

Assessing the chemical-induced estrogenicity using *in silico* and *in vitro* methods

Elizabeth Goya-Jorge^{a,b,1}, Mazia Amber^{c,1}, Rafael Gozalbes^{a,d}, Lisa Connolly^c,
Stephen J. Barigye^{a,d,*}

^a ProtoQSAR SL., CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia, 12 Av. Benjamin Franklin, 46980, Paterna, Valencia, Spain

^b Department of Food Science, Faculty of Veterinary Medicine-FARAH, University of Liège, 10 Av. Cureghem, 4000, Sart-Tilman, Liège, Belgium

^c The Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, BT9 5DL, Belfast, Northern Ireland, United Kingdom

^d MolDrug AI Systems SL, 45 Olimpia Arozena Torres, 46018, Valencia, Spain

ARTICLE INFO

Edited by Dr. M.D. Coleman

Keywords:

Estrogen receptor
Endocrine disruptor
Persistent organic pollutant
QSAR
Reporter gene assay
Predictive toxicology

ABSTRACT

Multiple substances are considered endocrine disrupting chemicals (EDCs). However, there is a significant gap in the early prioritization of EDC's effects. In this work, *in silico* and *in vitro* methods were used to model estrogenicity. Two Quantitative Structure-Activity Relationship (QSAR) models based on Logistic Regression and REPTree algorithms were built using a large and diverse database of estrogen receptor (ESR) agonism. A 10-fold external validation demonstrated their robustness and predictive capacity. Mechanistic interpretations of the molecular descriptors (C-026, nArOH, PW5, B06[Br-Br]) used for modelling suggested that the heteroatomic fragments, aromatic hydroxyls, and bromines, and the relative bond accessibility areas of molecules, are structural determinants in estrogenicity. As validation of the QSARs, ESR transactivity of thirteen persistent organic pollutants (POPs) and suspected EDCs was tested *in vitro* using the MMV-Luc cell line. A good correspondence between predictions and experimental bioassays demonstrated the value of the QSARs for prioritization of ESR agonist compounds.

1. Introduction

The hormone estrogen plays an important role in growth, male and female reproduction, cardiovascular health, cognition, and behaviour. Any disruption in levels of estrogen can lead to an imbalance in hormonal homeostasis and an alteration of the expression of steroid hormone receptors (Hamilton et al., 2017; WHO/UNEP, 2013).

The European Commission (EC) and the World Health Organization

(WHO) define an EDC as every substance that causes endocrine-related adverse health effects in an organism, its progeny, or subpopulations (European Commission, 2011). EDCs have been associated with several diseases and conditions in humans and metazoans and are considered a challenge for the traditional risk assessment (Futran Fuhrman et al., 2015). Three main classes of EDCs may be identified according to their chemical nature and potential applications: *i*) synthetic chemicals used in agriculture and industry/consumer products *ii*) synthetic compounds

Abbreviations: AD, applicability domain; DDT, dichlorodiphenyltrichloroethane; DMSO, dimethylsulfoxide; DMEM, Dulbecco's modified eagle medium; EDCs, endocrine disrupting chemicals; ESR, estrogen receptor; HCB, hexachlorobenzene; k-NN, k-nearest neighbours; LReg, logistic regression; LDA, linear discriminant analysis; MC, media control; MTT, 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide; MLR, multiple linear regression; ANOVA, one-way analysis of variance; OECD, Organisation for Economic Co-operation and Development; *o*-chlordane, oxy-chlordane; PBS, phosphate buffered saline; POPs, persistent organic pollutants; PC, positive control; PCB, polychlorinated biphenyl; PRC, Precision-Recall Curve; PFOA, perfluorooctanoic acid; PFDA, perfluorodecanoic acid; PFNA, perfluorononanoic acid; PFUnDA, perfluoroundecanoic acid; PLSR, partial least-squares regression; QSAR, Quantitative Structure-Activity Relationship; RGA, reporter gene assay; ROC, Receiver Operating Characteristics; RBA, relative binding affinity; SC, solvent control; SEM, standard error of the mean; SVM, support vector machines; t-nonachlor, trans-nonachlor.

* Corresponding author at: ProtoQSAR SL., CEEI (Centro Europeo de Empresas Innovadoras), Parque Tecnológico de Valencia, 12 Av. Benjamin Franklin, 46980, Paterna, Valencia, Spain.

E-mail addresses: egoya@uliege.be (E. Goya-Jorge), m.amber@qub.ac.uk (M. Amber), rgozalbes@protoqsar.com (R. Gozalbes), l.connolly@qub.ac.uk (L. Connolly), sjbarigye@gmail.com (S.J. Barigye).

¹ Equally contributed as first authors.

<https://doi.org/10.1016/j.etap.2021.103688>

Received 4 March 2021; Received in revised form 7 June 2021; Accepted 8 June 2021

Available online 10 June 2021

1382-6689/© 2021 Elsevier B.V. All rights reserved.

used in the pharmacology field, and *iii*) natural compounds or phytoestrogens that display estrogen-like activity (Gregoraszczyk and Ptak, 2013).

Persistent organic pollutants (POPs) include natural or synthetic chemical compounds resistant to degradation and persist in the environment over long periods of time, biomagnifying in the food chain, and bioaccumulating in the exposed body (European Commission, 2006). POPs are widely distributed in environment and in our daily lives, as many of them are still used in the industry. Food is considered the main source of human exposure to POPs (Guo et al., 2019; Rashid and Alshahrani, 2020). Some POPs such as dichlorodiphenyltrichloroethane (DDT), heptachlor, hexachlorobenzene (HCB), and polychlorinated biphenyl (PCB) derivatives are proven or suspected EDCs, presenting a growing global health and environmental concern (Ren et al., 2017).

Perfluorinated POPs may be involved in causing tumours and toxicity in the liver, kidneys, and major endocrine organs (Steenland et al., 2010). Meanwhile, early-life exposure to polychlorinated diphenyl ethers can affect learning and memory functions in adult animals (Eriksson et al., 2001). Chlorinated POPs can also affect the expression of ESRs. POPs such as DDT, PCBs, and their derivatives stimulate the *in vitro* proliferation of ESR-positive MCF-7 breast cancer cells (Shekhar et al., 1997). Furthermore, even low dose exposure (0.1 μ M) to PCBs during critical windows of development may lead to the downregulation of ESR β -positive cells in the anteroventral periventricular nucleus, a key brain area controlling reproductive function in mice (Salama et al., 2003).

The available experimental data on POPs acting as EDCs is still limited due to the time and the number of resources needed to conduct a detailed toxicological characterization. Furthermore, assessing the endocrine disruptive potential of all suspected individual chemicals and their mixtures using animal models is not an achievable goal due to practical and ethical reasons. Alternatives to animal experimentation including *in silico* and *in vitro* methods provide relevant and reproducible information to identify possible endocrine bioactive compounds (Goya-Jorge et al., 2020; Ruiz et al., 2017). These methodologies save time, resources and provide valuable insight into the molecular mechanisms, which are advantageous when compared to *in vivo* techniques. Furthermore, they have the regulatory endorsement of the EC and the Organisation for Economic Co-operation and Development (OECD) (OECD, 2014).

Reporter gene assays (RGA) are a commonly used tool to study gene expression of transcription factors such as ESR, androgen receptor, and aryl hydrocarbon receptor. These methods are convenient, relatively inexpensive and they give reliable measurements (Connolly et al., 2011). The associated mechanism is that agonist ligands bind to the transcription factor (e.g. ESR) to form a complex that is translocated to the nucleus. When this complex dissociates in the nucleus, the receptor dimerizes and binds to DNA on specific steroid response sequences which initiate the transcription of target genes. RGAs monitor this transcriptional process of steroid receptors using a recombinant reporter gene, allowing the endocrine disrupting potential of compounds to be assessed (Frizzell et al., 2011).

In the present work, experimental results derived from ESR-RGA were used for the *in silico* modelling the estrogenicity of chemicals, with the goal of studying suspected endocrine disruptors and POPs. Therefore, the mapped chemical space was mainly comprised of potentially toxic chemicals. Many of the analysed compounds could represent a high risk for wildlife of fish and birds (Chen and Hale, 2010; Christou et al., 2020), and their presence has been demonstrated in the blood and breast milk of the Scandinavian population (Berntsen et al., 2017).

We design herein chemoinformatic models to predict estrogenicity based on Quantitative Structure-Activity Relationship (QSAR) analysis. The predictive capacity of the built models was assessed according to OECD guidelines and experimentally validated by means of a cell-based *in vitro* RGA using a set of POPs.

