



CrossMark  
click for updates

Cite this: DOI: 10.1039/c7em00034k

# A review of quantitative structure–property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps†

Tom M. Nolte\* and Ad M. J. Ragas

Many organic chemicals are ionizable by nature. After use and release into the environment, various fate processes determine their concentrations, and hence exposure to aquatic organisms. In the absence of suitable data, such fate processes can be estimated using Quantitative Structure–Property Relationships (QSPRs). In this review we compiled available QSPRs from the open literature and assessed their applicability towards ionizable organic chemicals. Using quantitative and qualitative criteria we selected the 'best' QSPRs for sorption, (a)biotic degradation, and bioconcentration. The results indicate that many suitable QSPRs exist, but some critical knowledge gaps remain. Specifically, future focus should be directed towards the development of QSPR models for biodegradation in wastewater and sediment systems, direct photolysis and reaction with singlet oxygen, as well as additional reactive intermediates. Adequate QSPRs for bioconcentration in fish exist, but more accurate assessments can be achieved using pharmacologically based toxicokinetic (PBTK) models. No adequate QSPRs exist for bioconcentration in non-fish species. Due to the high variability of chemical and biological species as well as environmental conditions in QSPR datasets, accurate predictions for specific systems and inter-dataset conversions are problematic, for which standardization is needed. For all QSPR endpoints, additional data requirements involve supplementing the current chemical space covered and accurately characterizing the test systems used.

Received 19th January 2017  
Accepted 20th February 2017

DOI: 10.1039/c7em00034k

rsc.li/process-impacts

## Environmental impact

Current study by our research group deals with the prediction of the fate properties of ionizable organic chemicals in the environment. In preparation of new modeling results, we have written a critical review of the current state of knowledge and knowledge gaps in the field on which our subsequent work will build. Specifically, we have identified biodegradation, sorption to sediment, direct photolysis, reaction with singlet oxygen and secondary reactive intermediates of ionizable chemicals as crucial endpoints which are not adequately covered by current fate QSPRs.

## Introduction

After anthropogenic chemicals have been utilized, they are disposed of and can enter the environment with possible detrimental effects on human and ecosystem health. The environmental fate of chemicals is relevant not only for existing (legacy) compounds,<sup>1</sup> but also for the development of new compounds.<sup>2</sup> Benign by design is a major building block on the road to green and sustainable chemistry and is of interest to general public, regulatory agencies, municipalities, companies and institutions that are responsible for wastewater treatment and clean drinking water.<sup>3</sup> The main processes governing the

fate of chemicals can be grouped into sorption, abiotic degradation, bioaccumulation and biodegradation which depend on the chemical and biological species involved, and environmental conditions applied (Fig. 1). Of these fate processes, biodegradation is often the most important factor in removal, and hence, is considered important in design, development and registration of a chemical.<sup>4</sup> Ionizable organic chemicals (IOCs) such as pharmaceuticals and herbicides represent a class of emerging pollutants of which many are inherently non-biodegradable.<sup>5–9</sup> However, assessments focus on apolar chemicals instead.<sup>10</sup> It is crucial that parameters governing the environmental fate of new compounds can already be estimated accurately during a chemical's development phase, *e.g.* to aid the development of new compounds that are non-persistent and/or non-bioaccumulative<sup>11</sup> or to predict their removal during advanced water treatment. When predicting environmental fate, it is important to differentiate factors of the test system, *i.e.*

Department of Environmental Science, Institute for Water and Wetland Research, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.  
E-mail: tom.m.nolte@gmail.com

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7em00034k

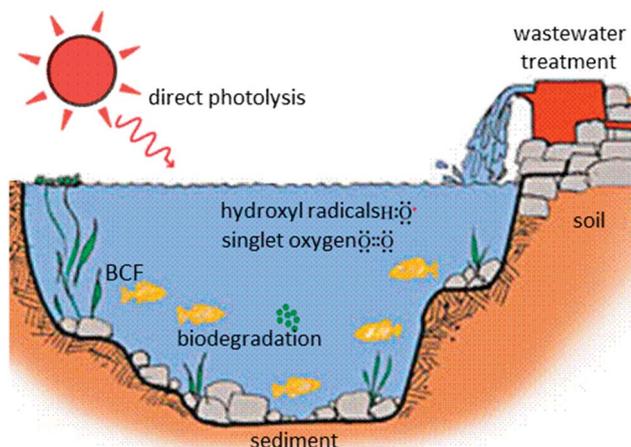


Fig. 1 Environmental fate processes of chemicals in aquatic environments can be distinguished into biodegradation, abiotic degradation, sorption and bioconcentration. Fate processes during wastewater treatment (red figure) are included in this review and include ozonation, chlorination and photolysis as well as reaction with hydroxyl radicals and singlet oxygen.

environment, from factors of the chemical. Whereas variability in biological species and environmental conditions may introduce uncertainty in the data, molecular properties form the basis of their intrinsic fate properties. In the absence of a homogeneous dataset, variability in conditions can greatly influence the reliability of predictions. This is especially the case for IOCs since their fate properties are heavily dependent on speciation and interdependencies may be involved.<sup>12,13</sup>

Quantitative Structure–Property Relationships (QSPRs) can be used to predict environmental fate by representing the chemical characteristics responsible for differences between diverse families of compounds.<sup>14,15</sup> QSPRs rely on the generation of molecular descriptors (MDs). Popular MD generation packages include Dragon,<sup>16</sup> PaDEL<sup>17</sup> and Corina<sup>18</sup> but also specialized methods exist.<sup>19,20</sup> The determination of potentially relevant MDs is done by methods such as principal component analysis (PCA) and genetic algorithms (GAs), after which the selected MDs are related to a property. Many statistical and machine learning algorithms are available such as regression, genetic algorithms, neural networks and rule-based decision making and their relative success depends on the inherent relationships between the descriptors and endpoint. Though useful methods exist (*e.g.* PBT indices),<sup>21</sup> it is well known that not one QSPR can fully describe the diverse range of fate and exposure processes a chemical may exert in environmental compartments.<sup>10,22</sup> In order to apply the optimal QSPR in risk assessment, it can be convenient to have a comprehensive database of models from which an end-user can choose. A distinction can be made between general, sub-general and specific QSPRs which are developed using general datasets, sub-general datasets and specific chemical classes, respectively. General models tend to have a wide range of applicability but relatively low accuracy, whereas class-specific models tend to have a narrow AD and high accuracy. The range of applicability

and the level of accuracy of sub-general models lie somewhere in between.<sup>23,24</sup> Due to their low cost and ease of use, QSPRs can seem very attractive from an end-user perspective. It is however vital that an end-user is able to judge whether the proposed application of the QSPR is within the limits of its design. The evaluation of QSPR models is part of good practice of model development and exploitation, and requires mechanistic understanding.<sup>25–27</sup> Evaluation commonly involves internal evaluation and external evaluation as well as an assessment of the applicability domain (AD). Internal and external evaluation is mostly performed by QSPR developers themselves using various methods,<sup>25,28,29</sup> as developers normally split their data into a training and a test set to detect overfitting. Without the characterization of a QSPR's AD, the reliability of its predictions is indeterminable; hence AD assessment is highly recommended. Not one method is best for defining the AD of QSPRs but several quantitative methods exist.<sup>30–33</sup>

The aim of this review was to evaluate the currently available public fate/exposure QSPR models for their applicability towards IOCs in aquatic environments. Doing so, we also identified the main experimental (environmental) conditions influencing the fate of IOCs and propose strategies for standardizing and homogenizing QSPR datasets to mitigate the influences of experimental conditions on internal dataset variability and predictions. In order to do this, we made an overview of aquatic QSPR models covering sorption, abiotic degradation, biodegradation and bioaccumulation. As preliminary investigation showed that this scope is very large we confined the review to generic media only, *e.g.* for general sediments and soils (not individual fractions) and degradation in representative water matrices (*e.g.* lakes, rivers, and wastewater). Only degradation models for individual transformation mechanisms (*e.g.*, biodegradation, direct and indirect photolysis, as shown in Fig. 1) were included. This was done because of the influence of environmental conditions on the overall removal<sup>34,35</sup> and inhomogeneity in existing QSPR 'meta-data', especially relevant for IOCs due to complexation and dissociation processes.<sup>13,36</sup> QSPRs compiled described the endpoint's relationship to a broad range of molecular descriptors, including calculated fingerprints, measured properties (*e.g.* water solubility, octanol ( $K_{ow}$ ) and lipid ( $K_{lipw}$ ) partitioning), substituent constants, 2D/3D descriptors, QC/QM and thermodynamic parameters.

## Methodology

### QSPR search

QSPRs for sorption, abiotic degradation, biodegradation, and bioconcentration were extracted from publications using Web of Science, PubMed and Google Scholar. Search terms are given in Table ESI0.† Searches were conducted before December 2016. QSPRs for individual reaction processes were preferred over multiple simultaneously occurring processes, *e.g.* both direct and indirect photolysis instead of total photolysis. We filtered the initial pool of 1000+ papers by manual selection of studies stating a combination of keywords in either the title or abstract using common text search methodology and expert judgement. In addition, we supplemented this study with QSPRs from

literature reviews such as Mamy *et al.* (2015) and Doucette *et al.* (2003)<sup>10,37–39</sup> and databases. The databases used were the Joint Research Centre QSAR Model Database,<sup>40</sup> the QsarDB,<sup>41</sup> the QSPR-THESAURUS database<sup>42</sup> and the LFER database from the Helmholtz Centre for Environmental Research.<sup>43</sup> Though some were compiled (Table ESI10†), no searches were performed for QSPRs predicting water treatment efficiencies using *e.g.* TiO<sub>2</sub>, ferrite, goethite, activated carbon fiber, sand/micro/nano-filtration or reverse osmosis as such methods are used only sporadically. Abiotic degradation processes resulting from reaction with relevant reactive intermediates (RIs) were identified, *i.e.* reaction with hydroxyl radicals (through UV-mediated lysis of H<sub>2</sub>O<sub>2</sub>, or from HOCl, ClO<sub>2</sub> and O<sub>3</sub>) and singlet oxygen. For biodegradation, we focused on QSPRs predicting biodegradation rate constants (or equivalent) for the parent chemical (*i.e.* primary biodegradation). Main qualitative information extracted from the model (if applicable) is as follows: (1) the model equation, (2) the types of compounds in the training and test series, and (3) the method and/or algorithm(s) used.

### QSPR selection

A total of 857 QSPR models were acquired and evaluated qualitatively based on  $r^2$ ,  $r_{\text{adj}}^2$ ,  $q^2$ , and  $N$  (train/test/ext.). Though the  $r^2$  and related coefficients are sub-optimal for comparing heterogeneous models and standardized reporting formats exist,<sup>40</sup> they are part of standard procedures, making them the most practical measure for comparison. From the compilation, we extracted models (Tables 1–6), using as inclusion criteria: (i) a training series with more than 10 compounds, and (ii) an  $r^2$  of higher than 0.6 for the training set. In cases where only  $r_{\text{adj}}^2$  or  $q^2$  values were available, we used  $r_{\text{adj}}^2 > 0.6$  and  $q^2 > 0.5$  as inclusion criteria.<sup>25–27</sup> In case multiple models were developed for the same dataset, only the ‘best performing’ global model is given.

Subsequently, we manually excluded QSPRs for sorption and bioconcentration which had a descriptive applicability domain (AD) of “hydrophobics”, “PAHs”, or “neutral”. (A)biotic degradation processes are mostly influenced by the relative presence of functional groups and electronic factors, rather than polarity;<sup>10,44,45</sup> hence, we selected QSPRs developed for ionizable functional groups.<sup>46–48</sup> We supplemented the selection with QSPRs commonly used in environmental regulation and risk assessment. These were the Sabljic *et al.* sorption models using the molecular connectivity method with improved correction factors<sup>49,50</sup> as implemented in EPI Suite,<sup>51</sup> the Franco and Trapp (2008) sorption QSPRs for IOCs<sup>52</sup> as implemented in recent SimpleTreat versions,<sup>53</sup> the Arnot-Gobas method to calculate the BCF in fish from mechanistic first principles,<sup>54</sup> and EPI Suite’s BIOWIN 5 classification QSPRs for biodegradation.<sup>51</sup> The models included were developed using a range of computing techniques such as Partial Least Squares (PLS), Multiple Linear Regression (MLR), Artificial Neural Networks (ANNs), Linear Free Energy Relationships (LFERs), allowing comparison between classical and more recent techniques. QSPRs that did not meet the selection criteria along with reporting information are given in the ESI.† As required by the OECD guidelines, we

interpret mechanistically the QSPRs and discuss the influences of the experimental (environmental) media. To this end, influencing processes were distinguished *e.g.* speciation as a function of  $\text{p}K_{\text{a}}$  and pH, binding mechanisms, temperature-dependences (Arrhenius; van’t Hoff) and medium chemistry.

## Results and discussion

### Sorption

**General.** A total of 333 QSPRs were found predicting sorption to sediments, soils, and suspended solids. QSPR correlation coefficients are highly variable depending on the compound class and on the number of compounds used to develop the relationship (Tables 1, ESI1†). Due to the variability in type, concentration and low saturation of OCs in sediments and soils, QSPRs are generally developed for  $K_{\text{oc}}$  (adsorption coefficient normalized to the OC content), but QSPRs also exist for  $K_{\text{d}}$  (linear adsorption coefficient),  $K_{\text{OM}}$  (normalized to the soil OM content), and  $K_{\text{L}}$  (Langmuir adsorption coefficient).<sup>10</sup> There were few QSPRs available that were specifically developed for IOCs. Rather, IOCs are included in QSPR datasets to develop global QSPRs with broad applicability. 17 QSPRs for  $K_{\text{oc}}$  and 4 QSPRs for  $K_{\text{d}}$  passed the selection criteria and are given in Table 1. Importantly, no appropriate QSPRs exist for sorption of IOCs specifically to sediment systems.

**Descriptors.** QSPRs were developed using heterogeneous descriptors *e.g.* experimental  $K_{\text{ow}}$  and  $\text{p}K_{\text{a}}$ ,<sup>52</sup> computed phys-chem. properties, topological fingerprints<sup>60,64</sup> and pharmacological volume of distribution (drugs).<sup>76</sup> The results indicate that  $K_{\text{ow}}$  (or  $D_{\text{ow}}$ ) alone is not an accurate predictor for sorption<sup>77,78</sup> (Table ESI1†). Notably, the QSPR  $K_{\text{oc}}$  model derived by Franco and Trapp (2008) for 93 acids ( $r^2 = 0.54$ ; 0.44) widely used in risk assessment did not pass the qualitative cut-off criteria, whereas the model for 65 bases partially met the criteria ( $r^2 = 0.76$ ; 0.55).<sup>52</sup> This indicates that uniparameter  $K_{\text{ow}}/D_{\text{ow}}$  models should be avoided.<sup>52,53,79,80</sup> The class-specific LR QSPRs by Matter-Muller *et al.* (1980)<sup>73</sup> and Sabljic *et al.* (1995)<sup>66</sup> are an exception to this rule, but these have a very limited AD. The poor predictions by  $K_{\text{ow}}$ -based QSPRs for IOCs can be attributed to specific interactions between IOCs and sorbents which are not covered by hydrophobicity.<sup>52,74</sup> The inclusion of polar and ionizable compounds in global QSPRs leads to a decrease in the correlation, confirming that processes other than hydrophobic bonding play a role<sup>56,66,74</sup> (Table 1). In order to improve predictions for ionizable chemicals, other molecular descriptors which reflect other interaction types need to be included. For example, scanning 1457 Dragon<sup>16</sup> descriptors, Goudarzi *et al.*<sup>81</sup> built MLR and ANN QSPRs for the prediction of  $K_{\text{oc}}$  for pesticides.

The FCM methods for sorption ( $K_{\text{oc}}$ ) to soil by Tao *et al.* (1999)<sup>55</sup> and Meylan *et al.* (1992)<sup>56</sup> used structural correction factors such as the number of aliphatic carbon bonds to two H-polar fragments (expressing hydrogen bonding), the number of primary nitrogens and branching. Similarly, Tao *et al.* (1999)<sup>63</sup> developed a QSPR (covering 6.4  $K_{\text{oc}}$  log-units) using topological indices and polarity factors ( $N = 543$  and  $r^2 = 0.86$ ). The more recent study by Wang *et al.* (2015)<sup>64</sup> screened a broad range of

Table 1 Compilation of QSPR models for sorption to environmental matrices with  $r^2 \geq 0.6$ ,  $N \geq 10$  and appropriate chemical space<sup>a</sup>

Endpoint	QSPR equation(s), algorithm(s), and descriptor(s) <sup>b</sup>	$r^2$ <sup>b</sup>	Training/test set size and description of chemical space	Matrix <sup>c</sup>	Reference
$K_{oc}$	$\log(K_{oc}) = \sum n_i F_i + \sum m_j F_j$ , fragment contribution method (FCM)	0.97	$N = 430/162$ . Miscellaneous organic compounds including 30 acids and 50 amines, ( $-0.31 \leq \log K_{oc} \leq 7.34$ )	Soil	55
$K_{oc}$	$\log(K_{oc}) = 0.53^1 \chi + \sum (N_p F_i) + 0.62$ , fragment contribution method (FCM)	0.955	$N = 189/205$ . Training set contained 64 non-polar and 125 polar organic compounds	Soil	56
$K_{oc}$	$\log(K_{oc}) = 2.12V_x + 0.72R_2 - 0.23 \sum a_2^H - 2.33 \sum b_2^O + 0.19$ , MLR	0.954	$N = 119$ . Miscellaneous organic compounds including acids and amines	Soil	57
$D_{oc}^c$	$\log(D_{oc}) = 0.0119V_c + 0.04875A_c - 0.726$ , MLR	0.94	$N = 29$ . C <sub>8</sub> H <sub>17</sub> N model amines (pH 6.8), >90% of species ionized	Pahokee peat soil	58
$K_{oc}$	$\log(K_{oc}) = 1.6004 - 7.7176(1.3781)I_{\text{new}} + 6.9497(1.1959)I_{\text{step}} + 0.3995(0.1472)^2 \chi + 0.7115(0.2310)^2 \chi^2$ , MLR	0.9365	$N = 49$ . Substituted anilines and phenols	Soil	59
$K_{oc}$	Least-squares support vector machine (LS-SVM), descriptors: $M_w$ , e-state, bond connectivity, and 24 fragments	0.910, (0.929; 0.874; 0.835)	$N = 571$ . Miscellaneous organic compounds, mostly non-polar	Soil	60
$K_{oc}$	$\log(K_{oc}) = aMR + b$ , LR	0.52 to 0.92	$N = 36$ . 10 esters, 8 acids, 10 amides, 8 amines	Podzol; alfisol; sediment	61
$K_{oc}$	Artificial Neural Network (ANN), descriptors: $K_{ow}$ , $\chi^V$ , $\alpha$ , $E_{\text{LUMO}}$ , $E_{\text{HOMO}}$ , $q^{\text{N}}/q^{\text{O}}$ , $\mu$ , $M_w$	0.865	$N = 42$ . Substituted anilines and phenols	Soil	62
$K_{oc}$	$\log(K_{oc}) = 0.188^1 \chi^V - 0.336^2 \chi + 0.755^6 \chi_{\text{ch}} + \sum n_i F_i + 0.922$ , MLR	0.86	$N = 543$ (400/143). Miscellaneous organic compounds. 85 nonpolar and 315 polar chemicals in the training set; 22 nonpolar and 121 polar chemicals in the validation set	Soil	63
$K_{oc}$	$\log(K_{oc}) = 0.039MLOGP2 + 0.010\alpha - 0.342O-058 - 0.069ATSC8v - 0.123n_N - 0.368n_{\text{ROH}} - 0.473P-117 + 2.335\text{pMaxA\_G/D} + 0.302\text{Mor16u} - 1.612$ , MLR	0.854	$N = 618$ . Including anilines and organic acids	Soil	64
$K_{oc}$	$\log(K_{oc}) = -2.19(\pm 0.30) + 2.10(\pm 0.14)\text{VED1} - 0.34(\pm 0.04)n_{\text{HACC}} - 0.31(\pm 0.05)\text{MAXDP} - 0.33(\pm 0.12)\text{ClC0}$ , MLR	0.82	$N = 93/550$ . Mostly non-ionic organic chemicals	Soil	65
$K_{oc}$	$\log(K_{oc}) = 0.52 \log(K_{ow}) + 1.02$ , $\log(K_{oc}) = 0.60 \log(K_{ow}) + 0.32$ , $\log(K_{oc}) = 0.62 \log(K_{ow}) + 0.85$ , LR	0.63; 0.75; 0.81	$N = 390$ . Non hydrophobics including acids and amines; 23 organic acids; 20 anilines	Soil	66
$K_{oc}$	$\log(K_{oc}) = 0.340 + 0.294\text{ch10} + 9.865 \times 10^{-2} E_{\text{minex}}^{\text{C-O}} + 12.897\text{RNBR} - 0.1407^E_2$ , MLR	0.806	$N = 322$ . Wide variety of organic chemicals	Soil	41 and 67
$K_{oc}$	LR and Artificial Neural Network (ANN), descriptors: $\chi^V$ , $K_{ow}^d$	0.77	$N = 777$ . Mostly non-ionic organic chemicals. Tested using polybrominated diphenyl ethers	Soil	51, 66 and 68
$K_{oc}$	$\log(K_{oc}) = 2.1560 + 0.42399 \log K_{ow} + 0.0027206\text{PNSA-1[Zefirov]} - 0.2413 \eta[\text{AM1}] - 0.40438\text{P}_{\text{TC-TM}}^{\text{max}}[\text{AM1}]$ , MLR	0.76	$N = 68/276$ . Including anilines and organic acids	Soil	69
$K_{oc}$	$\log(K_{oc}) = 0.0168M_0 - 0.017M_2 - 0.040M_3 + 0.19M_{\text{acc}} - 0.27M_{\text{don}} + 0.37$ , MLR	0.71	$N = 387/53$ . Miscellaneous organic compounds, (mostly $1.5 < \log(K_{oc}) < 5$ )	Soil	70

Table 1 (Contd.)

