



Cite this: *Green Chem.*, 2020, **22**, 3558

## Transition-state rate theory sheds light on 'black-box' biodegradation algorithms†

T. M. Nolte,<sup>id</sup>\*<sup>a,b</sup> W. J. G. M. Peijnenburg,<sup>id</sup><sup>c,d</sup> T. J. H. M. van Bergen<sup>id</sup><sup>a</sup> and A. J. Hendriks<sup>a</sup>

Biodegradation is a predominant removal mechanism for organic pollutants in the aquatic and terrestrial environment and needs to be determined to design 'green chemicals' amongst an increasingly large set of industrial chemicals. Decades of research have been dedicated to producing biodegradation models, though improving those models has become problematic due to 'black box' models driven by incomparable or conflicting experimental results. In this study, we tested the plausibility and applicability of an intuitive algebraic formula stemming from transition-state rate theory. The formula is overarching, describing the pseudo first-order biodegradation rate constant in terms of computationally easily obtainable electronic, steric/geometrical, energetic and thermodynamic properties. Surprisingly, statistical evaluation using experimental data shows that the formula performs equal to or better than established 'black-box' models. We interpret the properties used, highlight the precise (inter)dependencies and discuss reaction- and diffusion-limiting mechanisms. Altogether, the work shows the potential to improve our understanding of biodegradation *via* 'first principles': it helps to unravel the causal mechanisms of the chemical fate in complex matrices. Amongst potential ramifications, this will enable a more precise and comprehensive environmental risk assessment.

Received 27th January 2020,  
Accepted 14th April 2020

DOI: 10.1039/d0gc00337a

[rsc.li/greenchem](http://rsc.li/greenchem)

## 1. Introduction

### 1.1. Relevance

Understanding the fate and transformation of organic pollutants is vital to evaluate their hazards posed by unwanted exposure. Following the principles of Green Chemistry, design of "less hazardous syntheses", "benign chemicals" and "degradation routes",<sup>1</sup> manufacturers evaluate potential biotransformations of their chemicals during early development.<sup>2–4</sup> Since biodegradation is a predominant removal mechanism for organic pollutants,<sup>5</sup> chemicals manufactured or imported in quantities over one ton per year can only be registered<sup>6</sup> if their 'ready biodegradability' is evaluated.<sup>7</sup> Laboratory tests<sup>8</sup> provide such information.

However, testing is time- and resource-intensive,<sup>9</sup> as evidenced by the body of existing and 'to-be-registered' chemicals.<sup>10–12</sup>

As an alternative, *in silico* methods such as quantitative structure–biodegradation relationships (QSBRs)<sup>3,13–16</sup> infer biodegradation from molecular characteristics. As such, QSBRs are helpful (as alternative data<sup>6</sup>) to handle chemical libraries and enable screening.<sup>10,17</sup> Many empirical models predict the biodegradability in aquatic media<sup>15</sup> such as Biowin, CATALOGIC and VEGA.<sup>16,18–20</sup> Undoubtedly, they contributed to our understanding, but drawbacks remain. QSBRs may apply to specific media/inocula,<sup>21</sup> *e.g.* wastewater/sludge,<sup>22,23</sup> sediment or surface water,<sup>15</sup> only. Low precisions<sup>24–26</sup> (*e.g.* a factor 5 error<sup>27</sup>) according to the OECD (The Organisation for Economic Co-operation and Development) standards<sup>28,29</sup> or semi-quantitative/categorical (*e.g.* 28-day pass) predictions can be involved. QSBRs have fundamentally limited applicability<sup>30</sup> (*e.g.* hydrocarbons only<sup>31</sup>) and become less comprehensive as the number of descriptors increases and when the (non-linear) algorithm does not visibly correspond to an underlying process.<sup>27</sup> Conversely, for only substituted benzenes<sup>31</sup> or hydrocarbons,<sup>32</sup> interpretation is more straightforward. Apparently, more precise values relate to 'first principles' modelling of *in situ* biodegradation kinetics.

<sup>a</sup>Radboud University Nijmegen, Department of Environmental Science, Institute for Water and Wetland Research, 6500 GL Nijmegen, The Netherlands.

E-mail: [t.nolte@science.ru.nl](mailto:t.nolte@science.ru.nl), [nolte@inorg.chem.ethz.ch](mailto:nolte@inorg.chem.ethz.ch), [tom.m.nolte@gmail.com](mailto:tom.m.nolte@gmail.com)

<sup>b</sup>Eidgenössische Technische Hochschule (ETH) Zurich, Laboratory of Inorganic Chemistry, Vladimir Prelog-Weg 1, 8093 Zurich, Switzerland

<sup>c</sup>Institute of Environmental Sciences, Leiden University, 2300 RA Leiden, The Netherlands

<sup>d</sup>National Institute of Public Health and the Environment, PO Box 1, 3720 BA Bilthoven, The Netherlands

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0gc00337a

## 1.2. Theory

Many processes affect the potential and the kinetics of biodegradation,<sup>23,33,34</sup> wherefore tests are often ‘inadequately’ standardized. Indeed, limited information on the media and inocula may be available<sup>3,21,35</sup> which hampers model improvement.<sup>3,22,36–38</sup> We can expect that after accounting for many such test-specific phenomena, the ‘intrinsic biodegradability’ of chemicals becomes apparent.<sup>23,27,39–42</sup>

Such an ‘intrinsic biodegradation rate’ might not entail information on microbial processes: adaptation, active transport,<sup>43,44</sup> the involvement of medium (pH; organic carbon), nutrients,<sup>45</sup> and back-transformation<sup>46</sup> (*i.e.* the factors are either non-existent, or constant for all chemicals). We might also assume no toxicity (*e.g.* under  $\mu\text{g L}^{-1}$  concentrations<sup>27</sup>), meaning a constant active (catalyzing) biomass.<sup>33</sup> We can also consider a constant influence of temperature on transport and (facilitated) diffusion.<sup>47,48</sup> If the data meet these conditions, the biodegradation process is pseudo first-order:

$$\text{biodegradation rate} = k_b \cdot [\text{active biomass}] \cdot [\text{substrate}] \quad (1)$$

The biodegradation rate constant  $k_b$  (eqn (1), taken in L per cell per h) relates to *e.g.* the size, and the steric and electronic aspects of the substrate chemical.<sup>3,27,49</sup> Qualitatively, such molecular characteristics are embedded in  $k_b$  *via* ‘some’ function  $f()$ :

$$k_b = f(\text{availability, accessibility, affinity, reactivity, ...}) \quad (2)$$

This function  $f()$  is notoriously complex. For example, it is difficult and laborious to determine *a-priori* the details of the biotransformation energy landscape (Fig. 1) for structurally diverse chemicals.<sup>50–53</sup> Since  $f()$  is too complex, we simplify.

Biodegradation rate constants may be near the diffusion limit in water or viscous cytosols in case microbes can evolve and/or adapt.<sup>54,55</sup> If so, rate constants for diffusion-limited and reaction-limited reactions are in the same order of magnitude.<sup>‡</sup> For low concentrations, *e.g.*  $\leq 0.1$  mM, substrate–substrate interactions are negligible.<sup>56</sup> Also, various enzymes have similar accessibilities, *e.g.* in case the chemical is much smaller than the active site.<sup>57</sup> Thus, we can treat the diffusion frequency<sup>58</sup> as a uniform factor for each chemical based on its dimensions, taking a certain (constant) diffusion pathway length. Then, the diffusion frequency becomes a product ( $\cdot$ ) function<sup>59–61</sup> and we can quantify  $f()$  and characteristics (in eqn (2)) to:<sup>60</sup>

$$k_b \propto \sum_{i=1}^i \left( D_{i-j} \cdot (d_{i-j}\lambda)^{-1} \cdot P_{i-j} \cdot e^{-\frac{\Delta G_{i-j}^\ddagger}{RT}} \right) \quad (3)$$

‡ A median  $k_{\text{cat}}/K_M \sim 5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{cat}}/K_M$  (“diffusion-controlled”) =  $10^6$ – $10^9 \text{ M}^{-1} \text{ s}^{-1}$  (Bar-Even *et al.*, 2011) and diffusion (Smoluchowski) limit  $\sim 10^{10} \text{ M}^{-1} \text{ s}^{-1}$  (Schurr and Schmitz, 1976) have been reported at room temperature. This amounts to thermodynamic (reaction) and diffusion barriers of 2–19 kJ mol<sup>-1</sup> and 6–23 kJ mol<sup>-1</sup>, respectively.

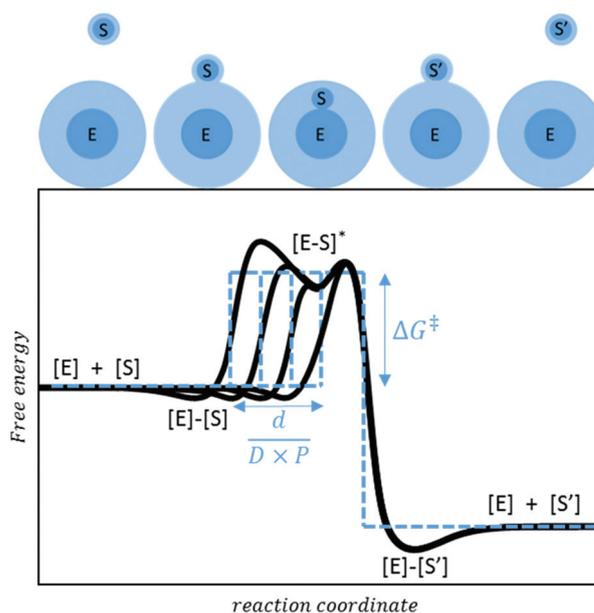


Fig. 1 Schematic representation (dashed lines) of the biodegradation mechanism (solid lines) in energetic and spatial dimension. The term  $\frac{d \cdot P}{D}$  describes the size of the dashed blue boxes and characterizes the inverse diffusion frequency factor (the size of the circle's light blue areas), and  $\Delta G^\ddagger$  is the activation energy. Dark blue circles denote the chemical substrate S and enzyme E (not to scale).

