Journal of Hazardous Materials 190 (2011) 106–112

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

QSAR classification models for the prediction of endocrine disrupting activity of brominated flame retardants

Simona Kovarich, Ester Papa*, Paola Gramatica

QSAR Research Unit in Environmental Chemistry and Ecotoxicology, DBSF, University of Insubria, Via J.H. Dunant 3, 21100 Varese, Italy

ARTICLE INFO

Article history:
Received 16 November 2010
Received in revised form 1 March 2011
Accepted 2 March 2011
Available online 9 March 2011

Keywords:
QSAR
Brominated flame retardants
Endocrine disruptors
SVHC
REACH

ABSTRACT

The identification of potential endocrine disrupting (ED) chemicals is an important task for the scientific community due to their diffusion in the environment; the production and use of such compounds will be strictly regulated through the authorization process of the REACH regulation. To overcome the problem of insufficient experimental data, the quantitative structure–activity relationship (QSAR) approach is applied to predict the ED activity of new chemicals. In the present study QSAR classification models are developed, according to the OECD principles, to predict the ED potency for a class of emerging ubiquitary pollutants, viz. brominated flame retardants (BFRs). Different endpoints related to ED activity (i.e. aryl hydrocarbon receptor agonism and antagonism, estrogen receptor agonism and antagonism, androgen and progesterone receptor antagonism, T4-TTR competition, E2SULT inhibition) are modeled using the *k*-NN classification method. The best models are selected by maximizing the sensitivity and external predictive ability. We propose simple QSARs (based on few descriptors) characterized by internal stability, good predictive power and with a verified applicability domain. These models are simple tools that are applicable to screen BFRs in relation to their ED activity, and also to design safer alternatives, in agreement with the requirements of REACH regulation at the authorization step.

1. Introduction

Increasing concern is being shown by the scientific community, regulators and the public about endocrine-disrupting chemicals (EDCs) that, in the environment, are adversely affecting human and wildlife health. There are different mechanisms through which these chemicals can exert their effects on the endocrine system: (i) agonistic effect by binding to the cellular receptor of a hormone, activating normal cell response at the wrong time or to an excessive extent; (ii) antagonistic effect by binding to the receptor, preventing natural hormonal binding and activation of the receptor; (iii) alteration of hormonal blood levels by binding to hormone transport proteins; (iv) interference with metabolic processes by affecting the synthesis, or elimination rate, of hormones. All these can lead to alterations in the maintenance of homeostasis, and in the reproduction, development and behaviour of the organism [1]. In the EU REACH regulation [2], endocrine disrupting chemicals are included in Title VII (Article 57-f), which deals with the authorization of substances of very high concern (SVHC).

Among the suspected EDCs, brominated flame retardants (BFRs) are an emerging class of ubiquitary pollutants that can act as endocrine disrupters.

BFRs are industrial products incorporated into combustible materials, such as plastics, wood and textiles, to increase their fire resistance. Brominated flame retardants include a structurally heterogeneous group of chemicals, and, among these, the most commercialized are polybrominated diphenyl ethers (PBDEs), tetrabromobisphenol A (TBBPA) and hexabromocyclododecane (HBCD). The wide dispersion of BFRs in the environment, their high lipophilicity, persistence and bioaccumulation potential, has led