Endpoint	QSPR equation(s), algorithm(s), and descriptor(s) <sup>e</sup>	$r^{2b}$	Training/test set size and description of chemical space	Matrix <sup>f</sup>	Reference
$K_{oc}$	$\log(K_{oc}) = \log(\theta_n \times 10 \exp(0.54 \log(P_n + 1.11)) + \theta_{on} \times 10 \exp(0.11 \log(P_n + 1.54)))$ , MLR Artificial Neural Network (ANN), descriptors: $\log P$ , Mlog $P$ , $pK_{a3}$ , HF, DB	0.44, (0.54) 0.8866; 0.9367	$N = 93$ . Organic acids	Soil	52
$K_d$	Stepwise regression, partial least squares and Bayesian artificial neural networks (ANNs), descriptors: $a\_base$ , $vs\_hyd$ , $PC^+$ , $a\_don$ ; $a\_acc$ , $r_{grv}$ , $vs\_pol$ , rings, $weight^h$ $\log(K_d) = 2.61 + (0.67)\log(K_{ow})$ , LR	0.62–0.69	$N = 148$ . 44 uncharged, 60 positively and 28 negatively charged, and 16 zwitterions	Agricultural soil; digested sludge–water mixtures Suspended solids from biological treatment Suspended solids from biological treatment	71 72 73
$K_d$	$\log(K_d) = (1.74 \pm 0.46) + (0.22 \pm 0.03)\log(D) + (0.92 \pm 0.10)\alpha + (0.99 \pm 0.28)\log(Pi \times energy) - (0.85 \pm 0.17)\log(TPSA) + (0.14 \pm 0.05)\eta_{HBD} + [(0.08 \pm 0.03)\eta_{HBA}]$ , MLR	0.59	$N = 217$ . 54 pharmaceuticals	Suspended solids from biological treatment	74

<sup>a</sup>  $n_i$ ;  $m_j$ : number of the  $i$ th fragment and the  $j$ th structural factor of the chemical,  $F_i$ ;  $F_j$ : fragment constant and structural factor for the  $i$ th fragment of the chemical,  $\chi$ : molecular connectivity index,  $N_p$ : number of the fragment occurring in the chemical,  $P_r$ : fragment correction factor,  $V_s$ : McGowan's characteristic volume,  $R_3$ : excess molar refraction,  $\sum d_2^H$ ;  $\sum b_2^O$ : effective hydrogen-bond acidity and basicity,  $V_c$ : turbomole-optimized molecular volumes,  $A_i$ : surface area with an electron density of 0.2 to 0.31,  $J_{neq}$ ;  $J_{neq}$ : Balaban type parameters,  $\chi$ : Randic index,  $\chi^*$ ;  $\chi^+$ ;  $\chi^-$ ;  $\chi_c$ : molecular connectivity indices,  $M_w$ ; weight: molecular weight, e-state: Kier–Hall electrotopological state atom index,  $K_{ow}$ ;  $\log P$ ;  $\log D$ ; MLOGP2: (squared Moriguchi) octanol–water partition (distribution) coefficients,  $\alpha$ : molecular polarizability,  $E_{LUMO}$ ;  $E_{HOMO}$ : energy of highest/lowest unoccupied/occupied molecular orbitals,  $q_{si}$ : net negative atomic charges on N and O atoms,  $\mu$ : dipole moment, O-058: number of =O groups, ATSC8v: centred Broto–Moreau autocorrelation of lag 8 weighted by van der Waals volume,  $n_N$ : number of nitrogen atoms,  $n_{ROH}$ : number of hydroxyl groups, P-117: X3-P = X group (phosphate), SpMaxA G/D: normalized leading eigenvalue from the distance/distance matrix (folding degree index), Mor16u: 3D-MORSE – signal 16/unweighted,  $n_{HACC}$ ; a. acc;  $n_{HBA}$ : number of acceptor atoms for H-bonds,  $chi0$ : Kier & Hall connectivity index (order 0),  $E_{min}^{C-O}$ : lowest exchange energy (AMI) for C–O bonds, RNBR; rings: (relative) number of (benzene) rings,  $T_2^E$ : topographic electronic index (AMI) over all bonds, PNSA-1[Zefirov]: partial negative surface area,  $P_{\pi-\pi}^{max}$ [AMI]: the maximum  $\pi-\pi$  bond order,  $\eta$ [AM1]: absolute hardness,  $M_1$ ,  $M_2$ ,  $M_3$ ;  $M_{acc}$   $M_{don}$ : s-moments (surface polarization charge density), and hydrogen-bond moments,  $P_n$ ;  $K_{ow}$  of the neutral molecule,  $\theta_{n3}$ ;  $\theta$ ;  $\alpha$ : neutral, ionic and cationic fractions,  $pK_a$ : proton dissociation constant, HF: hydrophilic factor, DB: number of double bonds,  $a\_base$ : number of basic atoms,  $vs\_hyd$ : sum of the VDW surface areas of hydrophobic atoms,  $PC^+$ : total positive partial charge,  $a\_don$ ;  $\eta_{HBD}$ : number of donor atoms for H-bonds,  $r_{grv}$ : radius of gyration,  $vs\_pol$ : VDW surface areas of polar atoms,  $Pi \times energy$ : total energy associated with aromatic rings, TPSA: topological polar surface area. <sup>b</sup>  $r^2$  denotes the Pearson correlation coefficient for the training set; numbers in brackets denote results for data subsets. <sup>c</sup> Organic carbon distribution coefficient, covering an environmentally relevant pH range of 4.5–7. <sup>d</sup> EPISUITE (ver. 2009) includes the Sabijic molecular connectivity method with improved correction factors and linear regression. <sup>e</sup> Number in brackets denotes SDS. <sup>f</sup> Only MOE descriptors<sup>75</sup> are given. <sup>g</sup> Because sorption differs greatly between matrices used, large datasets including both soils and sediments are standardized using  $K_{oc} = K_d \times f_{oc}$ .

chemical descriptors but did not improve results indicating that information-rich structural fragments are best able to capture non-specific interactions between IOCs and organic carbon. Additionally, Klamt *et al.* (2002)<sup>70</sup> and Poole and Poole (1999)<sup>57</sup> developed general-purpose QSPRs for  $K_{oc}$  using quantum-chemical calculations and density functional theory (DFT) combined with a continuum solvation method ( $r^2 = 0.71$ ), and solvation parameters ( $r^2 = 0.954$ ) characterizing the contributions of different intermolecular interactions (*e.g.* hydrogen bonding, cavity formation, and dipole-dipole). Von Oepen<sup>61</sup> developed models for sorption ( $K_{oc}$ ) of esters, acids, amides and amines to sediment, podzol and alfisol soils using electronic, geometric and topological descriptors ( $r^2 = 0.52$ – $0.92$ ).

Sorption to wastewater solids as described by  $K_d$  was related by Barron *et al.* (2009)<sup>71</sup> to  $\log P$ ,  $pK_a$ , a hydrophilic factor and the number of double bonds and rings. Analysis of an inhomogeneous dataset using ANNs by Berthod *et al.* (2016)<sup>72</sup> showed the non-linear influence of hydrophobicity, charge and molecular shape using polarizability and electrostatic potential and field-related descriptors. Though the MLR  $K_d$  model for wastewater solids by Sathyamoorthy and Ramsburg (2013)<sup>74</sup> was inferior ( $r^2 = 0.59$ ) to the Barron *et al.* (2009) ( $r^2 = 0.94$ ) and Berthod *et al.* (2016) ( $r^2 = 0.62$ – $0.69$ ) QSPRs, it showed the relative importance of involved sorption processes, with  $D_{ow}$ , TPSA (topological polar surface area),  $n_{HBD}$ ,  $n_{HBA}$  (number of hydrogen donors and acceptors) and Pi.Energy (total  $\pi$  energy associated with the aromatic rings) as most important chemical descriptors. The model by Bronner and Goss (2011)<sup>82</sup> showed the relevance of van der Waals interactions, cavity formation and H-donor properties of the sorbate when sorbing to Pahokee peat. The combination of studies indicates that interaction type-specific descriptors improve predictions when homogeneous datasets are used (Table 1). Sorption is favorable when the free energy of interaction between the sorbent and sorbate is negative with the driving force of the interaction being either enthalpy or entropy.<sup>83</sup> The intermolecular interactions involved in sorption are diverse, *e.g.* van der Waals interactions, hydrophobic bonding, hydrogen bonding, charge transfer, ligand exchange, direct and induced ion-dipole and dipole-dipole interactions and chemisorption. Since many of the selected global QSPRs (Table 1) use a combination of MDs for these interaction types, their relative performance may be dependent more on the algorithm relating the MDs to the sorption endpoint.

**QSPR algorithms.** The majority of QSPRs (16) were developed using linear or multi-linear regression (LR or MLR), 3 were developed using ANNs, 2 using a FCM and 1 using a Least-Squares Support Vector Machine (LS-SVM). For sorption standardized to the fraction organic carbon ( $K_{oc}$ ), various algorithms produce satisfactory results. Evaluation of the relative performances of algorithms is difficult due to the inhomogeneity of datasets and metrics used. However, different techniques have been used on similar  $K_{oc}$  datasets<sup>56,63,66,70</sup> and  $r^2$  is often reported. The FCM QSPR for soils by Meylan *et al.* (1992)<sup>56</sup> for polar and non-polar organic compounds was tested using 11 acids and 10 anilines in the validation set ( $N = 205$ ;  $r^2 = 0.856$ ), and a similar dataset ( $N = 592$ ) was used to develop a  $K_{oc}$  model

using a FCM ( $r^2 = 0.969$ ) for miscellaneous chemicals.<sup>63</sup> It seems that the FCMs applied by Meylan *et al.* (1992) and Tao *et al.* (1999)<sup>55,56</sup> perform best, though the  $r^2$  of the validation results was significantly lower than that of the training results (0.955) in the case of Meylan *et al.* The datasets for soils and sediments originally compiled by Meylan *et al.*<sup>56</sup> and Sabljic *et al.*<sup>66</sup> were used multiple times after their original publication, using different methods. Using a SVM and MLR, Wang (2009) and Wang *et al.* (2015) developed QSPRs for organic compounds<sup>60,64</sup> with inferior predictions compared to those of the previous methods ( $r^2 = 0.835$ – $0.929$  and  $0.854$ , resp.), potentially due to the broad AD (an updated EPI Suite dataset). An entry in the QSPR thesaurus database described EPISUITE (Ver. 4.0, 2009), which was developed using molecular connectivity indices (MCI) ( $r^2 = 0.77$ ).<sup>41</sup> The datasets were also re-used by Kahn *et al.* (2005) and Gramatica *et al.* (2006) for 344 ( $r^2 = 0.76$ ) and 643 ( $r^2 = 0.82$ ) compounds, both using MLR.<sup>65,69</sup> Furthermore, Tulp *et al.* (2010) developed a MLR model for  $K_{oc}$  ( $r^2 = 0.806$ ) for 322 organic chemicals<sup>67</sup> and class-specific QSPRs were developed by Liu and Yu (2005) and Mishra *et al.* (2012) using ANNs and MLR for substituted anilines and phenols.<sup>59,62</sup> For predictions of  $K_{oc}$  by global QSPRs, FCMs generally perform better than regression methods.

The best performing QSPR predicting adsorption coefficients  $K_d$  (unstandardized to the organic carbon fraction) were the ANNs by Barron (2009) for homogeneous datasets of wastewater sludge and agricultural soil.<sup>71</sup> In contrast to ANNs, linear regressions and free energy relationships (LFERs) are based on a mechanistic understanding of the partition process.<sup>10</sup> However, different interaction types (*i.e.* descriptors) relate to the sorption endpoint in a non-linear way *via e.g.* speciation, complexation, conformation, and reorientations in OM complexes.<sup>85</sup> ANNs outperform regression-based QSPRs and LFERs<sup>60,62,71,72,81</sup> (Table 1) because their structure allows for such complex non-linear relationships between molecular descriptors and endpoints. In ANNs, the weight assigned to each neuron is not predetermined, but depends on the data and training method. Though mechanistic information can be obtained, a drawback of ANNs is their lower interpretability.<sup>72</sup> Similarly, LS-SVMs also outperformed MLR in the case of  $K_d$  prediction. Considering the similarity in the sorption media used, interaction types and descriptors selected, ANNs are likely among the most effective methods to model sorption in sediment.

**Dataset homogenization.** Variability in environmental conditions can alter the outcome of a sorption experiment, especially in the case of IOCs. Many of the developed sorption QSPRs do not account for variations in environmental conditions such as pH, OC content, CEC (cation exchange capacity), ionic strength, temperature, *etc.*<sup>86–88</sup> Many of these conditions influence the outcome of a sorption experiment, and moreover, their influence may be non-linear and involved processes may be interrelated. As a result, QSPR predictions for specific environmental matrices may be misleading. Illustratively, the ANN QSPR ( $N = 49$ ;  $r^2 = 0.89$ ) by Barron *et al.* (2009) for sorption of PhACs onto relatively homogeneous wastewater suspended solids<sup>71,74</sup> had better statistics compared to the ANN by Berthod

*et al.* (2016) for a more heterogeneous dataset of wastewater solids ( $N = 148$ ;  $r^2 = 0.62-0.69$ ).<sup>72</sup>

Various approaches are available to homogenize datasets to environmental conditions. For example, to predict  $K_d$  and  $K_{oc}$  as a function of pH, a species-specific approach can be taken weighing the contributions of the neutral and ionic fractions separately. Sorption data can be standardized using *e.g.* the Henderson–Hasselbalch equation for monovalent acids and bases:

$$D_d(\text{pH}) = \frac{K_{d,n}}{1 + 10^{a(\text{pH}-\text{p}K_a)}} + \frac{K_{d,\text{ion}}}{1 + 10^{a(\text{pH}-\text{p}K_a)}} \quad (1)$$

where  $K_{d,n}$  and  $K_{d,\text{ion}}$  are the sorption coefficients of the neutral and ionized compounds and the parameter  $a$  is 1 and  $-1$  for acids and bases, respectively. The relationship can be used to

discriminate between an observed  $K_d$  (at a specific pH), and a speciation-specific  $K_d$ , both for which QSPRs can be built.<sup>12</sup> Not only  $\text{H}^+$ , but also the presence of  $\text{Al}^{3+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ , and  $\text{Na}^+$  cations influences sorption as they determine whether binding positions in the sorbate matrix are saturated.<sup>89</sup> To account for such effects the CEC could either be (1) standardized in sorption experiments,<sup>90</sup> (2) applied in competitive cation exchange-type corrections,<sup>91</sup> or (3) included as a QSPR descriptor itself.

Normally,  $K_d$  is corrected to  $K_{oc}$  accounting for the variable fraction of organic carbon ( $f_{oc}$ ) in sediment or soil (see  $K_{oc}$  QSPRs in Table 1). However, for low values of  $f_{oc}$  ( $\sim 10^{-4}$  in some aquifers), direct sorption onto the mineral matrix becomes important.<sup>92</sup> Also, since the concentration and composition of OMs varies according to depth, nonlinear relationships between  $K_d$  and  $f_{oc}$  should be anticipated. Illustratively, the conversion of  $K_{oc}$  to  $K_d$  for 78 acidic

Table 2 Compilation of QSPR models for reaction with hydroxyl radicals (\*OH) with  $r^2 \geq 0.6$ ,  $N \geq 10$  and appropriate chemical space<sup>a</sup>

Endpoint	QSPR equation(s), algorithm(s) and descriptor(s)	$r^2$ ( $r_{\text{adj}}^2$ ) <sup>c</sup>	$q^2$ <sup>c</sup>	Training/test set size and description of chemical space	References
$\log(k_{\text{OH}})$	$\log(k_{\text{OH}}) = 2.153 + 0.895\text{DBE} - 0.134\text{WPSA}$ PCA and MLR	0.918 (0.915)	0.856	$N = 55/28/28$ . Pharmaceuticals, personal care products or organic solvents	10 and 101
$\log(k_{\text{OH}})$	Artificial Neural Network (ANN) descriptors: $n_{\text{sp}^3\text{C}}$ , $n_{\text{C}=\text{C}_{\text{ex.in}}}$ , $n_{\text{OH}_{\text{prim,sec,ter}}}$ , $n_{\text{COOH,COO}^-}$ , $n_{\text{RC}(=\text{O})\text{R}'}$ , $n_{\text{CH}=\text{O}}$ , $n_{\text{R}-\text{O}-\text{R}'}$ , $n_{\text{RCOOR}'}$ , $n_{\text{C}=\text{C}/\text{C}=\text{O}}$ , aromaticity, $n_{\text{ortho}}$ , $n_{\text{meta}}$ , $n_{\text{para}}$	0.9	0.81	$N = 146/31$ . Polar (84%) and non-polar (16%) compounds, qualitative screening of the Buxton <i>et al.</i> (1988) database	102 and 103
$\log(k_{\text{OH}})$	$k_{\text{OH}} = \sum [k_i \times \text{PIF}(\alpha\text{-group}) \times \text{PIG}(\beta\text{-group})]$ improved fragment contribution method (FCM)	0.89	—	$N = 72$ . Alkanes, alcohols, organic acids and bases	104 and 105
$k_{\text{OH}}^b$	$1000k_{\text{OH}} = 15.315 + 291.290E_{\text{HOMO}} + 0.310C_v^0 + 113.117q_{\text{H}^+}$ stepwise MLR	0.878	0.805	$N = 17$ . Chlorophenols and antimicrobial agents in a Fenton-like system (mild $\text{Cu}^{2+}/\text{H}_2\text{O}_2$ conditions)	106
$\log(k_{\text{OH}})$	$\log(k_{\text{OH}}) = 16.451 - 6.932M_e + 0.159n_{\text{DB}} - 0.679\text{CH}_2\text{RX} + 0.401n_{\text{HAcc}} - 0.460V_{\text{index}} - 0.363\text{MATS2m} - 0.362\text{Mor27p}$ ; $\log(k_{\text{OH}}) = 17.215 - 7.564M_e + 0.160n_{\text{DB}} - 0.625\text{CH}_2\text{RX} + 0.310n_{\text{HAcc}} - 0.563V_{\text{index}} - 0.362\text{MATS2m} - 0.427\text{Mor27p}$ . Forward MLRs	(0.846), (0.823)	0.804; 0.773	$N = 88/28$ ; $N = 89/29$ . Various chemical classes (phenols, polycyclic aromatic hydrocarbons, alkanes, halogenated aromatic compounds, organophosphorus compounds, <i>etc.</i> )	107
$\log(k_{\text{OH}})$	$\log(k_{\text{OH}}) = -0.628(\pm 0.408)M_w - 2.150(\pm 0.633)\text{piPC09} + 2.249(\pm 0.596)\text{Mor02e} - 0.694(\pm 0.213)\text{Mor26p} + 3.444(\pm 0.999)E_{\text{HOMO}} + 11.905(\pm 0.978)\text{GA}$ and MLR	0.803	0.66	$N = 60$ . Benzene structures with different substitute groups and other heterocycles, containing nitrogen or sulphur	108
$\log(k_{\text{OH}})$	$\log(k_{\text{OH}}) = 0.99 - 0.56q_{\text{C}_{\text{mean}}} - 0.34\text{BELV1} - 0.42\text{ATS2m} + 0.66\text{piPC03}$ MLR	0.774	0.584	$N = 122/25$ . Various organic micropollutants including solvents, pharmaceuticals and pesticides in a UV/ $\text{H}_2\text{O}_2$ process	100