2. Materials and methods

2.1. *In silico*

2.1.1. Database and workflow

A database of chemical compounds with estrogenicity measurements through different RGAs was built from literature in the context of ProtectED project (<http://protected.eu.com/>) (SI-1). The cut-off of active or inactive estrogenic activity was established following OECD guidelines for ESR stably transfected transactivation *in vitro* bioassays (OECD, 2016). The compiled substances include environmental and persistent pollutants, pesticides, industrial chemicals, pharmaceuticals and phytochemicals. Structural curation of the database was carried out using standard protocols provided by the Standardizer and StructureChecker tools from ChemAxon (<https://www.chemaxon.com>). Mixtures, inorganic substances, metal complexes and duplicate compounds were discarded. Fig. 1 represents the general workflow followed to build the *in silico* classification models of estrogenic activity.

2.1.2. Molecular descriptors and feature selection

The structures of all chemicals in the dataset were parametrized using 2030 (0–2D) molecular descriptors from DRAGON software (Todeschini et al., 2007). The calculated parameters correspond to constitutional descriptors, topological descriptors, functional group counts, atom-centred fragments, molecular properties, 2D binary and 2D frequency fingerprints. Details on the definition of all molecular descriptors are provided elsewhere (Todeschini and Consonni, 2009).

Relevant molecular descriptors for QSAR modelling were selected using the 'CfsSubsetEval' attribute evaluator and the 'BestFirst' search method as implemented in Weka software (Hall et al., 2009). Full training and cross-validation modes were applied in a supervised attribute selection.

2.1.3. QSAR modelling

k-Means and hierarchical clustering methods as implemented in STATISTICA software (Weiß, 2007) assisted in the rational division of 10 different combinations of training (80 %) and test (20 %) sets. Complete linkage was used as the linkage rule and the squared Euclidean distance as the distance measure.

QSAR classification models were built to relate the chemical structures to their estrogenic effects using the linear LReg method and the non-linear fast decision tree learner REPTree, as implemented in WEKA software (Hall et al., 2009). The applicability domain (AD) of the built models was estimated using leverage metric, which is based on the Euclidean distance to the centre of the models' descriptor space. The AD refers to chemical space for which the models' predictions may be deemed reliable and is a critical component of QSAR workflows, since built models will offer predictions even for molecules outside the respective chemical spaces. Any inferences based on such predictions must be treated with extreme caution.

Several classification metrics were considered to evaluate the performance of the trained QSAR models. The accuracy percentage indicated the proportion of correctly predicted activity, while precision and sensitivity (or recall) parameters indicated the positive prediction values and the true positive rate, respectively. The true negative rate (specificity) and the area under a Receiver Operating Characteristic (ROC) curve were also analysed. The obtained classifiers were used to screen potential endocrine disrupting substances from a chemical library of The Institute for Global Food Security from Queen's University Belfast and thirteen compounds were selected for *in vitro* experimental validation.

2.2. *In vitro* bioassays

RGA based on the MMV-Luc cell line (developed from MCF-7 human mammary tumour cells that express the endogenous ESR) was used for the evaluation of estrogenic *in vitro* activity of 13 POPs. This cell line was

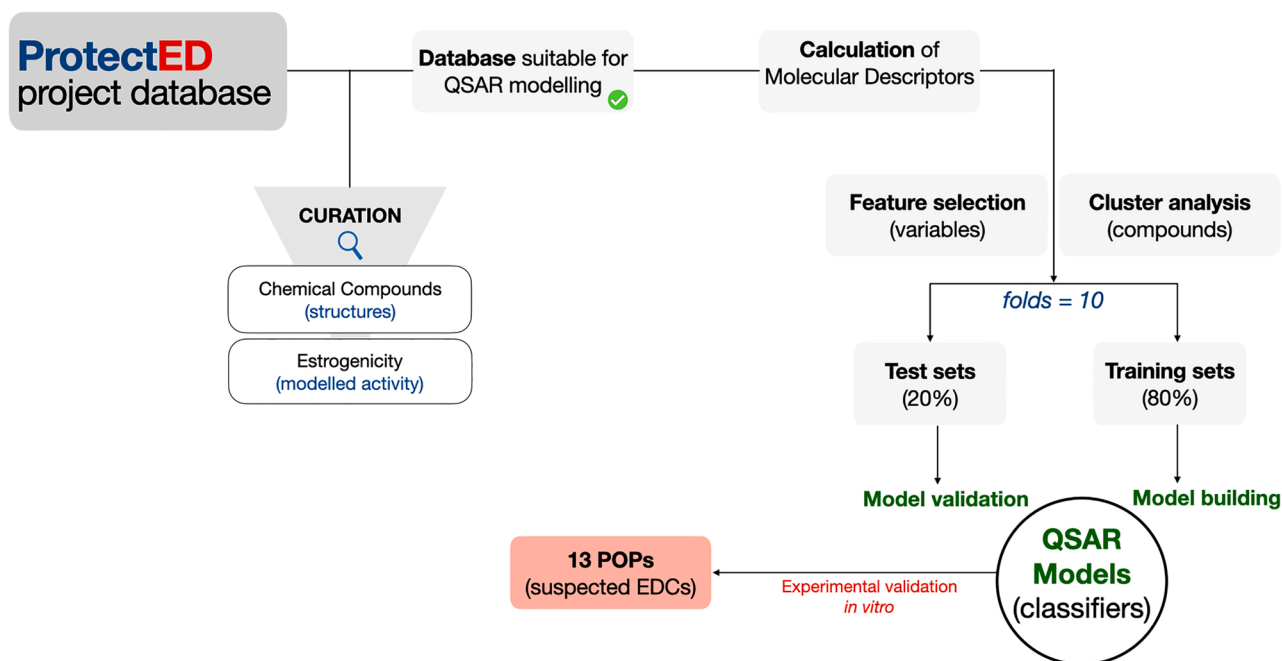


Fig. 1. Flowchart for the building and validation of QSAR classifiers of estrogenic activity.

developed and kindly provided by the Department of Food Science of the University of Liège, Belgium. Details on the development of the cell line can be found elsewhere (Willemssen et al., 2004).

2.2.1. Chemicals

A set of thirteen POPs was selected for the *in vitro* validation of the predictive capacity of QSAR models. The list of POPs and the concentration ranges tested are provided in Table 1. These compounds were obtained from Sigma Aldrich (Missouri, USA) except PCB-118 provided by Dr. Ehrenstorfer (Augsburg, Germany). All chemicals were dissolved in dimethylsulfoxide (DMSO) (Acros Organics, Molinons, France), except HCB in hexane (Merck, Massachusetts, USA), and stored at -20°C .

2.2.2. Cell culture

MMV- Luc cells were cultured in a 75 cm^2 tissue culture flasks (BD Biosciences, Bedford, MA, USA) at 37°C with 5% CO_2 and 95 % humidity. Dulbecco's Modified Eagle Medium (DMEM) without phenol red was used as cell culture medium and supplemented with 10 % foetal bovine serum, 1% penicillin/streptomycin, and L-glutamine (2 mM) to support growth. Cells were transferred two passages prior to RGA analysis into assay media, which was composed of DMEM, 10 % hormone depleted serum, 1% penicillin/streptomycin, and L-glutamine (2 mM). All RGA cell culture reagents were supplied by Invitrogen, Paisley, UK.

2.2.3. Reporter gene assays

MMV- Luc cells were seeded at a density of 40,000 cells per well, into white-walled 96 well plates with clear flat bottoms (Greiner Bio-One, Frickenhausen, Germany). Plates were incubated at 37°C and 5% CO_2 overnight prior to exposure to the tested compounds.

2.2.3.1. Estrogen receptor agonism. A dose-response curve of 17β -estradiol prepared in assay media was optimised with concentrations ranging from 0.002 nM to 4.0 nM (Fig. 2). Subsequently, cells were exposed to each of the 13 POPs at the concentration range provided in Table 1. Assay media consisting of DMEM, 10 % hormone depleted serum, 1% penicillin/streptomycin, and L-glutamine (2 mM), was used as a media control (MC). DMSO was used as a solvent control (SC) (0.2 % v:v in

assay media). Following 48 -h exposure, the supernatant was discarded, and the cells washed twice with Phosphate-Buffered Saline (PBS) prior to lysis with $1 \times$ cell lysis reagent facilitated by agitation (Promega, Southampton, UK). Finally, 100 μL luciferase (Promega, Southampton, UK) was injected into each well and luciferase activity was measured using the Mithras Multimode Reader (Berthold, Germany). The response of the cell line to the tested compounds was compared with the solvent control (SC).