In eqn (3),  $D$  is the diffusion coefficient ( $\text{\AA}^2 \text{ s}^{-1}$ ),  $d$  is a characteristic distance ( $\text{\AA}$ ) that depends on the trajectory of the substrate chemical towards the active transformation site and  $\lambda$  is the DeBroglie wavelength ( $\text{\AA}$ ).  $\Delta G^\ddagger$  is the activation energy ( $\text{J mol}^{-1}$ ),  $T$  is the temperature (K),  $R$  is the ideal gas constant ( $\text{J mol}^{-1} \text{ K}^{-1}$ ) and  $P$  is a partition function describing thermodynamic equilibrium (dimensionless). The indices  $i$  and  $j$  are the active sites of the biotransformation in a functional group (see the ESI<sup>†</sup>) of the molecule, and bacteria/enzymes, respectively.

Taken together, the pre-exponential term in eqn (3) may be regarded as the frequency factor  $A_{i-j} = D_{i-j}(d_{i-j}\lambda)$  (in  $\text{s}^{-1}$  for the 1<sup>st</sup> order reaction) in collision theory.<sup>58</sup> Similar derivations exist for abiotic reactions<sup>60</sup> and for conversion of biomass.<sup>62</sup> The terms in eqn (3) broadly correspond to the qualitative characteristics in eqn (2). The dashed blue lines in Fig. 1 visualize the terms as related to the complex biodegradation mechanism (solid black lines). The blue circles illustrate the effective distances between the substrate chemical (S) and enzyme (E).

The aim of this paper is to test the plausibility of eqn (3) and the assumptions associated. We do this by evaluating the individual contribution of the terms in eqn (3) to  $k_b$  and for this purpose selected characteristics describing the diffusion process and the thermodynamics. We combined these into frequency factors  $A$  and activation energies  $\Delta G^\ddagger$  and independently correlated them with experimental  $k_b$  (section 2.1) and (in the case of gaps) QSBR-derived  $k_b$  values (section 2.2).

## 2. Methods

### 2.1. Collection and curation of experimental $k_b$ values

To maximize the validity of eqn (1) and the likelihood of a constant (active) biomass, we made a selection of biodegradation data. Details on data selection are described in S0† and elsewhere.<sup>27</sup> Focusing on primary aerobic biodegradation, we excluded tests and chemicals implying partial or full reductive biodegradation (*e.g.* inocula obtained from anaerobic sediments incubated with nitro ( $-\text{NO}_2$ ) group chemicals). We excluded hydrolytically (abiotic) unstable chemicals (Fig. S9†), with the exception of sterically hindered and aromatic esters.

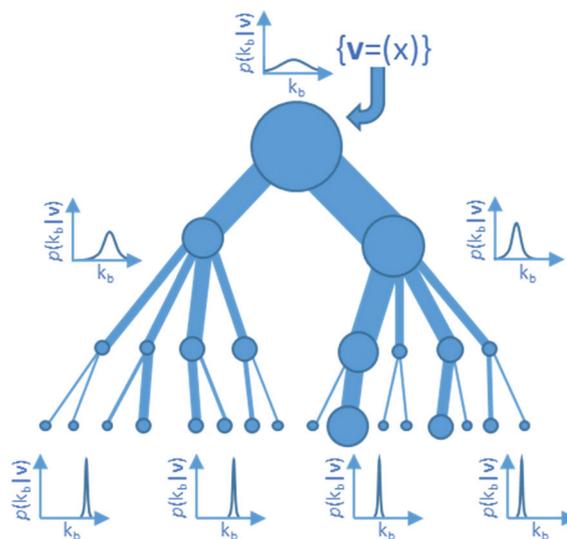
We pooled the dataset with that of previous studies<sup>27</sup> to widen the chemical diversity by including molecules entailing the steric and electronic aspects known to affect biodegradation in general.<sup>3,27,49,63</sup> Noteworthy new inclusions are herbicides and pharmaceuticals. The chemical structures presented a wide range of properties (Table 1); the structural diversity of the molecules is high, *e.g.*  $\sim 8$  and  $>1$  orders of magnitude in  $K_{\text{OW}}$  and volume, respectively (Fig. S1†).

The dataset contained a set of 550 unique chemicals, Fig. S1.† The data were corrected for sorption to organic carbon and converted to a pseudo first-order rate constant (eqn (1)) taking a biomass of  $\sim 10^8$  cells per L.<sup>27</sup> A unit of L per cell per h was chosen to allow implementation for different biomasses. The values for  $\log(k_b)$  ( $k_b$  in L per cell per h) range from  $-13$  (perfluorooctanoic acid) to  $-8.5$  (acetaldehyde) and the median value for  $\log(k_b)$  is  $-10.5$  (equivalent to a  $\text{DT}_{50} \sim 2$  weeks in surface water).

### 2.2. Generation of QSBR-predicted $k_b$ values

Speciation states were manually adjusted taking  $\text{pH} = 7$  and experimental (Drugbank; Pubchem) or predicted<sup>66</sup>  $\text{p}K_a$  as the starting point. The 3D structures of the molecules were optimized using OpenBabel. 5000+ chemical descriptors were calculated by common packages such as Chemopy, PaDEL<sup>65,67-70</sup> and RDKit<sup>71</sup> *via* SMILES input to the web-based platform Online Chemical Modeling Environment.<sup>65</sup> As descriptors can capture similar molecular characteristics (*i.e.* are intercorrelated), we grouped highly intercorrelated descriptors ( $R^2 > 0.95$ ).<sup>72</sup>

We split the dataset randomly between training and testing sets (3:1). Based on previous studies,<sup>27,73,74</sup> we selected a random forest regression (RF-R) algorithm<sup>75</sup> with 10 trees with infinite tree depth for the development of a QSBR model. The RF method links descriptors to an endpoint in a non-linear



**Fig. 2** Simplified representation of the random forest regression tree. The input feature space is given by  $v = (x)$ .  $x$  are the descriptors and  $k_b$  is the dependent variable.  $v$  is used during training to optimize parameters (nodes + connections) in the tree. Confidences associated with different nodes increase from the root (top) to the leaves (bottom). Mathematical details in ref. 75.

way, suitable in the case of heterogeneous data and complex endpoints,<sup>76</sup> Fig. 2:

We evaluated the accuracy of the QSBR-predicted values *via* (1) external testing, (2) leave-5-out cross validation,<sup>28,77</sup> and (3) comparison with CATABOL<sup>78</sup> predictions. We evaluated the stability using different (random) training/testing data splits.

### 2.3. Characterisation of $A$ and $\Delta G^\ddagger$

Custom descriptors were calculated *via* MOPAC, Chemopy, ChemAxon and Molinspiration<sup>65,67-70</sup>: surface area, accessible surface area, volume,<sup>79,80</sup> the octanol–water partitioning coefficient ( $K_{\text{OW}}$ ) and parameters encoding electronic aspects:  $E_{\text{HOMO}}$  and atom-specific ( $i$ , eqn (3)) delocalizability ( $\delta$ ) indices.<sup>23,81-85</sup> S2† describes detailed calculations of custom descriptors. We first screened the descriptors for their individual relevance, and then used them to calculate  $D$ ,  $d$ ,  $P$  (which in turn are used to calculate  $A$ ) and  $\Delta G^\ddagger$  (eqn (3)). S2† provides a theoretical underpinning of eqn (3) and its terms.

Changing a single term in eqn (3) can automatically affect one or more of the other terms (inter-correlations). For this reason, we evaluate the (contribution of) characteristics in eqn (3) with respect to  $k_b$  orthogonally (independently): vary one characteristic while keeping the other constant. For this purpose, we defined ‘similarly reactive chemicals’ and ‘similarly diffusive chemicals’ here as chemicals for which there is no variation in the terms  $e^{(-\Delta G_{i-j}^\ddagger/RT)}$  and  $A$ , respectively (eqn (3)). We also defined ‘electron-rich’ chemicals as those containing only C, H or O atoms, or a combination thereof (excluding *e.g.* N or halogen atoms).

**Table 1** The diversity of the chemicals considered in this study's QSBR

Property	$\log(K_{\text{OW}}, \text{pH} = 7)$	$\log(V)/\log(\text{\AA}^3)$	$E_{\text{HOMO}}/\text{eV}$	$\log(S)^b$	Formal charge
Range ( $N = 550$ )	$-4$ to $4^a$	1.4 to 2.8	$-13$ to $-8$	$-6$ to $1$	$-3$ to $2$

<sup>a</sup> For 95% of the data (Fig. S1†).<sup>64</sup> <sup>b</sup> No units provided.<sup>65</sup>  $K_{\text{OW}}$  = octanol water partitioning coefficient;  $V$  = volume;  $E_{\text{HOMO}}$  = the energy of the highest occupied molecular orbital;  $S$  = water solubility.

