Disentanglement of the chemical, physical, and biological processes aids the development of quantitative structure-biodegradation relationships for aerobic wastewater treatment

Tom M. Nolte [a,b,*,1], Guangchao Chen [a,c,1], Coen S. van Schayk [a], Kevin Pinto-Gil [d], A. Jan Hendriks [a], Willie J.G.M. Peijnenburg [c,e], Ad M.J. Ragas [a,f]

[a] Department of Environmental Science, Institute for Water and Wetland Research, Radboud University, PO Box 9010, 6500 GL Nijmegen, the Netherlands
[b] Eidgenossische Technische Hochschule (ETH) Zurich, Laboratory of Inorganic Chemistry, Vladimir-Prelog-Weg 1, 8093 Zurich, Switzerland
[c] Institute of Environmental Sciences, Leiden University, 2300 RA Leiden, the Netherlands
[d] Research Programme on Biomedical Informatics (GRIB), Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Dept. of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain
[e] National Institute of Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, the Netherlands
[f] Department of Science, Faculty of Management, Science & Technology, Open University, Valkenburgerweg 177, 6419 AT Heerlen, the Netherlands

HIGHLIGHTS

- Data for micropollutant concentrations in Dutch STPs were collected and curated.
- Accounting for chemical and physical processes, rate constants for primary biodegradation were calculated.
- Global and class-specific QSBRs were developed.
- Primary biodegradation relates to energetic and electrostatic properties and structural complexity.

GRAPHICAL ABSTRACT
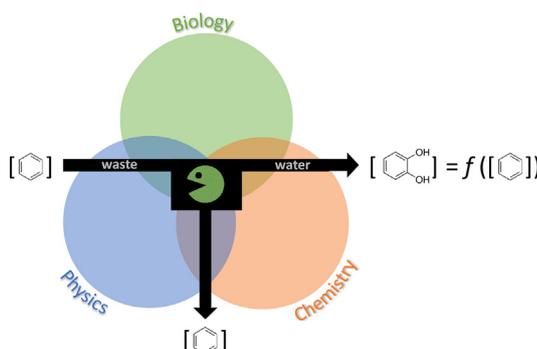
ARTICLE INFO

ABSTRACT

Attenuation of organic compounds in sewage treatment plants (STPs) is affected by a complex interplay between chemical (e.g. ionization, hydrolysis), physical (e.g. sorption, volatilization), and biological (e.g. biodegradation, microbial acclimation) processes. These effects should be accounted for individually, in order to develop predictive cheminformatics tools for STPs. Using measured data from 70 STPs in the Netherlands for 69 chemicals (pharmaceuticals, herbicides, etc.), we highlighted the influences of 1) chemical ionization, 2) sorption to sludge, and 3) acclimation of the microbial consortia on the primary removal of chemicals. We used semi-empirical corrections for each of these influences to deduce biodegradation rate constants upon which quantitative structure-biodegradation relationships (QSBRs) were developed. As shown by a global QSBR, biodegradation in STPs generally relates to structural complexity, size, energetics, and charge distribution. Statistics of the global QSBR were reasonable, being $R^2_{training}$ = 0.69 (training set of 51 compounds) and $R^2_{validation}$ = 0.50 (validation set of 18 compounds). Class-specific QSBRs utilized electronic properties potentially relating to rate-limiting enzymatic steps. For class-specific QSBRs, values of R^2 of in between 0.7 and 0.8 were obtained. With caution, environmental risk assessment methodologies may apply these models to estimate biodegradation rates for 'data-poor'

* Corresponding author.
  E-mail address: t.nolte@science.ru.nl (T.M. Nolte).
[1] These authors contributed equally to this paper.

compounds. The approach also highlights 'meta data' on STP operational parameters needed to develop QSBRs of better predictability in the future.

1. Introduction

Organic pollutants such as detergents, personal care products, and excreted or improperly disposed pharmaceuticals may be present in the influent of municipal STPs (Yuan et al., 2017). In STPs, organic compounds may undergo dissipation processes like abiotic degradation (hydrolysis and photolysis), volatilization, sorption to suspended solids, and primary/ultimate biodegradation (Nolte and Ragas, 2017). After incomplete removal in STPs, organic chemicals can enter the environment via effluent discharge and sludge disposal (Lautz et al., 2017). Understanding the fate and transformation profiles of chemicals of concern during wastewater treatment has become one of the major challenges when evaluating the hazards posed by potential environmental contaminants. This understanding is also relevant for developing new "green chemicals" which are less persistent in the environment in accordance with the principle of 'benign-by-design' (Rucker and Kummerer, 2012). Amongst the fate processes, biodegradation is often the predominant removal mechanism for organic pollutants in wastewater and, hence, is considered important for the design, development, and registration of a chemical. Under EU's Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) legislation, chemicals manufactured or imported in quantities over one ton per year must be evaluated for their ready biodegradability (EC, 2006). The common way of obtaining such information is to perform standardized laboratory tests, for example, in accordance with the guidelines of the Organization for Economic Co-operation and Development (OECD) (OECD 303A, 2001; OECD 302A, 1981; OECD 302B, 1992; OECD 302C, 2009; OECD 301, 1992; OECD 310, 2014; OECD 311, 2006; OECD 314, 2008). Carrying out experiments with the large amount of existing and new chemicals to be tested, is time-consuming and costly. Recent approaches (EU, 2006) also do not fully meet the benign-by-design concept aiming to estimate the environmental fate profile of newly developed chemicals from structural characteristics only (Leder et al., 2015). Using appropriate mathematical and statistical methods, we may infer biodegradability from chemical characteristics i.e. via quantitative structure–biodegradation relationships (QSBRs). As such, alternative in silico methods such as QSBRs are considered a potential help (e.g. suitable alternative data sources in REACH) to handle large compound libraries and enable rapid screening (OECD, 2004).

Many studies have demonstrated the possibility of predicting the (bio)degradability of diverse chemical families in aquatic media (Nolte and Ragas, 2017; Acharya et al., 2019a; Acharya et al., 2019b; Lee and von Gunten, 2012). Popular tools for this purpose include EPI Suite BIOWIN™, CATALOGIC, VEGA, TOPKAT, and START (Pizzo et al., 2013; Dimitrov et al., 2011). Such systems often focus on semi-quantitative, i.e. categorical, predictions of half-lives. Systems such as the EAWAG-BBD Pathway Prediction System (EAWAG PPS) and its successor system EnviPath aim to predict plausible biotransformation pathways and products (Wicker et al., 2016). Similarly, a tool named XenoSite serves to estimate susceptible atomic sites of molecules to cytochrome P450-mediated modifications (Zaretzki et al., 2013). These platforms help to unravel the mechanisms of biodegradation of organic compounds in complex aquatic systems. Many more QSBRs are available, but these often have a limited applicability domain, e.g. substituted benzenes (Lu et al., 2011) or hydrocarbons

(Howard et al., 2005) only, or apply to specific media such as surface water (Nolte et al., 2018). Given the importance in assessing exposure levels of chemicals in the environment (Seth et al., 2008), chemical biodegradation in STPs is of specific interest. Thus, programs like STPWIN™ in EPI Suite emerged specifically for STPs but most predictions are semi-quantitative. In turn, Burgis (2012) sought to develop a quantitative model for biological removal in wastewater sludge, based on OECD tests and other laboratory assays. The authors found that test outcomes varied substantially between experiments, greatly hampering model development. Undoubtedly, these studies highly contributed to the understanding of degradation of chemicals in STPs. However, the quantitative estimation of the biodegradation rate of a chemical during wastewater treatment remains problematic. A deeper understanding of the relevant mechanisms is needed in order to develop and improve QSBRs. Apart from the commonly considered biological processes (i.e. biodegradation), chemical and physical processes significantly affect the removal of a chemical, and hence, significantly complicate the modeling (Nolte and Ragas, 2017; Burgis, 2012). In STPs, organic chemicals can undergo processes such as (de)protonation (pK_a/pH), volatilization, (abiotic) hydrolysis, and sorption/desorption to sludge. Furthermore, the acclimation of microbial consortia may play a role. Depending on chemical functionality, these processes may exhibit a significant influence on the attenuation when compared to biodegradation (Nolte and Ragas, 2017).

Addressing the aforementioned difficulties, this study aimed to produce a quantitative, robust estimate for the pseudo-first order biodegradation rate constant of heterogeneous, multifunctional organic chemicals in STPs. Based on influent and effluent data from STPs throughout the Netherlands, we disentangled chemical, physical, and biological processes. The resulting data were used to develop global and class-specific QSBR models. The models account for various chemical characteristics and associated enzymatic steps and can quantitatively predict the primary biodegradation step for organic compounds in aerobic wastewater treatment.