<sup>a</sup> DBE: double bond equivalence, WPSA: weakly polar component of solvent accessible surface area,  $n_{\text{sp}^3\text{C}}$ : number of  $\text{sp}^3$  carbons,  $n_{\text{C}=\text{C}_{\text{ex.in}}}$ : number of external and internal  $\text{C}=\text{C}$  double bonds,  $n_{\text{OH}_{\text{prim,sec,ter}}}$ : number of primary, secondary and tertiary alcohols,  $n_{\text{COOH,COO}^-}$ : number of carboxylic acids and ions,  $n_{\text{RC}(=\text{O})\text{R}'}$ ;  $n_{\text{CH}=\text{O}}$ : number of ketones and aldehydes,  $n_{\text{R}-\text{O}-\text{R}'}$ : number of ethers,  $n_{\text{RCOOR}'}$ : number of esters,  $n_{\text{C}=\text{C}/\text{C}=\text{O}}$ : number of non-aromatic  $\text{C}=\text{C}$  and  $\text{C}=\text{O}$  conjugated bonds,  $n_{\text{ortho}}$ ;  $n_{\text{meta}}$ ;  $n_{\text{para}}$ : number of couples of functional groups in *ortho*, *meta* and *para* positions,  $k_i$ : group kinetic rate constant,  $F(\alpha\text{-group})$ ;  $F(\beta\text{-group})$ : neighboring constants of  $\alpha$ - and  $\beta$ -position groups,  $E_{\text{HOMO}}$ : energy of the highest occupied molecular orbital,  $q_{\text{H}^+}$ : most positive net charge of hydrogen atoms,  $C_v^0$ : molar heat capacity at constant volume,  $M_e$ : mean atomic Sanderson electronegativity,  $n_{\text{DB}}$ : number of double bonds,  $n_{\text{CH}_2\text{RX}}$ : number of  $\text{CH}_2\text{RX}$  (primary alkyl halides) functional groups,  $n_{\text{HAcc}}$ : number of acceptor atoms for H-bonds,  $V_{\text{index}}$ : Bataan  $V_{\text{index}}$ , MATS2m: Moran autocorrelation of lag 2 weighted by mass, Mor27p; Mor26p: 3D-Morse signal 27 and 26/weighted by polarizability,  $M_w$ : molecular weight, piPC09: topological molecular path count of order 09, Mor02e: 3D-Morse signal 02/weighted by atomic Sanderson electronegativities,  $q_{\text{C}_{\text{mean}}}$ : electronic charge on a C atom, BELV1: lowest eigenvalue no. 1 of Burden matrix/weighted by atomic van der Waals volumes, ATS2m: Broto-Moreau autocorrelation of a topological structure – lag 2/weighted by atomic masses 2D autocorrelations, and piPC03: molecular multiple path count of order 3. <sup>b</sup> In contrast to the other models (second order), the QSPR by Peng (2016) was developed for pseudo-first order kinetics. <sup>c</sup>  $r^2$ ,  $r_{\text{adj}}^2$  and  $q^2$  denote the Pearson correlation coefficient, adjusted Pearson correlation coefficient (in brackets) for the training set and  $q^2$  for the testing set, resp.

pesticides in 36 temperate soils by Kah and Brown (2007) ( $r^2 = 0.721$ )<sup>93</sup> leads to a loss in the prediction accuracy. Given that the  $f_{oc}$  of ordinary sediments is 0.001–0.025 the calculated variation is a factor of  $\sim 25$ , in agreement with the variation in the  $K_d$  data.<sup>87,94</sup> Moreover, IOCs may have affinities towards sediments and soils attributable to specific OC types.<sup>94,95</sup> For example, Bronner and Goss (2011)<sup>82</sup> and Droge *et al.* (2013)<sup>58</sup> recently developed pp-LFERs and LR QSPRs for the sorption of 129 chemicals including pesticides and PhACs ( $r^2 = 0.93$ ) and 29 highly ionized amines ( $r^2 = 0.94$ ) to Pahokee peat. Compared to normal soil, peat has relatively more carboxylic groups.<sup>96</sup> Corrections for sorption based on the OM-type are possible,<sup>93</sup> but require appropriate characterization of the matrix *e.g.* by distinguishing between different OM fractions in sediment and soil.<sup>61,95,97</sup>

### Reaction with hydroxyl radicals (\*OH)

**General.** Hydroxyl radicals (\*OH) are generated chemically during water treatment, as well as through photochemical production from DOM and  $\text{NO}_3^-$  in surface water,<sup>98,99</sup> the latter with steady state concentrations [ $^*\text{OH}$ ]<sub>ss</sub> of  $10^{-18}$  to  $10^{-16}$  M in summers. In total, 58 QSPRs were found for reaction with hydroxyl radicals (through  $\text{H}_2\text{O}_2/\text{UV}$  and Fenton-type oxidation). From these, 7 satisfactory QSPRs are given in Table 2 and discussed below. With the exception of 1 study, all QSPRs were developed for the second-order rate constant. No QSPR was specifically developed for IOCs but all partially covered the required chemical space. Though the MLR QSPR by Wols and Vries (2012)<sup>100</sup> had significantly lower performance, the dataset included the highest proportion of IOCs and may be preferred for robust predictions.

**Descriptors.** MLR results by Sudhakaran and Amy (2013)<sup>101</sup> give information on the degradation process, with descriptors for \*OH reactions being double bond equivalence (DBE), ionization potential (IP), electron affinity (EA) and weakly polar component of solvent accessible surface area (WPSA). A more specific QSPR was developed by Kusic *et al.* (2009)<sup>108</sup> for benzene structures with different substitute groups and other heterocycles containing nitrogen or sulfur ( $N = 60$ ;  $r^2 = 0.803$ ) using Dragon and QC descriptors (semi-empirical and using DFT at the B3LYP theory level and 6-31G(d,p) basis set). Peng *et al.*<sup>106</sup> used physicochemical and quantum chemical (Gaussian) descriptors to predict first-order degradation rate constants for antimicrobial agents ( $r^2 = 0.878$ ) in a Fenton system using  $\text{Cu}^{2+}/\text{H}_2\text{O}_2$ . Though the descriptors used were heterogeneous, the energy of the highest occupied molecular orbital ( $E_{\text{HOMO}}$ ) recurred in two QSPRs as well as QSPRs for other abiotic reactions.<sup>104–106,108</sup> All used descriptors relate to the relative importance of rate limiting steps of the reactions with hydroxyl radicals (\*OH), *e.g.* bond breaking, H-abstraction, transition state formation, steric interactions and diffusion.

**Algorithms.** The majority of QSPRs for \*OH reaction (5) was developed using MLR, 1 was developed using ANNs, and 1 using a FCM.<sup>100–102,107</sup> The oldest QSPR for \*OH reactions with polar and non-polar compounds was developed by Dutot *et al.* (2003)<sup>102,103</sup> using multi-layer perceptron ANNs ( $r^2 = 0.90$ ). Jin *et al.* (2015)<sup>107</sup> and Sudhakaran and Amy (2013)<sup>101</sup> used forward

MLR and PCA/MLR for QSPRs, predicting rate constants for various functional groups ( $r_{\text{adj}}^2 = 0.846$ ) and PPCPs and organic solvents ( $r^2 = 0.918$ ). A reaction specific SAR strategy was applied by Monod and Doussin (2008)<sup>104</sup> who developed a FCM QSPR for alkanes, alcohols, organic acids and bases ( $N = 72$ ;  $r^2 = 0.89$ ) allowing the user to determine not only the overall removal rate constant but also the partial reactivity of chemical functionalities following \*OH attack bringing mechanistic information on transformation products. All methods resulted in comparable performance, with discrepancies in  $r^2$  caused largely by the descriptors selected and chemical space covered. For example, using also MLR, Wols and Vries<sup>100</sup> acquired inferior predictions for rate constants of 147 different solvents, pharmaceuticals and pesticides ( $r^2 = 0.774$ ). Solvation is expected to affect the reactions for polar and ionizable functional groups due to hydrogen bonding. This may cause over- or underestimation because of invalid thermochemical additivity. A more sophisticated approach (*e.g.*, quantum mechanical calculation) may be required to investigate these effects. Illustratively, Minakata and Crittenden (2011), Minakata *et al.* (2011) and Minakata *et al.* (2014)<sup>109–111</sup> used FCMs with LFERs and transition state theory (TST) to distinguish between different reactions by \*OH ( $N = 477$ ;  $r^2 = 0.699–0.882$ ), but multifunctional group compounds were omitted.

### Chlorination (HOCl/OCl<sup>-</sup> and ClO<sub>2</sub>)

**General.** Commonly, chlorination refers to reaction with either HOCl/OCl<sup>-</sup>, which is added in the form of *e.g.* gaseous chlorine ( $\text{Cl}_2$ ), sodium hypochlorite (NaOCl) or calcium hypochlorite ( $\text{Ca}(\text{ClO})_2$ ), or reaction with chlorine dioxide ( $\text{ClO}_2$ ). The influence of pH on reaction rates through (de)protonation of the oxidant (HOCl/OCl<sup>-</sup>,  $\text{p}K_a = 7.53$ ) has been documented.<sup>112,113</sup> For example, Deborde and von Gunten (2008) described mechanistically the relationship between the apparent rate constant and the individual HOCl and OCl<sup>-</sup> reagents.<sup>114</sup> Since HOCl is a more powerful oxidant than OCl<sup>-</sup>, measurements often correspond to the HOCl reaction,<sup>114</sup> hence, all QSPRs have been developed for circumneutral pH ( $\text{pH} \sim 7$ ). A total of 26 and 17 QSPRs were found for chlorination (HOCl<sub>2</sub> and ClO<sub>2</sub>, resp.), of which 5 applicable studies are given in Table 3 and are discussed below. All QSPRs for the reaction with HOCl and ClO<sub>2</sub> were developed for the second-order rate constant.

**Descriptors.** Very recently, Lee and von Gunten (2012)<sup>115</sup> developed empirical QSPRs for reaction with HOCl/OCl<sup>-</sup> using LFERs and Taft, Hammett and empirical substituent descriptor constants for dissociated phenols, tertiary amines, primary amines, amine derivatives and aromatics ( $r^2 = 0.86–0.97$ ). Gallard and von Gunten (2002)<sup>113</sup> developed similar QSPRs for anionic compounds (Table ES13†). Class-specific models for reaction with chlorine dioxide ( $\text{ClO}_2$ ) have very recently been developed by Lee and von Gunten (2012)<sup>115</sup> using also Hammett and Taft substituent descriptors, and have good statistics ( $r^2 = 0.79–0.99$ ). Similar models were developed earlier by Tratnyek (1994) (ES14†).<sup>118</sup> For two aromatic reactants differing only in substituent(s), the change in the free energy of activation is proportional to the change in Gibbs free energy. In this case, the

Table 3 Compilation of QSPR models for reaction with hypochlorite (HOCl & OCl<sup>-</sup>) and chlorine dioxide (ClO<sub>2</sub>) with  $r^2 \geq 0.6$ ,  $N \geq 10$  and appropriate chemical space<sup>a</sup>

Endpoint	QSPR equation(s), algorithm(s), and descriptor(s)	$r^2$ <sup>b</sup>	Training/test set size and description of chemical space	Reference
$\log(k_{\text{HOCl}})$	$\log(k_{\text{HOCl}/\text{PhO}^-}) = 4.46(\pm 0.15) - 4.90(\pm 0.44) \sum (\sigma_{o,m,p}^-) \log(k_{\text{HOCl}/\text{amine}})$ $\log(k_{\text{HOCl}/\text{amine}}) = 4.79(\pm 0.29) - 1.96(\pm 0.80) \sum (\sigma^*) \log(k_{\text{HOCl}/\text{amine}})$ $9.70(\pm 0.32) - 2.10(\pm 0.18) \sum (\sigma^*)$ LRS	0.94; 0.86; 0.97	$N = 65$ . 35 dissociated phenols; 8 tertiary amines; 22 primary amines & amine derivatives	115
$\log(k_{\text{HOCl}})$	$\log(k_{\text{HOCl}}) = 0.78 \text{p}K_a - 3.63 \text{LR}$	0.93	$N = 10$ . Monovalent anionic endocrine disruptors. Rate constant for ED + HOCl > products	116
$\log(k_{\text{ClO}_2})$	$\log(k_{\text{ClO}_2/\text{PhOH}}) = 0.41(\pm 0.50) - 4.69(\pm 1.13) \sum (\sigma_{o,m,p}^+)$ ; $\log(k_{\text{ClO}_2/\text{PhO}^-}) = 8.03(\pm 0.13) - 3.24(\pm 0.28) \sum (\sigma_{o,m,p}^-)$ ; $\log(k_{\text{ClO}_2/\text{ANH}_2}) = 5.34(\pm 0.67) - 2.47(\pm 0.94) \sum (\sigma_{o,m,p}^+)$ ; $\log(k_{\text{ClO}_2/\text{amine}}) = 4.82(\pm 0.13) - 1.72(\pm 0.40) \sum (\sigma^*)$ ; $\log(k_{\text{ClO}_2/\text{amine}}) = 4.53(\pm 1.23) - 4.51(\pm 1.64) \sum (\sigma^*)$ LRS	0.86; 0.95; 0.99; 0.86; 0.79	$N = 79$ . 15 non-dissociated phenols; 32 dissociated phenols; 4 anilines; 16 tertiary amines; 12 primary and secondary amines	115
$\log(k_{\text{HOCl}})$	$\log(k_{\text{HOCl}}) = 1.14(\pm 0.26) \sigma^* + 7.24(\pm 0.17)$ ; $\log(k_{\text{HOCl}}) = -2.24(\pm 0.82) \sigma^* + 4.92(\pm 0.32)$ ; $\log(k_{\text{HOCl}}) = -11.82(\pm 1.26) \sigma_1 - 5.12(\pm 0.72)$ ; $\log(k_{\text{HOCl}}) = -3.58(\pm 1.14) \sum (\sigma_{o,p,m}) + 6.83(\pm 0.42)$ ; $\log(k_{\text{HOCl}}) = -3.87(\pm 0.66) \sum (\sigma_{o,p,m}) + 1.93(\pm 0.55)$ ; $\log(k_{\text{HOCl}}) = -3.77(\pm 0.44) \sum (\sigma_{o,p,m}) + 1.88(\pm 0.37)$ LRS	0.884; 0.907; 0.997; 0.908; 0.882; 0.943	$N = 56$ . 14 secondary and primary amines; 7 tertiary amines; 5 phenoxide ions, phenol, anisole and butylphenylether; 8 substituted phenoxide ions and 1,3-di-hydroxy-benzene anions; 22 phenoxide ions, 1,3-di-hydroxy-benzene anions; 21 phenoxide ions, 1,3-di-hydroxy-benzene anions without Cl <sub>2</sub> BO <sub>2</sub> <sup>-</sup>	114
$\log(k_{\text{ClO}_2})$	$\log(k_{\text{ClO}_2}) = 0.466E_{[2]} - 0.174E_{[4]} - 2.998 \text{MLR}$	0.8545	$N = 10$ . Anilines	117

<sup>a</sup>  $\sigma_{o,m,p}^-$ ;  $\sigma_{o,m,p}^+$ ; Hammett substituent constants,  $\sigma^*$ ; Taft substituent constants,  $\sigma_i$ ; estimated Hammett constants of the substituents on benzene (O<sup>-</sup>, OH, OCH<sub>3</sub> and OC<sub>4</sub>H<sub>9</sub>) ( $T$  22–25 °C),  $\text{p}K_a$ ; proton dissociation constant,  $E_{[2]}$ ;  $E_{[4]}$ : molecular autocorrelation topological indices. <sup>b</sup>  $r^2$  denotes the Pearson correlation coefficient for the training set.

electronic effect of the substituents directly influences the rate of the reaction. However, if a substituent is located either near or far from the reaction center (for *e.g.* multifunctional IOCs), steric influence may mask or attenuate the electronic effect. In such cases additional descriptors are required.

A different approach was taken by Lei and Snyder (2007)<sup>119</sup> who used ionization potential (IP) and 3D molecular descriptors to predict removal of endocrine disrupting compounds, pharmaceuticals and personal care products ( $N = 73$ ;  $r^2 = 0.705$ ) by chlorine (in %), see ESI3.† Deborde *et al.* (2004)<sup>116</sup> developed a QSPR for reaction with HOCl/OCl<sup>-</sup> for monovalent anionic endocrine disruptors based on  $\text{p}K_a$  ( $N = 10$ ;  $r^2 = 0.93$ ) and a similar approach was taken earlier by Rebenne (1996)<sup>120</sup> for monovalent and divalent anionic compounds. Interestingly, Gerritsen and Margerum (1990) developed a QSPR ( $r^2 = 0.993$ ) based on a Swain–Scott plot using nucleophilicity data from Hine (1962), but it was applied to only 7 inorganic anions (ESI3†).<sup>121</sup> Fan *et al.* (2005) used molecular autocorrelation topological indices (ATIs) with rate constants (ClO<sub>2</sub>) of anilines ( $N = 10$ ;  $r^2 = 0.9244$ ).<sup>117</sup> Rorije (1996) found that  $E_{\text{HOMO}}$  correlated with the kinetic rate constant for phenols, whereas Gramatica *et al.* (2005) found similar results using an additional leverage-weighted autocorrelation descriptor (ESI4†).<sup>122,130</sup>

**Algorithms.** All selected QSPRs for chlorination were class-specific regression-type QSPRs. For both HOCl and ClO<sub>2</sub>, the LFERs from Lee and von Gunten (2012)<sup>115</sup> are considered best due to their dataset size and chemical space covered, though alternative methods such as ANNs or FCMs may either optimize predictions, or enable the development of global QSPRs. For this, quantum mechanical descriptors may be needed.

### Ozonation (O<sub>3</sub>)

**General.** Ozonation of wastewater can achieve significant abatement of many organic micropollutants since ozone is a strong oxidant. A total of 89 ozonation QSPRs were found in the open literature. Following earlier work by Lee and von Gunten (2012),<sup>115</sup> a large dataset ( $N = 216$ ) for ozonation has recently been used by Lee *et al.* (2015)<sup>48</sup> to develop QSPRs for phenols ( $N = 35$ ;  $r^2 = 0.94$ – $0.95$ ), anilines ( $N = 16$ ;  $r^2 = 0.82$ – $0.86$ ), mono- and dialkoxybenzenes ( $N = 17$ ;  $r^2 = 0.80$ – $0.90$ ), trimethoxybenzenes ( $N = 4$ ;  $r^2 = 0.950$ – $0.997$ ), benzene derivatives ( $N = 40$ ;  $r^2 = 0.65$ – $0.82$ ), olefins ( $N = 45$ ;  $r^2 = 0.82$ – $0.85$ ) and aliphatic amines ( $N = 59$ ;  $r^2 = 0.67$ – $0.83$ ). Of the 6 QSPRs selected, only 1 QSPR predicts ozonation under acidic conditions (pH = 4) developed by Zhu *et al.* (2015)<sup>123</sup> ( $N = 28$ ;  $r^2 = 0.596$ – $0.803$ ). Satisfactory ozonation QSPRs are given in Table 4 and were developed for second-order reaction rate constants.

**Descriptors.** The class-specific QSPRs by Lee *et al.* (2015) were based on one descriptor only, *i.e.* the energy of the highest occupied molecular orbital ( $E_{\text{HOMO}}$ ) using Hartree–Fock and DFT calculations at the B3LYP theory level with 6-31G and 6-311++G\*\* basis sets,<sup>48</sup> representing intrinsic reactivity. Though heavily dependent on the compound class, the uni-parameter ( $E_{\text{HOMO}}$ ) ozonation QSPRs had large applicability and performed best. While most O<sub>3</sub> models used electronic and quantum-chemical descriptors, only Lee and von Gunten (2012)

Table 4 Compilation of QSPR models for reaction with ozone (O<sub>3</sub>) with  $r^2 \geq 0.6$ ,  $N \geq 10$  and appropriate chemical space<sup>a</sup>

Endpoint	QSPR equation(s), algorithm(s) and descriptor(s)	$r^{2b}$	$q^{2b}$	Training/test set size and description of chemical space	References
$\log(k_{O_3})$	$\log(k_{O_3}) (<2) = 2.327 - 2.876 \log(\text{AMW});$ $\log(k_{O_3}) (\geq 2) = 7.747 - 4.171 \log(\text{AMW}) +$ $2.382n_{\text{ArOH}}$ stepwise MLR, classification and PLR	0.964	—	$N = 22/33$ . Pharmaceutically active compounds, hormones, EDCs, pesticides, flame retardants, <i>etc.</i>	124
$\log(k_{O_3})$	$\log(k_{O_3}) = b + a \times \sum \sigma_{o,m,p}^+ \log(k_{O_3}) = b +$ $a \times \sum \sigma_{o,m,p}^- \log(k_{O_3}) = b + a \times \sum \sigma^*$ LRs	0.81–0.96	—	$N = 243$ . 24 non-dissociated phenols; 13 dissociated phenols; 14 anilines; 50 benzene derivatives; 48 olefins; 54 amines and amine derivatives; 16 aromatics; 24 aromatics	115
$\log(k_{O_3})$	$\log(k_{O_3}) = a \times E_{\text{HOMO}} + b$ LRs	0.65–0.997	—	$N = 216$ . 35 phenols; 16 anilines; 17 mono- and dialkoxybenzenes; 4 trimethoxybenzenes; 40 benzene derivatives; 45 olefins; 59 aliphatic amines	48
$\ln(k_{O_3})$	$\ln(k_{O_3}) = 43.765 + 0.195\text{DBE} - 0.28\text{WPSA}$ $- 0.855\text{IP}$ MLR	0.832	0.823	$N = 55/28$ . Pharmaceuticals, personal care products and organic solvents	10 and 101
$\ln(k_{O_3})$	$\ln(k_{O_3}) = 2.489 + 21.117f(0)_n -$ $126.519f(0)_x - 3.852q(\text{C})_{\text{min}} + 48.501f(+)_x$ MLR	0.802	0.658	$N = 28/5$ . Various compounds among dyes, fertilizers, pharmaceuticals in refinery wastewater (pH = 4)	123
$\ln(k_{O_3})$	$\ln(k_{O_3}) = 2.452 - 46.811f(+)_n + 3.517q(\text{C})_x$ $+ 11.734E_{\text{HOMO}}$ MLR	0.723	0.604	$N = 27/5$ . Various compounds among dyes, fertilizers, pharmaceuticals in refinery wastewater (pH = 7)	125

<sup>a</sup> AMW: average molecular weight,  $n_{\text{ArOH}}$ : number of phenolic groups (aromatic hydroxyls),  $\sigma_{o,m,p}^+$ ;  $\sigma_{o,m,p}^-$ ; Hammett substituent constants,  $\sigma^*$ : Taft substituent constants,  $E_{\text{HOMO}}$ : energy of the highest occupied molecular orbital, DBE: double bond equivalence, WPSA: weakly polar component of solvent accessible surface area, IP: ionisation potential,  $f(0)_n$ ;  $f(0)_x$ ;  $f(+)_n$ ;  $f(+)_x$ : Fukui reactivity indices,  $q(\text{C})_{\text{min}}$ ;  $q(\text{C})_x$ : minimum and maximum of the most negative partial charge on a main-chain carbon and carbon/nitrogen atom (non-uniformity of electric charge). <sup>b</sup>  $r^2$  and  $q^2$  denote the Pearson correlation coefficient for the training and the testing set, resp.

used Taft and Hammett substituent coefficients.<sup>115</sup> MLR results from Sudhakaran and Amy (2013)<sup>101</sup> showed the relevance of double bond equivalence (DBE), ionization potential (IP), and weakly polar component of solvent accessible surface area (WPSA) descriptors.