### 3. Results and discussion

Frequency factors  $A$  and activation energies  $\Delta G^\ddagger$  were successfully related to molecular characteristics. Via  $A$  and  $\Delta G^\ddagger$  (eqn (3)), we can differentiate between diffusion-limited and reaction-limited biodegradation. The differentiation has previously been made theoretically<sup>86</sup> and used to describe abiotic reactions.<sup>87</sup> Fig. S7 and S8† depict substances with higher  $k_b$  values than expected only on the basis of reactivity and  $A$ . These chemicals are polar and charged; charge inhomogeneity is known to affect *e.g.* active uptake.<sup>27,88</sup> The regressions with  $k_b$  values show appreciable statistical significance: depending on the degradation regime, eqn (3) explains 58–76% of the variance in the  $\log(k_b)$  values. This is comparable to the results for ‘conventional’ or ‘black-box’ QSBR methods (50–80%) and meets regulatory requirements ( $R^2 > 0.5$ <sup>10,16,89,90</sup>). This shows that eqn (3) can compete with more complex methods, and implies that eqn (3) can be used as a basis to further improve our understanding of biodegradation in environmental matrices. Details are described below.

#### 3.1-1. Diffusion-limited biodegradation (A)

A regression was performed for electron-rich, ‘similarly reactive chemicals’ for which the term  $e^{(-\Delta G_{i,j}^\ddagger/RT)} \approx 1$  (eqn (3)). Then, we obtained a reasonable fit between  $k_b$  and  $A$  (Fig. 3). Fig. S7† shows additional relationships. The regression coefficient ( $R^2$ ) for ‘similarly reactive chemicals’ (closed black symbols) is higher as compared to the regression coefficient for all chemicals considered simultaneously, Fig. 3. Fig. 4 depicts similar results. The offset of the regression is lower for heterogeneous chemicals (entailing diverse functionalities, *e.g.* multivalent

charges and heavy atoms): we observe structurally higher  $k_b$  values *e.g.* ‘electron-rich’ linear ethylene glycol oligomers (Fig. 4; S6B†). Ribose analogs and amino acids (*i.e.* natural substances) have even higher  $k_b$  values than on the basis of  $A$  (Fig. 3; S7† and regressions with individual characteristics in Fig. 4). The inclusion of multiplicity (number of equivalent functional groups,  $\Sigma i$  term in eqn (3)) did not visibly contribute to a better fit.

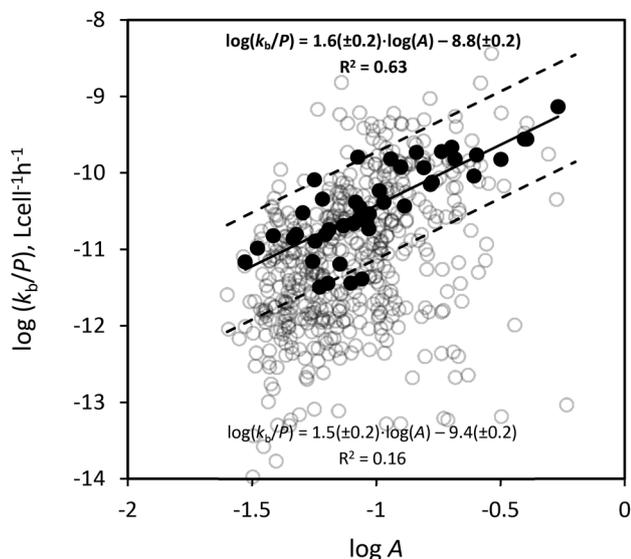
The relationship between  $k_b/P$  (excluding ‘natural substances’) and the frequency factor  $A$  indicates (Fig. 3) that the biodegradation of ‘similarly reactive’, electron-rich chemicals is diffusion-limited (eqn (3)). In other words, diffusion establishes the upper limit for  $k_b$ . We note that an upper limit can be set also by active uptake.

#### 3.1-2. Individual geometrical characteristics

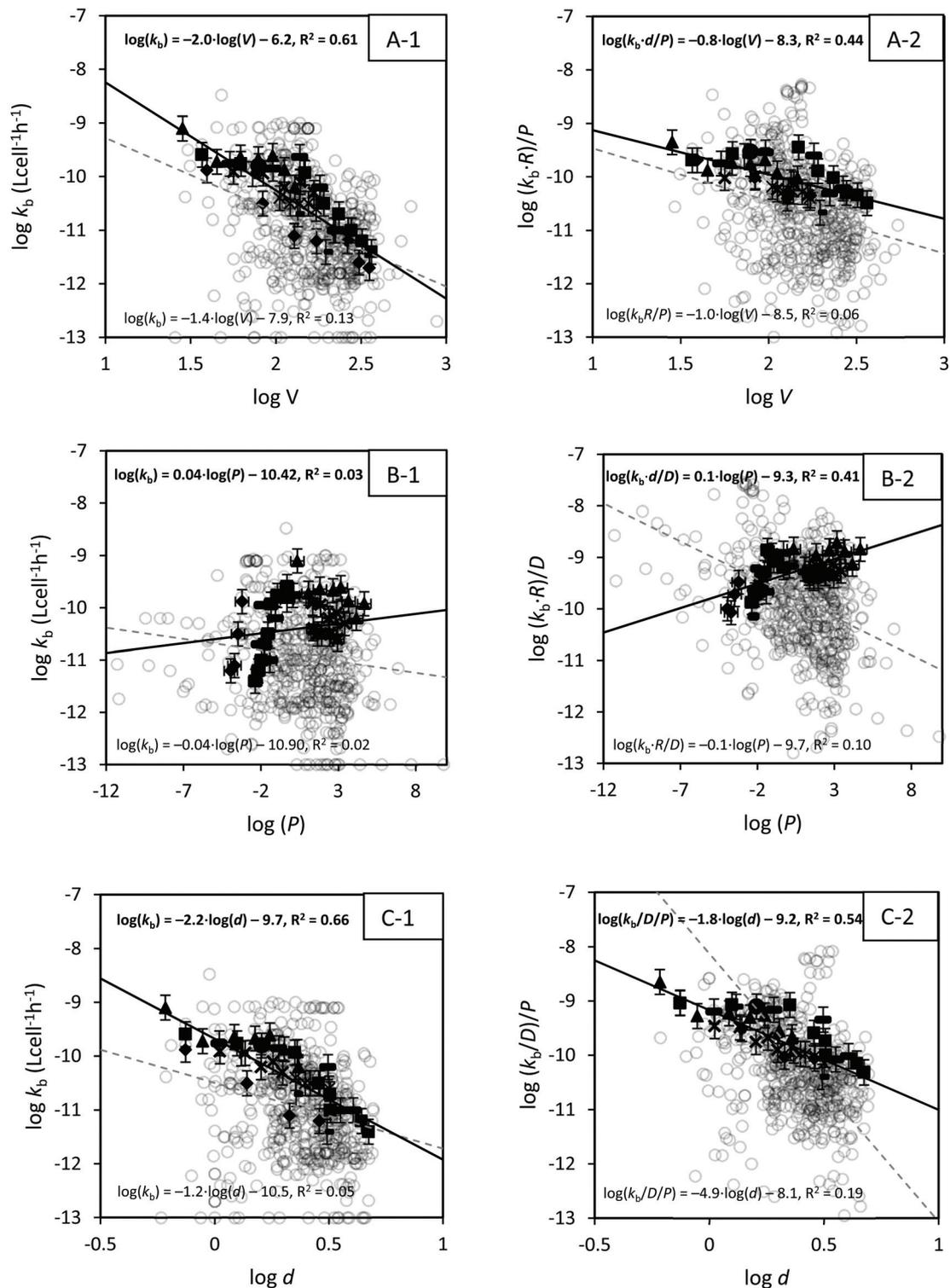
According to eqn (3), the frequency factor  $A$  consists of the diffusion coefficient (*i.e. via*  $V$ ), the partitioning function ( $\sim K_{OW}$ ) and accessibility (*i.e. d*) (eqn (3); SI2†). The results from the regression of  $k_b$  values with individual geometric descriptors are given in Fig. 4-1. Based on the results in Fig. 4-1, we transformed the  $k_b$  values according to eqn (3). For example, the ‘scaling exponent’ for  $V$  involving the transformed values is  $-0.8 \pm 0.1$  ( $k_b \cdot d \cdot P^{-1} \propto V^{-0.8 \pm 0.1}$ , Fig. 4A-2), which agrees with previous studies.<sup>91–93</sup> In analogy, when corrected for hydrophobicity, the permeabilities for larger penetrants (diameter  $\geq 0.6$  nm) follow the Stokes–Einstein relation for diffusion ( $D \sim cV^{-1/3}$  with  $c$  being a geometry dependent factor).<sup>93</sup> The ‘pure’ scaling exponent for volume is  $-2.0 \pm 0.1$  ( $k_b \propto V^{-2.0 \pm 0.1}$ ) for ‘similarly reactive chemicals’ ( $-1.4 \pm 0.4$ , *i.e.*  $k_b \propto V^{-1.4 \pm 0.4}$  for all chemicals). Since diffusion constants scale to  $V$  with exponents of  $\sim -0.8$  ( $k_b \propto V^{-0.8}$ ) depending on the penetrant and the barrier,<sup>61,79,93–95</sup> the regression between volume and  $k_b$  suggests a double relationship ( $k_b \propto (V \cdot V)^{-x}$ ).