2. Materials and methods

2.1. Data compilation and curation

We extracted influent and effluent data from the Watson database (Wassenaar et al., 2017; Pieters and Mol-Jansen, 2015) covering observed attenuation of concentrations of chemicals as a consequence of passing an STP. The database entails STPs throughout the Netherlands (see Fig. S1). We applied strict criteria of data curation to ensure the quality of the data used (see Supplementary information, S). Screening of an initial set of 34,000 records of the raw data from the Watson database left 1292 measurements (70 STPs in the Netherlands) for a total of 69 compounds. Data for 51 compounds was reserved (S1) for development of the global QSBR model (i.e. the training set, see Table 1). The training data consist mainly of pharmaceuticals (28) and polycyclic aromatic hydrocarbons (PAHs, 11), along with herbicides, insecticides, fungicides, etc. Initial evaluation of the obtained data revealed that for 28 compounds (Table S2), the influent concentration (C_i) influences the attenuation, as depicted in Fig. S2A.

In order to distinguish between the total concentration and the biologically available concentration of chemicals (available for

Table 1
List of compounds within the training set of the developed global model, including name, structure, and "measured" (with errors in 1 standard deviation, σ, in parenthesis) and QSBR-calculated values of $k_{b,s}$ (by the global model), in units of 1/d. The chemical structures shown indicate the major speciation state as function of pK_a/pH (pH = 7).

Compound	Structural speciation (pKa/pH)	Measured $k_{b,s}$	Calculated $k_{b,s}$
2,6-Dichlorophenol		1.6 (±0.7)	1.9
2-Nonylphenol		0.9 (±0.9)	0.5
Acenaphthene		1.8 (±0.8)	1.7
Aminomethanesulfonic acid		0.10 (±0.04)	0.7
Anthracene		1.7 (±1.0)	1.8
Atenolol		1.0 (±0.4)	0.5
Benz[a]anthracene		1.9 (±1.4)	1.8
Benzo[a]pyrene		1.8 (±1.6)	1.4
Benzo[b]fluoranthene		1.5 (±1.4)	1.4
Benzo[k]fluoranthene		1.6 (±1.5)	1.9
Bezafibrate		2.4 (±1.0)	3.0
Bisphenol A		1.9 (±0.9)	1.2
Caffeine		3.9 (±1.7)	4.3
Carbamazepine		0.7 (±0.3)	0.9
Clozapine		1.2 (±0.5)	1.3
Diclofenac		0.9 (±0.4)	0.8

Table 1 (continued)

Compound	Structural speciation (pKa/pH)	Measured $k_{b,s}$	Calculated $k_{b,s}$
Dipyridamole		3.1 (±1.3)	2.5
Diuron		1.4 (±0.6)	1.6
Estrone		3.8 (±1.6)	3.8
Fluoranthene		1.8 (±1.1)	1.8
Fluorene		1.7 (±0.8)	1.6
Gabapentin		0.13 (±0.05)	0.2
Gemfibrozil		1.4 (±0.6)	2.4
Glyphosate		0.4 (±0.2)	0.4
Hydrochlorothiazide		0.5 (±0.2)	0.5
Ibuprofen		3.1 (±1.3)	2.2
Imazalil		1.9 (±0.9)	1.2
Imidacloprid		1.7 (±0.7)	1.4
Iomeprol		1.1 (±0.5)	1.4
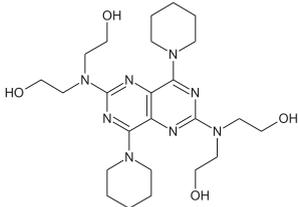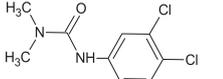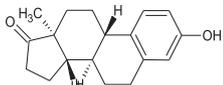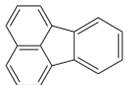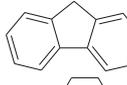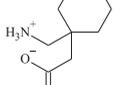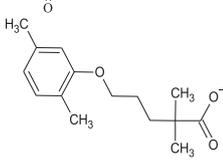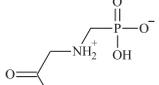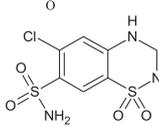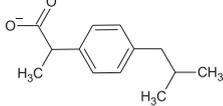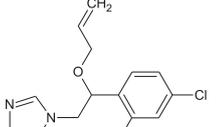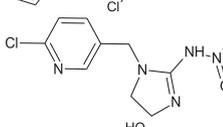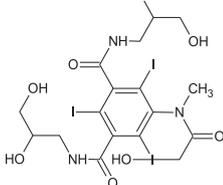
Table 1 (*continued*)

Compound	Structural speciation (pKa/pH)	Measured $k_{b,s}$	Calculated $k_{b,s}$
Ioxitalamic acid		1.9 (±0.8)	1.8
Irbesartan		0.5 (±0.4)	0.8
Ketoprofen		1.4 (±0.6)	2.6
Levetiracetam		0.7 (±0.3)	0.7
Lidocaine		1.2 (±0.5)	2.0
Lindane		2.2 (±1.1)	2.3
Linuron		1.4 (±0.6)	1.0
MCPA		0.9 (±0.4)	1.0
Metformin		1.8 (±0.8)	1.4
Metoprolol		0.4 (±0.2)	0.6
N,N-Diethyl-m-toluamide		2.0 (±0.9)	2.5
Naphthalene		2.5 (±1.2)	1.4
Naproxen		2.3 (±1.0)	2.3
Oxazepam		0.6 (±0.3)	0.9
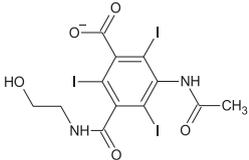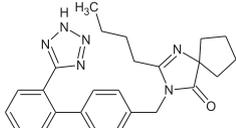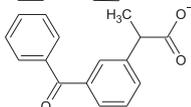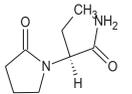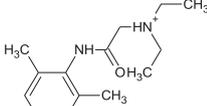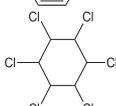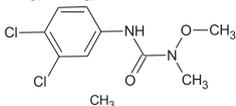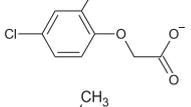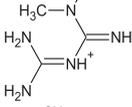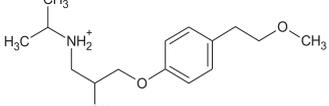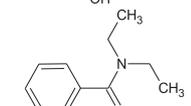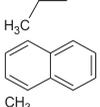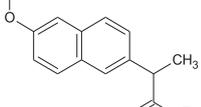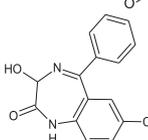
Table 1 (*continued*)

Compound	Structural speciation (pKa/pH)	Measured $k_{b,s}$	Calculated $k_{b,s}$
Phenanthrene		2.2 (±1.3)	1.8
Pipamperone		1.6 (±0.7)	1.4
Pyrene		2.0 (±1.3)	1.7
Simazine		1.5 (±0.7)	1.4
Sotalol		0.6 (±0.2)	0.5
Sulfamethoxazole		1.5 (±0.6)	1.1
Trimethoprim		1.2 (±0.5)	1.2
Valsartan		2.1 (±0.9)	1.4
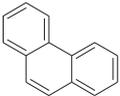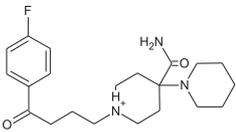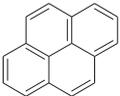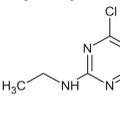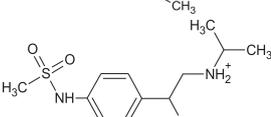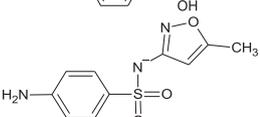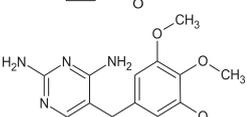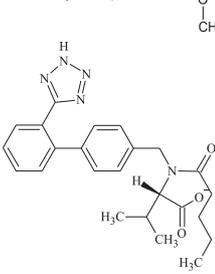
biodegradation), we accounted for the removal via the sorption to STP sludge before deriving k_b values (Section 2.2). The removal via sorption to sludge was estimated via the sorption partition coefficient (K_d, in L/kg). Values of $\log(K_d)$ were collected from the study of Berthod et al.(2017) and the iPiE Sum database (iPiE*Sum, 2018), and were arithmetically averaged if multiple values were available for a single compound. If an experimental value for K_d was unavailable, STPWIN (EPI Suite) was directly used for estimating the percentage adsorbed to sludge (Seth et al., 2008). The proton dissociation constant pK_a was obtained from the open literature e.g. DrugBank (DrugBank, 2019) and used to calculate the percentage ionization at pH = 7 and the apparent K_d. If a value for pK_a was not available, it was estimated via Chemaxon (https://chemaxon.com/marvin-archive/5_2_0/marvin/). The percentage of chemical removal via volatilization was also estimated by EPI Suite and applied, if applicable.