Lee *et al.* (2014),<sup>126</sup> Lei and Snyder (2007)<sup>119</sup> and Sudhakaran (2012)<sup>127</sup> developed QSPRs for the removal (in %) of diverse chemicals such as PPCPs, endocrine-disrupting compounds (EDCs) and solvents (Table ES14†). From this, the WPSA (the weakly polar component of surface area) appeared most critical in determining removal, as it explained 70% of variance in the data. Subsequent descriptors used were PISA (carbon  $\pi$  solvent accessible surface area), #metab (number of bonds amenable to electrophilic attack), #rtvFG (bonds unstable to nucleophilic attack).<sup>119</sup> The descriptors selected verify that electron-donating groups favor ozone reactions while electron-withdrawing groups tend to decrease the reaction rate of organic compounds.

**Algorithms.** All satisfactory QSPRs for reaction with ozone (O<sub>3</sub>) were regressions, such as the MLR QSPRs by Jin *et al.* (2012) ( $r^2 = 0.964$ ),<sup>124</sup> Sudhakaran and Amy (2013) ( $N = 55$ ;  $r^2 = 0.832$ )<sup>101</sup> and Zhu *et al.* (2014) ( $N = 27$ ;  $r^2 = 0.562$ – $0.723$ )<sup>125</sup> for diverse compounds. Inter-dataset correlations have been found between QSPR endpoints, including the correlation found by Lee and von Gunten (2012) for O<sub>3</sub>/ClO<sub>2</sub> and O<sub>3</sub>/HOCl for aromatics (both  $r^2 = 0.94$ ).<sup>115</sup> It is therefore likely that techniques other than regression may help prediction optimizations and development of global QSPRs.

### Reaction with singlet oxygen (<sup>1</sup>O<sub>2</sub>)

**General.** Singlet oxygen (<sup>1</sup>O<sub>2</sub>) can be generated from (solar) irradiation of DOM-rich water<sup>35</sup> as well as from reactions between ozone and biomolecules<sup>128</sup> and can reach steady state concentrations [<sup>1</sup>O<sub>2</sub>]<sub>ss</sub> in surface waters of 7 to 11 × 10<sup>-14</sup> M in summers. Calculations show that <sup>1</sup>O<sub>2</sub> contributes significantly to long-term degradation in surface water, making QSPRs essential for fate predictions.<sup>35</sup> From the literature, 11 QSPRs were found for oxidation by <sup>1</sup>O<sub>2</sub>. Importantly, the most recent MLR QSPR by Arnold *et al.* (2017)<sup>129</sup> does not cover cationic organic functional groups, for which models are yet to be developed. Though no QSPRs exist for a heterogeneous (multifunctional anionic and cationic) set of compounds, recent methods are discussed below. Full information about <sup>1</sup>O<sub>2</sub> QSPRs can be found in the ES19.†

**Descriptors.** The recent QSPR by Arnold *et al.* (2017)<sup>129</sup> for phenols and phenolates ( $N = 87$ ;  $r^2 = 0.71$ ) used the one-electron oxidation potential from DFT calculations. Earlier, Gramatica *et al.* (2005)<sup>122</sup> developed a similar model ( $N = 19$ ;  $r^2 = 0.935$ ) using  $E_{\text{HOMO}}$  and an autocorrelation descriptor (R3e+) using MOPAC and Dragon (ES1†). The oldest model by Rorijje and Peijnenburg (1996)<sup>130</sup> for phenols used electronic properties from MO calculations and other physicochemical properties ( $N = 21$ ;  $r^2 = 0.871$ ). These studies indicate that a larger batch of descriptors should be screened to predict <sup>1</sup>O<sub>2</sub> reaction with broader chemical functionalities.

**Algorithms.** All QSPRs found were regressions. Useful inter-dataset correlations have been found between QSPR endpoints,

including the Canonica *et al.* (2003) regression between  $^1\text{O}_2$  and  $\text{ClO}_2$  ( $r^2 = 0.95$ ) for substituted phenols,<sup>131</sup> indicating that similar algorithms for modeling  $^1\text{O}_2$  and  $\text{ClO}_2$  reactions may apply.

### Direct photolysis (quantum yield) and molar absorption

**General.** Upon irradiation using UV or sunlight, direct and indirect photolysis can occur. Direct photolysis, *i.e.* loss of pollutants due to the direct solar or ultraviolet (UV) irradiation, is usually described through molar absorbance ( $\epsilon_\lambda$ ) and corresponding quantum yield ( $\Phi_\lambda$ ), whereas indirect photolysis involves reaction with radiation-generated RIs. Most models collected for quantum yield were developed only for hydrophobic aromatics (Table ESI7†). The only model partially applicable to IOCs was the MLR quantum yield QSPR developed by Juretic (2013)<sup>132</sup> for single-benzene ring compounds with various types (hydroxy-, chloro-, nitro-, methyl-, carboxyl-, sulfo-, amino-, *etc.*) and varying numbers (1–4) of substituents ( $N = 23$ ;  $r^2 = 0.823$ ).

**Descriptors.** Juretic *et al.* (2013) showed that the molecular symmetry relative orientation and asymmetry of hydrophobic groups were the two most influential descriptors for quantum yield.<sup>132</sup>  $\epsilon_\lambda$  depended on the overall number of bonds, and particular atom-centered fragment related to substituents possessing double and single bonded heteroatoms (O or N) to C atoms, as well as the shape and size of the molecule and polarizability. For the unpublished QSPRs by Wols and Vries (2012) 5 descriptors were selected but their significance was relatively low.<sup>100</sup> The results show that a larger batch of descriptors should be screened for potential relevance to the quantum yield and molar absorption of multifunctional IOCs.

**Algorithms.** Interestingly, Wols and Vries (2012) attempted to develop QSPRs for molar absorption ( $N = 123$ ;  $r^2 = 0.53$ ) and quantum yield for diverse emerging micropollutants ( $N = 108$ ;  $r^2 = 0.49$ ) using GAs and MLR, and found that predictions were improved by classifying the compounds into different groups.<sup>100</sup> These studies indicate that alternative QSPR algorithms such as ANNs and random forests (RFs) may be needed to accurately predict  $\Phi_\lambda$  and  $\epsilon_\lambda$ .

### Abiotic hydrolysis

**General.** Chemical hydrolysis involves ‘attack’ by the electron-rich oxygen atom (nucleophile) of water molecules on an electron-poor (electrophilic) central atom. A total of 67 QSPRs were found for abiotic hydrolysis, of which satisfactory QSPRs are discussed here (for full information, see the ESI†). Firstly, the HYDROWIN program of the US EPA’s EPI Suite<sup>51,133</sup> calculates second-order acid or base-catalyzed hydrolysis rate constants at 25 °C in water for various chemical classes using LFERs and data from Mill *et al.* (1987).<sup>134</sup> Five class-specific hydrolysis models exist for brominated alkanes ( $N = 16$ ;  $r^2 = 0.770$ ), esters ( $N = 103$ ;  $r^2 = 0.974$ ), carbamates ( $N = 62$ ;  $r^2 = 0.973$ ), alkyl/phenyl-carbamates ( $N = 18$ ;  $r^2 = 0.903$ ) and (p-subst.) benzonitriles ( $N = 14$ ;  $r^2 = 0.858$ ),<sup>78</sup> see Table ESI8.† Secondly, the SPARC program<sup>135</sup> calculates hydrolysis for acid and base-catalyzed transformation and neutral catalysis under variable conditions (temperature, solvent or mixture, and catalyst) for carboxylic acid ( $N = 654$ ;  $r^2 = 0.96$  (base-catalyzed) and  $N = 667$ ;  $r^2 = 0.97$  (acid-

catalyzed)) and phosphorous acid ester ( $N = 83$ ;  $r^2 = 0.93$  (base-catalyzed)) hydrolysis.<sup>78</sup>

An additional QSPR was extracted from the QSARDB repository (ID: Q8-10-30-265)<sup>84</sup> for heterogeneous chemicals ( $N = 166$ ;  $r^2 = 0.715$ ). The data were gathered from the report by Mackay *et al.* (2006)<sup>136</sup> including quantitative and semi-quantitative data from different sources, some lacking meta-data. Because experimental protocols were not provided, ‘hydrolysis’ could imply reaction by both  $\text{H}_2\text{O}$  and  $^*\text{OH}$ . Furthermore, a PLSR model by Tsuyuri *et al.*<sup>137</sup> predicts hydrolysis for various kinds of chemicals ( $N = 103$ , of which 32 are stable against hydrolysis) including carbamate, triazine, chloroacetanilide and organophosphate pesticides ( $q^2 = 0.615$ ). Finally, a MLR model by Narasimhan *et al.* (2006)<sup>138</sup> predicted hydrolysis rate constants of 37 benzoglycolamide ester prodrugs ( $r^2 = 0.823$ ). The PLSR and MLR models by Tsuyuri *et al.*<sup>137</sup> and Narasimhan *et al.*<sup>138</sup> are not recommended due to their low performance.

**Descriptors.** EPI Suite’s hydrolysis LFERs were constructed using Hammett, Taft,  $\text{p}K_a$  and the net charge on phosphorus (for phosphate esters), whereas SPARC is commercial software and descriptors are unknown.<sup>51,135</sup> Though descriptors were unknown, comparison of 124 ester compounds showed that SPARC performs slightly better than EPI Suite based on correlation ( $r^2 = 0.98$  and  $0.965$ , resp.) and mean unsigned error, but has a lower frequency of potential significant outliers.<sup>78,139</sup>

**QSPR algorithms.** HYDROWIN uses the LFER methodology,<sup>134</sup> whereas SPARC algorithms are based on fundamental chemical structure theory by parametrization of intra- and intermolecular interactions linked by appropriate thermodynamic relationships.<sup>78</sup> Hydrolysis depends on the presence of a hydrolysable chemical group (*e.g.* esters, amides, imides, carbamates, sulfonamides, ureas, nitriles, epoxides, *etc.*) with different relative reactivities, depending on steric, electronic and resonance effects of the chemical and the acidity/alkalinity of the medium. For example, esters are hydrolyzed more readily than amides because the nitrogen atom of amides is a better electron donor by resonance which decreases the reactivity of amide carbonyl to hydrolytic attack. The HYDROWIN LFERs are strictly limited to homologous compounds because of the class-specific sensitivity of the chemical reaction, as indicated by the varying slopes and intercepts of the regressions.

### Homogenization of datasets for abiotic degradation

Reviews exist that list environmentally relevant oxidants in surface-, waste- and drinking water.<sup>140</sup> Chemical treatment of wastewater includes chlorination,<sup>141</sup> ozonation and UV treatment (standalone or combined with peroxidation), whereas other chemical additives are relatively rare (*e.g.* sulfate radical and chromium).<sup>142</sup> RIs formed from chemical additives include hydroxyl radicals ( $^*\text{OH}$ ), singlet oxygen ( $^1\text{O}_2$ ) and hypochlorous acid ( $\text{HClO}/\text{ClO}^-$ ), whereas chlorine dioxide ( $\text{ClO}_2$ ), ozone ( $\text{O}_3$ ) and hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) can react directly with the chemical. Other relevant reactive intermediates include carbonate radicals ( $\text{CO}_3^{*-}$ ), triplet organic matter ( $^3\text{NOM}$ ), superoxide ( $\text{O}_2^{*-}$ ) and sulfate radicals ( $\text{SO}_4^{*-}$ ). QSPRs have been applied to modeling the reactivity of pollutants towards

chemical additives and RIs, though research has mostly focused on modeling advanced oxidation processes including ozonation, chlorination and UV/H<sub>2</sub>O<sub>2</sub> treatment and direct photolysis (*i.e.* quantum yield  $\Phi_\lambda$ ). Available QSPRs have varying degrees of applicability towards IOCs (ESI<sup>†</sup>).<sup>129,143,144</sup> ‘Abiotic degradation’ is a summation of all non-microbial processes, *i.e.* chemical water treatment, direct/indirect photolysis, abiotic hydrolysis and thermal degradation. Combining all terms for all abiotic reaction processes results in the following expression:

$$k = \sum_1^n k_{nm}^{2nd} [CA_n] + 2.303 \int_0^\lambda (\Phi_\lambda \varepsilon_\lambda I_\lambda) d\lambda + \sum_1^j k_{ij}^{2nd} [RI_j] + k_{hydr}^{1st} \quad (2)$$

where the first, second, third and fourth terms represent the removal rate through chemical additives (CAs), direct photolysis, reactive intermediates (RIs, *i.e.* indirect photolysis and by chemical treatment), and abiotic hydrolysis, respectively.<sup>145</sup>

During wastewater treatment, water chemistry, organic matter and natural/artificial light contribute towards the generation of RIs including \*OH and <sup>1</sup>O<sub>2</sub>.<sup>146</sup> Due to their influence on generation of these RIs, environmental conditions can greatly influence the overall removal. The result is that a model built using a homogeneous but non-representative dataset may falsely predict high removal, *e.g.* because a different RI acts as the primary reagent.<sup>35</sup> Apart from modeling reactivity towards specific RIs, other strategies focus on total reactivity. Illustratively, in the ‘total photolysis’ QSPRs by Beiknejad *et al.* (2014)<sup>147</sup> and Ioel *et al.* (2009)<sup>148</sup> for azo dyes ( $N = 18$ ;  $r^2 = 0.9231$ ) and 1,4 dihydropyridine drugs ( $N = 19$ ;  $r^2 = 0.87$ ) degradation mechanisms are ‘summed’, and the same may apply for QSPRs predicting removal efficiencies chlorination and ozonation (ESI<sup>†</sup>). Moreover, Sudhakaran *et al.* (2012) showed that 4 different descriptors (MON, EA, #X, and #ringatoms) could be applied for ozonation and AOP depending on the river water type. Such QSPRs may have very narrow applicability since they do not consider individual RIs that form during the degradation process, but rather the total removal resulting from the chemical additive. Careful selection of such datasets is needed based on the methodology and medium applied to assure correct combination of QSPRs in degradation frameworks.

Perhaps the most studied influencing parameters for abiotic degradation processes are temperature and pH. The only QSPR model available for ozonation under acidic conditions (pH = 4) was from Zhu *et al.* (2015) and hence should be used under such conditions.<sup>123</sup> Whereas the effect of pH can be accounted for by the Henderson–Hasselbalch equation (eqn (1)), temperature dependence is associated with every degradation process. For most organic pollutants, the effect of temperature is accounted for through the Arrhenius relationship and its derivations. The Arrhenius relationship can be used to standardize QSPR datasets by applying temperature corrections as was done by Li *et al.* (2014) for the reaction with \*OH standardized to 298 K.<sup>149</sup>

$$k = A \exp\left(-\frac{E_a}{RT}\right) \quad (3)$$

$$\frac{k_2}{k_1} = \frac{A \exp\left(-\frac{E_a}{RT_2}\right)}{A \exp\left(-\frac{E_a}{RT_1}\right)} = \exp\left(\frac{E_a}{R} \left[\frac{1}{T_2} - \frac{1}{T_1}\right]\right) \quad (4)$$

where  $k_1$  and  $k_2$  are the reaction rate constants at temperatures  $T_1$  and  $T_2$ ,  $R$  is the universal gas constant, and  $E_a$  is the activation energy. It has been shown that the Arrhenius equation is applicable for biological treatment systems.<sup>150</sup>  $E_a$  of most organic decomposition reactions fall in the range of 50 to 100 kJ mol<sup>-1</sup> (12 to 24 kcal mol<sup>-1</sup>), with a typical value of about 79 to 84 kJ mol<sup>-1</sup> (19 to 20 kcal mol<sup>-1</sup>).<sup>151</sup> However, activation energies are reagent specific, with  $E_a$  of aqueous O<sub>3</sub>, \*OH and <sup>1</sup>O<sub>2</sub> reactions in the range of 35–50,<sup>152</sup> 10–15 (ref. 153) and 25–55 kJ mol<sup>-1</sup>.<sup>154</sup> Using such ranges, it is possible to estimate temperature effects on compound decomposition when activation energies are unknown. Illustratively, modeling efforts using specific temperature ranges ( $T = 22$ – $25$  °C;  $T = 25$  °C) yielded constructive results.<sup>144</sup>

### Biotic degradation

**General.** For most organic chemicals, biodegradation is the principal abatement process in aquatic environments and influences their behavior and associated toxicity in aquatic ecosystems.<sup>155,156</sup> Biodegradation can be distinguished in *e.g.* (1) mineralization (of *e.g.* H<sub>2</sub>O, CO<sub>2</sub>, CH<sub>4</sub> and urea), or (2) primary biodegradation (*i.e.* removal of the parent).<sup>45,157</sup> Importantly, values can differ greatly between biodegradation definitions, especially in the case of incomplete mineralization.<sup>158</sup> Aerobic aquatic mineralization has been investigated best, with standardized tests such as the OECD test series available.<sup>159,160</sup> In such tests, unspecified microorganisms interact with the test compound and its (mostly unknown) degradation products. Biodegradation is monitored by O<sub>2</sub> consumption, CO<sub>2</sub> evolution, radioactivity measurement, or DOC removal. Results for mineralization (in %) are often categorized into ‘readily’ or ‘not readily’ biodegradable. The field of biodegradation modeling was repeatedly reviewed,<sup>3,4,10,40,45,161–168</sup> and evaluation of classification methods has been done before.<sup>29,169</sup> Popular models include VEGA, EPI Suite’s BIOWIN, TOPKAT and START which were recently evaluated for personal care products, and multi-functional organics ( $N = 722$ ) by Boethling *et al.* and Pizzo *et al.*<sup>170,171</sup> Based on prediction accuracy, VEGA performed best.<sup>171</sup> Various other classification QSPRs exist but have not been externally evaluated.<sup>172</sup> Most of the current classification models use structural fragments of a variety of organic compounds. In fact, organic compounds are structurally highly diverse, so a chemical with new structural features may not be treatable by a classification QSPR, especially since many transformation reactions occur in the mineralization pathway. A different line of research tries to describe and predict products and pathways of biodegradation without focusing on quantification, such as EnviPath and EAWAG-BBD.<sup>173,174</sup> Better understanding and more precise mineralization predictions require pathway prediction and biotransformation rates to be combined. To our knowledge, the only model accounting for both is CATALOGIC which performed better than Biowin6,

Multicase and the PLS model of Loonen *et al.*<sup>164,175–177</sup> Many of the available biodegradation QSPRs are not within the scope of this review as they either predict mineralization, or give biodegradation classifications or probabilities.<sup>4,51</sup> For example, Yin *et al.* (2007) and Greaves *et al.* (2001) developed QSPRs for the removal (%) of acid dyestuffs ( $N = 20$ ;  $r^2 = 0.81$ )<sup>178</sup> and anionic water soluble dyes ( $N = 103$ ;  $r^2 = 0.593$ )<sup>179</sup> (ESI2†). In contrast, QSPRs built on primary biodegradation may be more reproducible since their described process is less complex.

A total of 33 QSPRs were found for primary biodegradation. However, no QSPRs fully passed the selection criteria; hence, the best QSPRs predicting (pseudo-)first order or second-order rate constants (or equivalent) for biodegradation of the parent compound in aqueous media are given in Table 5. The largest dataset on principal biodegradation was used by Dimitrov *et al.* (2011) who developed the CATALOGIC framework predicting biodegradation pathways and giving semi-quantitative estimations for the removal of the parent compound ( $N > 500$  individual transformations;  $r^2 = 0.79$ ).<sup>177</sup> Rapid biotransformation was predicted to occur for oxiranes, ketenes, acyl halides, thiocarboxylic acids, hydroperoxide, nitrenes and geminal diols. Apart from primary half-life, CATALOGIC provides results for BOD (%), ultimate half-life (days), quantities of parent and biodegradation products, and AD details. Media included in CATALOGIC datasets were wastewater, surface water and soils. A different QSPR by Tamm *et al.* (2010) predicts the biodegradation half-life of organic pollutants including aromatic, aliphatic and cyclic amines, ketones, alcohols and esters ( $N = 146$ ;  $r^2 = 0.87$ )<sup>41,180</sup> using 1D, 2D, and 3D theoretical and QC descriptors (at AM1 level). Data for this QSPR were extracted from the report Mackay *et al.* (2006),<sup>136</sup> containing primary biodegradation data from both mixed and pure bacterial cultures, heterogeneous test media, and estimations from die-away tests. Burgis (2012) developed QSPRs<sup>181</sup> for the pseudo first-order biodegradation rate constant in activated sludge, using nine different literature studies ( $N = 65$ ;  $r^2 = 0.49$ ), and divided into three smaller subsets ( $r^2 = 0.97$ ;  $0.88$ ;  $0.90$ ). Tebes-Stevens and Jones (2004) developed a QSPR<sup>182</sup> for microbial reductive dehalogenation of chlorinated benzenes and phenols in sediment systems ( $N = 24$ ;  $r^2 = 0.67$ ).