On the basis of  $V$  only, we observe certain outliers. Their ‘low’  $k_b$  values are explained likely by steric inhibition (bulky group, *e.g.* neopentane). However, when alternative sites for microbial attack are present, the relative values for  $k_b$  increase as a function of distance from the bulky group (Fig. S4†). In other words, when the molecule is longer (*e.g.* having a linear alkane/glycol tail group), the inhibition is less pronounced (Fig. S4†). The calculated values for ‘relative accessibilities’, *i.e.*  $d$  values, range from 1 Å to 4 Å. Fig. 4C-1 and C-2 show the results of regressions. Strictly,  $d$  is a characteristic distance that depends on the trajectory of the substrate chemical towards the active site (eqn (3) and Fig. 1). Relevant distances vary between the type of enzyme and substrate: rate-limiting forming/breaking of bonds occurs over distances of 1.5 Å to 3.5 Å (ref. 96–98) and cutoff values for diffusion through membranes and cell wall pores are 1–2 Å (ref. 99) and 4–20 Å,<sup>100,101</sup> respectively. One expects that, if a reaction is possible at all, the distance  $d$  is larger upon incorporation of a bulky group (confirmed by the negative slope in Fig. 4C).

$V$  is also interrelated with  $K_{OW}$ , Fig. S5.† Thus, the incorporation of extra functional groups affects  $P$ . When corrected for the diffusion coefficient  $D$  and ‘accessibility’  $d$ , we find

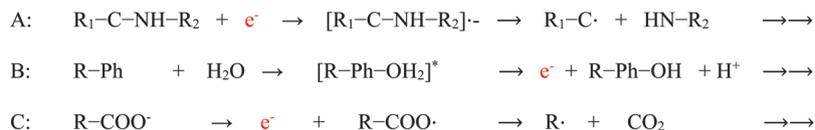


**Fig. 3** The relationship between  $\log(k_b/P)$  and the frequency factor,  $\log(A)$ . Solid black circles are ‘similarly reactive’, electron rich chemicals  $e^{(-\Delta G_{i,j}^\ddagger/RT)} \sim \text{constant} \sim 1$ ,  $R^2 = 0.63$ . Open gray circles are all chemicals considered in this study ( $R^2 = 0.16$ ). Dashed lines indicate 95% confidence intervals ( $2\sigma$ ).

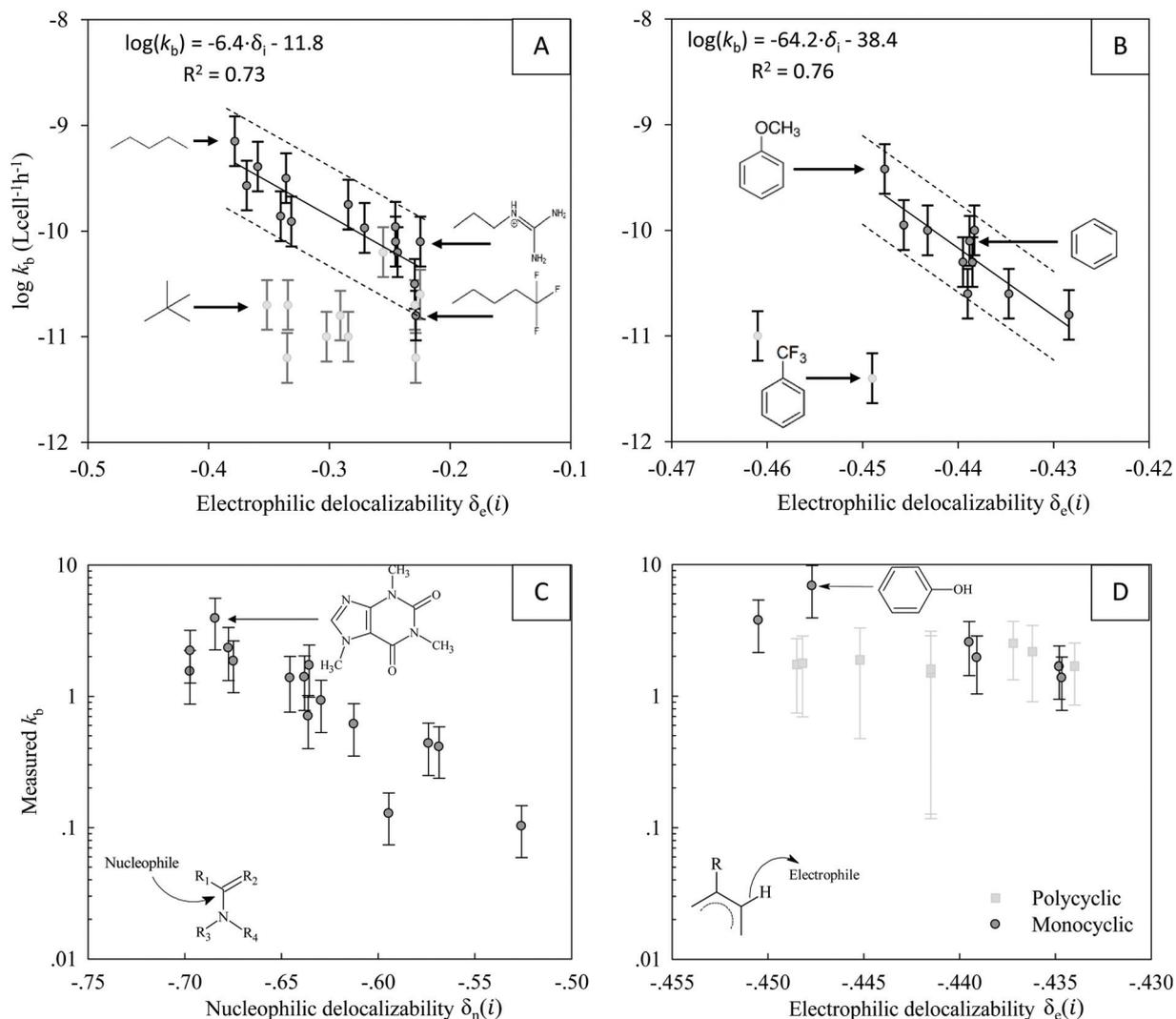


**Fig. 4** Biodegradation versus  $\log V$ ,  $\log P$  (i.e.  $\log K_{OW}$ ) and logarithm of the characteristic distance  $\log d$  for non-transformed (4-1) and transformed  $\log k_b$  (4-2) values. Solid black datapoints denote 'similarly reactive' chemicals.  $P$  is calculated via  $K_{OW}$  for the species at pH = 7. Error bars denote prediction uncertainty ( $1\sigma$ ). Symbols denote different 'families of structurally similar chemicals' (Fig. S6†), e.g. ethylene glycol oligomers.  $R^2 = 0.7$ – $0.8$  (class-specific regressions).





**Scheme 1** Simplified reaction schemes for amine dehydrogenase (A, carbon–nitrogen bond), dioxygenation (B,  $\text{sp}^2$  carbon–hydrogen bond) and decarboxylation (C, carbon–carbon bond). For each arrow in the schemes an enzymatic step is involved.

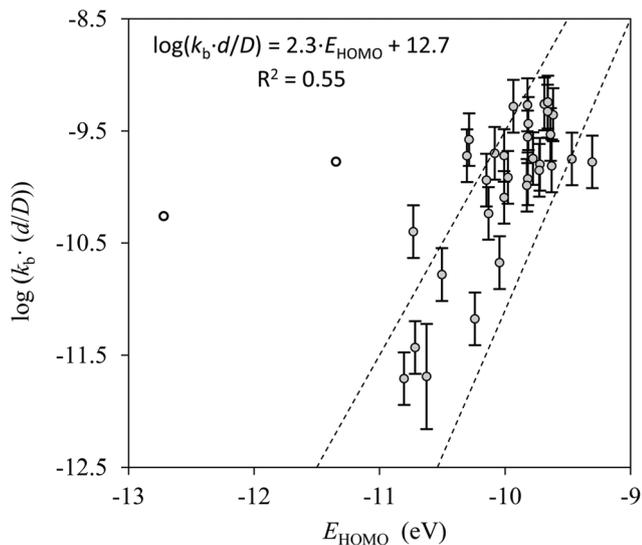


**Fig. 6** Biodegradation rate constants versus delocalizability  $\delta(i)$  ( $\delta(i) \propto \Delta G^\ddagger(i-j)$ ). Error bars indicate uncertainty associated with QSBR-predictions (top) and data conversion (bottom). Top:  $\log(k_b)$  for surface water vs. electrophilic delocalizability for functionalized linear alkanes (A) and benzene analogs (B). We took delocalizabilities as minimum values on aliphatic (A) and maximum values on aromatic carbons (B),  $i$ . Dashed lines indicate 95% confidence intervals ( $2\sigma$ ). Outliers include trifluorotoluene and trichlorotoluene (B). Bottom:  $k_b$  for wastewater (N-containing chemicals in C and monocyclic aromatics in D) vs. nucleophilic and electrophilic delocalizability, figure reproduced with permission ref. 23.

$\Delta G_{\delta(i)}$ , in the transition state is more likely  $[(E-S)]^*$  in Scheme 1B (left arrow) for aromatics. Thus, the relative influence of  $\delta(i)$  on  $\Delta G^\ddagger$  appears larger.