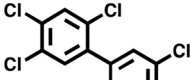
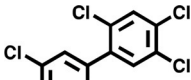
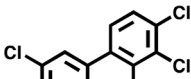
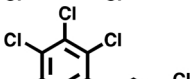
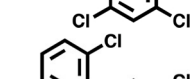
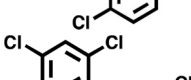
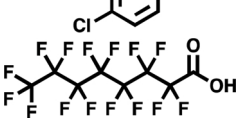



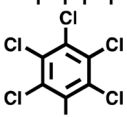
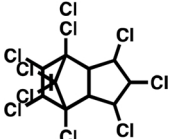
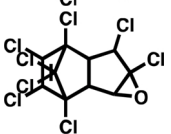
2.2.3.2. Estrogen receptor antagonism. All POPs were tested in the presence of 0.04 nM 17β -estradiol used as a positive control (PC), calculated as the EC_{50} concentration of 17β -estradiol standard curve (Fig. 2). Cells were exposed to the concentration range of POPs given in Table 1. DMSO was used as a solvent control (SC) (0.2 % v:v in assay media). Following 48 h exposure, the supernatant was discarded, and the cells washed twice with PBS, lysed with $1 \times$ cell lysis reagent facilitated by agitation (Promega, Southampton, UK). Then, 100 μL of Luciferase (Promega, Southampton, UK) was injected into each well, and luciferase activity was measured using the Mithras Multimode Reader (Berthold, Germany). The response of the cell line to the POPs was measured and compared with the PC (0.04 nM 17β -estradiol).

2.2.4. Cell viability assay

Cellular metabolism, an indicator of cell viability and cytotoxicity, was monitored by a colorimetric assay using 3-(4,5-dimethylthiazolyl-2)-2,5-diphenyltetrazolium bromide (MTT) (Sigma-Aldrich, St Louis, MO, USA). Cells were seeded at 40,000 cells per well and incubated for 24 h. Then, cells were treated at a concentration range (see Table 1) for each tested compound, and plates were incubated for 48 h. The supernatant was then discarded, and the cells washed once with PBS. Lastly, 50 μL of MTT solution (2 mg/mL stock in PBS diluted 1:6 in assay media) was added to each well and the cells were incubated for 3 h. Viable cells convert the soluble yellow MTT to insoluble purple formazan.

After incubation, the supernatant was once again removed and 200 μL /well of DMSO was added to dissolve the formazan crystals by agitation at 37°C for 10 min. Optical density was measured at 570 nm with a reference filter at 630 nm. Viability was calculated as the absorbance percentage of the sample when compared with the

Table 1
List of POPs *in vitro* evaluated: structures, identifiers, and range of tested concentration.

Structure	Common name* and CAS number	IUPAC name	Tested [μ M]	
			Min	Max
	PCB 118 31508-00-6	1,2,4-trichloro-5-(3,4-dichlorophenyl) benzene	0.01	2.50
	PCB 153 35065-27-1	2,2',4,4',5,5'-Hexachlorobiphenyl	0.01	2.50
	PCB 138 35065-28-2	1,2,3-trichloro-4-(2,4,5-trichlorophenyl) benzene	0.01	2.50
	PCB 180 35065-29-3	1,2,3,4-tetrachloro-5-(2,4,5-trichlorophenyl) benzene	0.01	2.50
	PCB 52 35693-99-3	1,4-dichloro-2-(2,5-dichlorophenyl) benzene	0.01	2.50
	PCB 101 37680-73-2	1,2,4-trichloro-5-(2,5-dichlorophenyl) benzene	0.01	2.50
	PFOA 335-67-1	2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-pentadecafluorooctanoic acid	0.50	25.00
	PFDA 335-76-2	2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,10-nonadecafluorodecanoic acid	0.50	25.00
	PFNA 375-95-1	2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9,9-heptadecafluorononanoic acid	0.50	25.00
	PFUnDA 2058-94-8	2,2,3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,11,11,11-henicosafuoroundecanoic acid	0.50	25.00
	HCB 118-74-1	1,2,3,4,5,6-hexachlorobenzene	0.08	15.00
	t-nonachlor 39765-80-5	(1R,2S,3R,5S,6R,7S)-1,3,4,5,7,8,9,10,10-nonachlorotricyclo [5.2.1.0(2,6)] dec-8-ene	0.50	25.00
	oxychlorodane 27304-13-8	1,5,6,8,9,10,11,11-octachloro-4-oxatetracyclo [6.2.1.0(2,7).0(3,5)] undec-9-ene	0.40	20.00

*PCB: polychlorinated biphenyl, PFOA: perfluorooctanoic acid, PFDA: perfluorodecanoic acid, PFNA, perfluorononanoic acid, PFUnDA, perfluoroundecanoic acid, HCB: hexachlorobenzene.

absorbance of untreated cells.

2.2.5. Statistical analysis

All *in vitro* experiments were performed in triplicate and repeated in three independent exposures. Data were analysed using GraphPad Prism v5.01 (San Diego, CA). Values are expressed as mean \pm standard error of

the mean (SEM) and in the percentage of solvent control (0.2 % DMSO) for each parameter. The comparison analysis was done using a one-way analysis of variance (ANOVA) followed by Dunnett's multiple comparisons test. Mean concentrations were tested for significant difference at 95 % confidence level. A *p* value of < 0.05 was considered statistically significant, *p* \leq 0.05 (*), \leq 0.01 (**), and \leq 0.001 (***).

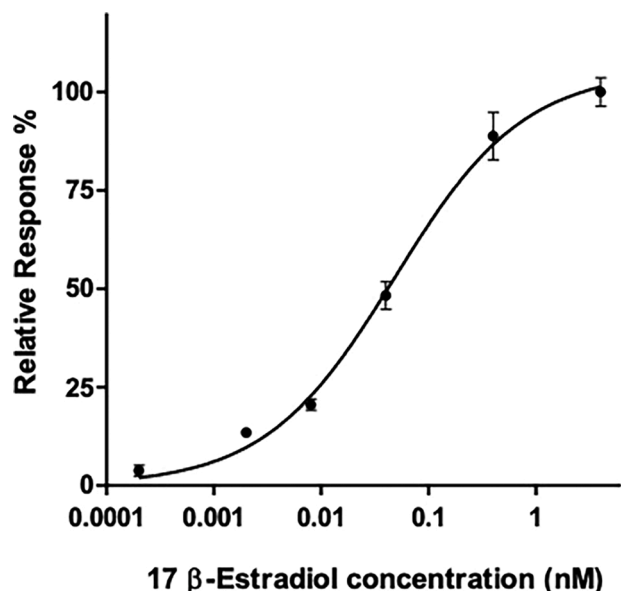


Fig. 2. Standard curve generated for 17 β -estradiol in the estrogen-responsive MMV-Luc cell line. Reporter gene transactivation was measured by luciferase activity as % response, $n = 3$. The solvent control (SC) 0.2 % DMSO was set at 0%. Maximum response (100 %) was detected at 4 nM. EC₅₀ response was observed at 0.04 nM.

3. Results

3.1. *In silico*

3.1.1. Database and molecular descriptors

The large and diverse ProtectED project database was constructed from literature containing 613 chemical compounds studied as ESR modulators in different RGAs (SI-1). After structure curation, 487 chemical compounds were retained for QSAR modelling purposes, comprising 389 inactive and 98 active estrogenic compounds. As previously mentioned, a total of 2030 descriptors were computed for all these compounds. After the removal of constant and nearly constant descriptors, as well as those with pair correlation coefficients greater than 0.9, 323 variables remained. The supervised feature selection procedure yielded four parameters (C-026, nArOH, PW5, B06[Br-Br]) to be used in the development of classification QSAR models.

3.1.2. Classification QSAR models for estrogenicity

Two classification QSAR models based on LReg and RPTree algorithms were obtained to predict the estrogenic potential of chemical compounds (see SI-2). The adopted configurations and validation metrics for these two models are presented in Table 2 and Table 3, respectively.

QSAR models were used to virtually screen an in-house library

Table 2
Configuration of the built QSAR models.