2.2. Calculation of k_b

We obtained the concentration of compounds available for biodegradation (C) in the aeration tank (Fig. 1) by:

$$C = P_p \cdot P_a \cdot C_i \qquad (1)$$

where P_p and P_a are the fractions of the remaining bioavailable concentrations of compounds in the primary and aeration tank, respectively, after sorption to sludge; C_i is the measured influent concentration as shown in Fig. 1. Values for C_i involving measurements for individual compounds were arithmetically averaged. P_p and P_a can be estimated by (Tebes-Stevens and Jones, 2004):

$$P = \frac{1}{K'_d + 1} \qquad (2)$$

$$K'_d = K_d \cdot TSS \qquad (3)$$

in which K_d is the solid/water partition coefficient in L/kg, and K'_d is the dimensionless suspended matter/water partition coefficient (i.e., corrected for the amount of suspended matter in the water). We assumed a total suspended solids (*TSS*) content of 220 mg/L for the primary tank. For the aeration tank, the *TSS* was assumed to be 800 mg/L (i.e. 20% of the total amount of 4000 mg/L because 20% is renewed daily; personal communication (E. Marsman, water board Rivierenland, 2018) (Fig. 1). The prediction of STPWIN (i.e. the percentage sorption, in %) corresponds to the value of $P_p \cdot P_a$ in Eq. (1) and was directly used to calculate C.
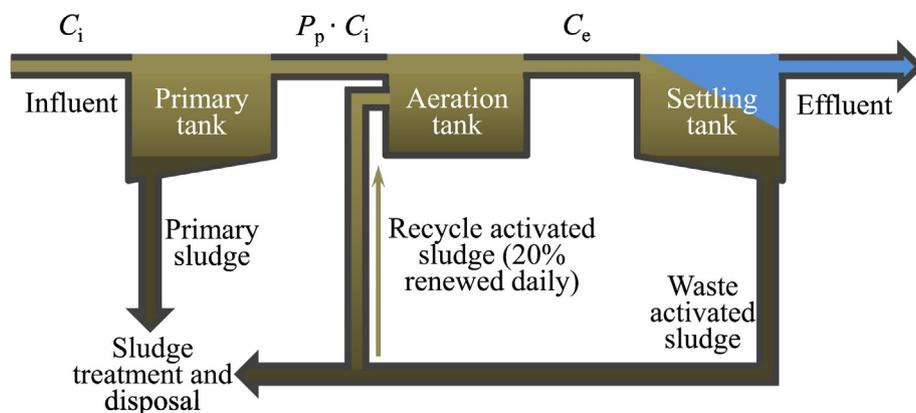
Fig. 1. Schematic illustration of the primary and secondary treatment of wastewater in a sewage treatment plant.

Next, we used the bioavailable concentration (in the aeration tank) and the effluent concentration (C and C_e, respectively, expressed as a molar concentration) to calculate the value of k_b (in units of 1/d) via:

$$k_b = -\frac{\ln\left(\frac{C_e}{C}\right)}{t_{HRT}} \tag{4}$$

where the hydraulic retention time (t_{HRT}) of the wastewater was assumed to equal 1 d (personal communication: E. Marsman, water board Rivierenland, 2018). We averaged the values for C_e involving measurements for individual compounds arithmetically.

2.3. Standardization of k_b

Evaluation of the obtained biodegradation data revealed that the biologically available concentration C influences the k_b for 28 compounds, as depicted in Fig. 2.

This relationship is often attributed to acclimation of the microbial consortia as the concentration of the contaminant increases (Nolte and Ragas, 2017; Nolte et al., 2018; Monod, 1942a). In the present study, we assumed that the dependency of k_b on the biologically available concentration (C) as shown in Fig. 2 applies equally for all compounds considered. Thus the biodegradation rate at concentration C (k_b) was corrected to produce a standard-

ized (s) biodegradation rate at a substrate concentration of 0.01 μmol/L ($k_{b,s}$) using interpolation and the empirical relationship shown in Fig. 2A and Eq. (5):

$$k_{b,s} = k_b + 0.7 \cdot (\log 0.01 - \log C) \tag{5}$$

This standardized pseudo-first order rate constant, $k_{b,s}$, was used as the endpoint for QSBR modeling throughout this study.

2.4. Characterization of errors

The corrections for HRT, sorption and acclimation, involved assumptions which introduced uncertainty in the obtained "measured" values for $k_{b,s}$ (Lee and von Gunten, 2012; Nolte et al., 2018; Kuo and Uppuluri, 1983). Errors in "measured" $k_{b,s}$ were quantified using semi-empirical error propagation. We did this via descriptions for: a) the uncertainty and variability in the free chemical concentration, and errors involved for b) the HRT and c) the microbial acclimation (for details see S2). The resulting errors in "measured" $k_{b,s}$, i.e. the ranges for $k_{b,s}$ (error margins σ in Table 1), were compared to the QSBR-predicted $k_{b,s}$ values (see Section 2.5). Then, we can evaluate the plausibility of the QSBRs in the light of the uncertainty in the training data.
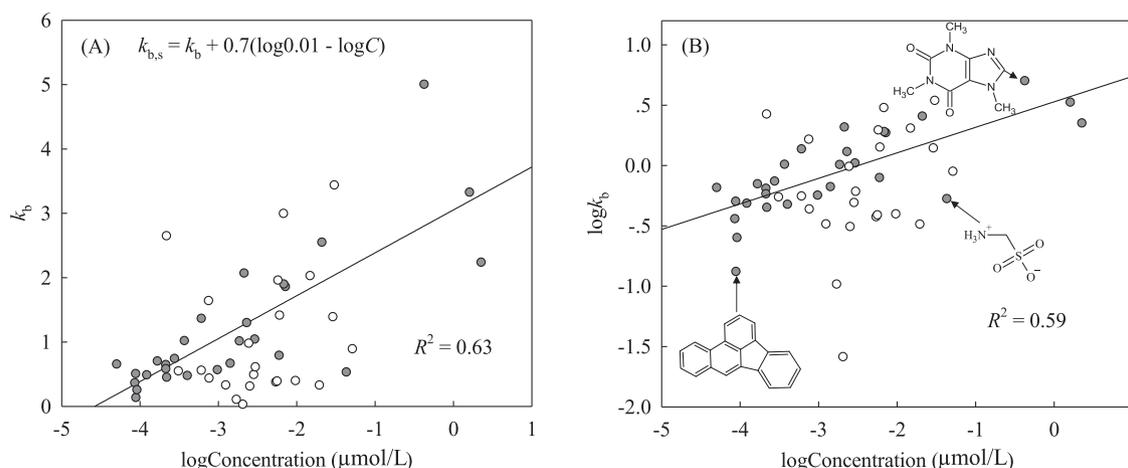
Fig. 2. Dependence of k_b on the biologically available concentration C, $n = 28$ (closed symbols), $R^2 = 0.63$ (log-normal regression, A) and $R^2 = 0.59$ (log-log regression, B), $p < 0.00001$ for both. The correlation suggests acclimation of the wastewater sludge to increasingly higher concentrations of chemicals. Open symbols denote compounds without a significant correlation between the influent concentration and removal ($n = 23$, $p < 0.05$ as cut-off criteria, Table S2). The relationship between removal for individual measurements and influent concentration is shown in Fig. S2.

Table 2

List of compounds within the validation set of the developed global model, including name, structure, and "measured" (with errors in 1 standard deviation, σ, in parenthesis) and QSBR-calculated values of $k_{b,s}$ (in units of 1/d). The chemical structures shown indicate the major speciation state as function of pK_a/pH (pH = 7).

Compound	Structural speciation	Measured $k_{b,s}$	Predicted $k_{b,s}$
2,6-Dichlorobenzamide		1.7 (±0.7)	1.0
Acetaminophen		4.8 (±2.0)	0.9
Atrazine		2.2 (±1.0)	1.1
Benzene		1.4 (±0.6)	1.0
Diatrizoic acid		1.3 (±0.6)	1.8
Ethylbenzene		4.6 (±2.0)	1.8
Flonicamid		0.4 (±0.2)	0.2
Iohexol		1.8 (±0.7)	1.3
Mecoprop		1.1 (±0.5)	2.3
Metolachlor		1.7 (±0.8)	0.6
m-Xylene		6.1 (±2.7)	1.5
o-Xylene		5.2 (±2.3)	3.5
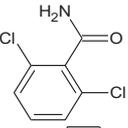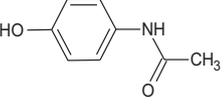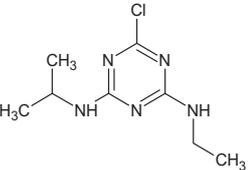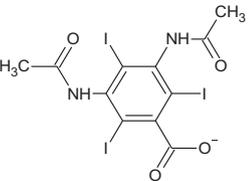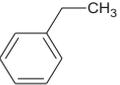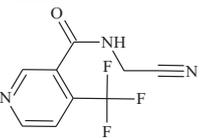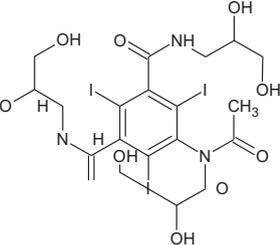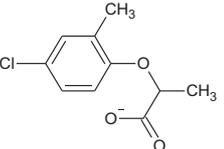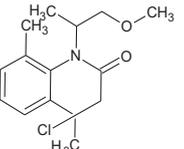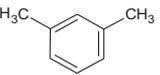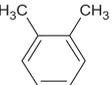
Table 2 (*continued*)

Compound	Structural speciation	Measured $k_{b,s}$	Predicted $k_{b,s}$
Pentachlorophenol		0.8 (±0.3)	0.6
Perfluorooctanoic acid*		0.3 (±0.3)	0.2
Phenol		6.9 (±3.0)	1.4
Pirimicarb		2.3 (±1.0)	3.8
Propoxur		1.4 (±0.6)	0.6
Toluene		2.6 (±1.1)	1.5
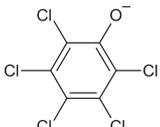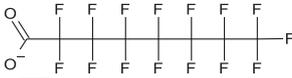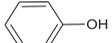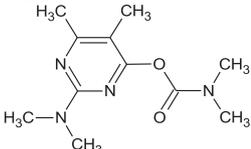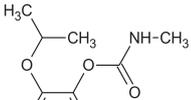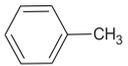
*Discrepancies in predicted and experimental k_d for PFOA have been reported, hence 90% sorption was taken based on *in situ* STP data [45], see S1 and S2.