The outliers in Fig. 6A include neopentane with lower values of  $k_b$  than expected on the basis of  $\delta_e(i)$  only. Outliers in Fig. 6A can be attributed to a low value for  $A$  (section 3.1). In Fig. 6B, outliers such as trifluorotoluene ( $k_b$  is 100 times lower

than expected, Fig. 6B) can also be explained by electronic factors, *i.e.* the ionization potential (IP) is  $0.5 \pm 0.3$  eV higher compared to the chemicals in the regression. Illustratively, we found a relationship between  $k_b$  and  $E_{\text{HOMO}} (\approx -\text{IP})$  for carboxylates, Fig. 7. The factor  $0.5 \pm 0.3$  eV implies a factor difference in  $k_b$  of  $10 \pm 5$  (based on non-phenomenological LFER behavior<sup>110</sup>).



**Fig. 7** Relationship between  $\log$ -transformed  $k_b$  values and the energy of the electron pair on the carboxylate group. In most cases, this is equivalent to  $E_{\text{HOMO}}$ , hence the x-axis title. Error bars denote prediction uncertainty. Dashed lines indicate expected values based on LFER.<sup>110</sup>  $k_b$  values were not corrected for  $P$  (S2†).

$E_{\text{HOMO}}$  explains 40–50% of the variance in  $k_b$  values for carboxylates (Fig. 7). On the basis of  $E_{\text{HOMO}}$ , trifluoroacetate and oxalate are expected to have a low and high  $k_b$ , respectively. In general, frontier orbital energies have been used widely to describe biodegradation,<sup>22,32,112</sup> as well as biotransformation.<sup>80,84</sup> The relation between  $k_b$  and  $E_{\text{HOMO}}$  characterizes a linear free-energy relationship (LFER) on the basis of  $\Delta G_{\text{r}}$ ,<sup>115,116</sup> in which the thermodynamic driving force is oxidative. Note that if the delocalization energy is constant (blue and green  $[E - S]^*$  parabola in Fig. 5), we return to LFER behavior based on  $\Delta G_{\text{r}}$  (e.g. via  $E_{\text{HOMO}}$ ), right black arrows.

In decarboxylation, the carboxylate can undergo electron transfer with a suitable partner (oxidant). During this process,

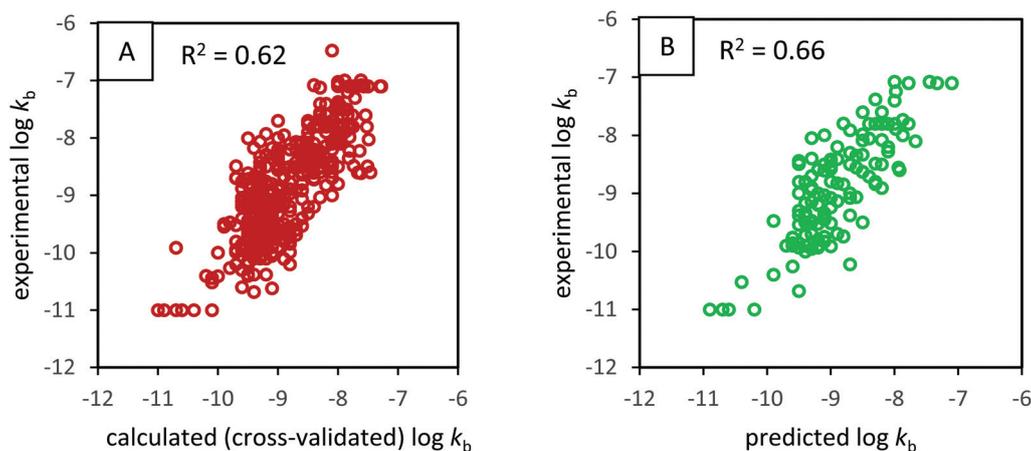
the carboxylate is oxidized to an acyloxy radical, which subsequently fragments to yield an alkyl (or alkylaryl) radical and  $\text{CO}_2$  (Scheme 1C). (Photo)chemical variants of this reaction have been examined.<sup>117,118</sup>  $E_{\text{HOMO}}$  (Fig. 7) illustrates the relevance of the first step of the mechanism. Tricyanoacetate did not adhere to the LFER (Fig. 7; S8A†). We note that free energies are susceptible to solvation,<sup>116,119</sup> which might refine relationships.<sup>110</sup> Outliers might be expected in the case  $k_b$  entails (also) a reductive pathway, such as for nitriles or amines (Scheme 1A).

### 3.3. Comparison of eqn (3) with existing methods

In an attempt to characterize the relative accuracy of the predictions *via* eqn (3), we applied additional techniques: statistical evaluation of an RF-QSBR model is given in Fig. 8; a validation *via* comparison with CATABOL (Fig. S2†) and Biowin (Table 2). RF-QSBR has  $R^2_{\text{ext}} = 0.66 \pm 0.05$ , and root-mean-squared error ( $\text{RMSE}_{\text{ext}} = 0.53 \pm 0.03$ ).

The explained variance of RF-QSBR ( $R^2 \sim 0.66$ ) is comparable to that by means of eqn (3) (0.58–0.76). In contrast, the RMSE differs significantly ( $\sim 0.53$  and  $0.22$ – $0.46$ , resp.). We attribute this to the inclusion of ‘exotic chemicals’ in the training set for the RF-QSBR, which virtually do not biodegrade: siloxanes, inorganics or chemicals with high degrees of halogenation/low carbon content. Previously, Arnot *et al.* correlated the output from different Biowin versions to aerobic half-lives of 40 chemicals, giving  $R^2 = 0.58$ – $0.78$ .<sup>120</sup> For CATABOL,  $R^2 = 0.69$ .<sup>121</sup> The values are similar to eqn (3) and the RF-QSBR (Table 2), though one might argue that there are differences in the applicability domain. To our knowledge, there is no consensus on which is the most reliable experimental dataset to develop/test biodegradation models.<sup>3,36,37,49,122</sup> Thus, statistical parameters reflect both prediction uncertainty and variability due to test conditions.

Many QSBRs use non-linear methods: analogous to a RF algorithm (Fig. 2), CATABOL simulates metabolism by a rule-based approach.<sup>78</sup> While little quantitative information can be



**Fig. 8** A: Cross-validated (5-fold) values for calculated  $\log(k_b)$  versus experimental  $k_b$  ( $R^2 = 0.62$ ). B: Predicted  $\log(k_b)$  versus experimental  $\log(k_b)$ ,  $R^2 = 0.66$ . For training (A) and evaluation (B), 75% and 25% of the total dataset were used, resp.

**Table 2** Statistical results regression of outcomes of eqn (3) with experimental biodegradation rate constants. The results of QSBRs are given as a comparison

Dataset	<i>N</i>	<i>R</i> <sup>2</sup>	<i>q</i> <sup>2</sup>	RMSE	MAE	<i>F</i>
Eqn (3) (diffusion, <i>A</i> )	44	0.63	—	0.36	—	—
Eqn (3) (reaction, $\Delta G^\ddagger$ ) <sup>a</sup>	14; 10; 30	0.73; 0.76; 0.58	—	0.22; 0.19; 0.46	—	—
Training set (Fig. 8A)	409	0.6 ± 0.04	0.6 ± 0.03	0.58 ± 0.02	0.46 ± 0.02	676
Test set (Fig. 8B)	131	0.66 ± 0.05	0.65 ± 0.05	0.53 ± 0.03	0.43 ± 0.03	250
Biowin	40	0.58–0.78	—	—	—	—
CATABOL	— <sup>b</sup>	0.69	—	—	—	—

<sup>a</sup> Separate datasets were considered since we distinguish between contributions from  $E_{\text{HOMO}}$  and  $d_e(i)$ . <sup>b</sup> The entire range (from readily degradable to difficult-to-degrade substances).

extracted, the more complex, non-linear structures might better capture intermediate cases (combined diffusion- and reaction-limited degradation) and additional phenomena. Phenomena not incorporated in eqn (3) whilst incorporated in the RF-QSBR and CATABOL may include differences in test conditions needed for heterogeneous (dissimilar) chemicals (*e.g.* toxicity or solubility), ion trapping,<sup>27,123,124</sup> non-linear sorption, co-metabolism, size exclusion<sup>100,101</sup> and a diversity of active uptake routes for heterogeneous (dissimilar) chemicals.

3D structures of the active sites and the trajectories of diffusion vary. This has implications for *d* and *P* (eqn (3)), whose precise values are determined by the nature of the optimal pathway towards the transition state, as influenced also by the matrix' geometric restrictions and interactions: the size, form, accessibility, and the nature of surrounding amino acid residues of a catalytic site are of greatest importance for the binding specificity.<sup>48</sup> Binding sites in related enzymes normally have related structures and cavities since evolution tends to conserve structural features that are of importance to biological function, activity and specificity. Thereby, the rates of transport and enzymatic transformation may be optimized towards each other, making 'intermediate cases' plausible. Variation in metabolic capacities seems reasonably small, even for strongly differing microbial communities.<sup>125,126</sup>

## 4. Practical application and outlook

Herein, we highlighted statistical as well as mechanistic limitations of the current empirical 'black-box' quantitative structure-biodegradation relationships. Such 'fitting' methods are subject to differences in testing setup/conditions interfering with the actual biodegradation 'signal', and allow 'noise' in the statistical model. Alternatively, we sought to adapt/test algebraic formulations describing biodegradation in terms of transition state theory. The results show that the algebraic formulations do not necessarily perform less well than statistical 'black-box' methods. Given biodegradation as a crucial parameter in environment assessment, risks for 'data-poor' chemicals can be assessed *via* these formulations.