Parameter	LReg Model	RPTree Model
Configuration	Coefficients -0.51521 (± 0.01) C-026 -0.55931 (± 0.03) nArOH -19.28679 (± 0.56) PW5 +3.27769 (± 0.05) intercept	Size of the trees: 15–41 (Details in SI-2)
Division of datasets (10-fold) clustering	<ul style="list-style-type: none"> Training (80 %): $n = 390$ compounds (20 %) 78 active, (80 %) 312 inactive Test (20 %): $n = 97$ compounds (20 %) 20 active (80 %) 77 inactive 	
Molecular descriptors Database	C-026, nArOH, PW5, B06[Br-Br] $n = 487$ compounds (98 active, 389 inactive)	

Table 3

Internal cross validation and 10-folds external validation of the built QSAR models.

Parameter *	LReg Model		RPTree Model	
	c.v. Train	External Test	c.v. Train	External Test
Balanced Accuracy	71.10 %	63.02 %	79.25 %	73.79 %
Precision	0.73	0.70 \pm 0.01	0.80	0.78 \pm 0.01
Sensitivity	0.73	0.72 \pm 0.01	0.80	0.78 \pm 0.01
Specificity	0.69	0.54 \pm 0.02	0.78	0.70 \pm 0.02
F-Measure	0.72	0.71 \pm 0.01	0.80	0.78 \pm 0.01
ROC Area	0.81	0.77 \pm 0.01	0.82	0.78 \pm 0.02
PRC Area	0.81	0.80 \pm 0.01	0.79	0.79 \pm 0.01

* Means from the 10-folds of Train or Test sets (\pm standard error). ROC: Receiver Operating Characteristics, PRC: Precision-Recall Curve.

(Queen's University Belfast) of potential EDCs. Thirteen selected compounds were tested *in vitro* as prospective experimental validation of the chemoinformatic models. Eight of them were predicted as active estrogenic modulators by one or both QSAR models (PCB 118, PCB 153, PCB 180, PCB 52, PCB 101, PFOA, *o*-chlordane, *t*-nonachlor), and five of them were predicted *in silico* as inactive (HCB, PCB 138, PFDA, PFNA, PFUnDA).

3.2. *In vitro* estrogenic activity measurements

The thirteen POPs were tested *in vitro* in the ESR-RGA. A dose-response curve was generated in the MMV-Luc cell line using the relevant steroid hormone standard 17 β -estradiol (Fig. 2).

In parallel, cellular metabolism as measured by MTT was used as an indicator for cytotoxicity. None of the 13 compounds were cytotoxic at any of the tested concentrations in the MMV-Luc cell line (Fig. 3).

Agonistic effects of the 13 tested POPs were expressed as a percentage relative to the maximal response seen in the 17 β -estradiol

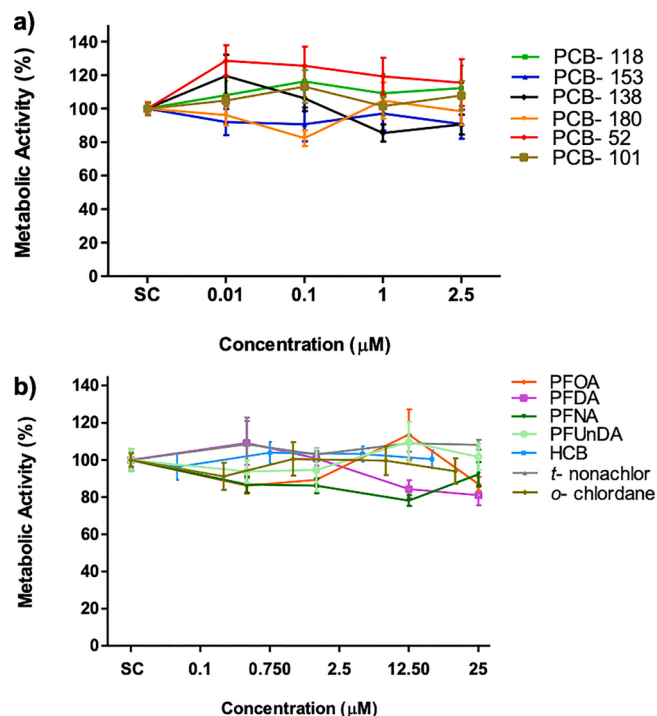


Fig. 3. Metabolic activity assessed by MTT conversion in the estrogen-responsive cell line MMV-Luc after 48 h exposure to a) PCB 118, PCB 153, PCB 138, PCB 180, PCB 52 and PCB 101, b) PFOA, PFDA, PFNA, PFUnDA, HCB, *t*-nonachlor and *o*-chlordane. Means \pm SEM % of untreated cells, $n = 3$. Significant differences from the solvent control (SC) are indicated with * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

standard curve (4 nM = 100 %). Five out of the thirteen compounds (PCB 180, PCB 101, PFOA, *o*-chlordane, *t*-nonachlor) showed statistically significant ESR transactivation at one or more tested concentrations as represented in Fig. 4 (a–d).

In the presence of 0.04 nM (*i.e.* EC₅₀) of 17β-estradiol, none of the tested compounds displayed significant synergistic nor antagonistic ESR transcriptional activity (Fig. 5).

4. Discussion

4.1. *In silico* models

The estrogenic activity models designed in this work complied with all the regulatory principles required by the OECD for the validation of QSAR models (OECD, 2016). First, the database used to model estrogenic activity was derived from a homogeneous and highly informative *in vitro* bioassay for ESR agonist modulation. Hence, the modelled

endpoint was defined as the estrogenic activity exhibited by chemicals in a RGA and transformed in a binary categorization of active or inactive estrogenic compounds (OECD, 2016). Simple and interpretable statistical algorithms (LReg and REPTree) were used to build QSAR models.

The models' AD was established based on the leverage values (SI-2) calculated from the corresponding molecular descriptors. All compounds employed in the modelling procedure and the virtual screening experiment were found to lie within the QSARs' AD.

The good statistical metrics in a strict 10-folds external validation demonstrated the predictive capacity of the classifiers. Therefore, it was verified that the QSAR models presented herein are useful to assess the estrogenic capacity of chemicals. Additionally, as detailed later, mechanistic interpretations of the molecular descriptors used for modelling are provided, as well as the *in vitro* experimental validation of the built classifiers.