**Descriptors.** QSPRs were generated using both chemical descriptors and Hammett/Taft substituent constants (Table 5). Descriptors selected by Tamm *et al.* relate to the charge distributions of the compound, *e.g.* fractional positive surface area (FPSA3), negatively charged surface area, and partially charged surface area of H atoms, molecular weight, the number of halogenide groups and the lowest resonance energy (AM1) for C–Cl bonds.<sup>41,180</sup> These descriptors are intuitive since biodegradation has been shown to be influenced more by electronic properties and functional groups (*e.g.* branching and halogenation), rather than by hydrophobicity.<sup>183–185</sup> Other descriptors used are  $pK_a$ , van der Waals radius, molecular weight, and sigma inductive constants. Though the QSPR by Damborsky *et al.* (1997)<sup>183</sup> for phenols and anilines was developed using data from cultured bacterial strains from surface water ( $N = 15$ ;  $r^2 = 0.955$ ), it provided mechanistic insight *via* its descriptors. Though their narrow AD limits their general applicability (Table 5), it is

known that large and highly charged compounds are often non-biodegradable, potentially due to low bio-uptake.<sup>5,183,186</sup>

**QSPR algorithms.** For biotic degradation various methods seem to be able to acquire reasonable results. Model techniques include regression-type QSPRs and ANNs. For surface waters, both the models by Tamm *et al.* (2010) and Dimitrov (2011)<sup>84,177,180</sup> may be applicable to IOCs due to their relatively large datasets. Though CATALOGIC likely has similar performance to the model by Tamm *et al.* (2010), there is little public information on the chemical space of the training set and on the model structure, making it impossible to evaluate the relative performance of various algorithms for primary biodegradation of IOCs.

**Dataset homogenization.** Much effort has been invested in standardizing biodegradation tests.<sup>159,160</sup> While in standardized tests such as the closed bottle test (CBT, OECD-301D) or the MITI-I test (OECD-301C) the amount of inoculum is prescribed (though not rigidly), its nature (the identity of bacteria) remains undefined.<sup>45</sup> From a chemical viewpoint, a compound's reaction rate is a function not only of its chemical structure and reaction conditions, but foremost of its reaction partner, *i.e.* bacteria and their relevant enzyme(s). The diversity of microorganisms potentially able to degrade chemicals and the diversity of their catabolic enzymes are large<sup>187</sup> but can be very specific. For example, Domaradzka *et al.* (2015) observed specific degradation and concluded that only a few microorganisms may be able to degrade drugs.<sup>188</sup> The set of bacteria present in a sediment, sewage treatment plant or groundwater varies from site to site and over time. Moreover, the respective physical and chemical conditions (including the test substance) can exert pressure to which species in a microorganism population react in different ways. The only QSPR for biodegradation in wastewater made available by Burgis (2012) did not pass the selection criteria when data from different WWTPs were pooled. Possible reasons for low performance were the lack of information in the literature describing the methods and conditions for sludge biodegradation kinetic experiments.<sup>181</sup> Clearly, biodegradation data are comparable only if obtained using the same microbial inoculum. This notion has been adopted by several authors stating that data, in particular those from BOD screening tests, are “poorly reproducible”.<sup>162,166,189–192</sup> A positive biodegradation result conflicting with negative results may not necessarily be wrong. Rather, it may be caused by the presence of specific bacteria that are rarely encountered. For many compounds only a single biodegradation result from an insufficiently characterized inoculum is available; hence, we cannot decide on its reliability. Since reproducibility is a minimum requirement for scientific data, most biodegradation tests cannot be considered scientifically rigorous, and models built thereon have little structural generality and are of questionable justification.<sup>3</sup> Future work should focus on the development of testing strategies to obtain robust biodegradation data, using more standardized inoculums and conditions. Subsequently, advanced data analysis may be used to disentangle the major factors involved.<sup>193</sup>

Apart from the microbial inoculum and its catabolic enzymes, biodegradation is dependent upon temperature, degree of acclimation, metabolic cofactors ( $O_2$ , nutrients, *etc.*), cellular

Table 5 Compilation of QSPR models for biodegradation in water matrices with  $r^2 \geq 0.6$ ,  $N \geq 10$  and appropriate chemical space<sup>e</sup>

Endpoint <sup>b</sup>	QSPR equation(s), algorithm(s), and descriptor(s)	$r^2$ <sup>c</sup>	$q^2$ <sup>c</sup>	Training/test set size and description of chemical space	References
$k_b$	$\log(k_b) = -11.233r_w + 0.31467pK_a - 12.738 \text{ PLSR}$	0.955	0.949	$N = 15$ . Phenols and anilines. Degradation of anilines by an unknown bacterial strain from the Oconee River.	183
$D_{T50}$	Artificial Neural Network (ANN) descriptors: FPSA3, $E_{\text{minHes}}^{\text{C-Cl}}$ , $M_w$ , NCSA, $n_{\text{Hal}}$ , PCSAH	0.87	0.7	$N = 126/20$ . Diverse set of organic pollutants (aromatic, aliphatic and cyclic amines, ketones, alcohols, esters, etc.). Biodegradation half-life ( $D_{T50}$ , hours) in water	41 and 180
$D_{T50}$	CATALOGIC (commercial) various algorithms and descriptors used	0.79 (half-lives); 0.84 (BOD)	0.88 (BOD)	$N > 500$ principal transformations; 109 + $x$ (2 datasets). Various organic chemicals. $\log(K_{\text{ow}}) \in [2.4, 20]$ . $M_w$ (Da) $\in [67, 700]$ . Biodegradation half-life ( $D_{T50}$ in days, assuming 1 <sup>st</sup> order kinetics) of the primary compound as well as a range of other relevant endpoints and pathways	164 and 177
$k_b$	$\log(k_b) = 0.935 \log(K_{\text{ow}}) - 1.17 \sum (\sigma) + 3.09 \sigma_{L,\text{ortho}} + 0.920 \sigma_{L,\text{meta}} - 4.61 \text{ MLR}$	0.67	—	$N = 24$ . Chlorinated benzenes and phenols (sorption-corrected biodegradation 1 <sup>st</sup> order rate constant)	182
$k_b$	$\log(k_b) = 5.254 - 0.00000029 \text{rotal } E + 0.4178E_{\text{HOMO}} - 0.07669 \text{dipole} - 0.00173 \text{HoF} - 0.01654A \text{ area} + 0.0315M \text{ area} - 0.01142M_w \text{ LR}$	0.49 (full dataset); 0.97, 0.88, 0.90 (subsets)	—	$N = 82$ (65 unique compounds). Various micropollutants including pharmaceuticals. Activated sludge media. Biodegradation pseudo-1 <sup>st</sup> order reaction rate constant	181

<sup>a</sup>  $r_w$ : van der Waal's radii;  $pK_a$ : proton dissociation constant,  $M_w$ : molecular weight, FPSA3: fractional PPSA (PPSA-3/TMSA) (Zefirov) unitless fractional positive surface area,  $E_{\text{minHes}}^{\text{C-Cl}}$  [AM1]: lowest resonance energy (AM1) of C-Cl bonds, NCSA: Negatively Charged Surface Area (Zefirov),  $n_{\text{Hal}}$ : number of halogenide groups, PCSAH: Partially Charged (Zefirov) Surface Area of H atoms,  $\sigma$ : Hammett sigma constants,  $K_{\text{ow}}$ : octanol-water partition coefficient,  $\sigma_{L,\text{ortho}}$ : sigma inductive constants for the *ortho* substituents,  $\sigma_{L,\text{meta}}$ : sigma inductive constants for the *meta* substituents, total  $E$ : total molecular energy,  $E_{\text{HOMO}}$ : energy of the highest occupied molecular orbital, dipole: dipole strength, HoF: heat of formation, A area: accessible surface area, M area: molecular surface area. <sup>b</sup>  $k_b$  and  $D_{T50}$  denote the second/pseudo-first order biodegradation rate constant and half-life, resp. <sup>c</sup>  $r^2$  and  $q^2$  denote the Pearson correlation coefficient for the training set and test set, resp. <sup>d</sup> Only considers the first oxidation step by monooxygenase (phenols) and dioxygenase (anilines). Phenols were degraded by *Pseudomonas putida* U.

transport, chemical partitioning, abiotic hydrolysis, *etc.*<sup>45</sup> Though QSPRs for overall biodegradation are useful, separation of biotransformation processes (membrane transport, enzymatic binding and reaction), and distinction between different transformation reactions, *i.e.* oxidation, conjugation, *etc.*, may improve predictions and provide further mechanistic insight.<sup>194</sup> However, these variables are difficult to control, and the structure of the resultant degradation products is often difficult to assess. Illustratively, Carballeira *et al.* (2006) and Carballeira *et al.* (2004) applied 3D-QSAR/CoMFA modelling to enzyme–ligand systems to predict for pure cultures the oxidation and reduction efficiencies (in %) of carbonyls and alcohols ( $N = 46$ ;  $q^2 = 0.549-0.798$ ),<sup>194,195</sup> which may be explored further to include mixed cultures as well. Additional shortcomings include the lack of disentanglement between simultaneous degradation and sorption, obscuring the mechanism and extent of removal. In the case of soil or sediment tests, variable water/solid ratios cause the formation of anaerobic areas, and influence diffusion-driven substance transport in the test system. As the datasets used by both Tamm *et al.* and Dimitrov *et al.* (Table 5)<sup>177,180</sup> included soil and sediment data, their QSPRs may be fitted to include sorption processes. Illustratively, the successful Tebes-Stevens and Jones (2004) QSPR for

biodegradation in sediment systems was developed by correcting the data for sorption.<sup>182</sup> Tests like OECD 309 with a low concentration of suspended solids may be valuable for assessing chemical fate in aquatic environments.<sup>196</sup>

### Bioconcentration factor

**General.** Here, the bioconcentration factor (BCF) is defined as the accumulation of a chemical in or on an organism when the source of the chemical is water.<sup>197</sup> Bioconcentration plays an important role in the regulatory assessment and is useful as a first estimate of internal concentration.<sup>198</sup> Tested organisms include various fish species as well as *Daphnia magna* and *Gammarus pulex*. A total of 154 QSPRs were found for bioconcentration, of which 96 were specifically for fish, 10 for water fleas, 9 for algae, and the remaining QSPRs for other organisms such as water louse, mussels, and aquatic plants. For prediction of the fish BCF, correlation coefficients are highly variable depending on the compound class and on the number of compounds used to develop the relationship (see the ESI11†). QSPRs for the BCF in fish applicable to IOCs are given in Table 6 and described below. Although none of the models for other aquatic species met our selection criteria, the ‘best performing’ models are also discussed.

Table 6 Compilation of QSPR models for bioconcentration from water matrices with  $r^2 \geq 0.6$ ,  $N \geq 10$  and appropriate chemical space<sup>a</sup>

Endpoint	QSPR equation(s), algorithm(s), and descriptor(s)	$r^{2b}$	$q^{2b}$	Compound class(es)	References
BCF <sub>fish</sub>	Integrated model relying on the predictions CAESAR and Meylan, together with the Applicability Domain Index (ADI) provided by VEGA	0.83	—	$N = 585$ . Diverse organic chemicals	200
BCF <sub>fish</sub>	CAESAR (hybrid QSPR) descriptors: Mlog $P$ , BEHp2, AEige, GATS5v, Cl-089, X0sol, MATS5v, SsCl	0.79	—	$N = 483$ . Diverse organic chemicals	200
BCF <sub>fish</sub>	$\log(\text{BCF}) = \log[10f_n \exp(0.62 \log K_{ow} - 0.15) + 10f_d \exp(0.28 \log K_{ow} - 0.07pK_a - 0.84)]$	0.76	0.75	$N = 44/21$ . Monovalent cations	202
BCF <sub>fish</sub>	CORAL descriptors: $n_{C_{arom}}$ , HALO, $\langle C_{branch} \rangle$ , 1RING, $n_{Branch}$ , $N$	0.76	0.60	$N = 851$ . Diverse organic chemicals	201
BCF <sub>fish</sub>	Artificial Neural Network (ANN) descriptors: CIQPlog $S$ , $M_w$ , QPlogHERG, #stars, QPPCaco, WPSA, IP, #noncon, #rtFG	0.731	0.635	$N = 608/152/76$ . Large variety of organic chemicals including 138 ionizable chemicals	200
BCF <sub>fish</sub>	$\log(\text{BCF}) = \log[10f_n \exp(0.64 \log(K_{ow}) - 0.12) + 10f_d \exp(0.37 \log(K_{ow}) + 0.06pK_a - 0.51)]$	0.69	0.53	$N = 49/24$ . Monovalent anions	202
BCF <sub>fish</sub>	Meylan $\log(\text{BCF}) = -1.37^b \log(K_{ow}) + 14.4 + \sum F_i$ LRs	0.62; 0.73	—	$N = 694$ . 610 non-ionic compounds; 84 ionic compounds	199
BCF <sub>fish</sub>	BCFBFAF v3.01 $\log(\text{BCF}) = a + b \log(K_{ow})$ LRs	—	—	$N = 527$ . 466 non-ionic and 61 ionic compounds (carboxylic acids, sulfonic acids, and quats)	54 and 207

<sup>a</sup>  $K_{ow}$ ; Mlog  $P$ : Moriguchi log of the octanol–water partition coefficient,  $pK_a$ : proton dissociation constant, BEHp2: highest eigenvalue no. 2 of Burden matrix/weighted by atomic polarizabilities, AEige: absolute eigenvalue sum from the electronegativity weighted distance matrix, GATS5v: Geary autocorrelation – lag 5/weighted by atomic van der Waals volumes, Cl-089: Cl atoms attached to C1(sp<sup>2</sup>), X0sol: solvation connectivity index chi-0, MATS5v: Moran autocorrelation – lag 5/weighted by atomic van der Waals volumes, SsCl: sum of all (–Cl) e-state values in the molecule,  $f_n$ : fraction of their neutral species,  $n_{C_{arom}}$ : carbon atom involved in the aromatic system, HALO: absence of fluorine, chlorine, and bromine,  $\langle C_{branch} \rangle$  branching from aromatic carbon connected to other carbon in the aromatic system, 1RING: presence of one ring,  $N$ : presence of nitrogen atoms, CIQPlog  $S$ : conformation-independent predicted aqueous solubility,  $M_w$ : molecular weight, QPlogHERG: IC<sub>50</sub> value for blockage of HERG K<sup>+</sup> channels, #stars: druglikeness, QPPCaco: apparent Caco-2 cell permeability, WPSA: weakly polar component of the SASA (halogens, P, and S), IP: ionization potential (PM3), #noncon: number of ring atoms unable to form conjugated aromatic systems (*e.g.* sp<sup>3</sup> C), #rtFG: number of reactive functional groups, and  $F_i$ : structure-based correction factor. <sup>b</sup>  $r^2$  and  $q^2$  denote the Pearson correlation coefficient for the training set and test set, resp.

All QSPRs for the BCF in fish were developed using large datasets for multiple fish species and exposure regimes and thus have broad applicability. QSPRs including a significant proportion of IOCs were the Meylan, CORAL and CAESAR models.<sup>199–201</sup> Building on this work, Gissi *et al.* (2013) developed a consensus model based on CAESAR and Meylan, and an ANN QSPR using a similar dataset.<sup>200</sup> A QSPR by Arnot and Gobas (2003), *i.e.* BCFBAF v3.01, included ionic compounds ( $N = 466$ , of which 61 ionic) but no statistical parameters could be found.<sup>54</sup> Fu *et al.* (2009) developed QSPRs specifically for monovalent anions ( $N = 73$ ;  $r^2 = 0.69$ ) and monovalent cations ( $N = 65$ ;  $r^2 = 0.76$ ).<sup>202</sup>

Seto and Handoh<sup>203</sup> provided an overview of algal- $K_{ow}$  relationships. Of the total 9 models gathered for the BCF in algae, the QSPR by Geyer *et al.* (1984) had the widest AD with  $0.6 \geq \log(K_{ow}) \geq 6.4$  ( $N = 41$ ;  $r^2 = 0.81$ ), and included pesticides, polar compounds and other organics.<sup>204</sup> Two QSPRs for the BCF in water fleas by Meredith-Williams *et al.* (2012) were developed for IOCs ( $N = 6$ ) using the liposome/water distribution coefficient ( $r^2 = 0.89$ ) and pharmacological volume of distribution ( $r^2 = 0.5$ ).<sup>205</sup> A QSPR for the BCF in aquatic plants by Lockhart *et al.* (1983) ( $N = 10$ ;  $r^2 = 0.91$ ) for *Lemna minor* (common duckweed) included hydrophilic pesticides and PCBs.<sup>206</sup> All the developed QSPRs for the BCF in crustaceans were for hydrophobic chemicals (PAHs, PCBs, and pesticides) only (ESI11†).

**Descriptors.** Bioconcentration is generally favorable when the free energy of interaction between the sorbent and sorbate is negative with interaction in various biological tissues driven by enthalpy and entropy. The classical method to predict the BCF involves its inherent relationship with hydrophobicity as expressed by  $K_{ow}$ . For IOCs however, very diverse molecular interactions play a role and recently developed QSPRs include such compounds (and descriptors) as well. The 9 QikProp descriptors selected in the most recent model by Gissi (2013) were molecular weight, solubility, hERG  $K^+$  channel blockage, druglikeness, Caco2 cell permeability, the weakly polar component solvent accessible surface area, ionization potential, the number of non-conjugated aromatic ring atoms, and the number of reactive functional groups which relate to passive and active uptake as well as pharmacokinetic and dynamic processes.<sup>200</sup> Both Meylan (1999) and Arnot and Gobas (2003) used  $K_{ow}$  in class-specific linear regressions. CORAL used SMILES-based descriptors<sup>208</sup> and CAESAR used Dragon descriptors but details were unknown.<sup>209</sup> Since Fu *et al.* acquired satisfactory regressions based only on the fraction ionized,  $pK_a$  and  $K_{ow}$ , alternative descriptors such as the lipid/water partitioning coefficient may be used to optimize predictions.<sup>202,210,211</sup> Though no adequate QSPRs exist for non-fish species, Qin *et al.* (2010) used regression with  $K_{ow}$  and a polarity descriptor to predict uptake in seven aquatic species, including *Aliivibrio fischeri*, river bacteria, algae, *Daphnia magna* and fish. The results indicated that anionic compounds can be easily absorbed into unicellular organisms, but this is not the case for multicellular *D. magna* and fish.<sup>212</sup> However, chemical uptake may be slow relative to phytoplankton growth, preventing thermodynamic equilibrium in algal cells.<sup>213</sup>

**QSPR algorithms.** The BCF in fish has been predicted using regression-type QSPRs and ANNs, both capable of producing

satisfactory results. The QSPRs of CORAL,<sup>201,214</sup> Meylan *et al.* (1999),<sup>199</sup> CAESAR,<sup>200,209</sup> BCFBAF<sup>54</sup> and Gissi *et al.* (2013)<sup>200</sup> had comparable performances, with slightly better statistics than recent QSPRs (Table 6). Given the diversity of chemicals and variability in biological (fish) species as well as algorithms used, predictions by these global QSPRs are likely close to optimal. The local LR models by Fu *et al.* (2009) for monovalent cations and anions are preferred in case the query compound is ionizable.<sup>202</sup> In the case of the Fu *et al.* dataset, algorithms and methods such as ANNs or FCMs in combination with a broader batch of chemical descriptors may be used to optimize predictions.

**Dataset homogenization.** Variability in experimental conditions can alter the outcome of a bioconcentration experiment with important factors and processes for IOCs in fish being pH, dissolved organic matter (DOM), the contaminant concentration, the presence of co-contaminants and temperature.<sup>215</sup> The relative importance of these factors in turn depends on the chemical and fish species used because dissociation/association, speciation and physiological characteristics influence bioavailability. Standardizing the BCF data for pH may involve determining separately the BCF for both ionized and unionized fractions, using the Henderson–Hasselbalch equation (eqn (1)) or additional experimentation.<sup>202</sup> Standardizing for sorption to DOM would involve calculating the *in situ* freely available fraction of the pollutant using  $K_{oc}$  or  $K_d$  (see previous sections) for the DOM analogous to similar methods.<sup>182</sup> However, as  $K_{oc}$  and  $K_d$  themselves are influenced by pH, a multispecies approach may be needed.<sup>216</sup>

Apart from experimental conditions there are apparent intra- and inter-species differences in ADME (absorption, distribution, metabolism and excretion) processes for fish as well as non-fish species.