Given the similarity in performances (Table 2), we hypothesize that nodes in rule-based decision models (*e.g.* Fig. 2) reflect chemical class-specific relationships, whose descriptors reflect the

terms in eqn (3). Mechanisms of biodegradation are not likely linearly dependent on (often inter-correlated) chemical characteristics; eqn (3) helps to rationalize why non-linear models perform satisfactorily. As the simultaneous presence of multiple bacteria, enzymes, metabolic steps, uptake and reaction pathways can substitute and supplement each other, validation *via* heterogenic (*e.g.* field) data will be challenging. Future study will need to adapt and test the formulations for increasingly structurally exotic chemicals as well as for anaerobic biodegradation.

Minimal parametrization of eqn (3) was necessary and therefore, we find it plausible that the formulations have better extrapolative capability. As a practical example, we expect perfluorooctanoic acid ( $E_{\text{HOMO, PFOA}} = -10.6$  eV;  $A_{\text{PFOA}} = 0.05$ ) and 'large' colloids such as humic acid ( $\delta_{(i), \text{HA}} = \sim -0.45$ ;  $A_{\text{HA}} = 0.001-0.01$ ) to be transformed with  $\log k_{\text{b, PFOA}} = -13.0$  ( $\text{DT}_{50} \sim 10$  years) and  $\log k_{\text{b, HA}} = -12.0-13.0$  (*i.e.* 1–10 years) in surface water, respectively. A larger extrapolation capacity would allow for a more robust 'green' metric to design 'new' or substitute chemicals or products. Therein, the potential for biodegradation can be assessed during design.

## Conflicts of interest

The authors declare no competing financial interest or other conflicts.

## Acknowledgements

This work is part of the research programme TTW financing the Contaminants of Emerging Concern in the Water Cycle (CERCEC) project number 15759, which is financed by the Dutch Research Council (NWO). Personal discussions with J. Koch and T. Nauser (ETH Zurich) aided the conceptualization and parametrization in this study. The discussions with G. Chen (RIVM) aided the interpretation of our results, and were greatly appreciated.

## References

- 1 *Handbook of Green Chemistry*, ed. P. T. Anastas, 2010, ISBN9783527628698.

- 2 C. Leder, T. Rastogi and K. Kummerer, Putting benign by design into practice—novel concepts for green and sustainable pharmacy: Designing green drug derivatives by non-targeted synthesis and screening for biodegradability, *Sustainable Chem. Pharm.*, 2015, **2**, 31–36.
- 3 C. Rucker and K. Kummerer, Modeling and predicting aquatic aerobic biodegradation - a review from a user's perspective, *Green Chem.*, 2012, **14**(4), 875–887.
- 4 K. Kummerer, Sustainable from the very beginning: rational design of molecules by life cycle engineering as an important approach for green pharmacy and green chemistry, *Green Chem.*, 2007, **9**(8), 899–907.
- 5 K. Fenner, S. Canonica, L. P. Wackett and M. Elsner, Evaluating Pesticide Degradation in the Environment: Blind Spots and Emerging Opportunities, *Science*, 2013, **341**(6147), 752–758.
- 6 European Chemicals Agency, *Guidance on Registration. Version 2.0. Guidance for the Implementation of REACH*, 2012.
- 7 E. Commission, Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency. 1907/2006. Official Journal of the European Union, 2006.
- 8 OECD, *OECD Guideline For The Testing Of Chemicals. Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test 309*, 2004.
- 9 A. J. Hendriks, How To Deal with 100,000+ Substances, Sites, and Species: Overarching Principles in Environmental Risk Assessment, *Environ. Sci. Technol.*, 2013, **47**(8), 3546–3547.
- 10 E. Benfenati, *The e-Book on QSAR and REACH: Theory, Guidance and Applications*, 2012.
- 11 ECHA Information on chemicals: Pre-registered substances. echa.europa.eu (accessed 20/07/2017).
- 12 ECHA, *ECHA's REACH 2018 Roadmap*, Helsinki, Finland, 2015.
- 13 U. Muhammad, A. Uzairu and D. E. Arthur, Review on: quantitative structure activity relationship (QSAR) modeling, *J. Anal. Pharm. Res.*, 2018, **7**(2), 240.
- 14 A. Sabljic and Y. Nakagawa, Biodegradation and Quantitative Structure-Activity Relationship (QSAR), in *In Non-First Order Degradation and Time-Dependent Sorption of Organic Chemicals in Soil*, 2014, ch. 4, pp. 57–84.
- 15 T. M. Nolte and A. M. J. Ragas, A review of quantitative structure-property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps, *Environ. Sci.: Processes Impacts*, 2017, **19**(3), 221–246.
- 16 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.
- 17 ECHA, *The use of alternatives to testing on animals for the REACH Regulation. Third report under Article 117(3) of the REACH Regulation. ECHA-17-R-02-EN*, 2017.
- 18 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**(7–8), 719–755.
- 19 R. S. Boethling and A. Sabljic, Screening-level model for aerobic biodegradability based on a survey of expert knowledge, *Environ. Sci. Technol.*, 1989, **23**, 672–679.
- 20 J. Jaworska, S. Dimitrov, N. Nikolova and O. Mekenyan, Probabilistic assessment of biodegradability based on metabolic pathways: catabol system, *SAR QSAR Environ. Res.*, 2002, **13**, 307–323.
- 21 C. Dick, S. Rey, A. Boschung, F. Miffon and M. Seyfried, Current limitations of biodegradation screening tests and prediction of biodegradability: A focus on fragrance substances, *Environ. Technol. Innov.*, 2016, **5**, 208–224.
- 22 C. Burgis, *Predicting Biological Removal of Contaminants in Wastewater Treatment: QSBR Modeling*, University of Virginia, Virginia, US, 2012.
- 23 T. M. Nolte, G. Chen, *et al.*, Quantitative Structure-Biodegradation Relationships for aerobic wastewater treatment: disentanglement of the chemical, physical, and biological effects, *STOTEN*, 2020, **708**, 133863.
- 24 R. Kuhne, R. U. Ebert and G. Schuurmann, Estimation of compartmental half-lives of organic compounds - Structural similarity versus EPI-Suite, *QSAR Comb. Sci.*, 2007, **26**(4), 542–549.
- 25 T. Gouin, I. Cousins and D. Mackay, Comparison of two methods for obtaining degradation half-lives, *Chemosphere*, 2004, **56**(6), 531–535.
- 26 K. Fenner, S. Canonica, B. I. Escher, L. Gasser, S. Spycher and H. C. Tulp, Developing methods to predict chemical fate and effect endpoints for use within REACH, *Chimia*, 2006, **60**(10), 683–690.
- 27 T. M. Nolte, K. Pinto-Gil, A. J. Hendriks, A. M. J. Ragas and M. Pastor, Quantitative structure-activity relationships for primary aerobic biodegradation of organic chemicals in pristine surface waters: starting points for predicting biodegradation under acclimatization, *Environ. Sci.: Processes Impacts*, 2018, **20**(1), 157–170.
- 28 A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Mol. Inf.*, 2010, **29**(6–7), 476–488.
- 29 ECHA, *Practical guide How to use and report (Q)SARs*, 2016.
- 30 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**(5), 1500–1511.
- 31 G. H. Lu, C. Wang and G. Z. Bao, Quantitative structure-biodegradation relationship study for biodegradation rates of substituted benzenes by river bacteria, *Environ. Toxicol. Chem.*, 2003, **22**(2), 272–275.
- 32 P. Howard, W. Meylan, D. Aronson, W. Stiteler, J. Tunkel, M. Comber and T. F. Parkerton, A new biodegradation pre-