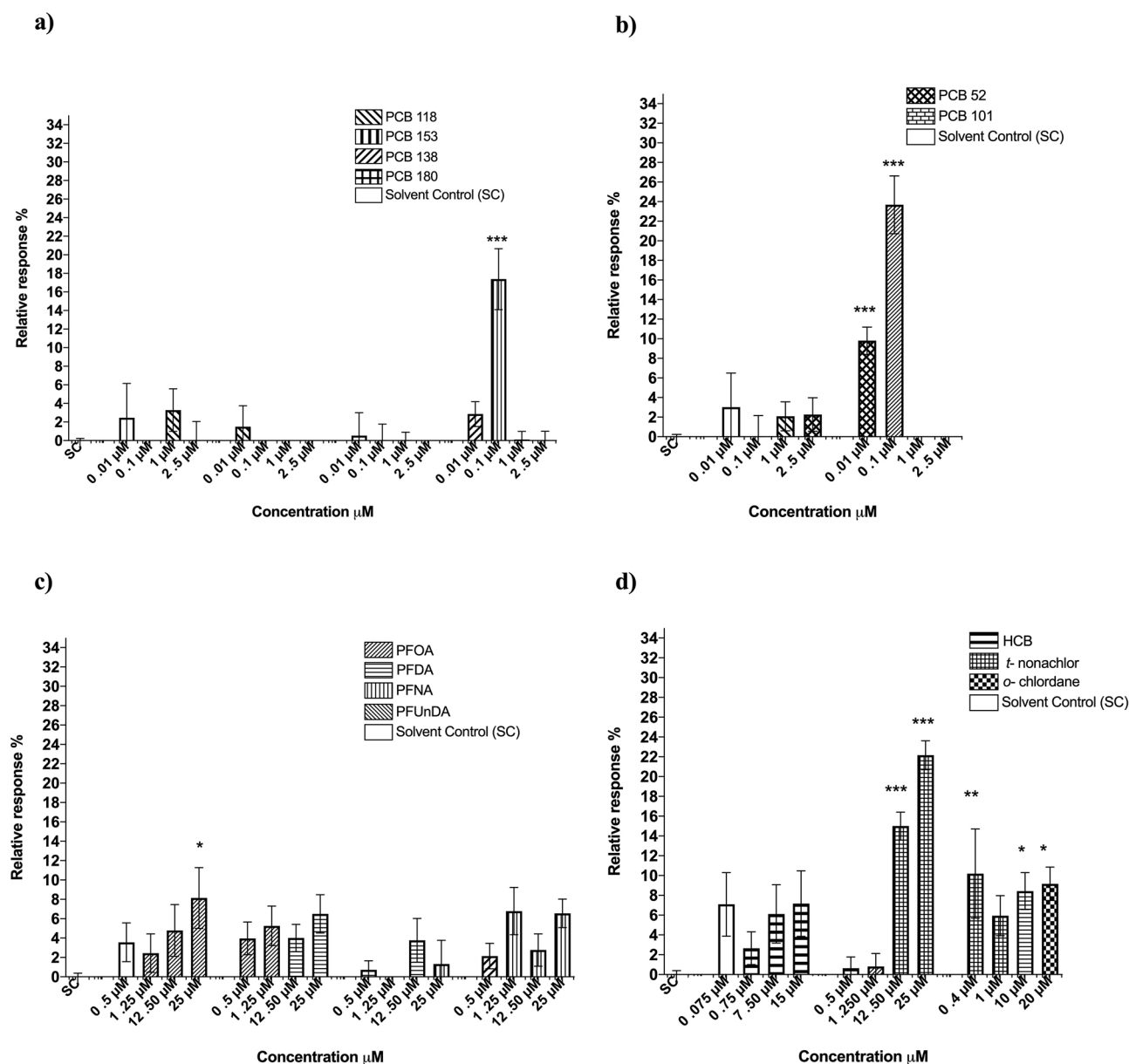


Fig. 4. The agonistic response of thirteen POPs in the estrogen-responsive cell line MMV-Luc. a) PCB 118, PCB 153, PCB 138, PCB 180; b) PCB 52 and PCB 101; c) PFOA, PFDA, PFNA and PFUnDA; d) HCB, *t*-nonachlor, *o*-chlordane. Responses are expressed as a % relative to 17β-estradiol response (4 nM = 100 %). DMSO 0.2 % was used as solvent control (SC). $n = 3$; Significant differences from the SC are indicated with * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

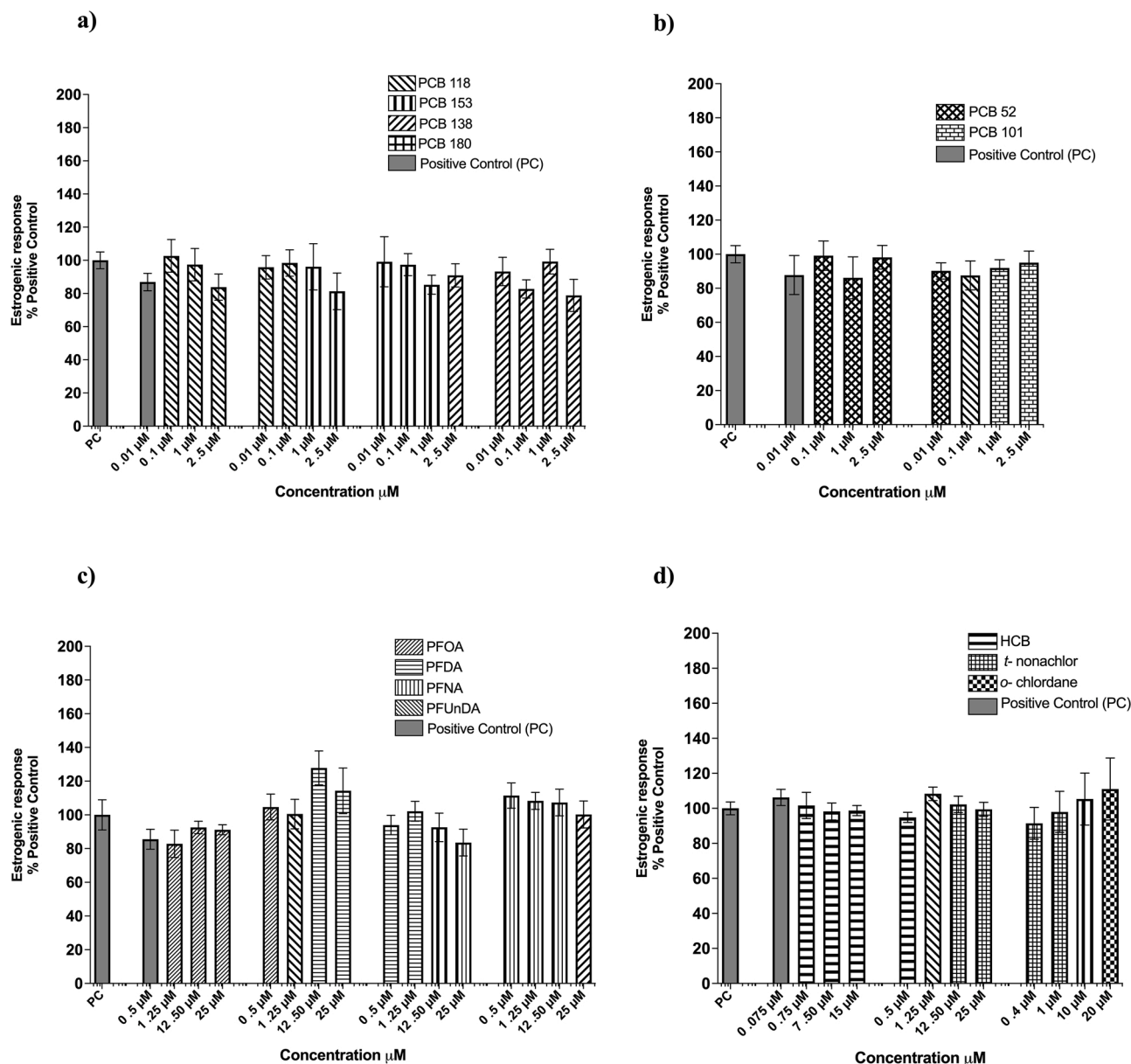


Fig. 5. The antagonist response in the estrogen-responsive cell line MMV-Luc. a) PCB 118, PCB 153, PCB 138, and PCB 180; b) PCB 52 and PCB 101; c) PFOA, PFDA, PFNA, and PFUnDA; d) HCB, *t*-nonachlor, and *o*-chlordane. All exposures were done in presence of 4 nM 17 β -estradiol as positive control (PC). Responses are expressed as % of the PC. $n = 3$; significant differences from the PC are indicated with * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$.

4.1.1. Molecular descriptors

Mechanistic interpretations of the relation between the variables (molecular descriptors) and the activity modelled, is highly recommended in chemoinformatics, as indicated by the fifth OECD validation principle for QSAR models (Meerts et al., 2001). The supervised algorithm used to select attributes in this work allowed to identify and rank the descriptors most relevant to the target variable (estrogenic activity). Furthermore, the four variables selected (C-026, nArOH, PW5, B06 [Br-Br]) are simple and interpretable parameters.

The atom-centred descriptor C-026 is defined as R-CX-R, where the - lines represent an aromatic bond that separates any group (R) from a C atom linked to an electronegative (X) atom (*i.e.* O, N, S, P, Se, halogens). Meanwhile, B06[Br-Br] descriptor codifies for the presence or absence of Br-Br at topological distance 6. Also closely related to C-026, the number of aromatic hydroxyls (nArOH) in molecules seems to be an important moiety for the ESR's modulation. Therefore, in general, heteroatoms in molecules (particularly O and Br), and the presence of aromatic bonds

are key structural features that may potentially influence the estrogenic activity of chemical compounds. In this sense, it is not surprising that diverse organobromine compounds such as brominated flame retardants and bisphenol A derivatives routinely added to consumer products (*e.g.* bottles, textile, food containers) have exhibited estrogenic activity (Randić and Basak, 1999).

On the other hand, PW5 is a topological index (Path/walk 5 - Randić shape index). The physicochemical weight of paths in this molecular descriptor is the so-called "Randić shape index" (Randić, 1995). Randić shape is a molecular shape profile defined as a sequence that represents the molecular shape and derived from the interatomic distances. All atoms at the molecular periphery contribute to the elements of the sequence (Estrada, 2002). As it has been previously highlighted, Randić index is an indication of the accessible perimeter of atoms and therefore, the relative bond accessibility areas of molecules (Saito and Rehmsmeier, 2015). Therefore, the selection of PW5 for modelling the estrogenic activity emphasizes the importance of considering the accessible areas

of molecules that could interact with ESRs.

4.1.2. QSAR models

The early prioritization of toxicant substances is crucial to guarantee human health and preserve the environment. In this work, simple and interpretable QSAR models based on LReg and REPTree were built to identify and prioritize chemical compounds as potential estrogenic disruptors.