2.5. Model development

To develop the global QSBR, 1D and 2D molecular descriptors were calculated using PaDel software v2.21 (Yap, 2011) after structure optimization of the chemicals in Avogadro v1.2.0 (Hanwell et al., 2012). These descriptors were selected by stepwise selection for the global model; given the sample size, a maximum of 10 descriptors was considered (Topliss and Costello, 1972; Tropsha, 2010) as a starting point, i.e. a criterion of $N_{chemicals} / N_{descriptor} \geq 5$ (Hermens et al., 1995). Mathematical analysis was performed by means of a partial least squares (PLS) algorithm to further limit the potential for over-fitting together with a 5-fold cross-validation and external validation (see Section 2.1). Data analysis tools used are embedded in the web-based platform Online Chemical Modeling Environment (Sushko et al., 2011).

Furthermore, we performed class-specific modeling to investigate the potential involvement of rate limiting steps adhering to specific chemical groups (see Section 3.2). We considered quantum-chemical descriptors which were obtained using MOPAC software (Stewart and MOPAC, 2016) and a previously developed method (Nolte and Peijnenburg, 2017; Rorije et al., 1995).

2.6. Model evaluation

Model performance was evaluated based on the coefficient of determination (R^2), the 5-fold cross-validated correlation coefficient (Q^2), the root mean square error ($RMSE$), and p values. A numeric evaluation of the applicability domain of the global QSBR is given in Fig. S5, whereas the structural diversity of compounds is shown in Tables 1 and 2. The class-specific models based on quantum-chemical descriptors were tested for known 'inert' compounds (DrugBank, 2019; Docherty et al., 2007; Aislabie et al., 1997).

Given the importance and necessity of evaluating the global QSBR, additional data from the Watson database (other than the training set) were selected using less strict criteria (details on the data selection can be found in S1), providing 18 more compounds as an external validation set (Table 2).

3. Results

3.1. Global QSBR

The compounds included in the training and validation sets are listed in Table 1 and Table 2 respectively along with the "measured" (with estimated errors) and predicted values for $k_{b,s}$. The measured values of $k_{b,s}$ of the training set ranges from 0.10 (±0.04) (aminomethanesulfonic acid) to 3.9 (±1.7) 1/d (caffeine). Measured values of $k_{b,s}$ of the validation set range from 0.3 (±0.3) (perfluorooctanoic acid) to 6.9 (±3.0) 1/d (phenol). Using the selected PaDel descriptors, we developed a global PLS model:

$$\begin{aligned}
\log k_{b,s} = {} & 26.7 - 0.149 \times \textbf{AATS5e} - 0.0112 \times \textbf{ATSC1i} \\
& + 0.771 \times \textbf{GATS4m} - 2.30 \times \textbf{BCUTw} - \textbf{1l} \\
& + 4.11 \times \textbf{ASP} - \textbf{4} + 0.00995 \times \textbf{SsOm} \\
& + 1.11 \times \textbf{ETA\_Shape\_Y} + 0.00930 \times \textbf{ZMIC3} \\
& + 0.200 \times \textbf{C3SP3} - 0.0493 \times \textbf{minHBint5}
\end{aligned} \quad (6)$$

$$n_{training} = 51, R^2_{training} = 0.69, Q^2_{training} = 0.69, RMSE_{training} = 0.18;$$
$$n_{validation} = 18, R^2_{validation} = 0.50, RMSE_{validation} = 0.37$$

where $\log k_{b,s}$ is the pseudo-first order rate constant (logarithmic unit) for primary aerobic biodegradation at an initial influent concentration (biologically available, i.e. aqueous) of 0.01 µmol/L. The coefficient of determination for the training set $R^2_{training}$ is 0.69, the value of the 5-fold cross-validated $R^2_{training}$ (i.e. $Q^2_{training}$) is 0.69. The global QSBR comprises the preset maximum of ten chemical descriptors (Eq. (6) and Table 3). Interpretation of these descriptors is presented in Section 4.3.2.

Two-thirds of the predictions of $k_{b,s}$ with regard to the validation compounds are within a factor of 2 from the "measured" values (see Table 2). The overall performance of the global model on

Table 3
Structural descriptors used in the global QSBR.

Symbol	Description
AATS5e	Average Broto-Moreau autocorrelation - lag 5/weighted by Sanderson electronegativities
ASP-4	Average simple path, order 4
ATSC1i	Centered Broto-Moreau autocorrelation - lag 1/weighted by first ionization potential
BCUTw-1l	nhigh lowest atom weighted BCUTS
C3SP3	Singly bound carbon bound to three other carbons
ETA_Shape_Y	Shape index Y
GATS4m	Geary autocorrelation - lag 4/weighted by mass
minHBint5	Minimum E-State descriptors of strength for potential Hydrogen Bonds of path length 5
SsOm	Sum of atom-type E-State: —O—
ZMIC3	Z-modified information content index (neighborhood symmetry of 3-order)

the validation set is expressed as $R^2_{\text{validation}} = 0.5$, indicating a reasonable predictive power of the model ($p = 0.01$, predictions are significant at $p < 0.05$).

The validation data include perfluorooctanoic acid (PFOA, Fig. 3B), which is dispersed globally in aquatic and terrestrial habitats, humans, and wildlife. PFOA is virtually non-biodegradable due to the presence of strong C—F bonds (Liou et al., 2010). According to EAWAG PSS (Wicker et al., 2016), there are no biodegradation pathways available for PFOA. For PFOA, the global model gave a fair prediction of $k_{b,s} = 0.2$ 1/d, versus a measured $k_{b,s} = 0.3$ (±0.3) 1/d as shown in Fig. 3B, although the error propagation might be underestimated because of surfactant-type sorption behavior (Arvaniti et al., 2014), see S2.

The global model performed well for benzene: an estimated value of $k_{b,s}$ of 1.0 is comparable to the "measured" value of 1.4 (±0.6) 1/d. The model performs suboptimal for acetaminophen and phenol (Fig. 3B), the latter with predicted $k_{b,s}$ of 1.4 1/d compared to a "measured" value of 6.9 (±3.0) 1/d. This might be attributed to the underestimation of the *ortho-para* directing effects of the hydroxyl group (this may also hold truth for estrone). In order to acquire a deeper mechanistic understanding of the biodegradation process, we developed class-specific QSBRs.

