops: A colored mine with nutritional benefits. *Plants* **7**, 1–22 (2018).

9. Durazzo, A. *et al.* Polyphenols: A concise overview on the chemistry, occurrence, and human health. *Phyther. Res.* **33**, 2221–2243 (2019).

10. Panche, A. N., Diwan, A. D. & Chandra, S. R. Flavonoids: an overview. *J. Nutr. Sci.* **5**, e47 (2016).

11. Pandey, R. P. & Sohng, J. K. Genetics of Flavonoids. in *Natural Products: Phytochemistry, Botany and Metabolism of Alkaloids, Phenolics and Terpenes* (eds. Ramawat, K. G. & Mérillon, J.-M.) 1617–1645 (Springer Berlin Heidelberg, 2013). doi:10.1007/978-3-642-22144-6_52

12. Abotaleb, M. *et al.* Flavonoids in cancer and apoptosis. *Cancers (Basel).* **11**, (2019).

13. Forkmann, G. & Heller, W. 1.26 - Biosynthesis of Flavonoids. in *Comprehensive Natural Products Chemistry* (eds. Barton, S. D., Nakanishi, K. & Meth-Cohn, O.) 713–748 (Pergamon, 1999). doi:https://doi.org/10.1016/B978-0-08-091283-7.00028-X

14. Teng, H. & Chen, L. Polyphenols and bioavailability: an update. *Crit. Rev. Food Sci. Nutr.* **59**, 2040–2051 (2019).

15. Cermak, R. *et al.* The influence of postharvest processing and storage of foodstuffs on the bioavailability of flavonoids and phenolic acids. *Mol. Nutr. Food Res.* **53**, S184–S193 (2009).

16. Mozaffarian, D. & Wu, J. H. Y. Flavonoids, Dairy Foods, and Cardiovascular and Metabolic Health: A Review of Emerging Biologic Pathways. *Circ. Res.* **122**, 369–384 (2018).

17. Scarmeas, N., Anastasiou, C. A. & Yannakoulia, M. Nutrition and prevention of cognitive impairment. *Lancet. Neurol.* **17**, 1006–1015 (2018).

18. Bousova, I. & Skalova, L. Inhibition and induction of glutathione S-transferases by flavonoids: possible pharmacological and toxicological consequences. *Drug Metab. Rev.* **44**, 267–286 (2012).

19. Cermak, R., Wein, S., Wolffram, S. & Langguth, P. Effects of the flavonol quercetin on the bioavailability of simvastatin in pigs. *Eur. J. Pharm. Sci.* **38**, 519–524 (2009).

20. de Souza Dos Santos, M. C., Goncalves, C. F. L., Vaisman, M., Ferreira, A. C. F. & de Carvalho, D. P. Impact of flavonoids on thyroid function. *Food Chem. Toxicol.* **49**, 2495–2502 (2011).

21. Navarro, V. J. *et al.* Catechins in dietary supplements and hepatotoxicity. *Dig. Dis. Sci.* **58**, 2682–2690 (2013).

22. Andres, S., Abraham, K., Appel, K. E. & Lampen, A. Risks and benefits of dietary isoflavones for cancer. *Crit. Rev. Toxicol.* **41**, 463–506 (2011).

23. Egert, S. & Rimbach, G. Which Sources of Flavonoids: Complex Diets or Dietary Supplements? *Adv. Nutr.* **2**, 8–14 (2011).

24. Tungmunnithum, D., Thongboonyou, A., Pholboon, A. & Yangsabai, A. Flavonoids and Other Phenolic Compounds from Medicinal Plants for Pharmaceutical and Medical Aspects: An

Overview. *Medicines* **5**, 93 (2018).

25. Xie, Y., Huang, S. & Su, Y. Dietary flavonols intake and risk of esophageal and gastric cancer: A meta-analysis of epidemiological studies. *Nutrients* **8**, (2016).

26. Romagnolo, D. F. & Selmin, O. I. Flavonoids and Cancer Prevention: A Review of the Evidence. *J. Nutr. Gerontol. Geriatr.* **31**, 206–238 (2012).

27. Martinez-Perez, C. *et al.* Novel flavonoids as anti-cancer agents: mechanisms of action and promise for their potential application in breast cancer. *Biochem. Soc. Trans.* **42**, 1017–1023 (2014).