The two chemoinformatic models of estrogenicity yielded good internal and external performance (Table 3), with balanced accuracies greater than 63 % for LReg and greater than 73 % for REPTree model. Precision, sensitivity and specificity parameters ranged from 0.54–0.73 for LReg and from 0.70–0.80 for REPTree. The F-measure of the binary classifiers was in all cases greater than 0.71. The area under the receiving operating characteristic curve (ROC) was 0.77 in the external validation of the LReg model, and 0.78 in the external validation of the REPTree model. The precision-recall curve (PRC) area, which is a reliable parameter especially in imbalanced datasets (Cherkasov et al., 2014), was close to 0.8 in the training and in the external validation of both QSARs. All these metrics statistically demonstrated the adequate robustness and the predictive capabilities of the obtained models (Li and Gramatica, 2010).

QSAR models for predicting the biological effects mediated by ESRs have been reported in the literature. The most relevant selected studies are summarized in Table 4 for comparison purposes.

In general, most QSAR models reported to study estrogenic effects have used the binding measurements to ESRs as an endpoint (as shown

Table 4
Comparative analysis of QSAR models of estrogenic activity.

QSAR Methods ^a	Database	Endpoints ^b	Accuracy Metrics	References
k-NN	245	RBA	0.62–0.77	(Asikainen et al., 2004)
BP-NN	125	RBA	0.81–0.85	(Ji et al., 2008)
MLR, BR-NN, PLSR	82	RBA (ESR α & ESR β)	ER α (0.63–0.92) ER β (0.75–0.98)	(Wang et al., 2010)
k-NN, IBk, RF	838	RBA	0.74–0.92	(Li et al., 2012a)
PLSR	36	RBA (ESR α)	0.58–0.94	(Li et al., 2012b)
CoMSIA (PLS)	44	REC ₂₀ (ESR α)	0.63–0.65	(Zang et al., 2013)
CART, LDA, SVM	1537	Binding (interaction score)	0.70–0.80	(Zhang et al., 2013)
STL & MTL (k-NN)	546	RBA (ESR α & ESR β)	0.53–0.73	(Ng et al., 2015)
	137	(ER β)		
DF	3308	RBA (ESR α)	0.70–0.89	(Zhang et al., 2017)
CART, LDA	440	Ternary classification (Agonism/Antagonism/Inactive)	0.64–0.73	(Zhang et al., 2017)
LReg REPTree	478	Agonism (ESRs)	0.63–0.74	This work

^a **QSAR methods.** k-nearest neighbours (k-NN), backpropagation neural networks (BP-NN), multiple linear regression (MLR), bayesian regularized neural network (BR-NN), partial least-squares regression (PLSR), local lazy method (IBk), RF (random forest), partial least-squares (PLS), comparative molecular similarity indices analysis (CoMSIA), classification and regression trees (CART), linear discriminant analysis (LDA), support vector machines (SVM), single-task learner (STL), multi-task learner (MTL), logistic regression (LReg), fast decision tree learner (REPTree).

^b **Endpoints.** relative binding affinity (RBA), 20 % relative effective concentration (REC₂₀, M). Estrogen Receptors (ESRs, ESR α , ESR β).

in Table 4). Such endpoints, although useful to identify the capacity of chemicals to bind and interact with the receptor's ligand binding domains, do not indicate the potential transcriptional modulation (agonism and/or antagonism) induced by chemicals. In this sense, most of these studies may not be directly compared with those obtained herein. Other authors have suggested ternary classification models classifying agonist, antagonist, or inactive estrogenic compounds (Zhang et al., 2017).

The binary classifiers of the present work are an important contribution due to the relevance of modelling ESR transactivation, different from some other endpoints used previously. Simple and interpretable statistical algorithms and variables were preferred in correspondence with the parsimony principle. In addition, good performance was obtained in the external validation for both LReg and REPTree models, and no significant differences were identified when compared with the available ternary classification models of estrogenic activity.

4.2. In vitro bioassay

The *in vitro* ESR-RGA provided rapid validation of the POPs as potential endocrine disruptors. Five out of thirteen test compounds were identified as ESR agonists. The effect responses varied for each compound and appeared independent of the exposure concentration. The test concentration used for PCBs ranged from 0.01 μ M up to 2.5 μ M. Two out of six tested PCBs showed a significant agonistic response in the MMV-Luc cell line. PCB 180 induced ESR transactivation at 0.1 μ M (Fig. 4a), while PCB 101 was active at 0.01 μ M and 0.1 μ M (Fig. 4b).

The selection of this set of POPs was based on recent studies that warned about the endocrine disruptive potential of these chemicals and their mixtures (McComb et al., 2020; Shannon et al., 2019; Yadav et al., 2021). Moreover, their demonstrated presence in water, food, blood, and breast milk, arouses great interest in their analysis (Berntsen et al., 2017; Collet et al., 2020; Doan et al., 2020; Thomsen et al., 2010). The range of concentrations tested was based on previous reports, and although higher than the estimated actual exposure, it could occur in highly exposed populations (Doan et al., 2019).

Many PCBs are known to interfere with ESR mediated responses and play a role in the initiation and progression of cancers (Rattenborg et al., 2002). One study demonstrated that PCB 180, PCB 138, and PCB 153 can alter estradiol induced expression of the key tumour suppressor gene BRCA1 (breast cancer type 1). The study used MCF-7 (BUS) (estrogen responsive) and MDA-MB-231 cell lines to study dose-response effects of POPs on BRCA1-promoter activity. In both cell lines, the estradiol-induced promoter activity was significantly reduced by PCBs at the maximum tested concentration of 10 μ M (Qiu et al., 2020). In our study, the PCB 180 and PCB 101 induced significant agonistic response in the estrogen responsive MMV-Luc cell line at much lower concentrations (Figure 4a & 4b). Both PCB 180 and PCB 101 were predicted as active estrogenic compounds by the *in silico* RepTree model (Table 5).

PFOA at 2.5 μ M induced a statistically significant agonist response as an activator of the ESR transcription (Fig. 4c). Consistent with these results, perfluorinated compounds such as PFOA and PFOS have been previously reported as activators of ESRs (Gao et al., 2013).

Chlordane derivatives such as *o*-chlordane and *t*-nonachlor are also found to interact with the ESRs. Chlordane and its derivatives were found to be agonists of the HeLa derived ESR α and ESR β at a concentration of 10 μ M. Both *o*-chlordane and *t*-nonachlor were predicted to be activators of ESR using our *in silico* model (Table 5). The experimental data indicated ESR agonistic response of *o*-chlordane and *t*-nonachlor at 0.4 μ M, 10 μ M, 20 μ M and 12.50 μ M, 25 μ M respectively (Fig. 4d).

4.3. In silico vs. In vitro

Table 5 presents results of the estrogenic activity (active/inactive) predicted by the QSAR models and the *in vitro* evaluation for the 13 POPs.

Table 5
In silico predictions and *in vitro* evaluation of the estrogenic activity of 13 POPs.

Experimental Dataset		<i>In silico</i> predictions		<i>In vitro</i> evaluation
Compounds	CAS number	LReg	RPTree	
HCB	118–74-1	✗	✗	✗
PCB 118	31508–00-6	✓	✓	✗
PCB 153	35065–27-1	✓	✗	✗
PCB 138	35065–28-2	✗	✗	✗
PCB 180	35065–29-3	✗	✓	+
PCB 52	35693–99-3	✓	✓	✗
PCB 101	37680–73-2	✓	✓	+
PFOA	335–67-1	✗	✓	+
PFDA	335–76-2	✗	✗	✗
PFNA	375–95-1	✗	✗	✗
PFUnDA	2058–94-8	✗	✗	✗
<i>o</i> -chlordane	27304–13-8	✓	✓	+
<i>t</i> -nonachlor	39765–80-5	✓	✓	+

✓ Active ✗ Inactive + Agonism

From Table 5, it is observed that *in vitro* analysis identified the five agonists (PCB 180, PCB 101, PFOA, *o*-chlordane and *t*-nonachlor) and that the RepTree model correctly predicted these same compounds as active estrogenic compounds. Three of the five *in vitro* identified estrogenic agonists were also predicted as active by the QSAR model based in LReg. However, two PCBs (PCB 118, PCB 52) were incorrectly predicted as active ESR modulators for both *in silico* models, while PCB 153 was incorrectly predicted as active by the LReg model. These incorrect predictions are likely due to the close structural similarity between PCBs and the presence of numerous active estrogenic PCBs in the *in silico* modelled compound data set.

