Chemosphere 256 (2020) 127068

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Predictive modeling of aryl hydrocarbon receptor (AhR) agonism

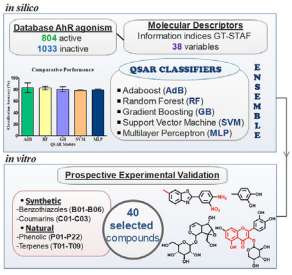
Elizabeth Goya-Jorge [a, b], Rosa M. Giner [b], Maité Sylla-Iyarreta Veitía [c], Rafael Gozalbes [a], Stephen J. Barigye [a, *]

[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

Article history:
Received 10 March 2020
Received in revised form
9 May 2020
Accepted 12 May 2020
Available online 17 May 2020

Handling Editor: Willie Peijnenburg

Keywords:
Aryl hydrocarbon receptor
Agonistic activity
QSAR
Computational modeling
Benzothiazoles
Flavonoids
Coumarins
Polyphenols
Triterpenes

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.

* Corresponding author.
E-mail address: sjbarigye@protoqsar.com (S.J. Barigye).

https://doi.org/10.1016/j.chemosphere.2020.127068
0045-6535/

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.

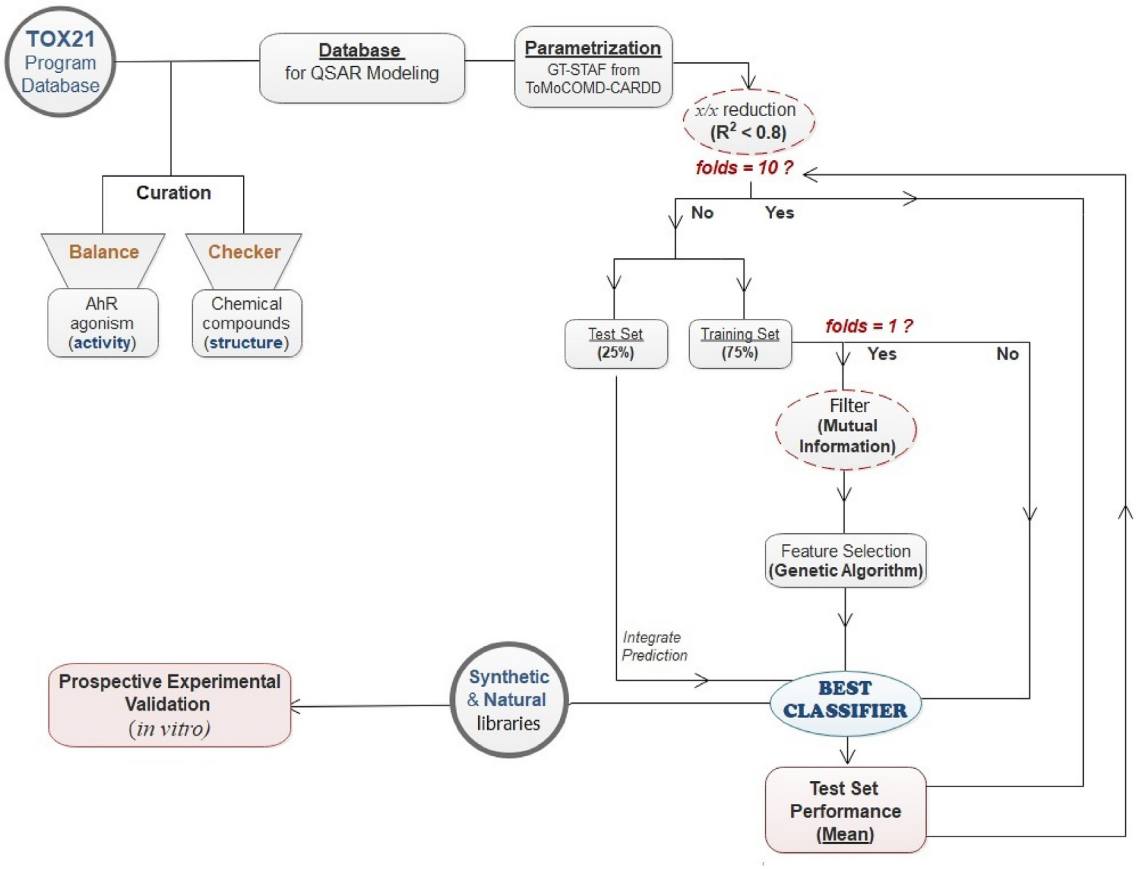


Fig. 1. Workflow employed in the development and validation of QSAR models for aryl hydrocarbon receptor (AhR) agonism.

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 *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.