In the case of IOCs in fish, distribution throughout the organism and excretion may be fast whereas biotransformation may be slow or has little influence on the BCF.<sup>217</sup> Pooled QSPRs developed for various fish species do not consider the physiology of the various organisms. As a result, such pooled QSPRs have high applicability, but their prediction accuracy is reduced (see the ESI11†). Important species characteristics influencing the BCF are body size, body surface/volume-ratio, elimination processes, lipid content, feeding and burrowing behavior and life history. Generally, exchange processes between the body and environment are relatively fast in small organisms with a low lipid content because of a relatively large surface/volume-ratio and an intensive blood circulation through the skin for oxygen uptake.<sup>218</sup> Also, fish in early life stages tend to have higher specific rates of feeding, metabolism and specific surface areas.<sup>219</sup> In contrast to QSPRs, PBTK models take into account ADME processes by mathematically describing anatomical, physiological, physical and chemical descriptions of these processes. For example, chemical uptake *via* the surface area of an organisms respiratory organs competes against chemical removal. In turn, chemical removal is a combination of loss *via* the respiratory surface, growth dilution, fecal excretion and metabolic biotransformation. Prediction methods exist for metabolic transformation<sup>220</sup> as well as for tissue, plasma, organs, protein<sup>221</sup> or lipid distribution.<sup>19</sup> Such parameters can

be considered (largely) independent of physiology and hence are used to parametrize multiple species-specific PBTK models.<sup>222-224</sup> Interestingly, Armitage *et al.* (2013) developed mechanistic BCF models specifically for anionic ( $N = 149$ ;  $r^2 = 0.68$ ) and cationic ( $N = 62$ ;  $r^2 = 0.75$ ) chemicals based on physiology (ESI11†).<sup>225</sup>

## Conclusion

Many chemical fate processes simultaneously occur in environmental water matrices. For many of these processes quality fate data are lacking, and hence, fate properties need to be estimated using QSPRs. When applying QSPRs in prediction workflows and multimedia models, it is important to understand the context in which QSPRs are developed.<sup>26,226,227</sup> Despite common validation by the developing authors, independent external evaluation and detailed AD assessment is performed only sporadically.<sup>65</sup> Quantitative assessment of each models' AD requires extraction, curation and reconstruction of the datasets used. Alternatively, a more efficient solution is a qualitative evaluation based on expert knowledge combined with *ad hoc* AD characterization. To our knowledge, this is the first study that qualitatively evaluates a large collection of chemical fate QSPRs using a uniform methodology and metric. Evaluation of the relative performance of methods is hampered due to inhomogeneity in the datasets and metrics used. For several of the models evaluated the source code, dataset, and/or evaluation metrics were not publically available, most often for commercial non-regression type models.<sup>177,228</sup> If such models are to be used in regulatory assessments, the models and datasets require reconstruction or third party quality assessment.

For the majority of the relevant fate processes identified, quality QSPRs ( $r^2 > 0.6$ ;  $q^2 > 0.5$ ) exist that are applicable to IOCs (Table 7). Applicable sorption models exist for soils and wastewater solids, whereas sorption to sediment systems was inadequately covered. For relevant abiotic degradation processes, models for reaction with hydroxyl radicals (\*OH), ozone (O<sub>3</sub>), hypochlorite and chlorine dioxide (HOCl and ClO<sub>2</sub>) were found applicable, whereas adequate models for direct photolysis/UV, reaction with singlet oxygen (<sup>1</sup>O<sub>2</sub>) and secondary reactive intermediates were lacking. However measuring reaction rate constants and/or half-lives of all relevant intermediates is nearly impossible, and hence, prioritization is needed. There is a need to improve existing methods of estimating environmental primary biodegradation half-lives to provide greater confidence in chemical assessments.<sup>229</sup> QSPRs for primary biodegradation in surface water exist. However, many ionizables are inherently non-biodegradable due to their charge and may constitute a boundary region of the chemical spaces included in these models. Adequate QSPRs for primary biodegradation of IOCs in wastewater sludge and sediment systems are lacking, but promising experimental, data curation, and modeling methods are available and need to be explored. Adequate bioconcentration QSPRs only exist for fish which the end-user should select based on the query compound and the chemical space of the QSPR. The success of algorithms used differed between the endpoints, but it seems that for some

Table 7 Summary of current strategies and knowledge gaps in QSPR modeling of fate properties for ionizable organic chemicals in aquatic matrices

Process	Chemical/biological species	QSPR ( $r^2 > 0.6$ ; $q^2 > 0.5$ and AD)	Algorithms ('best')	Descriptor types ('best')	Test homogenization conditions
Sorption	Organic carbon (soil/sediment, $K_{oc}$ )	17	FCM	Fragment constants	pH; CEC; OM type
	Wastewater solids and soil, $K_d$	4; 1	ANN	2D and 3D descriptors	pH; CEC; OM type; $f_{oc}$
Abiotic degradation (photolysis)	Percent removal by adsorption	1 (nano-filtration, SI) 1 (GAC adsorption, SI)	MLR	2D and 3D descriptors	N/A
	$\epsilon$ and $\phi$ (direct photolysis)	<b>Not available</b>	N/A	N/A	pH; temperature
	*OH (indirect photolysis; water treatment)	7	FCM; ANN; MLR	Fragment constants; 2D and 3D descriptors; QC parameters	pH; temperature
Abiotic degradation (hydrolysis)	<sup>1</sup> O <sub>2</sub> (indirect photolysis; water treatment)	<b>Not available</b>	N/A	N/A	pH; temperature
	Other indirect photolysis and water treatment	(AD partially covered) (CO <sub>3</sub> <sup>-*</sup> ; <sup>3</sup> NOM, SI)	LR; MLR	QC parameters; 2D and 3D descriptors	pH; temperature
Abiotic degradation (water treatment)	Water ( $k_{hydr}$ )	2	LFER; other	Substituent constants; thermodynamic parameters	pH; temperature
	HOCl; ClO <sub>2</sub> (chlorination)	3	LR; MLR	Substituent constants	pH; temperature
Primary biodegradation	O <sub>3</sub> (ozonation)	6	LR; MLR	Substituent constants; 2D and 3D descriptors; QC parameters	pH; temperature
	Surface water <i>etc.</i> , $k_b$ and $D_{T50}$	2 (AD partially covered)	ANN; PLSR; other	2D and 3D descriptors; QC parameters	pH; biomass; DOC; degree of acclimation, <i>etc.</i>
Bioconcentration	Fish (BCF)	8	ANN; MLR; LR	2D and 3D descriptors	pH; fish species; <i>etc.</i>
	Other (BCF)	<b>Not available</b>	N/A	N/A	pH; species; <i>etc.</i>

endpoints a specific algorithm applies (e.g. ANNs for  $K_d$ , and FCMs and LR for several abiotic degradation mechanisms). This can be explained by the relative non-linearity and interdependencies between predictors used and the endpoints.

Whereas the properties of chemicals are often the theoretical basis of QSPRs, environmental factors affect test systems and create variability in the data, often in non-linear ways. In the absence of a homogeneous dataset, variability in environmental factors can greatly influence model predictions, especially in the case of IOCs. Illustratively, photolysis and tertiary treatment QSPRs predict removal as a summation of reaction processes, but due to high sensitivities and variability in chemical/biological and environmental species involved, such QSPRs have narrow applicability. QSPRs should be developed for standardized matrix characteristics that are most relevant and representative of field conditions. Modeling persistence ideally involves determining with relative certainty the reactivity and affinity towards the most important environmental reagents and matrices. The results show that experimental matrices and datasets used for QSPR modeling should be subject to further standardization and homogenization. Removal frameworks built upon QSPRs may need to include also a kinetic component to account for spatial and temporal variations in the relative concentrations of all reagents, intermediates, metabolites, and matrices. This review shows that though much progress has been made to assess the fate of IOCs in aquatic environments, several knowledge gaps remain and further data generation and modeling techniques are needed.

## List of abbreviations

ACF	Activated carbon fiber
AD	Applicability domain
ANN	Artificial neural network
ATI	Autocorrelation topological index
BCF	Bioconcentration factor
BOD	Biological oxygen demand
CBT	Closed bottle test
CEC	Cation exchange capacity
DFT	Density functional theory
DOC	Dissolved organic carbon
$D_{ow}$	Octanol water distribution coefficient
GA	Genetic algorithm
FCM	Fragment contribution method
IOC	Ionizable organic chemical
$K_{lipw}$	Lipid water partitioning coefficient
$K_{oc}$	Organic carbon partitioning coefficient
$K_{ow}$	Octanol water partitioning coefficient
LFER	Linear free energy relationship
LR	Linear regression
LS-SVM	Least squares support vector machine
MCI	Molecular connectivity index
MLR	Multiple linear regression
MO	Molecular orbital
PAHs	Polyaromatic hydrocarbons
PBTK	Physiologically based toxicokinetic
PCA	Principal component analysis

PhACs	Pharmaceutically active compounds
PPCPs	Pharmaceuticals and personal care products
PLSR	Partial least squares regression
pp-LFER	poly parameter free energy relationship
QSPRs	Quantitative structure–property relationships
RI	Reactive intermediate
SVM	Support vector machine

## Acknowledgements

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under iPiE grant agreement no. 115735, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' kind contribution.

## References

- 1 EC amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy, Brussels, 2012.
- 2 EMA, *Guideline on the chemistry of active substances*, European Medicines Agency, London, United Kingdom, 2015.
- 3 C. Rucker and K. Kummerer, Modeling and predicting aquatic aerobic biodegradation – a review from a user's perspective, *Green Chem.*, 2012, **14**, 875–887.
- 4 M. Pavan and A. P. Worth, *Review of QSAR Models for Ready Biodegradation*, European Commission Directorate General Joint Research Centre, Ispra, VA, Italy, 2006.
- 5 C. Boix, M. Ibanez, J. V. Sancho, J. R. Parsons, P. de Voogt and F. Hernandez, Biotransformation of pharmaceuticals in surface water and during waste water treatment: identification and occurrence of transformation products, *J. Hazard. Mater.*, 2016, **302**, 175–187.
- 6 L. Toräng, *Biodegradation rates of chemicals in surface water and groundwater assessed in batch simulation tests*, Technical University of Denmark, Lyngby, Denmark, 2003.
- 7 P. D. Capel and S. J. Larson, A Chemodynamic Approach for Estimating Losses of Target Organic-Chemicals from Water during Sample Holding Time, *Chemosphere*, 1995, **30**, 1097–1107.
- 8 G. Gasser, I. Pankratov, S. Elhanany, P. Werner, J. Gun, F. Gelman and O. Lev, Field and laboratory studies of the fate and enantiomeric enrichment of venlafaxine and *O*-desmethylvenlafaxine under aerobic and anaerobic conditions, *Chemosphere*, 2012, **88**, 98–105.
- 9 Q. T. Liu, R. I. Cumming and A. D. Sharpe, Photo-induced environmental depletion processes of beta-blockers in river waters, *Photochem. Photobiol. Sci.*, 2009, **8**, 768–777.
- 10 L. Mamy, D. Patureau, E. Barriuso, C. Bedos, F. Bessac, X. Louchart, F. Martin-Laurent, C. Miege and P. Benoit, Prediction of the Fate of Organic Compounds in the Environment From Their Molecular Properties: A Review, *Crit. Rev. Environ. Sci. Technol.*, 2015, **45**, 1277–1377.

- 11 P. G. Jessop, F. Ahmadpour, M. A. Buczynski, T. J. Burns, N. B. Green, R. Korwin, D. Long, S. K. Massad, J. B. Manley, N. Omidbakhsh, R. Pearl, S. Pereira, R. A. Predale, P. G. Sliva, H. VanderBilt, S. Weller and M. H. Wolf, Opportunities for greener alternatives in chemical formulations, *Green Chem.*, 2015, **17**, 2664–2678.
- 12 A. Franco, W. Fu and S. Trapp, Influence of soil pH on the sorption of ionizable chemicals: modeling advances (vol. 28, pg. 458, 2009), *Environ. Toxicol. Chem.*, 2009, **28**, 2018.
- 13 N. Rioja, P. Benguria, F. J. Penas and S. Zorita, Competitive removal of pharmaceuticals from environmental waters by adsorption and photocatalytic degradation, *Environ. Sci. Pollut. Res.*, 2014, **21**, 11168–11177.
- 14 J. C. Dearden, The History and Development of Quantitative Structure–Activity Relationships (QSARs), *International Journal of Quantitative Structure-Property Relationships*, 2016, **1**, 1–44.
- 15 A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. H. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard and A. Tropsha, QSAR Modeling: Where Have You Been? Where Are You Going To?, *J. Med. Chem.*, 2014, **57**, 4977–5010.
- 16 R. Todeschini, V. Consonni, A. Mauri and M. Pavan, *DRAGON Software version 3.0*, Milano, Italy, 2003.
- 17 C. W. Yap, PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints, *J. Comput. Chem.*, 2011, **32**, 1466–1474.
- 18 J. Sadowski, J. Gasteiger and G. Klebe, Comparison of Automatic 3-Dimensional Model Builders Using 639 X-Ray Structures, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 1000–1008.
- 19 T. Ingram, S. Storm, L. Kloss, T. Mehling, S. Jakobtorweihen and I. Smirnova, Prediction of Micelle/Water and Liposome/Water Partition Coefficients Based on Molecular Dynamics Simulations, COSMO-RS, and COSMOmic, *Langmuir*, 2013, **29**, 3527–3537.
- 20 E. M. Larson, Chem3D, MOPAC & ChemDraw, *Chem. Listy*, 2006, **100**, A14–A15.
- 21 P. Gramatica, S. Cassani and A. Sangion, PBT assessment and prioritization by PBT index and consensus modeling: comparison of screening results from structural models, *Environ. Int.*, 2015, **77**, 25–34.
- 22 TWG (Quantitative) Structure Activity Relationship [(Q)SAR] Guidance Document 2012.
- 23 E. A. Helgee, L. Carlsson, S. Boyer and U. Norinder, Evaluation of Quantitative Structure–Activity Relationship Modeling Strategies: Local and Global Models, *J. Chem. Inf. Model.*, 2010, **50**, 677–689.
- 24 R. Cox, D. V. S. Green, C. N. Luscombe, N. Malcolm and S. D. Pickett, QSAR workbench: automating QSAR modeling to drive compound design, *J. Comput.-Aided Mol. Des.*, 2013, **27**, 321–336.
- 25 P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR Comb. Sci.*, 2007, **26**, 694–701.
- 26 A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Mol. Inf.*, 2010, **29**, 476–488.
- 27 H. R. Ravichandran Veerasamy, A. Jain, S. Sivadasan, C. P. Varghese and R. K. Agrawal, Validation of QSAR Models – Strategies and Importance, *Int. J. Drug Des. Discovery*, 2011, **2**, 511–519.
- 28 G. Schuurmann, R. U. Ebert, J. W. Chen, B. Wang and R. Kuhne, External Validation and Prediction Employing the Predictive Squared Correlation Coefficient – Test Set Activity Mean vs. Training Set Activity Mean, *J. Chem. Inf. Model.*, 2008, **48**, 2140–2145.
- 29 J. Devillers, P. Pandard and B. Richard, External validation of structure–biodegradation relationship (SBR) models for predicting the biodegradability of xenobiotics, *SAR QSAR Environ. Res.*, 2013, **24**, 979–993.
- 30 F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni and R. Todeschini, Comparison of Different Approaches to Define the Applicability Domain of QSAR Models, *Molecules*, 2012, **17**, 4791–4810.
- 31 P. Carrio, O. Lopez, F. Sanz and M. Pastor, eTOXlab, an open source modeling framework for implementing predictive models in production environments, *J. Cheminf.*, 2015, **7**, 8.
- 32 P. Carrio, M. Pinto, G. Ecker, F. Sanz and M. Pastor, Applicability Domain Analysis (ADAN): A Robust Method for Assessing the Reliability of Drug Property Predictions, *J. Chem. Inf. Model.*, 2014, **54**, 1500–1511.
- 33 D. M. Gadaleta, G. F. Mangiatordi, M. Catto, A. Carotti and O. Nicolotti, Applicability Domain for QSAR Models: Where Theory Meets Reality, *International Journal of Quantitative Structure-Property Relationships*, 2016, **1**, 45–63.
- 34 L. S. Lautz, J. Struijs, T. M. Nolte, A. M. Breure, E. van der Grinten, D. van de Meent and R. van Zelm, Evaluation of SimpleTreat 4.0: simulations of pharmaceutical removal in wastewater treatment plant facilities, *Chemosphere*, 2017, **168**, 870–876.
- 35 M. Minella, V. Maurino, C. Minero and D. Vione, Modelling photochemical transformation of emerging organic pollutants in surface waters: effect of water level fluctuations following outflow or evaporation, relevant to arid and semi-arid environments, *Int. J. Environ. Anal. Chem.*, 2013, **93**, 1698–1717.
- 36 M. P. S. Petrovic and D. Barcelo, *Analysis, Removal, Effects and Risk of Pharmaceuticals in the Water Cycle, Occurrence and Transformation in the Environment*, Elsevier, Oxford, UK, 2nd edn, 2013, vol. 62.
- 37 W. J. Doucette, Quantitative structure–activity relationships for predicting soil-sediment sorption coefficients for organic chemicals, *Environ. Toxicol. Chem.*, 2003, **22**, 1771–1788.
- 38 B. M. S. Gawlik, N. Sotiriou, E. A. Feicht, S. Schulte-Hostede and A. Kettrup, Alternatives for the determination of the soil adsorption coefficient,  $K_{oc}$ , of non-ioninorganic compounds: a review, *Chemosphere*, 1997, **34**, 2525–2551.
- 39 R. D. Y. S. Wauchope, J. B. H. J. Linders, R. Kloskowski, K. Tanaka, B. Rubin, A. Katayama, W. Kördel, Z. Gerstl, M. Lane and J. B. Unsworth, Pesticide soil sorption parameters: theory, measurement, uses, limitations and reliability, *Pest Manage. Sci.*, 2002, **58**, 419–445.

- 40 M. Pavan and A. P. Worth, Publicly-accessible QSAR software tools developed by the Joint Research Centre, *SAR QSAR Environ. Res.*, 2008, **19**, 785–799.
- 41 V. Ruusmann, S. Sild and U. Maran, QSAR DataBank repository: open and linked qualitative and quantitative structure–activity relationship models, *J. Cheminf.*, 2015, **7**, 32.
- 42 I. Sushko, S. Novotarskyi, R. Korner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, R. Todeschini, A. Varnek, G. Marcou, P. Ertl, V. Potemkin, M. Grishina, J. Gasteiger, C. Schwab, I. I. Baskin, V. A. Palyulin, E. V. Radchenko, W. J. Welsh, V. Kholodovych, D. Chekmarev, A. Cherkasov, J. Aires-de-Sousa, Q. Y. Zhang, A. Bender, F. Nigsch, L. Patiny, A. Williams, V. Tkachenko and I. V. Tetko, Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information, *J. Comput.-Aided Mol. Des.*, 2011, **25**, 533–554.
- 43 S. Endo, N. Watanabe, N. Ulrich, G. Bronner and K.-U. Goss, *UFZ-LSER database v 3.1*, Helmholtz Zentrum für Umweltforschung – UFZ, Leipzig, Germany, 2015.
- 44 J. C. Dearden and R. M. Nicholson, The Prediction of Biodegradability by the Use of Quantitative Structure–Activity–Relationships – Correlation of Biological Oxygen-Demand with Atomic Charge Difference, *Pestic. Sci.*, 1986, **17**, 305–310.
- 45 J. W. Raymond, T. N. Rogers, D. R. Shonnard and A. A. Kline, A review of structure-based biodegradation estimation methods, *J. Hazard. Mater.*, 2001, **84**, 189–215.
- 46 J. R. Kanofsky and P. Sima, Singlet Oxygen Production from the Reactions of Ozone with Biological Molecules, *J. Biol. Chem.*, 1991, **266**, 9039–9042.
- 47 K. Fenner, S. Canonica, L. P. Wackett and M. Elsner, Evaluating Pesticide Degradation in the Environment: Blind Spots and Emerging Opportunities, *Science*, 2013, **341**, 752–758.
- 48 M. Lee, S. G. Zimmermann-Steffens, J. S. Arey, K. Fenner and U. von Gunten, Development of Prediction Models for the Reactivity of Organic Compounds with Ozone in Aqueous Solution by Quantum Chemical Calculations: The Role of Delocalized and Localized Molecular Orbitals, *Environ. Sci. Technol.*, 2015, **49**, 9925–9935.
- 49 J. R. Baker, J. R. Mihelcic and A. Sabljic, Reliable QSAR for estimating  $K_{oc}$  for persistent organic pollutants: correlation with molecular connectivity indices, *Chemosphere*, 2001, **45**, 213–221.
- 50 A. Sabljic and M. Protic, Relationship between Molecular Connectivity Indexes and Soil Sorption Coefficients of Polycyclic Aromatic-Hydrocarbons, *Bull. Environ. Contam. Toxicol.*, 1982, **28**, 162–165.
- 51 USEPA, *Estimation Programs Interface Suite™ for Microsoft® Windows, v. 4.11*, Washington, DC, USA, 2012.
- 52 A. Franco and S. Trapp, Estimation of the soil–water partition coefficient normalized to organic carbon for ionizable organic chemicals, *Environ. Toxicol. Chem.*, 2008, **27**, 1995–2004.
- 53 A. Franco, J. Struijs, T. Gouin and O. R. Price, Evolution of the Sewage Treatment Plant Model SimpleTreat: Applicability Domain and Data Requirements, *Integr. Environ. Assess. Manage.*, 2013, **9**, 560–568.
- 54 J. A. Arnot and F. A. P. C. Gobas, A food web bioaccumulation model for organic chemicals in aquatic ecosystems, *Environ. Toxicol. Chem.*, 2004, **23**, 2343–2355.
- 55 S. Tao, H. S. Piao, R. Dawson, X. X. Lu and H. Y. Hu, Estimation of organic carbon normalized sorption coefficient ( $K_{oc}$ ) for soils using the fragment constant method, *Environ. Sci. Technol.*, 1999, **33**, 2719–2725.
- 56 W. Meylan, P. H. Howard and R. S. Boethling, Molecular Topology Fragment Contribution Method for Predicting Soil Sorption Coefficients, *Environ. Sci. Technol.*, 1992, **26**, 1560–1567.
- 57 S. K. Poole and C. F. Poole, Chromatographic models for the sorption of neutral organic compounds by soil from water and air, *J. Chromatogr. A*, 1999, **845**, 381–400.
- 58 S. T. J. Droge and K. U. Goss, Development and Evaluation of a New Sorption Model for Organic Cations in Soil: Contributions from Organic Matter and Clay Minerals, *Environ. Sci. Technol.*, 2013, **47**, 14233–14241.
- 59 M. S. S. Mishra, R. S. Nigam and V. Pandey, QSAR analysis of soil sorption coefficients for polar organic chemicals: substituted anilines & phenols, *Int. J. Pharm. Life Sci.*, 2012, **3**, 1660–1665.
- 60 B. Wang, J. W. Chen, X. H. Li, Y. N. Wang, L. Chen, M. Zhu, H. Y. Yu, R. Kuhne and G. Schuurmann, Estimation of Soil Organic Carbon Normalized Sorption Coefficient ( $K_{oc}$ ) Using Least Squares-Support Vector Machine, *QSAR Comb. Sci.*, 2009, **28**, 561–567.
- 61 B. Vonoepen, W. Kordel, W. Klein and G. Schuurmann, Predictive Qspr Models for Estimating Soil Sorption Coefficients – Potential and Limitations Based on Dominating Processes, *Sci. Total Environ.*, 1991, **109**, 343–354.
- 62 G. S. Liu and J. G. Yu, QSAR analysis of soil sorption coefficients for polar organic chemicals: substituted anilines and phenols, *Water Res.*, 2005, **39**, 2048–2055.
- 63 S. Tao and X. X. Lu, Estimation of organic carbon normalized sorption coefficient ( $K_{oc}$ ) for soils by topological indices and polarity factors', *Chemosphere*, 1999, **39**, 2019–2034.
- 64 Y. Wang, J. W. Chen, X. H. Yang, F. Lyakurwa, X. H. Li and X. L. Qiao, *In silico* model for predicting soil organic carbon normalized sorption coefficient ( $K_{oc}$ ) of organic chemicals, *Chemosphere*, 2015, **119**, 438–444.
- 65 P. Gramatica, E. Gianì and E. Papa, Statistical external validation and consensus modeling: a QSPR case study for  $K_{oc}$  prediction, *J. Mol. Graphics Modell.*, 2007, **25**, 755–766.
- 66 A. Sabljic, H. Gusten, H. Verhaar and J. Hermens, Qsar Modeling of Soil Sorption – Improvements and Systematics of  $\log K_{oc}$  vs.  $\log K_{ow}$  Correlations, *Chemosphere*, 1995, **31**, 4489–4514.
- 67 I. T. Tulp, G. Karelson, D. Dobchev, K. Tämm, J. Jänes, E. Härk and M. Karelson, *QSAR Model for Organic Carbon-*