3.2. Class-specific QSBRs

We studied the potential role of electronic properties in control of the transformations underlying biodegradation via quantum-
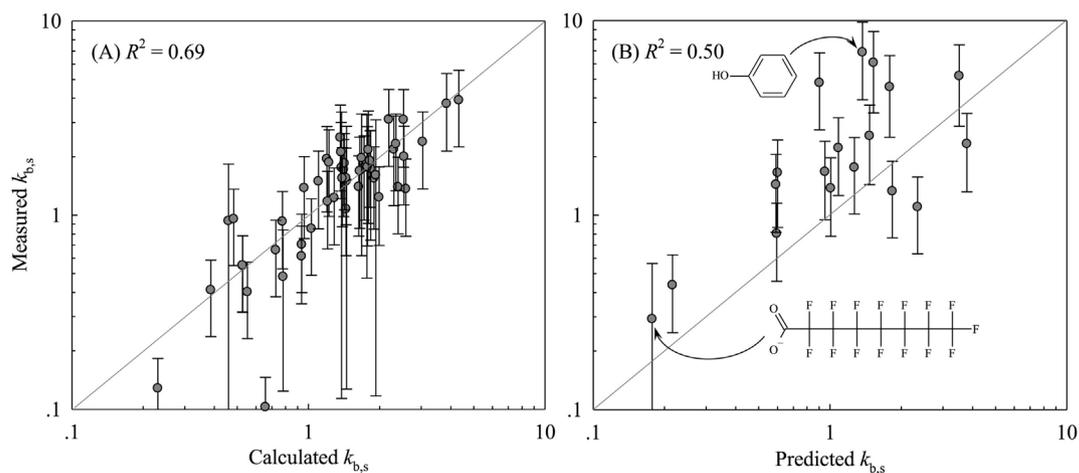
Fig. 3. Model fit of the derived global model on (A) training data, and (B) external validation data. Error bars indicate the uncertainty associated with "measured" $k_{b,s}$.
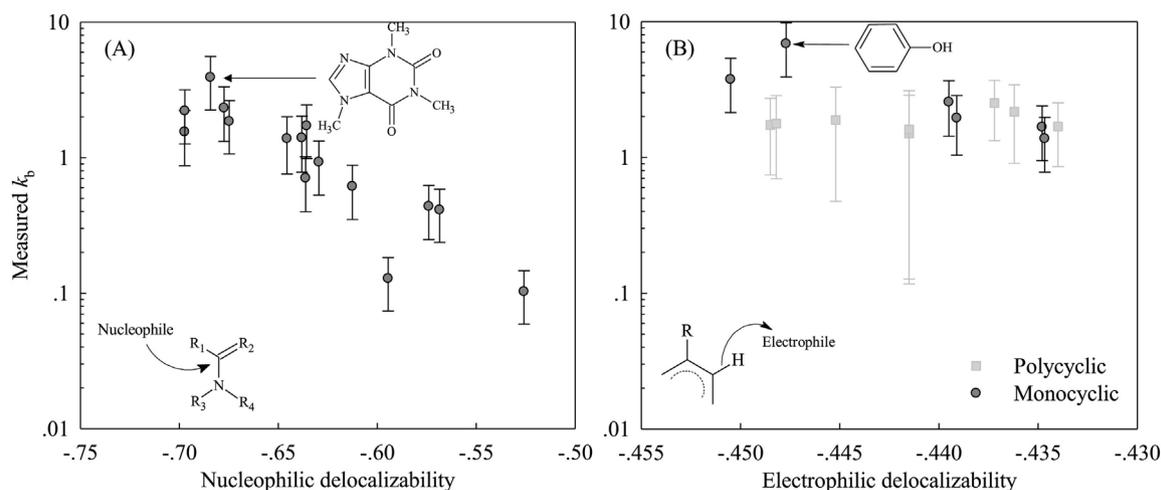
Fig. 4. Pseudo-first order biodegradation rate constant $k_{b,s}$ versus delocalizability.(A) $k_{b,s}$ versus nucleophilic delocalizability on the carbon atom of $R_1 = C(R_2)$-N(R_3)-R_4, predicted to undergo exclusively C—N cleavage according to EAWAG PPS (Wicker et al., 2016). Relationship I: log$k_{b,s, C—N}$ = 8.0 × dNr − 5.1 (R^2 = 0.79, n = 15, p < 0.00001). 2,6-Dichlorobenzamide was omitted as it undergoes oxidation according to Fig. 4B ($dEr = −0.43$, $k_{b,s, C—N} = 1.7$ 1/d).(B) $k_{b,s}$ versus electrophilic delocalizability on an aromatic carbon, predicted to undergo exclusively mono- and dioxygenation according to EAWAG PPS. Polycyclics are not included in correlation in view of the relatively high errors involved. Relationship II: log$k_{b,s, C—H}$ = 34.3 × dEr − 14.7 (R^2 = 0.78, n = 6, p = 0.02).

chemical descriptors. First, the energy gap between the highest occupied molecular orbital and the lowest unoccupied molecular orbital (E_{HOMO}-E_{LUMO}) explained 14% of the total variance in $k_{b,s}$ (p = 0.007), see Fig. S7. E_{HOMO}-E_{LUMO} alone did not possess satisfactory (e.g. $R^2 > 0.5$ (OECD, 2004)) predictive power. In turn, E_{HOMO} and E_{LUMO} separately were marginally significant ($p = 0.04$). Also, the presence of nitrogen-containing groups was important (Fig. S7, Table S3).

Frontier orbitals in multifunctional organic molecules may be part of different functional groups, which can provide a false indication of reactivity. Instead, delocalizability indices (Fukui et al., 1961) are normally computed for specific atoms. The use of delocalizability (taken as the maximum value within a molecule) indices gave satisfactory correlations for both nitrogen containing compounds ($R^2 = 0.79$, $n = 15$, $p < 0.00001$, Fig. 4A) and non-nitrogen containing compounds ($R^2 = 0.78$, $n = 6$, $p = 0.02$, Fig. 4B). Lists of these compounds are provided in Tables S3 and S4. As visualized in Fig. 4B, phenol ($k_{b,s} = 6.9 \pm 3.0$) is accompanied with a high electrophilic delocalizability (large negative value). In general, electrophilic delocalizability denotes donation of charge (e.g. by hydroxyl groups) into the delocalized aromatic ring and could enable a higher susceptibility to degradation of the Ar moieties. Fig. 4A shows that caffeine ($k_{b,s} = 3.9 \pm 1.7$) is accompanied by a more negative nucleophilic delocalizability.

Relationship I (Fig. 4A) was tested for pyridinium-like compounds (nucleophilic delocalizability ~ −0.51) and Relationship II (Fig. 4B) was tested for DDT (Aislabie et al., 1997) and mitotane (DrugBank, 2019) (electrophilic delocalizabilities ~ −0.43). The results corroborate the low aerobic biodegradability (predicted $k_{b,s} = 0.1$) for pyridiniums (Docherty et al., 2007), and medium-to-low aerobic biodegradability (predicted $k_{b,s} = 1.0$) for DDT and mitotane, Fig. 4B.

4. Discussion

4.1. Physical processes

Sorption is an important process affecting chemical removal, especially for highly hydrophobic compounds such as PAHs. Due to the absence of experimental data, the removal of chemicals via sludge needed to be estimated (via e.g. k_d). Chemical removal due to sorption on STP sludge was considered for modeling, meaning that the effect of sorption/desorption on bioavailability should have been excluded. However, due to the uncertainty involved (see S2) the influence of sorption may not be fully accounted for. For highly hydrophobic chemicals, the error in the bioavailable fractions (C) used to calculate $k_{b,s}$ (Eqs. 1–4)) was calculated to be relatively large (Fig. S4). A high extent of sorption limits the free concentration, but the error introduced by both uncertainty and variability increases as a function of K_{ow} (see Fig. S4). Highly hydrophobic chemicals and surfactants might show non-linear sorption isotherms (depending on the matrix), which is not anticipated by the current method (Eqs. (2), (3)) used for estimating sorption of chemicals on STP sludge. Additionally, the HRT of the wastewater (t_{HRT}, assumed to be 1d) can vary ~30% between STPs which as well affects the overall removal (Lautz et al., 2017). The errors in $k_{b,s}$ values were approximated for each chemical individually (see details in S2) with relatively large errors for highly hydrophobic chemicals, and chemicals with $pK_a \approx pH$.

4.2. Chemical processes

In STPs, chemicals may dissociate into components depending on the surrounding medium. The influence of abiotic hydrolysis on chemical attenuation in STPs was excluded by removing compounds with ester type of bond(s) from the data set (see S1). In the current study, the speciation state of the chemicals (pH = 7 assumed) was implemented. The compounds estimated to be present in their ionized form (as a function of pK_a and the pH) contain phenolic, carboxylic, amino, and sulfonamide groups. Since pH is not documented in the Watson database, an uncertainty is introduced. pH can affect sorption and bioavailability, especially when a compound has a pK_a around circumneutral pH (i.e. when the ionization state is uncertain).

The ionization state of a chemical affects its partitioning behavior (e.g. sorption) and bioavailability (for biodegradation) as the ionic species of compounds have lower lipophilicity (Hale and Abbey, 2017). We noted that sotalol ($k_{b,s} = 0.6 \pm 0.2$) and metoprolol ($k_{b,s} = 0.4 \pm 0.2$) should be susceptible to biotransformation based on their functional groups alone (having e.g. alcohol groups), however relatively low values of $k_{b,s}$ might instead reflect low cellular uptake due to their size and/or charge ($pK_a \approx 10$) of these chemicals. Smaller compounds are taken up by microorganisms, i.e. transferred through membranes more efficiently, whereas compounds with high hydrophobicity may prefer staying in the membrane rather than reaching the cell interior (Jing et al., 2009). In general, a change in size or hydrophobicity affects the bioavailability of compounds to microbes. The result may be a low biodegradation rate constant, which is the result of a low uptake by microbes, rather than an 'intrinsically' low biodegradability.

4.3. Biological processes

4.3.1. Microbial

Biodegradation assays are generally not easy to reproduce (Weiss and Cozzarelli, 2008; Amat et al., 2003). Microbial acclimation has long been identified as a factor influencing the attenuation of organic pollutants, but this process is difficult to control in STPs. Upon a higher chemical concentration, there exists a greater pressure for bacteria to excrete enzymes responsible for mitigating harmful effects of those pollutants. The pollutant might also exert an evolutionary pressure to enrich bacteria capable of using the pollutant as a source of energy.