28. Hooper, L. *et al.* Effects of chocolate, cocoa, and flavan-3-ols on cardiovascular health: a systematic review and meta-analysis of randomized trials. *Am. J. Clin. Nutr.* **95**, 740–751 (2012).

29. Wang, X., Ouyang, Y. Y., Liu, J. & Zhao, G. Flavonoid intake and risk of CVD: a systematic review and meta-analysis of prospective cohort studies. *Br. J. Nutr.* **111**, 1–11 (2014).

30. Rees, A., Dodd, G. F. & Spencer, J. P. E. The effects of flavonoids on cardiovascular health: A review of human intervention trials and implications for cerebrovascular function. *Nutrients* **10**, (2018).

31. Cheng, J. *et al.* Cyanidin Chloride inhibits ovariectomy-induced osteoporosis by suppressing RANKL-mediated osteoclastogenesis and associated signaling pathways. *J. Cell. Physiol.* **233**, 2502–2512 (2018).

32. Beking, K. & Vieira, A. Flavonoid intake and disability-adjusted life years due to Alzheimer's and related dementias: a population-based study involving twenty-three developed countries. *Public Health Nutr.* **13**, 1403–1409 (2010).

33. Ninfali, P., Antonini, E., Frati, A. & Scarpa, E.-S. C-Glycosyl Flavonoids from Beta vulgaris Cicla and Betalains from Beta vulgaris rubra: Antioxidant, Anticancer and Antiinflammatory Activities—A Review. *Phyther. Res.* **31**, 871–884 (2017).

34. Farzaei, M. H. *et al. Targeting inflammation by flavonoids: Novel therapeutic strategy for metabolic disorders. International Journal of Molecular Sciences* **20**, (2019).

35. Farhat, G., Drummond, S. & Al-Dujaili, E. A. S. Polyphenols and Their Role in Obesity Management: A Systematic Review of Randomized Clinical Trials. *Phyther. Res.* **31**, 1005–1018 (2017).

36. Al-Dosary, D. I., Alhomida, A. S. & Ola, M. S. Protective Effects of Dietary Flavonoids in Diabetic Induced Retinal Neurodegeneration. *Curr. Drug Targets* **18**, 1468–1476 (2017).

37. Belwal, T., Nabavi, S. F., Nabavi, S. M. & Habtemariam, S. Dietary anthocyanins and insulin resistance: When food becomes a medicine. *Nutrients* **9**, (2017).

38. Procházková, D., Boušová, I. & Wilhelmová, N. Antioxidant and prooxidant properties of flavonoids. *Fitoterapia* **82**, 513–523 (2011).

39. Chirumbolo, S. *et al.* Targeting cancer with phytochemicals via their fine tuning of the cell survival signaling pathways. *Int. J. Mol. Sci.* **19**, (2018).

40. Perez-Vizcaino, F. & Fraga, C. G. Research trends in flavonoids and health. *Arch. Biochem. Biophys.* **646**, 107–112 (2018).

41. Park, H. *et al.* Isoflavones as Ah Receptor Agonists in Colon-Derived Cell Lines: Structure-Activity Relationships. *Chem. Res. Toxicol.* **32**, 2353–2364 (2019).

42. Goya-Jorge, E., Giner, R. ., Veitía, M. S.-I., Gozalbes, R. & Barigye, S. J. Predictive modeling of aryl hydrocarbon receptor (AhR) agonism. *Chemosphere* **156**, 127068 (2020).

43. Jin, U. H. *et al.* Structure-dependent modulation of aryl hydrocarbon receptor-mediated activities by flavonoids. *Toxicol. Sci.* **164**, 205–217 (2018).

44. Kaur, M. & Badhan, R. K. S. Phytochemical mediated-modulation of the expression and transporter function of breast cancer resistance protein at the blood-brain barrier: An in-vitro study. *Brain Res.* **1654**, 9–23 (2017).

45. Ashida, H. *et al.* Luteolin suppresses TCDD-induced wasting syndrome in a cultured adipocyte model. *Pestic. Biochem. Physiol.* **120**, 14–20 (2015).

46. Ronnekleiv-Kelly, S. M. *et al.* Aryl hydrocarbon receptor-dependent apoptotic cell death

induced by the flavonoid chrysin in human colorectal cancer cells. *Cancer Lett.* **370**, 91–99 (2016).