The five agonists (PCB 180, PCB 101, PFOA, *o*-chlordane, and *t*-nonachlor) identified *in vitro* were correctly predicted as active estrogenic compounds for the RepTree model. The majority of agonists were also predicted as active by the QSAR model based in LReg. However, two PCBs (PCB 118, PCB 52) were incorrectly predicted as active ESR modulators for both *in silico* models, while PCB 153 was incorrectly predicted as active by the LReg model.

With 61.5 % and 84.6 % of classification accuracy for LReg and REPTree models, respectively, it may be inferred that both QSARs demonstrated good predictive power. Therefore, their use is highly recommended to prioritize chemicals as ESR-mediated endocrine disruptors.

5. Conclusions

In this study, *in vitro* estrogenic agonism and antagonism determinations are reported for a set of 13 POPs with demonstrated presence in water, food, blood, and breast milk. Two classification QSARs were obtained and validated to predict the agonist transactivation of estrogen receptors caused by chemical compounds. These chemoinformatic tools were built based on LReg and REPTree, which are simple and interpretable statistical methods. Internal and external validation procedures demonstrated the robustness and predictive power of the *in silico* methods. To the best of our knowledge, these are the first binary classifiers suggested from a large and diverse dataset of chemicals tested in a RGA, and with a prospective experimentally

validated performance (>60 % accuracy), for the early prioritization of chemical compounds with potential estrogenic activity. Besides, mechanistic interpretations revealed that structural characteristics of molecules such as the presence of heteroatoms, aromatic hydroxyls, and the relative bond accessibility areas are important for ESR's modulation and, consequently, for the endocrine disrupting activity.

CRedit authorship contribution statement

Elizabeth Goya-Jorge & Mazia Amber: Investigation, Conceptualization, Formal analysis, Writing - original draft. **Rafael Gozalbes & Lisa Connolly:** Supervision, Funding acquisition, Project administration, Writing - review. **Stephen J. Barigye:** Supervision, Conceptualization, Writing - review & editing. All authors approved the final version presented.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

This work was developed by the Innovative Training Network "ProtectED" (<http://protected.eu.com/>): "PROTECTION against Endocrine Disruptors; Detection, mixtures, health effects, risk assessment and communication". The project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie actions (MSCA) with grant agreement No. 722634. We would like to thank Dr. Steven Verhaegen for his careful reading and suggestions on this manuscript. Also, thanks to Laureano E. Carpio for his assistance in some graphical representations.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.etap.2021.103688>.

References

- Asikainen, A.H., Ruuskanen, J., Tuppurainen, K.A., 2004. Consensus kNN QSAR: A versatile method for predicting the estrogenic activity of organic compounds in silico. A comparative study with five estrogen receptors and a large, diverse set of ligands. *Environ. Sci. Technol.* 38, 6724–6729. <https://doi.org/10.1021/es049665h>.
- Berntsen, H.F., Berg, V., Thomsen, C., Ropstad, E., Zimmer, K.E., 2017. The design of an environmentally relevant mixture of persistent organic pollutants for use in vivo and in vitro studies. *J. Toxicol. Environ. Health Part A* 24, 1002–1016. <https://doi.org/10.1080/15287394.2017.1354439>.
- Chen, D., Hale, R.C., 2010. A global review of polybrominated diphenyl ether flame retardant contamination in birds. *Environ. Int.* 36, 800–811. <https://doi.org/10.1016/j.envint.2010.05.013>.
- 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>.
- Christou, M., Fraser, T.W.K., Berg, V., Ropstad, E., Kamstra, J.H., 2020. Calcium signaling as a possible mechanism behind increased locomotor response in zebrafish larvae exposed to a human relevant persistent organic pollutant mixture or PFOS. *Environ. Res.* 187, 109702. <https://doi.org/10.1016/j.envres.2020.109702>.
- Collet, B., van Vugt-Lussenburg, B.M.A., Swart, K., Helmus, R., Naderman, M., de Rijke, E., Eggesbø, M., Brouwer, A., van der Burg, B., 2020. Antagonistic activity towards the androgen receptor independent from natural sex hormones in human milk samples from the Norwegian HUMIS cohort. *Environ. Int.* 143, 105948. <https://doi.org/10.1016/j.envint.2020.105948>.
- Connolly, L., Ropstad, E., Verhaegen, S., 2011. In vitro bioassays for the study of endocrine-disrupting food additives and contaminants. *TRAC - Trends in Analytical Chemistry* 30, 227–238. <https://doi.org/10.1016/j.trac.2010.10.009>.
- Doan, T.Q., Berntsen, H.F., Verhaegen, S., Ropstad, E., Connolly, L., Igout, A., Muller, M., Scippo, M.L., 2019. A mixture of persistent organic pollutants relevant for human exposure inhibits the transactivation activity of the aryl hydrocarbon receptor in vitro. *Environ. Pollut.* 254. <https://doi.org/10.1016/j.envpol.2019.113098>.
- Doan, T.Q., Connolly, L., Igout, A., Nott, K., Muller, M., Scippo, M.L., 2020. In vitro profiling of the potential endocrine disrupting activities affecting steroid and aryl hydrocarbon receptors of compounds and mixtures prevalent in human drinking