In line with these notions, a relationship between the bioavailable concentration of organic compounds and the pseudo-first order biodegradation constant k_b was established empirically (Fig. 2). Though statistically significant, the normalization of $k_{b,s}$ to 0.01 μmol/L introduced an error, taken to be ~30% (S2). Eq. (5) is considered applicable to the concentration range investigated, but due to its empirical nature it should not be applied for chemicals present at concentrations $<10^{-4}$ μmol/L (which would result in a negative $k_{b,s}$) and $>10^{-1}$ μmol/L. Relationships between removal efficiency and initial concentration were observed for half of the data set (Table S2) but not for all compounds, even though the relationship was assumed to be applicable for all compounds considered.

It is clear that the concentration-k_b relationship (Fig. 2) does not apply equally well to all compounds. The tendency of microbial communities to adapt to specific chemicals is likely to depend on the chemical of interest (Fig. S2). These issues require further investigation. The absence of relationships for specific compounds might be due to either chemical/physical effects, co-metabolism, or the presence of specific biochemical pathways, cascades, and feedback loops (Nolte et al., 2018; Monod, 1942b; Qu and Vondriska, 2009; del Giorgio and J, 1998; Singh, 2012; Hazen, 2010). Additionally, compounds with antimicrobial properties might be removed via entirely different mechanisms. For antimicrobials, the QSBRs may predict high biodegradability whereas, in fact, the compound(s) might adversely affect the microbial consortia. Nevertheless, concentrations in STPs are generally at a sub-microgram level, i.e. inducing low toxicity. It could be worthwhile to distinguish

between specific (viable) populations responsible for biodegradation, using e.g. Monod kinetics (Nolte et al., 2018; Monod, 1942a).

Bacteria might metabolize complex multifunctional organic molecule in different ways: via e.g. amidohydrolase, oxidase, dehydrogenase, etc. The relative importance of such pathways depends on the 'background' enzymatic activity as well as the 'intrinsic' compatibility/reactivity of the corresponding chemical group. It is only the latter that can be quantified using QSBR parameters.

4.3.2. QSBR parameters

Following the principles established by the OECD (OECD, 2004), chemical descriptors used for the global QSBR (Fig. 3) were also interpreted to pinpoint key characteristics of the compounds to biodegradation. In the model, the Broto-Moreau autocorrelation descriptor describes how a property (e.g. Sanderson electronegativity, ionization potential) is distributed along the topological structure. These descriptors are weighted by the ionization potential and electronegativity, which could indicate that biodegradation is directly determined by the energetics (also through SsOm) and localization of electrons (also through minHBint5) in specific bonds. The role of ionization potential and electronegativity affecting biodegradation was also reported by Mansouri et al. (2013). GATS4m and ASP-4 describe distances within a molecule and indicate that size and shape (also characterized by ETA_Shape_Y and BCUTw-1l) are relevant. This was also confirmed by previous studies in which the Geary autocorrelation (weighted by mass) index (Chen et al., 2014), accessible surface area (Nolte et al., 2018), and molecular weight (Lu et al., 2011) were found to describe biodegradability. C3SP3 could characterize the influence of aromaticity, or steric effects. BCUTw-1l is an eigenvalue-based descriptor that describes the distribution of molecular weight (see also ZMIC3). Molecular weight is generally reported as a significant factor in affecting biodegradation (Boethling et al., 1994; Cheng et al., 2012; Tunkel et al., 2000). In line with these results, Nolte et al. (2018) noted that microbial consortia can degrade a wide range of pollutants, but a greater complexity in molecular structure generally inhibits biodegradation.

The combination of the aforementioned descriptors shows that predicting biodegradation of structurally diverse compounds in a complex system like an STP involves many chemical characteristics, such as structural complexity (size, shape, volume, molecular weight etc.) as well as energetic (ionization potential, etc.) and electrostatic (electronegativity, charge, etc.) properties. This may be a consequence of the existence of many pathways relating to numerous specialized bacteria and fungi capable of degrading chemicals. Therefore, overall a multitude of pathways are being exploited by the multitude of bacterial consortia present in STPs. Meanwhile, biodegradation of specific families of compounds may be possibly less complex as often very specific and well-defined pathways are reported. Limiting the chemical space of the chemicals could result in more specific QSBRs containing fewer descriptors that allow for more efficient mechanistic interpretation, as indicated by the developed class-specific QSBRs in Figs. 4 and S6.

The training set of the global model developed in this study contains 11 PAHs (Table 1). The biodegradation rate for PAHs is reported to be generally inversely proportional to the number of rings (Cerniglia, 1992). Based on the same approach of developing the global QSBR, a QSBR was derived for these 11 PAHs together with four other compounds that were predicted to undergo (di) oxygenation using the descriptor MATS4v as shown in Fig. S6. This model describes the decrease of biodegradation rate constant with the increase of ring number for the compounds, as previously reported (Cerniglia, 1992). However, the high uncertainty in the free concentrations for PAHs results in high uncertainties for $k_{b,s}$ which questions the plausibility of this QSBR. Reported elsewhere,

the variation in biodegradation for PAHs with two to four rings is generally only within one order of magnitude (Wammer and Peters, 2005). The decrease of biodegradability with increasing ring number of the aromatics was questioned elsewhere (Wammer and Peters, 2005; Xu and Li, 2012) since PAHs with more rings are also electron-richer and thus can be assumed more susceptible to electrophilic attack (the pathway shown in Fig. 4B). The recalcitrance of PAHs increases also with molecular weight, the octanol-water partition coefficient (K_{OW}) and reversely with water solubility (Cerniglia, 1992) which relate to bioavailability/uptake (Section 4.2). These issues might explain why PAHs do not adhere to the relationship between $k_{b,s}$ and delocalizability (Fig. 4B).

A high delocalizability generally corresponds to a low activation energy (Rorije et al., 1995). The presence of nitrogen atoms was also relevant, as was also reported by Nolte et al.(2018). Thus, relationship I in Fig. 4A could indicate that for nitrogen containing molecules, $k_{b,s}$ is related to the activation energy upon enzymatic attack. During the attack (here nucleophilic), one or more electrons are transferred within a complex between the compound and an enzyme. This relationship (Fig. 4A) characterizes site-specific biodegradability for compounds containing $R_1 = C(R_2)$-N(R_3)-R_4 moieties. However, the relationship is not considered suitable for nitrogen containing compounds containing also electron-rich or long aliphatic substructures, which undergo e.g. mono- and deoxygenation instead. As mentioned, the presence of electron donors (—OH in phenols) promotes the susceptibility of the ortho and para positions to electrophilic aromatic substitutions, and the reverse applies to electron-withdrawing groups (e.g. halogens). Again, the activation energy might be involved which relates to the strength of the weakest carbonaceous Ar—H bond (Fig. 4B). This may hold true for estrone for which the biodegradability rate constant was relatively high ($k_{b,s} = 3.8 \pm 1.6$). The ortho-para directing effect would also apply to 2-nonylphenol ($k_{b,s} = 0.9 \pm 0.9$) but its biodegradation may be inhibited by bioavailability (i.e. lack of uptake) due to the presence of the bulky nonyl chain. The relevance of electronic effects may also be expressed by descriptors such as ATSC1i (describing the distribution of ionization potential along the skeleton of a molecule) and AATS5e (characterizing the distribution of electronegativity) in the global QSBR.