- diction model specific to petroleum hydrocarbons, *Environ. Toxicol. Chem.*, 2005, **24**(8), 1847–1860.
- 33 J. Monod, *Recherches sur la croissance des Cultures Bactériennes*, Paris, 1942.
- 34 T. J. Martin, J. R. Snape, A. Bartram, A. Robson, K. Acharya and R. J. Davenport, Environmentally Relevant Inoculum Concentrations Improve the Reliability of Persistent Assessments in Biodegradation Screening Tests, *Environ. Sci. Technol.*, 2017, **51**(5), 3065–3073.
- 35 B. A. J. Poursat, R. J. M. van Spanning, P. de Voogt and J. R. Parsons, Implications of microbial adaptation for the assessment of environmental persistence of chemicals, *Crit. Rev. Environ. Sci. Technol.*, 2019, **49**(23), 2220–2255.
- 36 W. Peijnenburg and W. Karcher, in *Proceedings of the Workshop on Quantitative Structure Activity Relationships for Biodegradation*, No. 719101021, RIVM, 1995, pp. 1–138.
- 37 K. Fenner, U. Schenker and M. Scheringer, Modelling environmental exposure to transformation products of organic chemicals, in *Handbook of Environmental Chemistry*, Springer, Berlin, 2008, vol. 2P, pp. 121–149.
- 38 A. Sabljic and W. Peijnenburg, IUPAC Technical Report on modeling lifetime and degradability in air, soil and water systems: Modeling lifetime and degradability of organic compounds in air, soil and water systems, *Pure Appl. Chem.*, 2001, **73**(8), 1331–1348.
- 39 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**(7), 1600–1609.
- 40 C. E. Cowan, T. W. Federle, R. J. Larson and T. C. Feijtel, Impact of biodegradation test methods on the development and applicability of biodegradation QSARs, *SAR QSAR Environ. Res.*, 1996, **5**(1), 37–49.
- 41 R. W. Okey and H. D. Stensel, A QSAR-based biodegradability model - A QSBR, *Water Res.*, 1996, **30**(9), 2206–2214.
- 42 M. Honti, S. Hahn, D. Hennecke, T. Junker, P. Shrestha and K. Fenner, Bridging across OECD 308 and 309 Data in Search of a Robust Biotransformation Indicator, *Environ. Sci. Technol.*, 2016, **50**(13), 6865–6872.
- 43 S. Wilkens, Structure and mechanism of ABC transporters, *F1000Prime Rep.*, 2015, **7**, 14.
- 44 Y. Moriyama, M. Hiasa, T. Matsumoto and H. Omote, Multidrug and toxic compound extrusion (MATE)-type proteins as anchor transporters for the excretion of metabolic waste products and xenobiotics, *Xenobiotica*, 2008, **38**(7–8), 1107–1118.
- 45 D. F. Paris and J. E. Rogers, Kinetic Concepts for Measuring Microbial Rate Constants - Effects of Nutrients on Rate Constants, *Appl. Environ. Microbiol.*, 1986, **51**(2), 221–225.
- 46 M. Radke, C. Lauwigi, G. Heinkele, T. E. Murdter and M. Letzel, Fate of the Antibiotic Sulfamethoxazole and Its Two Major Human Metabolites in a Water Sediment Test, *Environ. Sci. Technol.*, 2009, **43**(9), 3135–3141.
- 47 P. B. Price and T. Sowers, Temperature dependence of metabolic rates for microbial growth, maintenance, and survival, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**(13), 4631–4636.
- 48 V. L. Arcus, E. J. Prentice, J. K. Hobbs, A. J. Mulholland, M. W. Van der Kamp, C. R. Pudney, E. J. Parker and L. A. Schipper, On the Temperature Dependence of Enzyme-Catalyzed Rates, *Biochemistry*, 2016, **55**(12), 1681–1688.
- 49 R. S. Boethling, P. H. Howard, W. Meylan, W. Stiteler, J. Beauman and N. Tirado, Group-Contribution Method for Predicting Probability and Rate of Aerobic Biodegradation, *Environ. Sci. Technol.*, 1994, **28**(3), 459–465.
- 50 S. Rinaldi, M. W. Van der Kamp, K. E. Ranaghan, A. J. Mulholland and G. Colombo, Understanding Complex Mechanisms of Enzyme Reactivity: The Case of Limonene-1,2-Epoxyde Hydrolases, *ACS Catal.*, 2018, **8**(7), 5698–5707.
- 51 A. Spinello, I. Ritacco and A. Magistrato, The Catalytic Mechanism of Steroidogenic Cytochromes P450 from All-Atom Simulations: Entwinement with Membrane Environment, Redox Partners, and Post-Transcriptional Regulation, *Catalysts*, 2019, **9**(1), 81.
- 52 L. Ji and G. Schuurmann, Computational Biotransformation Profile of Paracetamol Catalyzed by Cytochrome P450, *Chem. Res. Toxicol.*, 2015, **28**(4), 585–596.
- 53 K. Matzjasyewski and T. P. Davis, Theory of radical reactions, in *Handbook of radical polymerization*, Wiley-Interscience, Hoboken, 2002, ch. 1.
- 54 A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik and R. Milo, The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters, *Biochemistry*, 2011, **50**(21), 4402–4410.
- 55 S. D. Copley, Evolution of efficient pathways for degradation of anthropogenic chemicals, *Nat. Chem. Biol.*, 2009, **5**(8), 560–567.
- 56 S. Senapati, C. F. Wong and J. A. McCammon, Finite concentration effects on diffusion-controlled reactions, *J. Chem. Phys.*, 2004, **121**(16), 7896–7900.
- 57 M. T. Zumstein, D. Rechsteiner, N. Roduner, V. Perz, D. Ribitsch, G. M. Guebitz, H. P. E. Kohler, K. McNeill and M. Sander, Enzymatic Hydrolysis of Polyester Thin Films at the Nanoscale: Effects of Polyester Structure and Enzyme Active-Site Accessibility, *Environ. Sci. Technol.*, 2017, **51**(13), 7476–7485.
- 58 J. J. Kotz, P. M. Treichel and J. R. Townsend, A Microscopic View of Reaction Rates, in *Chemistry and Chemical Reactivity*, Brooks/Cole Cengage Learning, 2011, ch. 15.5.
- 59 J. M. Schurr and K. S. Schmitz, Orientation Constraints and Rotational Diffusion in Bimolecular Solution Kinetics. A Simplification, *J. Phys. Chem.*, 1976, **80**(17), 1934–1936.

- 60 T. L. Hill, Diffusion Frequency Factors in Some Simple Examples of Transition-State Rate Theory, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**(3), 679–683.
- 61 A. Walter and J. Gutknecht, Permeability of Small Nonelectrolytes through Lipid Bilayer-Membranes, *J. Membr. Biol.*, 1986, **90**(3), 207–217.
- 62 A. J. Hendriks, Organisms, in *Syllabus for Environmental and Ecological Modelling*, 2016, ch. 7, p. 79.
- 63 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*, 2014, vol. 1174, pp. 57–84.
- 64 Molinspiration Molinspiration Cheminformatics. <http://www.molinspiration.com/>.
- 65 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**(6), 533–554.
- 66 *ChemAxon Calculator Plugin for structure property prediction. Marvin Version 5.2.0.*
- 67 J. J. P. Stewart, *MOPAC*, Stewart Computational Chemistry: Colorado Springs, CO, USA, 2016.
- 68 J. Dong, D. S. Cao, H. Y. Miao, S. Liu, B. C. Deng, Y. H. Yun, N. N. Wang, A. P. Lu, W. B. Zeng and A. F. Chen, ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation, *J. Cheminf.*, 2015, **7**(60), DOI: 10.1186/s13321-015-0109-z.
- 69 D. S. Cao, Q. S. Xu, Q. N. Hu and Y. Z. Liang, ChemoPy: freely available python package for computational biology and cheminformatics, *Bioinformatics*, 2013, **29**(8), 1092–1094.
- 70 C. W. Yap, PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints, *J. Comput. Chem.*, 2011, **32**(7), 1466–1474.
- 71 RDKit Open-source cheminformatics; <http://www.rdkit.org>, 2016.
- 72 A. Racz, D. Bajusz and K. Heberger, Intercorrelation Limits in Molecular Descriptor Preselection for QSAR/QSPR, *Mol. Inf.*, 2019, **38**(8–9), 1800154.
- 73 J. Wicker, K. Fenner, L. Ellis, L. Wackett and S. Kramer, Predicting biodegradation products and pathways: a hybrid knowledge- and machine learning-based approach, *Bioinformatics*, 2010, **26**(6), 814–821.
- 74 D. Gamberger, D. Horvatic, S. Sekusak and A. Sabljic, Applications of experts' judgement to derive structure biodegradation relationships, *Environ. Sci. Pollut. Res.*, 1996, **3**(4), 224–228.
- 75 L. Breiman, Random forests, *Mach. Learn.*, 2001, **45**(1), 5–32.
- 76 A. Criminisi, J. Shotton and E. Konukoglu, *Decision Forests for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning; Microsoft Research technical report TR-2011-114*, 2011.
- 77 A. Golbraikh and A. Tropsha, Beware of  $q(2)!$ , *J. Mol. Graphics Modell.*, 2002, **20**(4), 269–276.
- 78 J. Jaworska, S. Dimitrov, N. Nikolova and O. Mekenyan, Probabilistic assessment of biodegradability based on metabolic pathways: Catabol system, *SAR QSAR Environ. Res.*, 2002, **13**(2), 307–323.
- 79 A. J. Hendriks, A. van der Linde, G. Cornelissen and D. T. H. M. Sijm, The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight, *Environ. Toxicol. Chem.*, 2001, **20**(7), 1399–1420.
- 80 A. Pirovano, M. A. J. Huijbregts, A. M. J. Ragas, K. Veltman and A. J. Hendriks, Mechanistically-based QSARs to Describe Metabolic Constants in Mammals, *ATLA, Altern. Lab. Anim.*, 2014, **42**(1), 59–69.
- 81 E. Rorije, J. H. Langenberg, J. Richter and W. J. Peijnenburg, Modeling reductive dehalogenation with quantum chemically derived descriptors, *SAR QSAR Environ. Res.*, 1995, **4**(4), 237–252.
- 82 J. L. Peng, J. Lu, Q. C. Shen, M. Y. Zheng, X. M. Luo, W. L. Zhu, H. L. Jiang and K. X. Chen, In silico site of metabolism prediction for human UGT-catalyzed reactions, *Bioinformatics*, 2014, **30**(3), 398–405.
- 83 J. Zaretski, M. Matlock and S. J. Swamidass, XenoSite: Accurately Predicting CYP-Mediated Sites of Metabolism with Neural Networks, *J. Chem. Inf. Model.*, 2013, **53**(12), 3373–3383.
- 84 D. F. V. Lewis, Frontier orbitals in chemical and biological activity: Quantitative relationships and mechanistic implications, *Drug Metab. Rev.*, 1999, **31**(3), 755–816.
- 85 M. Karelson, V. S. Lobanov and A. R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.*, 1996, **96**(3), 1027–1043.
- 86 D. A. Benson and M. M. Meerschaert, Simulation of chemical reaction via particle tracking: Diffusion-limited versus thermodynamic rate-limited regimes, *Water Resour. Res.*, 2008, **44**(12), W12201.
- 87 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**(23), 13925–13932.
- 88 E. Petzinger and J. Geyer, Drug transporters in pharmacokinetics, *Naunyn-Schmiedeberg's Arch. Pharmacol.*, 2006, **372**(6), 465–475.
- 89 OECD, *Document for Exposure Assessment based on Environmental Monitoring*, OECD Environment, Health and Safety Publications, Series on Testing and Assessment, 2013.
- 90 J. C. Dearden, P. Rotureau and G. Fayet, QSPR prediction of physico-chemical properties for REACH, *SAR QSAR Environ. Res.*, 2013, **24**(4), 545–584.