- sorption Partition Coefficient ( $\log K_{oc}$ ) QMRF Entry in QSARDB (JRC Inventory), 2010.
- 68 I. V. Tetko, P. Sopasakis, P. Kunwar, S. Brandmaier, S. Novoratskyi, L. Charochkina, V. Prokopenko and W. J. G. M. Peijnenburg, Prioritisation of Polybrominated Diphenyl Ethers (PBDEs) by Using the QSPR-THESAURUS Web Tool, *ATLA, Altern. Lab. Anim.*, 2013, **41**, 127–135.
- 69 I. Kahn, D. Fara, M. Karelson, U. Maran and P. L. Andersson, QSPR treatment of the soil sorption coefficients of organic pollutants, *J. Chem. Inf. Model.*, 2005, **45**, 94–105.
- 70 A. Klamt, F. Eckert and M. Diedenhofen, Prediction of soil sorption coefficients with a conductor-like screening model for real solvents, *Environ. Toxicol. Chem.*, 2002, **21**, 2562–2566.
- 71 L. Barron, J. Havel, M. Purcell, M. Szpak, B. Kelleher and B. Paull, Predicting sorption of pharmaceuticals and personal care products onto soil and digested sludge using artificial neural networks, *Analyst*, 2009, **134**, 663–670.
- 72 L. Berthod, D. C. Whitley, G. Roberts, A. Sharpe, R. Greenwood and G. A. Mills, Quantitative structure–property relationships for predicting sorption of pharmaceuticals to sewage sludge during waste water treatment processes, *Sci. Total Environ.*, 2017, **579**, 1512–1520.
- 73 C. Mattermuller, W. Gujer, W. Giger and W. Stumm, Non-Biological Elimination Mechanisms in a Biological Sewage-Treatment Plant, *Prog. Water Technol.*, 1980, **12**, 299–314.
- 74 S. Sathyamoorthy and C. A. Ramsburg, Assessment of quantitative structural property relationships for prediction of pharmaceutical sorption during biological wastewater treatment, *Chemosphere*, 2013, **92**, 639–646.
- 75 Inc., C. C. G, *Molecular Operating Environment (MOE)*, Montreal, QC, Canada, 2013.
- 76 M. Williams, P. L. Ong, D. B. Williams and R. S. Kookana, Estimating the Sorption of Pharmaceuticals Based on Their Pharmacological Distribution, *Environ. Toxicol. Chem.*, 2009, **28**, 2572–2579.
- 77 J. Tolls, Sorption of veterinary pharmaceuticals in soils: a review, *Environ. Sci. Technol.*, 2001, **35**, 3397–3406.
- 78 C. J. van Leeuwen and T. G. Vermeire, *Risk Assessment of Chemicals: An Introduction*, Springer, 2 edn, 2007.
- 79 R. Oldenkamp, M. A. J. Huijbregts, A. Hollander, A. Versporten, H. Goossens and A. M. J. Ragas, Spatially explicit prioritization of human antibiotics and antineoplastics in Europe, *Environ. Int.*, 2013, **51**, 13–26.
- 80 B. Roig, *Pharmaceuticals in the Environment: Current Knowledge and Need Assessment to Reduce Presence and Impact*, IWA Publishing, London, UK, 2010.
- 81 N. Goudarzi, M. Goodarzi, M. C. U. Araujo and R. K. Galvao, QSPR Modeling of Soil Sorption Coefficients ( $K_{oc}$ ) of Pesticides Using SPA-ANN and SPA-MLR, *J. Agric. Food Chem.*, 2009, **57**, 7153–7158.
- 82 G. Bronner and K. U. Goss, Predicting Sorption of Pesticides and Other Multifunctional Organic Chemicals to Soil Organic Carbon, *Environ. Sci. Technol.*, 2011, **45**, 1313–1319.
- 83 J. W. Hamaker and J. M. Thompson, Adsorption, *In Organic Chemicals in the Soil Environment*, Dekker, New York, 1972, vol. 1.
- 84 V. Ruusmann, S. Sild and U. Maran, QSAR DataBank repository: open and linked qualitative and quantitative structure–activity relationship models, *J. Cheminf.*, 2015, **7**, 32.
- 85 J. T.-B. S. Schwarz, K. U. Eckhardt and H. R. Schulten, Sorption of Sulfonamide Antibiotics to Soil Organic Sorbents: Batch Experiments with Model Compounds and Computational Chemistry, *ISRN Soil Sci.*, 2012, 159189.
- 86 R. Hance, Observations on the relationship between the adsorption of diuron and the nature of the adsorbent, *Weed Res.*, 1965, **5**, 108–114.
- 87 A. Delle Site, Factors affecting sorption of organic compounds in natural sorbent/water systems and sorption coefficients for selected pollutants. A review, *J. Phys. Chem. Ref. Data*, 2001, **30**, 187–439.
- 88 D. Panagopoulos, *Partitioning and Persistence of Volatile Methylsiloxanes in Aquatic Environments*, Stockholm University, Stockholm, 2016.
- 89 D. S. Burden and J. L. Sims, *Ground water issue: fundamentals of soil science as applicable to management of hazardous wastes*, United States Environmental Protection Agency, 1999.
- 90 S. E. M. Oufqir, M. Alaoui El Belghiti, A. Zrineh and M. El Azzouzi, Adsorption of imazethapyr on six agricultural soils of Morocco: evaluation of the impact of soil properties, *Arabian J. Chem.*, 2013, **36**, DOI: 10.1016/j.arabc.2013.11.030.
- 91 J. R. Fabrega, C. T. Jafvert, H. Li and L. S. Lee, Modeling competitive cation exchange of aromatic amines in water-saturated soils, *Environ. Sci. Technol.*, 2001, **35**, 2727–2733.
- 92 H. F. Hemond and E. J. Fechner, *Chemical Fate and Transport in the Environment*, Academic Press, San Diego, USA, 1994.
- 93 M. Kah and C. D. Brown, Prediction of the adsorption of ionizable pesticides in soils, *J. Agric. Food Chem.*, 2007, **55**, 2312–2322.
- 94 J. J. Hassett, J. C. Means, W. L. Banwart and S. G. Wood, *Sorption Properties of Sediment and Energy-Related Pollutants*, USEPA, Athens GA, 1980.
- 95 D. R. Garbarini and L. W. Lion, Influence of the Nature of Soil Organics on the Sorption of Toluene and Trichloroethylene, *Environ. Sci. Technol.*, 1986, **20**, 1263–1269.
- 96 M. Csubák, *Characterization of humic acids of different main type of soils*, Department of Soil Science, Faculty of Agriculture, University of Debrecen, Debrecen, Hungary, accessed online on 07-03-2017 at [http://www.date.hu/~csubak/Publikaciok/geoderma\\_CsM.pdf](http://www.date.hu/~csubak/Publikaciok/geoderma_CsM.pdf).
- 97 A. A. Ahmed, S. Thiele-Bruhn, P. Leinweber and O. Kuhn, Towards a molecular level understanding of the sulfanilamide–soil organic matter–interaction, *Sci. Total Environ.*, 2016, **559**, 347–355.

- 98 J. D. Willey, R. J. Kieber and R. D. Lancaster, Coastal rainwater hydrogen peroxide: concentration and deposition, *J. Atmos. Chem.*, 1996, **25**, 149–165.
- 99 R. J. Kieber, R. D. Lancaster and J. D. Willey, Variations in Hydrogen-Peroxide Concentrations in Coastal North-Carolina Rainwater, *Abstr. Pap. Am. Chem. Soc.*, 1995, **210**, 175.
- 100 B. A. Wols and D. Vries, Modelling the water treatment efficiency of emerging contaminants by QSARs, 2012, unpublished results.
- 101 S. Sudhakaran and G. L. Amy, QSAR models for oxidation of organic micropollutants in water based on ozone and hydroxyl radical rate constants and their chemical classification, *Water Res.*, 2013, **47**, 1111–1122.
- 102 A. L. Dutot, J. Rude and B. Aumont, Neural network method to estimate the aqueous rate constants for the OH reactions with organic compounds, *Atmos. Environ.*, 2003, **37**, 269–276.
- 103 G. V. G. Buxton, C. L. Greenstock, W. P. Helman and A. B. Ross, Critical Review of Rate Constants for Reactions of Hydrated Electrons, Hydrogen Atoms and Hydroxyl Radicals ( $\cdot\text{OH}/\text{O}^-$ ) in Aqueous Solution, *J. Phys. Chem. Ref. Data*, 1988, **17**, 513.
- 104 A. Monod and J. F. Doussin, Structure–activity relationship for the estimation of OH-oxidation rate constants of aliphatic organic compounds in the aqueous phase: alkanes, alcohols, organic acids and bases, *Atmos. Environ.*, 2008, **42**, 7611–7622.
- 105 Y. N. Wang, J. W. Chen, X. H. Li, S. Y. Zhang and X. L. Qiao, Estimation of Aqueous-Phase Reaction Rate Constants of Hydroxyl Radical with Phenols, Alkanes and Alcohols, *QSAR Comb. Sci.*, 2009, **28**, 1309–1316.
- 106 J. Peng, J. Li, H. Shi, Z. Wang and S. Gao, Oxidation of disinfectants with Cl-substituted structure by a Fenton-like system  $\text{Cu}^{(2+)}/\text{H}_2\text{O}_2$  and analysis on their structure–reactivity relationship, *Environ. Sci. Pollut. Res. Int.*, 2016, **23**, 1898–1904.
- 107 X. H. Jin, S. Peldszus and P. M. Huck, Predicting the reaction rate constants of micropollutants with hydroxyl radicals in water using QSPR modeling, *Chemosphere*, 2015, **138**, 1–9.
- 108 H. Kusic, B. Rasulev, D. Leszczynska, J. Leszczynski and N. Koprivanac, Prediction of rate constants for radical degradation of aromatic pollutants in water matrix: a QSAR study, *Chemosphere*, 2009, **75**, 1128–1134.
- 109 D. Minakata and J. C. Crittenden, Theoretical study in the aqueous-phase subsequent free-radical reactions induced by hydroxyl radicals, *ACS National Meeting in Denver CO. August 29*, 2011.
- 110 D. Minakata, S. P. Mezyk, J. W. Jones, B. R. Daws and J. C. Crittenden, Development of Linear Free Energy Relationships for Aqueous Phase Radical-Involved Chemical Reactions, *Environ. Sci. Technol.*, 2014, **48**, 13925–13932.
- 111 D. Minakata, W. H. Song and J. Crittenden, Reactivity of Aqueous Phase Hydroxyl Radical with Halogenated Carboxylate Anions: Experimental and Theoretical Studies, *Environ. Sci. Technol.*, 2011, **45**, 6057–6065.
- 112 B. Y. Chen, T. Zhang, T. Bond and Y. Q. Gan, Development of quantitative structure activity relationship (QSAR) model for disinfection byproduct (DBP) research: a review of methods and resources, *J. Hazard. Mater.*, 2015, **299**, 260–279.
- 113 H. Gallard and U. von Gunten, Chlorination of phenols: kinetics and formation of chloroform, *Environ. Sci. Technol.*, 2002, **36**, 884–890.
- 114 M. Deborde and U. von Gunten, Reactions of chlorine with inorganic and organic compounds during water treatment – kinetics and mechanisms: a critical review, *Water Res.*, 2008, **42**, 13–51.
- 115 Y. Lee and U. von Gunten, Quantitative structure–activity relationships (QSARs) for the transformation of organic micropollutants during oxidative water treatment, *Water Res.*, 2012, **46**, 6177–6195.
- 116 M. Deborde, S. Rabouan, H. Gallard and B. Legube, Aqueous chlorination kinetics of some endocrine disruptors, *Environ. Sci. Technol.*, 2004, **38**, 5577–5583.
- 117 W. P. Fan Zhi-yun, H. Jun-li, S. Li-qiang, Z. Yjiong-jie and L. Ying-jie, QSAR between ATI and Kinetic Rate Constants of Anilines Oxidation with  $\text{ClO}_2$ , *J. Environ. Sci.*, 2004, **16**, 238–241.
- 118 P. G. Tratnyek and J. Hoigne, Kinetics of Reactions of Chlorine Dioxide (Oclo) in Water. 2. Quantitative Structure–Activity–Relationships for Phenolic-Compounds, *Water Res.*, 1994, **28**, 57–66.
- 119 H. X. Lei and S. A. Snyder, 3D QSPR models for the removal of trace organic contaminants by ozone and free chlorine, *Water Res.*, 2007, **41**, 4051–4060.
- 120 L. M. Rebenne, A. C. Gonzalez and T. M. Olson, Aqueous chlorination kinetics and mechanism of substituted dihydroxybenzenes, *Environ. Sci. Technol.*, 1996, **30**, 2235–2242.
- 121 C. M. Gerritsen and D. W. Margerum, Nonmetal Redox Kinetics – Hypochlorite and Hypochlorous Acid Reactions with Cyanide, *Inorg. Chem.*, 1990, **29**, 2757–2762.
- 122 P. Gramatica, P. Pilutti and E. Papa, Ranking of phenols for abiotic oxidation in aqueous environment: a QSPR approach, *Ann. Chim.*, 2005, **95**, 199–209.
- 123 H. C. Zhu, W. M. Guo, Z. M. Shen, Q. L. Tang, W. C. Ji and L. J. Jia, QSAR models for degradation of organic pollutants in ozonation process under acidic condition, *Chemosphere*, 2015, **119**, 65–71.
- 124 X. Jin, *Quantitative Structure–Property Relationships Modeling of Rate Constants of Selected Micropollutants in Drinking Water Treatment Using Ozonation and UV/H<sub>2</sub>O<sub>2</sub>*, University of Waterloo, 2012.
- 125 H. C. Zhu, Z. M. Shen, Q. L. Tang, W. C. Ji and L. J. Jia, Degradation mechanism study of organic pollutants in ozonation process by QSAR analysis, *Chem. Eng. J.*, 2014, **255**, 431–436.
- 126 Y. Lee, L. Kovalova, C. S. McCardell and U. von Gunten, Prediction of micropollutant elimination during

- ozonation of a hospital wastewater effluent, *Water Res.*, 2014, **64**, 134–148.
- 127 S. Sudhakaran, J. Calvin and G. L. Amy, QSAR models for the removal of organic micropollutants in four different river water matrices, *Chemosphere*, 2012, **87**, 144–150.
- 128 J. R. Kanofsky and P. Sima, Singlet Oxygen Production from the Reactions of Ozone with Biological Molecules, *J. Biol. Chem.*, 1991, **266**, 9039–9042.
- 129 W. A. Arnold, Y. Oueis, M. O'Connor, J. E. Rinaman, M. G. Taggart, R. E. McCarthy, K. A. Foster and D. E. Latch, QSARs for phenols and phenolates: oxidation potential as a predictor of reaction rate constants with photochemically produced oxidants, *Environ. Sci.: Processes Impacts*, 2017, DOI: 10.1039/c6em00580b.
- 130 E. Rorije and W. J. G. M. Peijnenburg, QSARs for oxidation of phenols in the aqueous environment, suitable for risk assessment, *J. Chemom.*, 1996, **10**, 79–93.
- 131 S. Canonica and P. G. Tratnyek, Quantitative structure–activity relationships for oxidation reactions of organic chemicals in water, *Environ. Toxicol. Chem.*, 2003, **22**, 1743–1754.
- 132 D. Juretic, H. Kusic, A. Papic, M. Smidt, O. Jezovita, I. Peternel and A. L. Bozic, Modeling of photodegradation kinetics of aromatic pollutants in water matrix, *J. Photochem. Photobiol., A*, 2013, **271**, 65–76.
- 133 USEPA Science Advisory Board (SAB), *Review of the Estimation Programs Interface Suite (EPI Suite)*, Washington, DC, 2007.
- 134 T. Mill, W. Haag, P. Penwell, T. Pettit and H. Johnson, *Environmental Fate and Exposure Studies Development of a PC-SAR for Hydrolysis: Esters, Alkyl Halides and Epoxides*, Menlo Park, CA, 1987.
- 135 AR Chem, *SPARC: Automated Reasoning in Chemistry*, 2010.
- 136 D. S. Mackay, W. Y. Shiu, K. C. Ma and S. C. Lee, *Handbook of Physical–Chemical Properties and Environmental Fate for Organic Chemicals*, CRC Press, Boca Raton, New York, 2006.
- 137 S. O. N. Tsuyuri, T. Fujimoto, H. Yamasaki, S. Hidaka, K. Okamoto, N. Kawashita and T. Takagi, Prediction of hydrolysis reaction of chemical compounds by multivariate analysis, in *Abstracts of Symposium on Structure Activity Relations, 2009*, 2009, vol. 37, pp. 37–38, 116.
- 138 B. Narasimhan, A. M. Ansari, N. Singh, V. Mourya and A. S. Dhake, A QSAR approach for the prediction of stability of benzoglycolamide ester prodrugs, *Chem. Pharm. Bull.*, 2006, **54**, 1067–1071.
- 139 S. H. Hilal, S. W. Karickhoff, L. A. Carreira and B. P. Shrestha, Estimation of carboxylic acid ester hydrolysis rate constants, *QSAR Comb. Sci.*, 2004, **22**, 917–925.
- 140 B. Sulzberger, S. Canonica, T. Egli, W. Giger, J. Klausen and U. von Gunten, Oxidative transformations of contaminants in natural and in technical systems, *Chimia*, 1997, **51**, 900–907.
- 141 J. A. Wojtowicz, Dichlorine Monoxide, Hypochlorous Acid, and Hypochlorites, *Kirk-Othmer Encycl. Chem. Technol.*, Wiley, New York, USA, 2004, DOI: 10.1002/0471238961.0409030823151020.a01.pub2.
- 142 EMA, Guideline on the chemistry of active substances, Committee for Medicinal Products for Human Use (CHMP), London, United Kingdom, EMA/CHMP/QWP/96664/2015, 2015.
- 143 R. Y. Xiao, T. T. Ye, Z. S. Wei, S. Luo, Z. H. Yang and R. Spinney, Quantitative Structure–Activity Relationship (QSAR) for the Oxidation of Trace Organic Contaminants by Sulfate Radical, *Environ. Sci. Technol.*, 2015, **49**, 13394–13402.
- 144 C. Perez, J. Sanchez, F. Marmol, P. Puig-Parellada and R. Pouplana, Reactivity of biologically important NSAID compounds with superoxide (O<sup>2</sup>( $\cdot\cdot$ )-), nitric oxide ((NO)-N( $\cdot\cdot$ )) and cyclooxygenase inhibition, *QSAR Comb. Sci.*, 2007, **26**, 368–377.
- 145 M. Petrovic and D. Barceló, *Analysis, fate and removal of pharmaceuticals in the water cycle*, Wilson & Wilson, 2007, vol. 50.
- 146 B. M. Peterson, A. M. McNally, R. M. Cory, J. D. Thoemke, J. B. Cotner and K. McNeill, Spatial and Temporal Distribution of Singlet Oxygen in Lake Superior, *Environ. Sci. Technol.*, 2012, **46**, 7222–7229.
- 147 D. Beiknejad and M. J. Chaichi, Estimation of photolysis half-lives of dyes in a continuous-flow system with the aid of quantitative structure–property relationship, *Front. Environ. Sci. Eng.*, 2014, **8**, 683–692.
- 148 G. Ioele, M. De Luca, F. Oliverio and G. Ragno, Prediction of photosensitivity of 1,4-dihydropyridine antihypertensives by quantitative structure–property relationship, *Talanta*, 2009, **79**, 1418–1424.
- 149 C. Li, X. H. Yang, X. H. Li, J. W. Chen and X. L. Qiao, Development of a model for predicting hydroxyl radical reaction rate constants of organic chemicals at different temperatures, *Chemosphere*, 2014, **95**, 613–618.
- 150 K. Thompson, J. Y. Zhang and C. L. Zhang, Use of fugacity model to analyze temperature-dependent removal of micro-contaminants in sewage treatment plants, *Chemosphere*, 2011, **84**, 1066–1071.
- 151 N. K. Pandit and R. P. Soltis, in *Introduction to the Pharmaceutical Sciences: An Integrated Approach*, Wolters Kluwer, 2007, ch. 4, Rates of Pharmaceutical processes.
- 152 J. Hoigne and H. Bader, Kinetics of Reactions of Chlorine Dioxide (Oclo) in Water. 1. Rate Constants for Inorganic and Organic-Compounds, *Water Res.*, 1994, **28**, 45–55.
- 153 N. K. Richards-Henderson, A. K. Hansel, K. T. Valsaraj and C. Anastasio, Aqueous oxidation of green leaf volatiles by hydroxyl radical as a source of SOA: kinetics and SOA yields, *Atmos. Environ.*, 2014, **95**, 105–112.
- 154 M. Stratakis and M. Orfanopoulos, Regioselectivity in the ene reaction of singlet oxygen with alkenes, *Tetrahedron*, 2000, **56**, 1595–1615.
- 155 P. Kuenemann, P. V. and J. Devillers Structure biodegradability relationships, in *Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, Kluwer Academic Publishers, Dordrecht, 1990.