Abbreviations: ED, endocrine disrupting; REACH, Registration, Evaluation, Authorization and Restriction of Chemicals; QSAR, quantitative structure–activity relationship; OECD, Organization for Economic Cooperation and Development; BFRs, brominated flame retardants; T4-TTR, thyroxin-transthyretin; E2SULT, estradiolsulfotransferase; *k*-NN, *k*-nearest neighbor; SVHC, substances of very high concern; EDC, endocrine disrupting chemicals; PBDEs, polybrominated diphenyl ethers; TBBPA, tetrabromobisphenol A; HBCD, hexabromocyclododecane; AhR, Aryl hydrocarbon receptor; RBA, AhR relative binding affinity; OH-PBDE, hydroxylated PBDE; 246-TBP, 2,4,6-tribromophenol; TBBPA-DBPE, tetrabromobisphenol-A-bis(2,3)dibromopropyl ether; DR_{ag}, AhR agonism; DR_{ant}, AhR antagonism; ER_{ag}, estrogen receptor agonism; ER_{ant}, estrogen receptor antagonism; AR_{ant}, androgen receptor antagonism; PR_{ant}, progesterone receptor antagonism; $T4\text{-}TTR_{comp}$, T4-TTR competing potency; $E2SULT_{inh}$, E2SULT inhibiting potency; CH_3O-PBDE, methoxylated PBDE; DBDE, decabromodiphenyl ethane; EBTPI, ethylene bistetrabromo phthalimide; TBE, 1,2-bis(2,4,6-tribromophenoxy) ethane; NER, non error rate; Sn, sensitivity; Sp, specificity; TP, true positive; TN, true negative; FP, false positive; FN, false negative; NER_{EXT}, external non error rate; AD, applicability domain; TSET, training set; PSET, prediction set; PBP, pentabromophenol; TCDD, 2,3,7,8-tetrachloro-dibenzo-p-dioxin; PCB, polychlorinated biphenyls; E2, estradiol; DHT, dihydrotestosterone; MPA, medroxyprogesterone acetate; MLR, multilinear regression.

* Corresponding author.
E-mail address: ester.papa@uninsubria.it (E. Papa).

0304-3894/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2011.03.008

to increasing concentrations in wildlife and humans [3–6]. Thus, a better understanding of the risk represented by these emerging pollutants is required.

Experimental evidence shows that BFRs are endocrine-active compounds with the potential to interfere with thyroid hormone homeostasis, as well as to interact with steroid receptors (e.g. estrogens, androgens) and aryl hydrocarbon receptors (dioxin-like-activity) [7–12].

Parallel with experimental studies, *in silico* strategies like QSARs (quantitative structure–activity relationships) represent an important tool to fill the gap of information on BFRs. In fact, QSAR models, recommended for use under REACH regulation, can be applied to predict lacking experimental data and to screen and prioritize chemicals, thus reducing costs and the number of tested animals. Furthermore, QSAR approaches can be successfully applied in procedures of "safe Chemical Design" as in the Drug Design process. In fact, safe molecule design is the earliest phase in the long process of placement of new safe substances onto the market. To date, several QSARs and 3D-QSARs predicting ED potency of BFRs have been published, most of them being regression models (linear and non-linear) for AhR relative binding affinity (RBA), anti-androgenic and anti-estrogenic activity [13–19].

Furthermore, the development and application of *in silico* approaches is being financially supported by the European Commission, through the 7th Framework Programme for Research, in order to predict lacking experimental data as well as to perform risk assessment of four classes of compounds of interest, including, among others, BFRs (CADASTER FP 7 PROJECT [20]). In this context, the present study has developed, according to the OECD principles [21], classification QSARs for different endpoints related to brominated flame retardant ED activity. The models were built on small and heterogeneous data sets, and were applied to predict the activity of 243 BFRs, including three alternatives to BFRs, listed in the EU-regulations, for which no experimental data are yet available.

2. Materials and methods

2.1. Data sets and classes

The experimental data sets, obtained from two studies of Hamers and co-workers [22,23], include a heterogeneous group of 29 brominated flame retardants, in particular some PBDEs and hydroxy-BDE congeners (OH-PBDEs), TBBPA, 2,4,6-tribromophenol (246-TBP), HBCDγ, and tetrabromobisphenol A-bis(2,3)dibromopropyl ether (TBBPA-DBPE). The modeled endpoints are Aryl hydrocarbon (dioxin) Receptor agonism (DR_{ag}) and antagonism (DR_{ant}), Estrogen Receptor agonism (ER_{ag}) and antagonism (ER_{ant}), Androgen Receptor antagonism (AR_{ant}), Progesterone Receptor antagonism (PR_{ant}), T4-TTR Competing Potency ($T4\text{-}TTR_{comp}$) and E2SULT Inhibiting Potency ($E2SULT_{inh}$).