- 91 W. Hayduk and H. Laudie, Prediction of diffusion coefficients for non-electrolytes in dilute aqueous solutions, *AIChE J.*, 1974, **20**(3), 611–615.
- 92 C. R. Wilke and P. C. Chang, Correlation of diffusion coefficients in dilute solutions, *AIChE J.*, 1955, **1**, 264–270.
- 93 S. J. Marrink and H. J. C. Berendsen, Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations, *J. Phys. Chem.*, 1996, **100**(41), 16729–16738.
- 94 T. M. Nolte, K. Kettler, J. A. J. Meesters, A. J. Hendriks and D. van de Meent, A Semi-Empirical Model for Transport of Inorganic Nanoparticles across a Lipid Bilayer: Implications for Uptake by Living Cells, *Environ. Toxicol. Chem.*, 2015, **34**(3), 488–496.
- 95 R. P. Schwarzenbach, P. M. Gschwend and D. M. Imboden, *Environmental Organic Chemistry*, Wiley, New York, 2003.
- 96 F. Planas, X. Sheng, M. J. McLeish and F. Himo, A Theoretical Study of the Benzoylformate Decarboxylase Reaction Mechanism, *Front. Chem.*, 2018, **6**, 205.
- 97 V. L. Schramm, Transition States and Transition State Analogue Interactions with Enzymes, *Acc. Chem. Res.*, 2015, **48**(4), 1032–1039.
- 98 K. E. Ranaghan, L. Ridder, B. Szeferczyk, W. A. Sokalski, J. C. Hermann and A. J. Mulholland, Transition state stabilization and substrate strain in enzyme catalysis: ab initio QM/MM modelling of the chorismate mutase reaction, *Org. Biomol. Chem.*, 2004, **2**(7), 968–980.
- 99 K. Kettler, K. Veltman, D. van de Meent, A. van Wezel and A. J. Hendriks, Cellular uptake of nanoparticles as determined by particle properties, experimental conditions, and cell type, *Environ. Toxicol. Chem.*, 2014, **33**(3), 481–492.
- 100 J. C. Gumbart, M. Beeby, G. J. Jensen and B. Roux, Escherichia coli Peptidoglycan Structure and Mechanics as Predicted by Atomic-Scale Simulations, *PLoS Comput. Biol.*, 2014, **10**(2), e100347.
- 101 R. E. Baynes and J. E. Riviere, *Hayes' Handbook of Pesticide Toxicology*, 37.2.2.1: Passive Transport, 3rd edn, 2010.
- 102 A. J. Hendriks, T. P. Traas and M. A. J. Huijbregts, Critical body residues linked to octanol-water partitioning, organism composition, and LC50 QSARs: Meta-analysis and model, *Environ. Sci. Technol.*, 2005, **39**(9), 1803–1808.
- 103 J. Damborsky and T. W. Schultz, Comparison of the QSAR models for toxicity and biodegradability of anilines and phenols, *Chemosphere*, 1997, **34**(2), 429–446.
- 104 Eawag Eawag Biocatalysis/Biodegradation Database and Pathway Prediction System. <http://eawag-bbd.ethz.ch/predict/> (accessed December 7).
- 105 J. Chakraborty, T. Jana, S. Saha and T. K. Dutta, Ring-Hydroxylating Oxygenase database: a database of bacterial aromatic ring-hydroxylating oxygenases in the management of bioremediation and biocatalysis of aromatic compounds, *Environ. Microbiol. Rep.*, 2014, **6**(5), 519–523.
- 106 W. P. Wang and Z. Z. Shao, Enzymes and genes involved in aerobic alkane degradation, *Front. Microbiol.*, 2013, **4**(1), 116.
- 107 T. M. Nolte and W. J. G. M. Peijnenburg, Aqueous-phase photooxygenation of enes, amines, sulfides and polycyclic aromatics by singlet ( $a^1\Delta_g$ ) oxygen: prediction of rate constants using orbital energies, substituent factors and quantitative structure–property relationships, *Environ. Chem.*, 2018, **14**(7), 442–450.
- 108 P. M. Morse, Diatomic molecules according to the wave mechanics. II. Vibrational levels, *Phys. Rev.*, 1929, **34**(1), 57–64.
- 109 K. Riemenschneider, H. M. Bartels, R. Dornow, E. Drechselgrau, W. Eichel, H. Luthe, Y. M. Matter, W. Michaelis and P. Boldt, Free-Radical Additions. 11. Kinetics of Free-Radical Additions of Bromomalononitrile to Alkenes and Pmo Treatment of the Reactivity and Regioselectivity, *J. Org. Chem.*, 1987, **52**(2), 205–212.
- 110 T. M. Nolte, T. Nauser, L. Gubler and W. J. G. M. Peijnenburg, Thermochemical unification of molecular descriptors to predict radical hydrogen abstraction with low computational cost, *Phys. Chem. Chem. Phys.*, 2019, submitted.
- 111 W. C. Steen and T. W. Collette, Microbial degradation of seven amides by suspended bacterial populations, *Appl. Environ. Microbiol.*, 1989, **55**(10), 2545–2549.
- 112 X. Xu and X. G. Li, QSAR for Predicting Biodegradation Rates of Polycyclic Aromatic Hydrocarbons in Aqueous Systems, *Chin. J. Struct. Chem.*, 2012, **31**(8), 1212–1221.
- 113 K. Fukui, H. Kato and T. Yonezawa, A New Quantum-Mechanical Reactivity Index for Saturated Compounds, *Bull. Chem. Soc. Jpn.*, 1961, **34**(8), 1111–1115.
- 114 X. Huang and J. T. Groves, Beyond ferryl-mediated hydroxylation: 40 years of the rebound mechanism and C-H activation, *J. Biol. Inorg. Chem.*, 2017, **22**(2–3), 185–207.
- 115 L. P. Hammett, The effect of structure upon the reactions of organic compounds benzene derivatives, *J. Am. Chem. Soc.*, 1937, **59**, 96–103.
- 116 R. W. Taft, J. L. M. Abboud, M. J. Kamlet and M. H. Abraham, Linear Solvation Energy Relations, *J. Solution Chem.*, 1985, **14**(3), 153–186.
- 117 C. Gong, X. M. Sun, C. X. Zhang, X. Zhang and J. F. Niu, Kinetics and Quantitative Structure-Activity Relationship Study on the Degradation Reaction from Perfluorooctanoic Acid to Trifluoroacetic Acid, *Int. J. Mol. Sci.*, 2014, **15**(8), 14153–14165.
- 118 J. W. Hilborn and J. A. Pincock, Rates of Decarboxylation of Acyloxy Radicals Formed in the Photocleavage of Substituted 1-Naphthylmethyl Alkanoates, *J. Am. Chem. Soc.*, 1991, **113**(7), 2683–2686.
- 119 M. J. Kamlet, J. L. M. Abboud, M. H. Abraham and R. W. Taft, Linear Solvation Energy Relationships. 23. A Comprehensive Collection of the Solvatochromic Parameters and Some Methods for Simplifying the Generalized Solvatochromic Equation, *J. Org. Chem.*, 1983, **48**(17), 2877–2887.

- 120 J. Arnot, T. Gouin and D. Mackay, *Development and Application of Models of Chemical Fate in Canada - Practical Methods for Estimating Environmental Biodegradation Rates*, Peterborough, Canada, 2005.
- 121 J. S. Jaworska, R. S. Boethling and P. H. Howard, Recent developments in broadly applicable structure-biodegradability relationships, *Environ. Toxicol. Chem.*, 2003, **22**(8), 1710–1723.
- 122 P. H. Howard, A. E. Hueber and R. S. Boethling, Biodegradation Data Evaluation for Structure Biodegradability Relations, *Environ. Toxicol. Chem.*, 1987, **6**(1), 1–10.
- 123 B. P. Mahoney, N. Raghunand, B. Baggett and R. J. Gillies, Tumor acidity, ion trapping and chemotherapeutics I. Acid pH affects the distribution of chemotherapeutic agents in vitro, *Biochem. Pharmacol.*, 2003, **66**(7), 1207–1218.
- 124 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**(4), 455–465.
- 125 R. Mendes and J. M. Raaijmakers, Cross-kingdom similarities in microbiome functions, *ISME J.*, 2015, **9**(9), 1905–1907.
- 126 S. T. Ramirez-Puebla, L. E. Servin-Garciduenas, B. Jimenez-Marin, L. M. Bolanos, M. Rosenblueth, J. Martinez, M. A. Rogel, E. Ormeno-Orrillo and E. Martinez-Romero, Gut and Root Microbiota Commonalities, *Appl. Environ. Microbiol.*, 2013, **79**(1), 2–9.