Food and Chemical Toxicology xxx (2017) 1–9

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

Development of human biotransformation QSARs and application for PBT assessment refinement

Ester Papa [a, *], Alessandro Sangion [a], Jon A. Arnot [b, c, d], Paola Gramatica [a]

[a] QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Theoretical and Applied Sciences, University of Insubria, Varese Italy
[b] ARC Arnot Research & Consulting, Toronto, ON Canada
[c] Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, ON Canada
[d] Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON Canada

ARTICLE INFO

Article history:
Received 24 November 2016
Received in revised form
3 April 2017
Accepted 10 April 2017
Available online xxx

Keywords:
In vivo biotransformation
Biotransformation half-life
QSAR
Hazard assessment
Refined PBT assessment

ABSTRACT

Toxicokinetics heavily influence chemical toxicity as the result of Absorption, Distribution, Metabolism (Biotransformation) and Elimination (ADME) processes. Biotransformation (metabolism) reactions can lead to detoxification or, in some cases, bioactivation of parent compounds to more toxic chemicals. Moreover, biotransformation has been recognized as a key process determining chemical half-life in an organism and is thus a key determinant for bioaccumulation assessment for many chemicals. This study addresses the development of QSAR models for the prediction of *in vivo* whole body human biotransformation (metabolism) half-lives measured or empirically-derived for over 1000 chemicals, mainly represented by pharmaceuticals. Models presented in this study meet regulatory standards for fitting, validation and applicability domain. These QSARs were used, in combination with literature models for the prediction of biotransformation half-lives in fish, to refine the screening of the potential PBT behaviour of over 1300 Pharmaceuticals and Personal Care Products (PPCPs). The refinement of the PBT screening allowed, among others, for the identification of PPCPs, which were predicted as PBTs on the basis of their chemical structure, but may be easily biotransformed. These compounds are of lower concern in comparison to potential PBTs characterized by large predicted biotransformation half-lives.

1. Introduction

The last few decades have seen an increasing interest of the scientific and regulatory communities to define correct procedures for the assessment of hazard and risks associated to chemical substances. Biotransformation reactions across species influence the fate, exposure and toxicity of substances present in the environment. The effect of biotransformation on the toxicological behaviour of chemicals has been largely investigated and described (Lech and Bend, 1980; Sijm et al., 2007; Walker et al., 2012). In organisms such as fish and mammals two phases of metabolism (i.e. Phase I and Phase II), which involve different enzymes and reactions, transform parent compounds into more polar metabolites which are more easily eliminated through the urinary system, by increasing their hydrophilicity. However, it is possible that Phase I reactions cause bio-activation of parent compounds which are transformed into toxic metabolites, such as in the case of benzo[a] pyrene and organophosphorous insecticides (Walker et al., 2012). Therefore, Phase I and Phase II reactions are clear examples of how the biotransformation processes can influence toxicity through the

Abbreviations: 1-CoPK, one-compartment pharmacokinetic model; AD, Applicability Domain; ADME, Absorption, Distribution, Metabolism and Elimination; B, Bioaccumulation; BCF, Bioconcentration Factor; CAS RN, Chemical Abstract Service Registry Number; CCC, Concordance Correlation Coefficient; CLINT, intrinsic clearance; E-State, Electrotopological-State; GA, Genetic Algorithm; h, hour; hi/i, leverage value; HL_B, whole-body primary biotransformation half-life; HL, half-life; HL_T, whole body total elimination half-life; I-State, Intrinsic-State; k_B, whole body primary biotransformation rate constant; K_m, Michaelis constant; k_T, whole body total elimination rate constant; k_x, sum of rate constants for passive elimination; MLR, Multiple Linear Regression; OLS, Ordinary Least Square method; PAH, polycyclic aromatic hydrocarbon; PBT, Persistent Bioaccumulative and Toxic; PC, Principal Component; PCA, Principal Component Analysis; PCB, polychlorinated biphenyl; PCDD, polychlorinated dibenzodioxin; PCDF, polychlorinated dibenzofuran; PPCP, Pharmaceutical and Personal Care Product; PT, Persistent and Toxic; Q^2_{LMO}, Q^2 Leave Many Out; Q^2_{LOO}, Q^2 leave-one-out; QSAR, Quantitative Structure-Activity Relationships; QUIK, Q Under the Influence of K rule; R^2_{Yscr}, R^2 Y-scrambling; RMSE, Residual Mean Squared Errors; RMSEP, Residual Mean Squared Errors Prediction; RMSET, Residual Mean Squared Errors Training; SMILES, Simplified Molecular Input Line Entry System; US-EPA, United State Environmental Protection Agency; Vm, maximum velocity of the reaction.