The homogeneous data sets used in our study are the result of an extended literature search specifically focused on ED properties of PBDEs and BFRs. Taking into account the complexity of the endpoints considered in this study, the decision to use only experimental data measured by one research group was made in order to guarantee a better quality and homogeneity of the input data, which were used for the development of our QSARs. In fact it is known that, mainly in case of small data sets, the use of heterogeneous experimental data from different sources and laboratories can affect the quality of QSAR models, by increasing the noise in the modeled response.

The definition of the classes of activity was based on the classification criteria proposed by Hamers and collaborators [22]. Due to the limited amount of data available for the levels of potency, from low to very high, suggested in literature [22], only binary classification models could be developed for the endpoints DR_{ag}, DR_{ant}, ER_{ag}, ER_{ant}, AR_{ant} and PR_{ant}, whose experimental data were available for 24 compounds (Class 1 = inactive (no ED potency) and Class 2 = active (any evidence of ED potency)). Three classes of ED potency were modeled for the endpoints $T4\text{-}TTR_{comp}$ and $E2SULT_{inh}$, for which a higher number of experimental data ($n_{obj} = 29$) were available (Class 1 = inactive (no ED potency), Class 2 = moderately active (low/moderate ED potency) and Class 3 = very active (high/very high ED potency)) (Table 1).

The developed models were then applied to predict the unknown ED potency for the remaining 209 PBDE congeners, several PBDE metabolites (OH-PBDEs and CH_3O-PBDEs), brominated phenols, brominated bisphenol A compounds (TBBPA analogs) and other BFRs on the market, including three alternative compounds to decaBDE, already listed in other regulations (i.e. decabromodiphenyl ethane – DBDE; ethylene bistetrabromo phthalimide – EBTPI; 1,2-bis(2,4,6-tribromophenoxy) ethane – TBE) [19]. The predicted classes of ED potency for all the BFRs considered in this study are available as Supplementary Data (Table S1).

2.2. Calculation of molecular descriptors

The chemical structures of BFRs were drawn using the Semi-empirical method AM1 in the HYPERCHEM program (ver. 7.03 for Windows, 2002) and were used as input files for descriptor calculations. The molecular descriptors, which lead to information on the mono-, bi- and tri-dimensional structure of the chemicals, were computed by the software DRAGON [24]. In a preliminary step, constant or near-constant values and descriptors with a high pair-wise correlation were excluded to reduce redundant and non-useful information. At the end of this procedure a final set of 701 descriptors was used as input variables in the model development.

2.3. QSAR modeling

Classification models quantify the relationship between one or more independent variables (the molecular descriptors) and a qualitative response variable, each representing the class of the corresponding sample (here the classes of ED potency). The classification model predicts the assignment of new compounds, for which the class is unknown, to one of the *a priori* defined classes. The *k*-nearest neighbor (*k*-NN) method was applied to predict the classes of ED potency. This classification method, based on the similarity of objects (chemicals), searches for the *k* nearest neighbors of each object in the data set. The assignment of a compound to a class is based on the class of the *k* most similar compounds, where similarity is defined by calculating the Euclidean distances between the descriptor vectors. The *k*-NN method was then applied to autoscaled data and the *a priori* probability of belonging to a class was set as proportional to the number of chemicals in the *a priori* classes of ED potency. The predictive power of the model was checked for *k* values between 1 and 10.

Due to the small dimensions of the training sets, we decided to take into account only models based on a maximum of two descriptors. Thus, all the mono- and bi-dimensional models from the 701 calculated molecular descriptors (all the possible combinations by the *All Subset Models* selection method, using in-house software) were explored by maximizing the overall percentage of correct assignments (percentage of non error rate – NER%) and the population of the best 100 models was analysed for each modeled endpoint. To compare the performances of the *k*-NN models selected in the population, NER% was also calculated separately for each class of activity [25].

Moreover, parameter sensitivity (Sn) and specificity (Sp) were calculated for the endpoints DR_{ag}, DR_{ant}, ER_{ag}, ER_{ant}, AR_{ant} and PR_{ant}