- water resources. *Chemosphere* 258, 127332. <https://doi.org/10.1016/j.chemosphere.2020.127332>.
- Eriksson, P., Jakobsson, E., Fredriksson, A., 2001. Brominated flame retardants: a novel class of developmental neurotoxicants in our environment? *Environ. Health Perspect.* 109, 903–908. <https://doi.org/10.1289/ehp.01109903>.
- Estrada, E., 2002. The structural interpretation of the randić index. *Internet Electron. J. Mol. Des.* 1, 360–366.
- European Commission, 2006. Council Decision of 14 October 2004 Concerning the Conclusion, on Behalf of the European Community, of the Stockholm Convention on Persistent Organic Pollutants (2006/507/EC).
- European Commission, 2011. 4th Report on the Implementation of the “Community Strategy for Endocrine Disrupters” a Range of Substances Suspected of Interfering With the Hormone Systems of Humans and Wildlife (COM (1999) 706).
- Frizzell, C., Ndossi, D., Verhaegen, S., Dahl, E., Eriksen, G., Sorlie, M., Ropstad, E., Muller, M., Elliott, C.T., Connolly, L., 2011. Endocrine disrupting effects of zearalenone, alpha- and beta-zearalenol at the level of nuclear receptor binding and steroidogenesis. *Toxicol. Lett.* 206, 210–217. <https://doi.org/10.1016/j.toxlet.2011.07.015>.
- Putran Fuhrman, V., Tal, A., Arnon, S., 2015. Why endocrine disrupting chemicals (EDCs) challenge traditional risk assessment and how to respond. *J. Hazard. Mater.* 286, 589–611. <https://doi.org/10.1016/j.jhazmat.2014.12.012>.
- Gao, Y., Li, X., Guo, L.-H., 2013. Assessment of estrogenic activity of perfluoroalkyl acids based on ligand-induced conformation state of human estrogen receptor. *Environ. Sci. Technol.* 47, 634–641. <https://doi.org/10.1021/es304030x>.
- Goya-Jorge, E., de Julián-Ortiz, J., Gosalbes, R., 2020. Protection against endocrine disruption through quantitative structure-activity relationship modelling. *Revista de Toxicología* 37, 55–68.
- Gregoraszczyk, E.L., Ptak, A., 2013. Endocrine-disrupting chemicals: some actions of POPs on female reproduction. *Int. J. Endocrinol.* 2013, 828532 <https://doi.org/10.1155/2013/828532>.
- Guo, W., Pan, B., Sakkiah, S., Yavas, G., Ge, W., Zou, W., Tong, W., Hong, H., 2019. Persistent organic pollutants in food: contamination sources, health effects and detection methods. *Int. J. Environ. Res. Public Health* 16. <https://doi.org/10.3390/ijerph16224361>.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. *SIGKDD Explor.* 11, 10–18. <https://doi.org/10.1145/1656274.1656278>.
- Hamilton, K.J., Hewitt, S.C., Arao, Y., Korach, K.S., 2017. Estrogen hormone biology. *Curr. Top. Dev. Biol.* 125, 109–146. <https://doi.org/10.1016/bs.ctdb.2016.12.005>.
- Ji, L., Wang, X., Yang, X., Liu, S., Wang, L., 2008. Back-propagation network improved by conjugate gradient based on genetic algorithm in QSAR study on endocrine disrupting chemicals. *Chinese Sci. Bull.* 53, 33–39. <https://doi.org/10.1007/s11434-007-0484-6>.
- Li, J., Gramatica, P., 2010. QSAR classification of estrogen receptor binders and pre-screening of potential pleiotropic EDCs. *SAR QSAR Environ. Res.* 21, 657–669. <https://doi.org/10.1080/1062936X.2010.528254>.
- Li, F., Wu, H., Li, L., Li, X., Zhao, J., Peijnenburg, W.J.G.M., 2012a. Docking and QSAR study on the binding interactions between polycyclic aromatic hydrocarbons and estrogen receptor. *Ecotoxicol. Environ. Saf.* 80, 273–279. <https://doi.org/10.1016/j.ecoenv.2012.03.009>.
- Li, X., Ye, L., Xiaoxiang, Wang, Xinzhou, Wang, Liu, H., Qian, X., Zhu, Y., Yu, H., 2012b. Molecular docking, molecular dynamics simulation, and structure-based 3D-QSAR studies on estrogenic activity of hydroxylated polychlorinated biphenyls. *Sci. Total Environ.* 441, 230–238. <https://doi.org/10.1016/j.scitotenv.2012.08.072>.
- McComb, J., Mills, I.G., Berntsen, H.F., Ropstad, E., Verhaegen, S., Connolly, L., 2020. Human-based exposure levels of perfluoroalkyl acids may induce harmful effects to health by disrupting major components of androgen receptor signalling in vitro. *Expo. Health* 12, 527–538. <https://doi.org/10.1007/s12403-019-00318-8>.
- Meerts, I.A., Letcher, R.J., Hoving, S., Marsh, G., Bergman, A., Lemmen, J.G., van der Burg, B., Brouwer, A., 2001. In vitro estrogenicity of polybrominated diphenyl ethers, hydroxylated PDBEs, and polybrominated bisphenol A compounds. *Environ. Health Perspect.* 109, 399–407. <https://doi.org/10.1289/ehp.01109399>.
- Ng, H.W., Doughty, S.W., Luo, H., Ye, H., Ge, W., Tong, W., Hong, H., 2015. Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets. *Chem. Res. Toxicol.* 28, 2343–2351. <https://doi.org/10.1021/acs.chemrestox.5b00358>.
- OECD, 2014. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment. OECD. <https://doi.org/10.1787/9789264085442-en>.
- 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>.
- Qiu, Z., Qu, K., Luan, F., Liu, Y., Zhu, Y., Yuan, Y., Li, H., Zhang, H., Hai, Y., Zhao, C., 2020. Binding specificities of estrogen receptor with perfluorinated compounds: a cross species comparison. *Environ. Int.* 134, 105284 <https://doi.org/10.1016/j.envint.2019.105284>.
- Randić, M., 1995. Molecular shape profiles. *J. Chem. Inf. Comput. Sci.* 35, 373–382. <https://doi.org/10.1021/ci00025a005>.
- Randić, M., Basak, S.C., 1999. Optimal molecular descriptors based on weighted path numbers. *J. Chem. Inf. Comput. Sci.* 39, 261–266. <https://doi.org/10.1021/ci9800763>.
- Rashid, H., Alshahrani, S.S.A., 2020. Diet: a source of endocrine disruptors. *Endocr. Metab. Immune Disord. - Drug Targets.* <https://doi.org/10.2174/1871530319666191022100141>.
- Rattenborg, T., Gjermansen, I., Bonfeld-Jørgensen, E.C., 2002. Inhibition of E2-induced expression of BRCA1 by persistent organochlorines. *Breast Cancer Res.* 4, R12. <https://doi.org/10.1186/bcr461>.
- Ren, J., Xiaoping, Wang, Wang, C., Gong, P., Wang, Xiruo, Yao, T., 2017. Biomagnification of persistent organic pollutants along a high-altitude aquatic food chain in the Tibetan Plateau: processes and mechanisms. *Environ. Pollut.* 220, 636–643. <https://doi.org/10.1016/j.envpol.2016.10.019>.
- 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>.
- Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One* 10, 1–21. <https://doi.org/10.1371/journal.pone.0118432>.
- Salama, J., Chakraborty, T.R., Ng, L., Gore, A.C., 2003. Effects of polychlorinated biphenyls on estrogen receptor-beta expression in the anteroventral periventricular nucleus. *Environ. Health Perspect.* 111, 1278–1282. <https://doi.org/10.1289/ehp.6126>.
- Shannon, M., Xie, Y., Verhaegen, S., Wilson, J., Berntsen, H.F., Zimmer, K.E., Ropstad, E., Green, B.D., Connolly, L., 2019. A human relevant defined mixture of persistent organic pollutants (POPs) affects in vitro secretion of glucagon-like peptide 1 (GLP-1), but does not affect translocation of its receptor. *Toxicol. Sci.* 172, 359–367. <https://doi.org/10.1093/toxsci/kfz192>.
- Shekhar, P.V.M., Werdell, J., Basur, V.S., 1997. Environmental Estrogen Stimulation of Growth and Estrogen Receptor Function in Preneoplastic and Cancerous Human Breast Cell Lines. *JNCI: Journal of the National Cancer Institute* 89, 1774–1782. <https://doi.org/10.1093/jnci/89.23.1774>.
- Steenland, K., Fletcher, T., Savitz, D.A., 2010. Epidemiologic evidence on the health effects of perfluorooctanoic acid (PFOA). *Environ. Health Perspect.* 118, 1100–1108. <https://doi.org/10.1289/ehp.0901827>.
- Thomsen, C., Stigum, H., Frøshaug, M., Broadwell, S.L., Becher, G., Eggesbø, M., 2010. Determinants of brominated flame retardants in breast milk from a large scale Norwegian study. *Environ. Int.* 36, 68–74. <https://doi.org/10.1016/j.envint.2009.10.002>.
- Todeschini, R., Consonni, V., 2009. *Handbook of Molecular Descriptors*. Wiley-VCH, Weinheim, Germany.
- Todeschini, R., Consonni, V., Mauri, A., Pavan, M., 2007. DRAGON—Software for the Calculation of Molecular Descriptors.
- Wang, Z., Li, Y., Ai, C., Wang, Y., 2010. In silico prediction of estrogen receptor subtype binding affinity and selectivity using statistical methods and molecular docking with 2-arylnaphthalenes and 2-arylquinolines. *Int. J. Mol. Sci.* 11, 3434–3458. <https://doi.org/10.3390/ijms11093434>.
- Weiß, C.H., 2007. StatSoft, Inc., Tulsa, OK.: STATISTICA, version 8. *Asta Adv. Stat. Anal.* 91, 339–341. <https://doi.org/10.1007/s10182-007-0038-x>.
- 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.*
- Willemsen, P., Scippo, M.-L., Kausel, G., Figueroa, J., Maghuin-Rogister, G., Martial, J.A., Muller, M., 2004. Use of reporter cell lines for detection of endocrine-disrupter activity. *Anal. Bioanal. Chem.* 378, 655–663. <https://doi.org/10.1007/s00216-003-2217-2>.
- Yadav, A., Amber, M., Zosen, D., Labba, N.A., Huijberts, E.H.W., Samulin Erdem, J., Haugen, F., Berntsen, H.F., Zienoldiny, S., Paulsen, R.E., Ropstad, E., Connolly, L., Verhaegen, S., 2021. A human relevant mixture of persistent organic pollutants (POPs) and perfluorooctane sulfonic acid (PFOS) enhance nerve growth factor (NGF)-induced neurite outgrowth in PC12 cells. *Toxicol. Lett.* 338, 85–96. <https://doi.org/10.1016/j.toxlet.2020.12.007>.
- Zang, Q., Rotroff, D.M., Judson, R.S., 2013. Binary classification of a large collection of environmental chemicals from estrogen receptor assays by quantitative structure-activity relationship and machine learning methods. *J. Chem. Inf. Model.* 53, 3244–3261. <https://doi.org/10.1021/ci400527b>.
- Zhang, L., Sedykh, A., Tripathi, A., Zhu, H., Afantitis, A., Mouchlis, V.D., Melagraki, G., Rusyn, I., Tropsha, A., 2013. Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches. *Toxicol. Appl. Pharmacol.* 272, 67–76. <https://doi.org/10.1016/j.taap.2013.04.032>.
- Zhang, Q., Yan, L., Wu, Y., Ji, L., Chen, Y., Zhao, M., Dong, X., 2017. A ternary classification using machine learning methods of distinct estrogen receptor activities within a large collection of environmental chemicals. *Sci. Total Environ.* 580, 1268–1275. <https://doi.org/10.1016/j.scitotenv.2016.12.088>.