* Corresponding author. QSAR Research Unit in Environmental Chemistry and Ecotoxicology, Department of Theoretical and Applied Sciences, University of Insubria, Via J.H. Dunant 3, 21100 Varese, Italy.
E-mail address: ester.papa@uninsubria.it (E. Papa).

http://dx.doi.org/10.1016/j.fct.2017.04.016
0278-6915/

generation of more hydrophilic and better excreted compounds (i.e. detoxification) or, at the opposite, the bio-activation of the parent compound.

In addition, biotransformation has been recognized as a key element which may influence bioaccumulation (Burkhard et al., 2012) i.e. the process by which chemicals are taken up by an organism including diet and all the possible routes of exposure. The identification of compounds characterized by long persistence in the organisms (i.e. slow biotransformation) is particularly important for the assessment of secondary poisoning due to exposure through the diet. Moreover, biotransformation rates are required to refine model calculated bioaccumulation hazard assessment metrics such as the Bioconcentration Factor (BCF) (Lech and Bend, 1980; Cowan-Ellsberry et al., 2008) and the total elimination half-life (HL_T) (Goss et al., 2013).

Furthermore, the reliable quantification of bioaccumulation is of particular interest in the current scientific and regulatory context since the assessment of substances potentially Persistent, Bioaccumulative and Toxic (PBT) is a main issue and a required procedure under several regulatory frameworks (Government Of Canada 1999; European Commission, 2006; Cowan-Ellsberry et al., 2008; UNEP 2008; European Chemicals Agency (ECHA) 2014; Lillicrap et al., 2016). Due to the high costs associated to experimental determination of bioaccumulation related parameters, the use of *in silico* models based on Quantitative Structure-Activity Relationships (QSAR) is now largely accepted in bioaccumulation science as a cost effective solution to be applied for the support and integration of hazard/risk assessment procedures in the absence of experimental data (e.g. European Commission, 2006; European Chemicals Agency, 2008; European Commission, 2009; European Commission, 2012).

Several tools are currently available to predict metabolic pathways and metabolites on the basis of the chemical structure of a compound, such as PASS (Borodina et al., 2003; Wilk-Zasadna et al., 2015), CATABOL and TIMES (Dimitrov et al., 2011a, 2011b, 2012; Karabunarliev et al., 2012; Mekenyan et al., 2012), METEOR™ (Marchant et al., 2008), META™ (MultiCASE, 2016), CORAL (Toropova and Toropov, 2014, 2017).

Other QSARs (e.g Long and Walker, 2003; Pirovano et al., 2015, 2016) predict kinetic parameters (i.e. intrinsic clearance (CLINT), Michaelis constant (Km) and maximum velocity of the reaction (Vm)) for different enzymatic reactions.

Furthermore, several QSAR models based on different types of molecular descriptors (e.g. molecular fragments, topological descriptors and Abraham descriptors) have been developed to predict the *in vivo* biotransformation constant (k_B) in fish from chemical structure (Arnot et al., 2009; Brown et al., 2012; Kuo and Di Toro, 2013; Arnot et al., 2014; Papa et al., 2014).