An intuitive way to determine whether a compound can be evaluated using a class-specific QSBR, is an assessment of the relative likelihood of branching metabolic routes. This however would require an in depth microbiological, bioinformatics, or thermodynamic study. As a first approximation, relative likelihood of pathways can be estimated using EAWAG PPS (Wicker et al., 2016). These estimations, affirm that relationships in Fig. 4A and B are strictly applicable only for homogeneous series of compounds. For example, in Fig. 4A for both aminomethanesulfonic acid ($k_{b,s} = 0.10 \pm 0.04$) and caffeine ($k_{b,s} = 3.9 \pm 1.7$) a very similar pathway was predicted (C—N cleavage, Fig. 5), albeit with different likelihoods. Although the initial attack is nucleophilic, the cleavage is overall an oxidative process. Depending on the substituent pattern of the carbon atom, ketones, quinones (caffeine), or aldehydes (aminomethane sulfonic acid) are produced. Interestingly, EAWAG PPS predicted a lower likelihood of C—N cleavage for urea groups C(=O)(NH$_2$)$_2$, as compared to the imidazolic methyl (Fig. 5). Accordingly, the C(=O)(NH$_2$)$_2$ urea group of caffeine was predicted to be only moderately biodegradable (nucleophilic delocalizability = −0.64, predicted $k_{b,s} = 1.0$). In EAWAG PPS, cleavage of C—N bonds in amides and urea derivatives is covered by a single rule since there is no chemical reason to divide them. It is also noted that cleavage of urea derivatives occurs between the N atom and the C atom with the most positive partial charge (i.e. donation of electron(s) into an easily accessible vacant low-lying orbital).
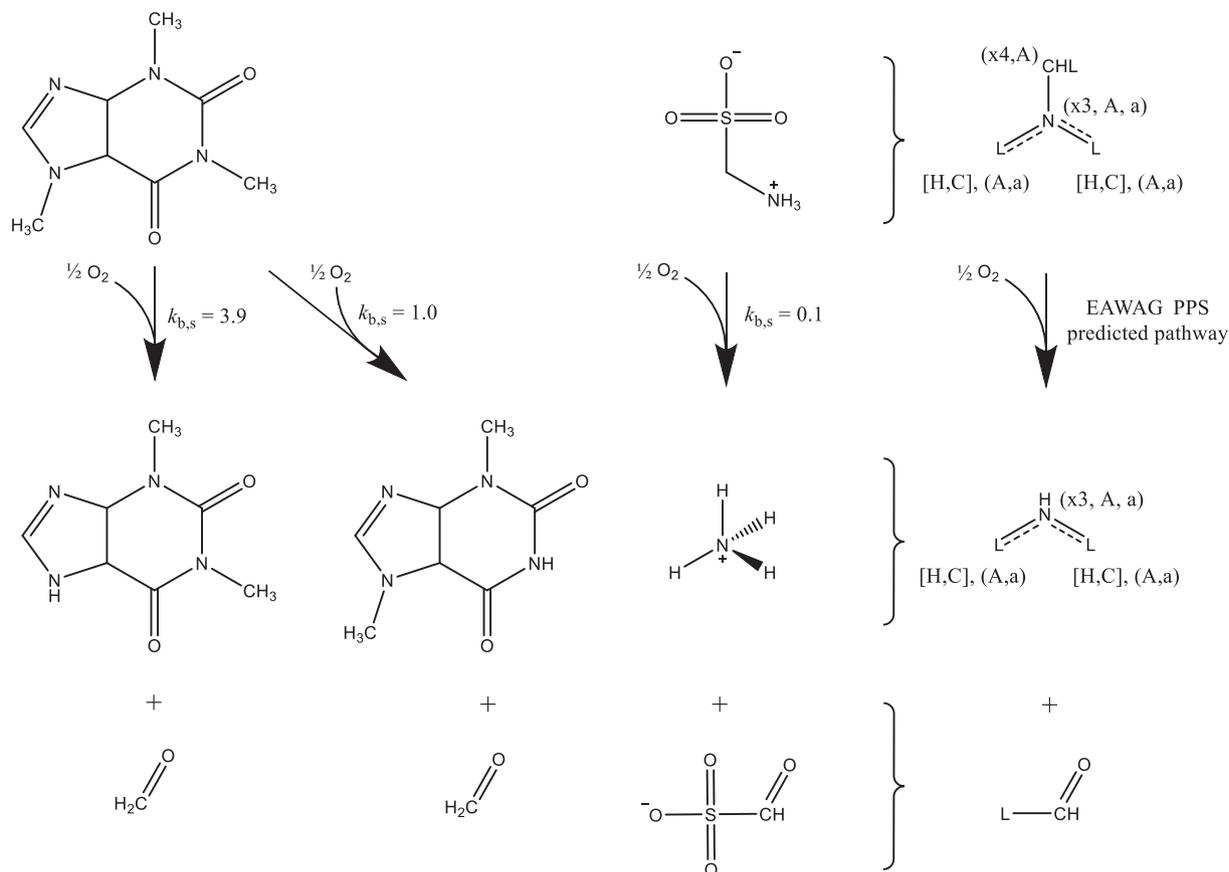
Fig. 5. Primary biotransformation of (A) caffeine, and (B) aminomethanesulfonic acid in STPs, and the overarching C—N cleavage pathway estimated by EAWAG PPS. $k_{b,s}$ values of relevant pathways are also provided.

In Fig. 4 multifunctional compounds containing hydroxyl groups were excluded, since neutral alcohols are relatively easily biodegradable with biodegradation proceeding via a different mechanism. Also, hydroxylation of benzylic carbon atoms and decarboxylation were not taken into account. It is worth noting that delocalizability indices have previously been used in P450 metabolic transformations in mammals (Zaretzki et al., 2013), as well as for the dehalogenation of simple benzene analogs in sediments systems (Rorije et al., 1995). Thus, the indices might prove useful for other classes, but the rate limiting step may take place elsewhere, e.g. on the alcoholic oxygen atom (Ji and Schüürmann, 2015).

4.4. Limitations and outlook

Limitations of this study include the limited amount of data that was available to derive and validate the empirical models. A difference in data quality between the training and validation set is anticipated (see S1). Although the Watson database contained entries of over 30,000 records of measured concentrations, rigorous selection criteria were used to obtain only the most reliable values for $k_{b,s}$. These criteria would need to be met also for new compounds for which additional biodegradation data could be used to strengthen the current QSBRs. The monitoring data in the Watson database were obtained during different seasons, under varying temperatures, STPs, influent concentrations, and by using different analytical techniques which may independently or collectively affect the removal efficiency of compounds in STPs. Excluding the records with effluent concentrations below the detection limit was suspected to produce bias towards readily biodegradable compounds, whereas predictions for "non-biodegradable" compounds are less certain. To better distinguish the variations in biodegradation caused by only structural heterogeneity of chemicals, accurate and site-specific sorption coefficients, especially for hydrophobic (e.g. PAHs) compounds, are required. Information on attenuation via sludge disposal is also of interest for this purpose. An understanding of the underlying mechanisms of microbial acclimation between different STPs and for different geographic locations is also important in this context. Such achievements could narrow the uncertainties and knowledge gaps that remain, and alleviate the current challenges in QSBR modeling. Archiving biodegradation data in a more comprehensive way seems fundamental for this field of research. It would highly benefit QSBR modelers if consistent input of data into on-line databases (e.g. Watson database) could be realized for scientific purposes. This would ensure the accessibility of biodegradation data as well as quantity and quality. Despite the limitations mentioned the statistical results for the QSBRs show the capability of predicting biodegradation, even though the predictions refer to a generic STP using "common" operating procedures only.

5. Conclusion

Removal of organic pollutants in STPs is a complex process under the influence of various chemical, physical, and biological processes. Biodegradation of compounds in wastewater might vary significantly with the changes of conditions (e.g. HRT and total suspended solid content). Therefore, estimating realistic biodegradation rates in STPs stemming from solely chemical structures seems fairly erratic. In this study, cheminformatics models were developed based on a series of assumptions to distinguish differ-

ences in chemical attenuation resulting from the structural divergence of the contaminants, including the approximation of chemical speciation after ionization (when applicable), exclusion of pollutant elimination via hydrolysis and sorption, and the consideration of acclimation of microbial consortia. The derived QSBRs generally have reasonable predictive power ($R^2_{\text{validation}} \geq 0.50$ and $0.7 < R^2 < 0.8$) to evaluate biodegradation of sets of heterogeneous organic chemicals during wastewater treatment (for the aqueous-phase only and standardized at an initial influent concentration of $0.01~\mu mol/L$). The modeling workflow enables the incorporation of STP-specific data (e.g. pH, total suspended solids, HRT, chemical concentration and microbial acclimation) when estimating chemical removal in STPs. In contrast, previous attempts of developing cheminformatics models for biodegradation in STPs often do not consider this. This study is the first of its kind attempting to unify categorical pathway prediction with quantitative predictions of biodegradation rates in STPs. Even though the derived models are only intended for primary biodegradation, they provide a starting point to evaluate full mineralization (through evaluation of individual metabolic steps) because specific chemical, physical, and biological processes can be better distinguished.

Declaration of competing interest

The authors have no conflicts of interests to disclose.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.133863.

References

Acharya, K. et al., 2019a. A quantitative structure-biodegradation relationship (QSBR) approach to predict biodegradation rates of aromatic chemicals. Water Res. 157, 181–190.

Acharya, K. et al., 2019b. The experimental determination of reliable biodegradation rates for mono-aromatics towards evaluating QSBR models. Water Res. 160, 278–287.

Aislabie, J.M., Richards, N.K., Boul, H.L., 1997. Microbial degradation of DDT and its residues - a review. N. Z. J. Agric. Res. 40 (2), 269–282.

Amat, A.M. et al., 2003. Ozonisation coupled with biological degradation for treatment of phenolic pollutants: a mechanistically based study. Chemosphere 53 (1), 79–86.

Arvaniti, O.S. et al., 2014. Sorption of perfluorinated compounds onto different types of sewage sludge and assessment of its importance during wastewater treatment. Chemosphere 111, 405–411.

Berthod, L. et al., 2017. Quantitative structure-property relationships for predicting sorption of pharmaceuticals to sewage sludge during waste water treatment processes. Sci. Total Environ. 579, 1512–1520.

Boethling, R.S. et al., 1994. Group-contribution method for predicting probability and rate of aerobic biodegradation. Environ. Sci. Technol. 28 (3), 459–465.

Burgis, C., 2012. Predicting Biological Removal of Contaminants in Wastewater Treatment QSBR Modeling.

Cerniglia, C.E., 1992. Biodegradation of polycyclic aromatic hydrocarbons. Biodegradation 3 (2–3), 351–368.

Chen, G. et al., 2014. Comparative study of biodegradability prediction of chemicals using decision trees, functional trees, and logistic regression. Environ. Toxicol. Chem. 33 (12), 2688–2693.

Cheng, F.X. et al., 2012. In silico assessment of chemical biodegradability. J. Chem. Inf. Model. 52 (3), 655–669.

del Giorgio, P.A., Cole, J.J., 1998. Bacterial growth efficiency in natural aquatic systems. Annu. Rev. Ecol. Syst. 29, 503–541.