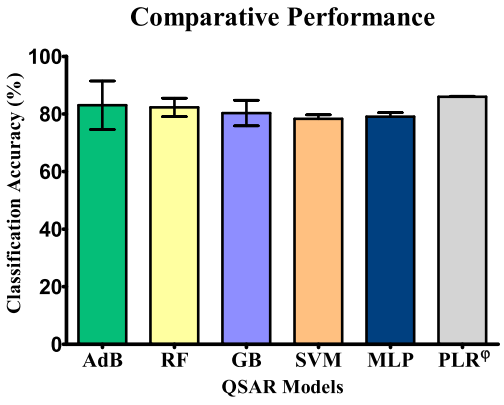


Fig. 2. Comparative Performance of the built QSAR models and a reference method, i.e. Partial Logistic Regression (PLR$^{\varphi}$) model from the literature named "QSAR 2:1" by Klimenko et al., (2019). The training performance and the 10-fold cross-validation of all QSAR models (represented as a range)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

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 $RPC_{max} \geq 10\%$ (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

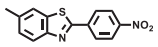
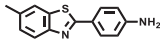
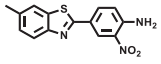
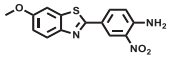
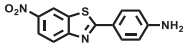
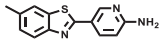
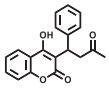
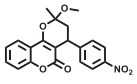
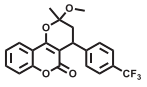
Table 2
Predictions of the QSAR models built for the AhR agonism of 40 selected chemical compounds (light blue: active, dark blue: inactive).

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]		RPC_{max} (%) ±SEM [e]	*in vitro* AhR agonism [f]	*in silico* Ensemble prediction [g]
			Min	Max			
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*)

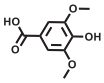
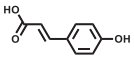
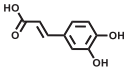
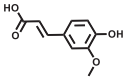
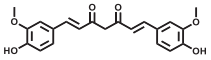
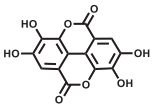
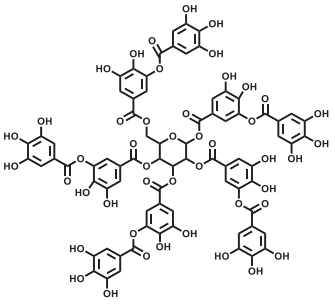
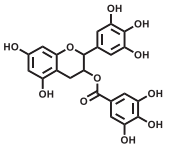
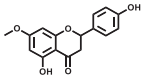
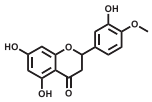
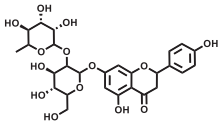
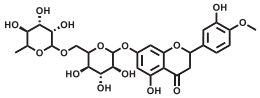
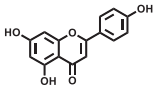
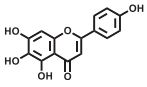
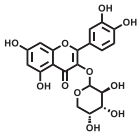
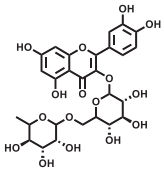
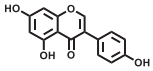
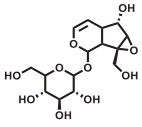
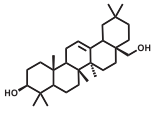
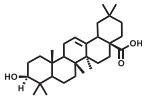
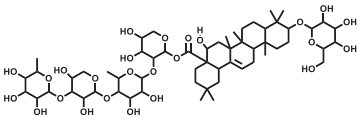
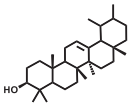
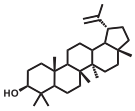
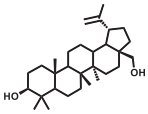
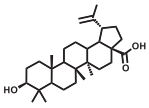
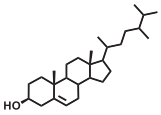
ID [a]	Origin [b]	Structure & CAS number [c]	[μM] [d]		RPC_{max} (%) ±SEM [e]	*in vitro* AhR agonism [f]	*in silico* Ensemble prediction [g]
			Min	Max			
P03	N	CAS 452-86-8 	50	100	9.09 ± 0.08	Inactive	TRUE
P04	N	CAS 2478-38-8 	50	100	7.19 ± 0.04	Inactive	TRUE
P05	N	CAS 119-36-8 	50	100	7.36 ± 0.03	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  CAS 2957-21-3	12.5	25	<5.0	Inactive	TRUE

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

(*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			
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-Iyarreta Veitía:** Supervision, Resources. **Rafael Gozalbes:** Supervision, Funding acquisition, Project administration. **Stephen J. Barigye:** Supervision, Conceptualization, Software, Writing - review & editing.

Acknowledgments

The authors acknowledge the financial support and the full funding for E. Goya-Jorge's PhD studies offered by the Innovative Training Network 'PROTECTED' (acronym of "protection against endocrine disruptors, detection, mixture, health effects, risk assessment and communication"), http://protected.eu.com/with grant agreement No. 722634 from the Marie Skłodowska-Curie actions (MSCA) of the European Union's Horizon 2020 framework. The authors thank to Prof. Dr. María Elisa Jorge Rodríguez, from the Universidad Central "Marta Abreu" de las Villas sited in Santa Clara, Cuba, who kindly provided some of the natural compounds tested *in vitro* in this work.

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.