This study addresses the development of QSAR models for the prediction of total whole body elimination half-lives (HL_T) and whole body biotransformation half-lives (HL_B), calculated for over 1000 chemicals (mainly represented by pharmaceuticals as well as by traditional contaminants) on the basis of five datasets, recently compiled by Arnot and colleagues (Arnot et al., 2014). The first aim of this work is the creation of statistically valid and predictive models, which meet regulatory standards, for the prediction of half-lives in human. The second aim is to show how QSAR predictions can be used for the refinement of chemical screening procedures for hazard assessment. In particular, an example is given on how predictions generated by several biological HL models can be used for the refinement of results from preliminary screening level assessments (i.e. the Persistent Bioacumulative and Toxic (PBT) behaviour of Pharmaceuticals and Personal Care Products (PPCPs), which were screened in former studies (Cassani and Gramatica, 2015; Sangion and Gramatica, 2016a)).

2. Material and methods

2.1. Biotransformation datasets

One dataset for the total elimination half-life (HL_T, hours) and four datasets for the primary biotransformation (metabolism) half-life (HL_B, hours) in humans were taken from literature (Arnot et al., 2014). The HL_T dataset was composed of 1105 heterogeneous organic compounds (composed of pharmaceuticals (80%) and traditional environmental contaminants (20%)) and consisted of measured and estimated adult total elimination half-lives (HLs) in human. All these data were derived from peer-reviewed sources and had been reported with quality assessment methods. The HL_S ranged between a minimum of 0.05 h for nitroglycerin and a maximum of 2×10^6 h for 2,3,4,5,2′,3′,5′,6'-octachlorobiphenyl. The four datasets of HL_B were derived on the basis of the HL_T dataset by a mass balance one-compartment pharmacokinetic model (1-CoPK) model for mammals. Basically the model estimates whole body primary biotransformation rate constants (k_B, h^{-1}) by subtracting the sum of rate constants for major processes of parent chemical passive elimination (referred to as k_X, h^{-1}) from the whole body total elimination rate constant (k_T, h^{-1}) (see Arnot et al., 2014 for details). Rate constants were converted to half-life values (HL_B, h) since half-lives (expressed in terms of time) are intuitively easier to understand and compare than rate constants (expressed in terms of inverse time). All the half-lives were converted to base 10 logarithmic form, prior to modelling. From uncertainty analysis and different parametrization of the 1-CoPK model, four different datasets for the biotransformation half-life (HL_{B1-4}) were obtained. The HL_{B1-4} datasets contain 1011, 1015, 935, 940 compounds, respectively. The data sets reflect highly complex and heterogeneous structures with molecular weight values ranging between 30.0 g/mol and 949.2 g/mol. Chemicals included in the datasets are mainly pharmaceuticals (80%), however other environmentally relevant compounds are also included such as halogenated organics (polychlorinated biphenyls (PCBs), polychlorinated dibenzodioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs), as well as aliphatic and aromatic hydrocarbons (polycyclic aromatic hydrocarbons (PAHs)) among others.

2.2. Molecular structures and descriptors

Simplified Molecular-Input Line Entry System (SMILES) notations were used to encode for 2D structural information for all the molecules in the five datasets. Canonical SMILES were derived by OpenBabel ver 2.3.2 (O'Boyle et al., 2011) software. Chemicals containing ionic salts were treated as discrete neutral organic chemical by removing the inorganic counterion in the SMILES notation. The SMILES strings were used to calculate mono-, and bi-dimensional molecular descriptors by the software PaDEL-Descriptor ver. 2.21 (Yap, 2011).

The use of three-dimensional descriptors was avoided since their calculation depends on 3D conformations of the studied chemicals and may give problems of reproducibility of the models. Constant descriptors, and descriptors found to be correlated pairwise (correlation greater than 0.98) were excluded from the total amount of descriptors generated by PaDEL-Descriptor (i.e. 1444 descriptors), to minimize redundant information. The procedure of cleaning and reduction of the dataset was performed by the software QSARINS (Gramatica et al., 2013, 2014). A final set of 425 descriptors was used as input variables for the variable subset selection procedure.

The list of the studied chemicals and the corresponding SMILES are reported in Supplementary Materials.

Please cite this article in press as: Papa, E., et al., Development of human biotransformation QSARs and application for PBT assessment refinement, Food and Chemical Toxicology (2017), http://dx.doi.org/10.1016/j.fct.2017.04.016