- 156 W. J. G. M. Peijnenburg, Structure–Activity–Relationships for Biodegradation – a Critical-Review, *Pure Appl. Chem.*, 1994, **66**, 1931–1941.
- 157 K. M. Scow. Rate of biodegradation, in *Handbook of Chemical Property Estimation Methods*, McGraw-Hill, New York, 1983.
- 158 S. Sasu, J. Metzger, M. Kranert and K. Kummerer, Biodegradation of the Antituberculosis Drug Isoniazid in the Aquatic Environment, *Clean: Soil, Air, Water*, 2015, **43**, 166–172.
- 159 OECD, OECD guidelines for the testing of chemicals, Guideline No. 301, Ready Biodegradability, Organisation for Economic Cooperation and Development, 1992.
- 160 Organisation for Economic Cooperation and Development, OECD Introduction to the OECD guidelines for testing of chemicals, section 3, Principles and strategies related to the testing of degradation of organic chemicals, Paris, 2006.
- 161 W. J. G. M. Peijnenburg, Structure–activity relationships for biodegradation: a critical review, *Pure Appl. Chem.*, 1994, **66**, 1931–1941.
- 162 P. H. Howard, Biodegradation, in *Handbook of Property Estimation Methods for Chemicals*, ed. R. S. Boethling and D. Mackay, Environmental and Health Sciences publishers, Lewis, Boca Raton, FL, 2000, pp. 281–310.
- 163 A. Sabljic, QSAR models for estimating properties of persistent organic pollutants required in evaluation of their environmental fate and risk, *Chemosphere*, 2001, **43**, 363–375.
- 164 J. S. Jaworska, R. S. Boethling and P. H. Howard, Recent developments in broadly applicable structure–biodegradability relationships, *Environ. Toxicol. Chem.*, 2003, **22**, 1710–1723.
- 165 J. R. Baker, D. Gamberger, J. R. Mihelcic and A. Sabljic, Evaluation of Artificial Intelligence Based Models for Chemical Biodegradability Prediction, *Molecules*, 2004, **9**, 989–1004.
- 166 M. Nendza, Prediction of Persistence, in *Predicting Chemical Toxicity and Fate*, CRC Press, Boca Raton, USA, 2004, pp. 315–331.
- 167 P. H. Howard. Predicting the persistence of organic compounds, in *Handbook of Environmental Chemistry*, Springer, Berlin, 2008.
- 168 A. Sabljic and Y. Nakagawa, Biodegradation and Quantitative Structure–Activity Relationship (QSAR), in *Non-First Order Degradation and Time-Dependent Sorption of Organic Chemicals in Soil*, American Chemical Society, 2014, ch. 4, pp. 57–84.
- 169 F. Sahigara, D. Ballabio, R. Todeschini and V. Consonni, Assessing the Validity of QSARs for Ready Biodegradability of Chemicals: An Applicability Domain Perspective, *Curr. Comput.-Aided Drug Des.*, 2014, **10**, 137–147.
- 170 R. Boethling, Comparison of ready biodegradation estimation methods for fragrance materials, *Sci. Total Environ.*, 2014, **497**, 60–67.
- 171 F. Pizzo, A. Lombardo, A. Manganaro and E. Benfenati, *In silico* models for predicting ready biodegradability under REACH: a comparative study, *Sci. Total Environ.*, 2013, **463**, 161–168.
- 172 K. Mansouri, T. Ringsted, D. Ballabio, R. Todeschini and V. Consonni, Quantitative Structure–Activity Relationship Models for Ready Biodegradability of Chemicals, *J. Chem. Inf. Model.*, 2013, **53**, 867–878.
- 173 Y. Tabei, Y. Yamanishi and M. Kotera, Simultaneous prediction of enzyme orthologs from chemical transformation patterns for *de novo* metabolic pathway reconstruction, *Bioinformatics*, 2016, **32**, 278–287.
- 174 J. Wickert, T. Lorschbach, M. Gutlein, E. Schmid, D. Latino, S. Kramer and K. Fenner, enviPath – The environmental contaminant biotransformation pathway resource, *Nucleic Acids Res.*, 2016, **44**, D502–D508.
- 175 H. Loonen, F. Lindgren, B. Hansen, W. Karcher, J. Niemela, K. Hiromatsu, M. Takatsuki, W. Peijnenburg, E. Rorije and J. Struijs, Prediction of biodegradability from chemical structure: modeling of ready biodegradation test data, *Environ. Toxicol. Chem.*, 1999, **18**, 1763–1768.
- 176 J. Tunkel, K. Mayo, C. Austin, A. Hickerson and P. Howard, Practical considerations on the use of predictive models for regulatory purposes, *Environ. Sci. Technol.*, 2005, **39**, 2188–2199.
- 177 S. Dimitrov, T. Pavlov, N. Dimitrova, D. Georgieva, D. Nedelcheva, A. Kesova, R. Vasilev and O. Mekenyan, Simulation of chemical metabolism for fate and hazard assessment. II CATALOGIC simulation of abiotic and microbial degradation, *SAR QSAR Environ. Res.*, 2011, **22**, 719–755.
- 178 L. Yin and X. Dan-li, Quantitative structure–activity relationship study on the biodegradation of acid dyestuffs, *J. Environ. Sci.*, 2007, **19**, 800–804.
- 179 A. J. Greaves, J. H. Churchley, M. G. Hutchings, D. A. Phillips and J. A. Taylor, A chemometric approach to understanding the bioelimination of anionic, water-soluble dyes by a biomass using empirical and semi-empirical molecular descriptors, *Water Res.*, 2001, **35**, 1225–1239.
- 180 T. Tamm, G. Karelson, D. Dobchev, K. Tämm, J. Jänes and M. Karelson, *QSAR ANN model for Persistence: biotic degradation in water; QSARDB entry in QMRF format*, 2012.
- 181 C. Burgis, *Predicting Biological Removal of Contaminants in Wastewater Treatment: QSBR Modeling*, University of Virginia, Virginia, US, 2012.
- 182 C. L. Tebes-Stevens and W. J. Jones, Estimation of microbial reductive transformation rates for chlorinated benzenes and phenols using a quantitative structure–activity relationship approach, *Environ. Toxicol. Chem.*, 2004, **23**, 1600–1609.
- 183 J. Damborsky and T. W. Schultz, Comparison of the QSAR models for toxicity and biodegradability of anilines and phenols, *Chemosphere*, 1997, **34**, 429–446.
- 184 A. Heitzer, Microbiological, molecular and biochemical methods for environmental risk analysis and bioprocess evaluation, *J. Microbiol. Methods*, 1998, **32**, 89–91.
- 185 NRC, Physicochemical Properties and Environmental Fate, in *A Framework to Guide Selection of Chemical Alternatives*,

- Substitutions*, Committee on the Design and Evaluation of Safer Chemical Substitutions, The National Academic Press, Washington, D. C., 2014, ch. 5.
- 186 A. D. Tappin, J. P. Loughnane, A. J. McCarthy and M. F. Fitzsimons, Unexpected removal of the most neutral cationic pharmaceutical in river waters, *Environ. Chem. Lett.*, 2016, **14**, 455–465.
- 187 L. P. Wackett, Questioning our perceptions about evolution of biodegradative enzymes, *Curr. Opin. Microbiol.*, 2009, **12**, 244–251.
- 188 D. Domaradzka, U. Guzik and D. Wojcieszynska, Biodegradation and biotransformation of polycyclic non-steroidal anti-inflammatory drugs, *Rev. Environ. Sci. Bio/Technol.*, 2015, **14**, 229–239.
- 189 P. H. Howard, A. E. Hueber and R. S. Boethling, Biodegradation Data Evaluation for Structure Biodegradability Relations, *Environ. Toxicol. Chem.*, 1987, **6**, 1–10.
- 190 J. R. Parsons and H. A. J. Govers, Quantitative Structure–Activity–Relationships for Biodegradation, *Ecotoxicol. Environ. Saf.*, 1990, **19**, 212–227.
- 191 J. H. P. Langenberg, W. J. G. M. Peijnenburg and E. Rorije, On the Usefulness and Reliability of Existing QSBRs for Risk Assessment and Priority Setting, *SAR QSAR Environ. Res.*, 1996, **5**, 1–16.
- 192 R. Posthumus, T. P. Traas, W. J. G. M. Peijnenburg and E. M. Hulzebos, External validation of EPIWIN biodegradation models, *SAR QSAR Environ. Res.*, 2005, **16**, 135–148.
- 193 IME Jahresbericht 2014/2015: Comparison and improvement of laboratory water/sediment test systems; Aachen, Germany, 2015; p 168.
- 194 J. D. Carballeira, M. A. Quezada, E. Alvarez and J. V. Sinisterra, High throughput screening and QSAR-3D/CoMFA: useful tools to design predictive models of substrate specificity for biocatalysts, *Molecules*, 2004, **9**, 673–693.
- 195 J. D. C. Rodriguez, E. A. Ruiz and J. V. S. Gago, 3D-QSAR/CoMFA models as a tool for biocatalysis and protein engineering, *Chem. Biochem. Eng. Q.*, 2006, **20**, 173–181.
- 196 S. Berkner and C. Thierbach, Biodegradability and transformation of human pharmaceutical active ingredients in environmentally relevant test systems, *Environ. Sci. Pollut. Res.*, 2014, **21**, 9461–9467.
- 197 W. G. S. Landis, R. M. Sofield and M. H. Yu, *Introduction to Environmental Toxicology: Molecular Structures to Ecological Landscapes*, CRC Press, Boca Raton, FL, 2011.
- 198 M. Adolfsson-Erici, G. Akerman and M. S. McLachlan, Measuring bioconcentration factors in fish using exposure to multiple chemicals and internal benchmarking to correct for growth dilution, *Environ. Toxicol. Chem.*, 2012, **31**, 1853–1860.
- 199 W. M. Meylan, P. H. Howard, R. S. Boethling, D. Aronson, H. Printup and S. Gouchie, Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient, *Environ. Toxicol. Chem.*, 1999, **18**, 664–672.
- 200 A. Gissi, O. Nicolotti, A. Carotti, D. Gadaleta, A. Lombardo and E. Benfenati, Integration of QSAR models for bioconcentration suitable for REACH, *Sci. Total Environ.*, 2013, **456**, 325–332.
- 201 A. Gissi, A. Lombardo, A. Roncaglioni, D. Gadaleta, G. F. Mangiatordi, O. Nicolotti and E. Benfenati, Evaluation and comparison of benchmark QSAR models to predict a relevant REACH endpoint: the bioconcentration factor (BCF), *Environ. Res.*, 2015, **137**, 398–409.
- 202 W. J. Fu, A. Franco and S. Trapp, Methods for Estimating the Bioconcentration Factor of Ionizable Organic Chemicals, *Environ. Toxicol. Chem.*, 2009, **28**, 1372–1379.
- 203 M. Seto and I. C. Handoh, Mathematical explanation for the non-linear hydrophobicity-dependent bioconcentration processes of persistent organic pollutants in phytoplankton, *Chemosphere*, 2009, **77**, 679–686.
- 204 H. Geyer, G. Politzki and D. Freitag, Prediction of Ecotoxicological Behavior of Chemicals – Relationship between Normal-Octanol–Water Partition-Coefficient and Bioaccumulation of Organic-Chemicals by Alga *Chlorella*, *Chemosphere*, 1984, **13**, 269–284.
- 205 M. Meredith-Williams, L. J. Carter, R. Fussell, D. Raffaelli, R. Ashauer and A. B. A. Boxall, Uptake and depuration of pharmaceuticals in aquatic invertebrates, *Environ. Pollut.*, 2012, **165**, 250–258.
- 206 W. L. Lockhart, B. B. and G. E. de March, Muir DCG uptake and toxicity of organic compounds: studies with an aquatic macrophyte (*Lemna minor*), *ASTM Spec. Tech. Publ.*, 1983, **802**, 460–468.
- 207 M. Nendza, S. Gabbert, R. Kuhne, A. Lombardo, A. Roncaglioni, E. Benfenati, R. Benigni, C. Bossa, S. Stempel, M. Scheringer, A. Fernandez, R. Rallo, F. Giralt, S. Dimitrov, O. Mekenyan, F. Bringezu and G. Schuurmann, A comparative survey of chemistry-driven *in silico* methods to identify hazardous substances under REACH, *Regul. Toxicol. Pharmacol.*, 2013, **66**, 301–314.
- 208 A. P. Toropova, A. A. Toropov, A. Lombardo, A. Roncaglioni, E. Benfenati and G. Gini, A new bioconcentration factor model based on SMILES and indices of presence of atoms, *Eur. J. Med. Chem.*, 2010, **45**, 4399–4402.
- 209 A. Lombardo, A. Roncaglioni, E. Boriani, C. Milan and E. Benfenati, Assessment and validation of the CAESAR predictive model for bioconcentration factor (BCF) in fish, *Chem. Cent. J.*, 2010, **4**(Suppl. 1), S1.
- 210 B. I. Escher, R. P. Schwarzenbach and J. C. Westall, Evaluation of liposome–water partitioning of organic acids and bases. 2. Comparison of experimental determination methods, *Environ. Sci. Technol.*, 2000, **34**, 3962–3968.
- 211 B. I. Escher, C. E. Cowan-Ellsberry, S. Dyer, M. R. Embry, S. Erhardt, M. Halder, J. H. Kwon, K. Johanning, M. T. T. Oosterwijk, S. Rutishauser, H. Segner and J. Nichols, Protein and Lipid Binding Parameters in Rainbow Trout (*Oncorhynchus mykiss*) Blood and Liver Fractions to Extrapolate from an *in Vitro* Metabolic Degradation Assay to *in Vivo* Bioaccumulation Potential of

- Hydrophobic Organic Chemicals, *Chem. Res. Toxicol.*, 2011, **24**, 1134–1143.
- 212 W. C. Qin, L. M. Su, X. J. Zhang, H. W. Qin, Y. Wen, Z. Guo, F. T. Sun, L. X. Sheng, Y. H. Zhao and M. H. Abraham, Toxicity of organic pollutants to seven aquatic organisms: effect of polarity and ionization, *SAR QSAR Environ. Res.*, 2010, **21**, 389–401.
- 213 D. L. Swackhamer and R. S. Skoglund, Bioaccumulation of Pcb's by Algae – Kinetics Versus Equilibrium, *Environ. Toxicol. Chem.*, 1993, **12**, 831–838.
- 214 A. P. Toropova, A. A. Toropov, S. E. Martyanov, E. Benfenati, G. Gini, D. Leszczynska and J. Leszczynski, CORAL: Monte Carlo Method as a Tool for the Prediction of the Bioconcentration Factor of Industrial Pollutants, *Mol. Inf.*, 2013, **32**, 145–154.
- 215 L. H. C. Nowell, P. D. Capel and P. D. Dileanis, *Pesticides in Stream Sediment and Aquatic Biota: Distribution, Trends, and Governing Factors*, CRC Press, Boca Raton, FL, 1999.
- 216 N. J. Diepens, M. J. Van den Heuvel-Greve and A. A. Koelmans, Modeling of Bioaccumulation in Marine Benthic Invertebrates Using a Multispecies Experimental Approach, *Environ. Sci. Technol.*, 2015, **49**, 13575–13585.
- 217 J. M. Armitage, R. J. Erickson, T. Luckenbach, C. A. Ng, R. S. Prosser, J. A. Arnot, K. Schirmer and J. W. Nichols, Assessing the bioaccumulation potential of ionizable organic compounds: current knowledge and research priorities, *Environ. Toxicol. Chem.*, 2016, DOI: 10.1002/etc.3680.
- 218 H. Loonen, D. C. G. Muir, J. R. Parsons and H. A. J. Govers, Bioaccumulation of polychlorinated dibenzo-*p*-dioxins in sediment by oligochaetes: influence of exposure pathway and contact time, *Environ. Toxicol. Chem.*, 1997, **16**, 1518–1525.
- 219 F. Y. Bois, M. Jamei and H. J. Clewell, PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals, *Toxicology*, 2010, **278**, 256–267.
- 220 M. L. Peach, A. V. Zakharov, R. F. Liu, A. Pugliese, G. Tawa, A. Wallqvist and M. C. Nicklaus, Computational tools and resources for metabolism-related property predictions. I. Overview of publicly available (free and commercial) databases and software, *Future Med. Chem.*, 2012, **4**, 1907–1932.
- 221 A. W. Mostrag-Szlichtyng, *A Review of QSAR Models and Software Tools for Predicting Biokinetic Properties*, Institute for Health and Consumer Protection, 2010.
- 222 T. Peyret and K. Krishnan, QSARs for PBPK modelling of environmental contaminants, *SAR QSAR Environ. Res.*, 2011, **22**, 129–169.
- 223 J. C. H. van Eijkeren, D. T. Jager and A. J. A. M. Sips, *Generic PBPK-modelling of lipophilic contaminants in the cow*, National Institute of public health and the environment (RVM), Bilthoven, Netherlands, 1998.
- 224 J. G. Bessems, G. Loizou, K. Krishnan, H. J. Clewell, C. Bernasconi, F. Bois, S. Coecke, E. M. Collnot, W. Diembeck, L. R. Farcal, L. Geraets, U. Gundert-Remy, N. Kramer, G. Kusters, S. B. Leite, O. R. Pelkonen, K. Schroder, E. Testai, I. Wilk-Zasadna and J. M. Zaldivar-Comenges, PBTK modelling platforms and parameter estimation tools to enable animal-free risk assessment Recommendations from a joint EPAA – EURL ECVAM ADME workshop, *Regul. Toxicol. Pharmacol.*, 2014, **68**, 119–139.
- 225 J. M. Armitage, J. A. Arnot, F. Wania and D. Mackay, Development and evaluation of a mechanistic bioconcentration model for ionogenic organic chemicals in fish, *Environ. Toxicol. Chem.*, 2013, **32**, 115–128.
- 226 N. N. M. Fjodorova, Integration of QSAR and RAS methods for the mechanistic interpretation of predictive models for carcinogenicity, *Comput. Struct. Biotechnol. J.*, 2012, **1**, 1–10.
- 227 A. P. Worth, A. Bassan, A. Gallegos, T. I. Netzeva, G. Patlewicz, M. Pavan, I. Tsakovska and M. Vracko, *The Characterisation of (Quantitative) Structure–Activity Relationships: Preliminary Guidance*, 2005.
- 228 A. A. Toropov, A. P. Toropova, A. Lombardo, A. Roncaglioni, E. Benfenati and G. Gini, CORAL: building up the model for bioconcentration factor and defining its applicability domain, *Eur. J. Med. Chem.*, 2011, **46**, 1400–1403.
- 229 J. G. T. Arnot and D. Mackay, *Development and Application of Models of Chemical Fate in Canada – Practical Methods for Estimating Environmental Biodegradation Rates*, Peterborough, Canada, 2005.