Dimitrov, S. et al., 2011. Simulation of chemical metabolism for fate and hazard assessment. II CATALOGIC simulation of abiotic and microbial degradation. SAR QSAR Environ. Res. 22 (7–8), 719–755.

Docherty, K.M., Dixon, J.K., Kulpa, C.F., 2007. Biodegradability of imidazolium and pyridinium ionic liquids by an activated sludge microbial community. Biodegradation 18 (4), 481–493.

DrugBank. 2019 [cited 2019 09-01]; Available from: https://www.drugbank.ca/drugs/DB00648.

European Commission, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency, in 1907/2006, E. Commission. Official Journal of the European Union, Editor.

European Union, Regulation (EC) 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH). Official Journal of the European Union, L 136/3–L 136/280. 2006.

Fukui, K., Kato, H., Yonezawa, T., 1961. A new quantum-mechanical reactivity index for saturated compounds. Bull. Chem. Soc. Jpn. 34 (8), 1111–1115.

Hale, T., Abbey, J., 2017. Drug transfer during breast-feeding. In: Fetal and Neonatal Physiology, pp. 239–248.e5.

Hanwell, M.D. et al., 2012. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminformatics 4.

Hazen, T.C., Chapter 7: cometabolic bioremediation, in Handbook of Hydrocarbon and Lipid Microbiology K.N. Timmis, Editor. 2010, Springer.

Hermens, J., B, S., Damborsky, J., Karcher, W., Mueller, M., Peijnenburg, W., Sabljic, A., Sjostrom, M., 1995. Assessment of QSARs for predicting fate and effects of chemicals in the environment: an international European project. Sar Qsar Environ. Res. 3, 223–236.

Howard, P. et al., 2005. A new biodegradation prediction model specific to petroleum hydrocarbons. Environ. Toxicol. Chem. 24 (8), 1847–1860.

iPiE*Sum, iPiE Summary Database Search. 2018.

Ji, L., Schüürmann, G., 2015. Computational biotransformation profile of paracetamol catalyzed by cytochrome P450. Chem. Res. Toxicol. 28 (4), 585–596.

Jing, P., Rodgers, P.J., Amemiya, S., 2009. High lipophilicity of perfluoroalkyl carboxylate and sulfonate: implications for their membrane permeability. J. Am. Chem. Soc. 131 (6), 2290–2296.

Kuo, W., Uppuluri, V., 1983. A review of error propagation analysis in systems. Microelectron. Reliab. 23 (2), 235–248.

Lautz, L.S. et al., 2017. Evaluation of SimpleTreat 4.0: simulations of pharmaceutical removal in wastewater treatment plant facilities. Chemosphere 168, 870–876.

Leder, C., Rastogi, T., Kümmerer, K., 2015. Putting benign by design into practice-novel concepts for green and sustainable pharmacy: designing green drug derivatives by non-targeted synthesis and screening for biodegradability. Sustain. Chem. Pharm. 2, 31–36.

Lee, Y., von Gunten, U., 2012. Quantitative structure-activity relationships (QSARs) for the transformation of organic micropollutants during oxidative water treatment. Water Res. 46 (19), 6177–6195.

Liou, J.S. et al., 2010. Investigating the biodegradability of perfluorooctanoic acid. Chemosphere 80 (2), 176–183.

Lu, G. et al., 2011. Correlation for the structure and biodegradability of substituted benzenes in Songhua river water. Chem. J. Internet 3 (7), 34.

Monod, J., 1942a. Recherches sur la croissance des Cultures Bactériennes. Hermann, Paris.

Monod, J., Recherches sur la croissance des Cultures Bactériennes. Hermann. 1942b, Paris.

Nolte, T.M., Peijnenburg, W.J.G.M., 2017. Aqueous-phase photooxygenation of enes, amines, sulfides and polycyclic aromatics by singlet (a(1)Delta(g)) oxygen: prediction of rate constants using orbital energies, substituent factors and quantitative structure-property relationships. Environ. Chem. 14 (7), 442–450.

Nolte, T.M., Ragas, A.M.J., 2017. A review of quantitative structure-property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps. Environmental Science-Processes & Impacts 19 (3), 221–246.

Nolte, T.M. et al., 2018. Quantitative structure-activity relationships for primary aerobic biodegradation of organic chemicals in pristine surface waters: starting points for predicting biodegradation under acclimatization. Environ. Sci. Processes Impacts 20 (1), 157–170.

OECD, 302A, Test No. 302A: Inherent Biodegradability: Modified SCAS Test. 1981, Organization for Economic Co-operation and Development.

OECD, 2004. OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models.

OECD 301, Test No. 301: Ready Biodegradability. 1992, Organization for Economic Co-operation and Development.

OECD 302B, Test No. 302B: Inherent Biodegradability: Zahn-Wellens/EVPA Test. 1992, Organization for Economic Co-operation and Development.

OECD 302C, Test No. 302C: Inherent Biodegradability: Modified MITI Test (II). 2009, Organization for Economic Co-operation and Development.

OECD 303A, OECD Guideline for Testing of Chemicals 303A. 2001, Organization for Economic Co-operation and Development.

OECD 310, Test No. 310: Ready Biodegradability - CO2 in Sealed Vessels (Headspace Test). 2014, Organization for Economic Co-operation and Development.

OECD 311, Test No. 311: Anaerobic Biodegradability of Organic Compounds in Digested Sludge: by Measurement of Gas Production. 2006, Organization for Economic Co-operation and Development.

OECD 314, Test No. 314: Simulation Tests to Assess the Biodegradability of Chemicals Discharged in Wastewater. 2008, Organization for Economic Co-operation and Development.

Pieters, B.J. and S. Mol-Jansen, Update Watson-Database. 2015, Netherlands National Institute for Public Health and the Environment.

Pizzo, F. et al., 2013. In silico models for predicting ready biodegradability under REACH: a comparative study. Sci. Total Environ. 463, 161–168.

Qu, Z., Vondriska, T.M., 2009. The effects of cascade length, kinetics and feedback loops on biological signal transduction dynamics in a simplified cascade model. Phys. Biol. 6 (1).

Rorije, E. et al., 1995. Modeling reductive dehalogenation with quantum chemically derived descriptors. SAR QSAR Environ. Res. 4 (4), 237–252.

Rucker, C., Kummerer, K., 2012. Modeling and predicting aquatic aerobic biodegradation - a review from a user's perspective. Green Chem. 14 (4), 875–887.

Seth, R., Webster, E., Mackay, D., 2008. Continued development of a mass balance model of chemical fate in a sewage treatment plant. Water Res. 42 (3), 595–604.

Singh, S.N., Chapter 7, in Microbial Degradation of Xenobiotics 2012, Springer.

Stewart, J.J.P., MOPAC, 2016. Stewart Computational Chemistry: Colorado Springs. CO, USA.

Sushko, I. et al., 2011. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J. Comput. Aided Mol. Des. 25 (6), 533–554.

Tebes-Stevens, C., Jones, W., 2004. Estimation of microbial reductive transformation rates for chlorinated benzenes and phenols using a quantitative structure-activity relationship approach. Environ. Toxicol. Chem. 23 (7), 1600–1609.

Topliss, J.G. and R.J. Costello, Chance correlations in structure-activity studies using multiple regression-analysis. Journal of Medicinal Chemistry, 1972. 15(10): p. 1066-&.

Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Mol. Inf. 29 (6–7), 476–488.

Tunkel, J. et al., 2000. Predicting ready biodegradability in the Japanese Ministry of International Trade and Industry test. Environ. Toxicol. Chem. 19 (10), 2478–2485.

Wammer, K.H., Peters, C.A., 2005. Polycyclic aromatic hydrocarbon biodegradation rates: a structure-based study. Environ. Sci. Technol. 39 (8), 2571–2578.

Wassenaar, P.N.H., et al., Substances of Very High Concern and the Transition to a Circular Economy: An Initial Inventory. 2017.

Weiss, J.V., Cozzarelli, I.M., 2008. Biodegradation in contaminated aquifers: incorporating microbial/molecular methods. Ground Water 46 (2), 305–322.

Wicker, J. et al., 2016. enviPath - the environmental contaminant biotransformation pathway resource. Nucleic Acids Res. 44 (D1), D502–D508.

Xu, X., Li, X.G., 2012. QSAR for predicting biodegradation rates of polycyclic aromatic hydrocarbons in aqueous systems. Chin. J. Struct. Chem. 31 (8), 1212–1221.

Yap, C.W., 2011. PaDEL-Descriptor: an open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem. 32 (7), 1466–1474.

Yuan, J.E., Van Dyke, M.I., Huck, P.M., 2017. Identification of critical contaminants in wastewater effluent for managed aquifer recharge. Chemosphere 172, 294–301.

Zaretzki, J., Matlock, M., Swamidass, S.J., 2013. XenoSite: accurately predicting CYP-mediated sites of metabolism with neural networks. J. Chem. Inf. Model. 53 (12), 3373–3383.