47. Sun, F. *et al.* A novel prostate cancer therapeutic strategy using icaritin-activated arylhydrocarbon-receptor to co-target androgen receptor and its splice variants. *Carcinogenesis* **36**, 757–768 (2015).

48. Arivazhagan, L. & Subramanian, S. P. Tangeretin, a citrus flavonoid attenuates oxidative stress and protects hepatocellular architecture in rats with 7, 12 - dimethylbenz(a)anthracene induced experimental mammary carcinoma. *J. Funct. Foods* **15**, 339–353 (2015).

49. Zhu, W. *et al.* Baicalin modulates the Treg/Teff balance to alleviate uveitis by activating the aryl hydrocarbon receptor. *Biochem. Pharmacol.* **154**, 18–27 (2018).

50. Xue, Y. *et al.* Baicalin inhibits inflammation and attenuates myocardial ischaemic injury by aryl hydrocarbon receptor. *J. Pharm. Pharmacol.* **67**, 1756–1764 (2015).

51. Bostikova, Z., Moserova, M., Pavek, P., Stiborova, M. & Hodek, P. Role of dihydromyricetin in cytochrome P450-mediated metabolism and carcinogen activation. *Neuroendocrinol. Lett.* **36**, 46–52 (2015).

52. Wang, S. *et al.* Hop (Humulus lupulus L.) Extract and 6-Prenylnaringenin Induce P450 1A1 Catalyzed Estrogen 2-Hydroxylation. *Chem. Res. Toxicol.* **29**, 1142–1150 (2016).

53. Tan, Y. Q., Chiu-Leung, L. C., Lin, S. mei & Leung, L. K. The citrus flavonone hesperetin attenuates the nuclear translocation of aryl hydrocarbon receptor. *Comp. Biochem. Physiol. Part - C Toxicol. Pharmacol.* **210**, 57–64 (2018).

54. Rotimi, S. O., Adelani, I. B., Bankole, G. E. & Rotimi, O. A. Naringin enhances reverse cholesterol transport in high fat/low streptozocin induced diabetic rats. *Biomed. Pharmacother.* **101**, 430–437 (2018).

55. Miao, Y. *et al.* Alpinetin improves intestinal barrier homeostasis via regulating AhR/suv39h1/TSC2/mTORC1/autophagy pathway. *Toxicol. Appl. Pharmacol.* **384**, 1–15 (2019).

56. Lv, Q. *et al.* Alpinetin exerts anti-colitis efficacy by activating AhR, regulating miR-302/DNMT- 1/CREB signals, and therefore promoting Treg differentiation. *Cell Death Dis.* **9**, 1–25 (2018).

57. Wang, K. *et al.* Cardamonin, a natural flavone, alleviates inflammatory bowel disease by the inhibition of NLRP3 inflammasome activation via an AhR/Nrf2/NQO1 pathway. *Biochem. Pharmacol.* **155**, 494–509 (2018).

58. Dunlap, T. L. *et al.* Red Clover Aryl Hydrocarbon Receptor (AhR) and Estrogen Receptor (ER) Agonists Enhance Genotoxic Estrogen Metabolism. *Chem. Res. Toxicol.* **30**, 2084–2092 (2017).

59. Bialesova, L., Novotna, A., Macejova, D., Brtko, J. & Dvorak, Z. Agonistic effect of selected isoflavones on arylhydrocarbon receptor in a novel AZ-AhR transgenic gene reporter human cell line. *Gen. Physiol. Biophys.* **24**, 331–334 (2015).

60. Froyen, E. B. & Steinberg, F. M. Genistein decreases basal hepatic cytochrome P450 1A1 protein expression and activity in Swiss Webster mice. *Nutr. Res.* **36**, 430–439 (2016).

61. Duke, J. Dr. Duke's Phytochemical and Ethnobotanical Databases. United States Department of Agriculture. *Agricultural Research Service* (2004). Available at: http://www.ars-grin.gov/duke/.

62. Zhou, J., Xie, G. & Yan, X. Molecular Structures, Pharmacological Activities, Natural Sources and Applications. *Encyclopedia of Traditional Chinese Medicines. Vol. 1: Isolated Compounds A-C* (2011).

63. Bhagwat, S., Haytowitz, D. B. & Holden, J. M. USDA Database for the Flavonoid Content of Selected Foods, Release 3.1. *Nutrient Data Laboratory B.H.N.R.C.* (2014).