

RESEARCH ARTICLE

Degradability of organic micropollutants with sonolysis—Quantification of the structural influence through QSPR modelling

Judith Glienke  ^{1,2}, Michael Stelter  ^{1,2,3}, Patrick Braeutigam  ^{1,2,3*}

1 Institute of Technical Chemistry and Environmental Chemistry, Friedrich Schiller University Jena, Jena, Germany, **2** Center of Energy and Environmental Chemistry (CEEC Jena), Friedrich Schiller University Jena, Jena, Germany, **3** Fraunhofer IKTS, Fraunhofer Institute for Ceramic Technologies and Systems, Hermsdorf, Germany

* patrick.braeutigam@uni-jena.de



OPEN ACCESS

Citation: Glienke J, Stelter M, Braeutigam P (2023) Degradability of organic micropollutants with sonolysis—Quantification of the structural influence through QSPR modelling. PLOS Water 2(1): e0000082. <https://doi.org/10.1371/journal.pwat.0000082>

Editor: Manpreet Singh Bhatti, Guru Nanak Dev University, INDIA

Received: October 4, 2022

Accepted: December 14, 2022

Published: January 30, 2023

Copyright: © 2023 Glienke et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All underlying data for this manuscript are included in the [Supporting Information](#) file.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Local and climate-driven challenges combined with an increasing anthropogenic pollution of the water compartment all around the world make a sustainable handling of wastewater imperative. New additional treatment methods are under examination, including cavitation-based advanced oxidation processes. To quantify structural influences on chemical processes, quantitative structure-property relationship (QSPR) modelling can be used, which calculates a correlation between a defined endpoint and structural properties expressed by molecular descriptors. In this study, QSPR modelling has been applied to investigate the structural influence on the degradability of organic micropollutants with high-frequency sonolysis. The dataset of a previous study on 32 phenol derivates was expanded by 60 mostly aromatic compounds, whose kinetic degradation constants were obtained in a standardized experimental setup. QSPR modelling was conducted using the software PaDEL for descriptor calculation and QSARINS for the modelling process using a multiple linear regression approach and genetic algorithm. All five OECD-requirements for applicable QSPR models were respected. The obtained model included 12 model descriptors, was evaluated with numerous statistical quality parameters, and shows good regression abilities as well as robustness and predictability ($R^2 = 0.8651$, $CCC_{tr} = 0.9277$, $Q^2_{loc} = 0.8010$, $R^2_{ext} = 0.7836$, $CCC_{ext} = 0.8838$, $Q^2_{F1} = 0.7697$). The interpretation of selected model descriptors showed interesting connections between the model results and the experimental background. A strong influence of the polarity of organic compounds on their degradability with high-frequency sonolysis could be quantified, as more nonpolar molecules are degraded faster. Additionally, the impact of specific fingerprints, including for example substituents with heteroatoms, the number of fused and non-fused aromatic rings as well as the numerical appearance of secondary carbon could be identified as relevant for this cavitation-based treatment method.

Introduction

Water is not only the foundation and a general need of life but also the world's most threatened resource. Natural local scarcity and the accelerating climatic-driven challenges are additionally exacerbated by the increasing anthropogenic pollution, which is not limited to drinking water, but affects all sources of fresh and sea water [1–3]. This consequently results in risks and harms not only for the natural environment and wildlife, but also for human health [4, 5]. Water pollutants are quite divers, including for example radioactive material, heavy metals and organic micropollutants [1]. The later have a wide structural variety and a huge range of toxicity. Among these emerging contaminants, special substance classes like pharmaceuticals, pesticides, herbicides, synthetic dyes, and cosmetics cause major concern around the world. Even though the concentration of these organic molecules is generally low at $\mu\text{g/L}$ or ng/L [6], possible human health risks include carcinogenic and mutagenic effects as well as acute and chronic toxicity [7, 8]. Endocrine disruptive chemicals (EDCs) additionally interfere with hormonal activities and therefore might disturb reproduction, development and behaviour [9]. In addition, most of these compounds are quite persistent and therefore only partially degraded by current installed water treatment plants or even bypass treatment altogether, emerging in receiving water, which is then again often used as human water source [10]. Despite this accumulating water usage cycle and although the production of these micropollutants steadily increases, a final solution for universally effective water treatment has yet to be found [7].

This is why additional water treatment methods are a crucial and growing research field. Within the development of such processes, advanced oxidation processes (AOPs) have been gaining more and more attention, as they present great potential for treating a large variety of organic micropollutants [11]. AOPs generally have in common, that highly reactive oxygen species (ROS) such as hydroxyl radicals (OH^\bullet) and superoxide anion radicals ($\text{O}_2^{\bullet-}$) are generated *in-situ*. With the low selectivity of the ROS, a conversion of micropollutants into non-toxic compounds or a full mineralization into CO_2 , H_2O and inorganic ions can be achieved [11].

Cavitation-based degradation processes has emerged as a promising oxidative technology for the degradation of organic compounds by utilizing the sonochemical phenomenon [12, 13]. Acoustic cavitation is produced when ultrasound frequencies between 20 and 1000 kHz are transmitted through a liquid [14]. The passing of the ultrasound wave through the medium results in a periodic movement of solvent molecules, creating compression and expansion cycles. If the local static pressure is equal or lower than the vapor pressure of the gas dissolved in the solvent, liquid voids and cavitation bubbles are formed [15]. The cavities expand during the cycles until they reach a critical size and collapse [16]. The implosion of these cavitation bubbles creates locally high temperatures and pressures of around 5000 K and 1000 atm, respectively. These so-called hot spots concentrate the ultrasonic energy and can be seen as micro-reactors within the liquid [12, 17].

Acoustic cavitation generally forms three different regions (Fig 1), where different reaction pathways can take place [14, 18].

Within the bubble interior, pyrolytic degradation reactions of volatile and hydrophobic molecules occur due to the high temperature and pressure. Additionally, vaporized water molecules undergo decomposition resulting in the formation of hydroxyl radicals. In the bubble-liquid interface, hydroxyl radicals react predominantly to form H_2O_2 and oxidize organic compounds. The bulk solution contains free reactive species, which migrated from the interface area. In the bulk, they can react with organic molecules in secondary sonochemical reactions [18].

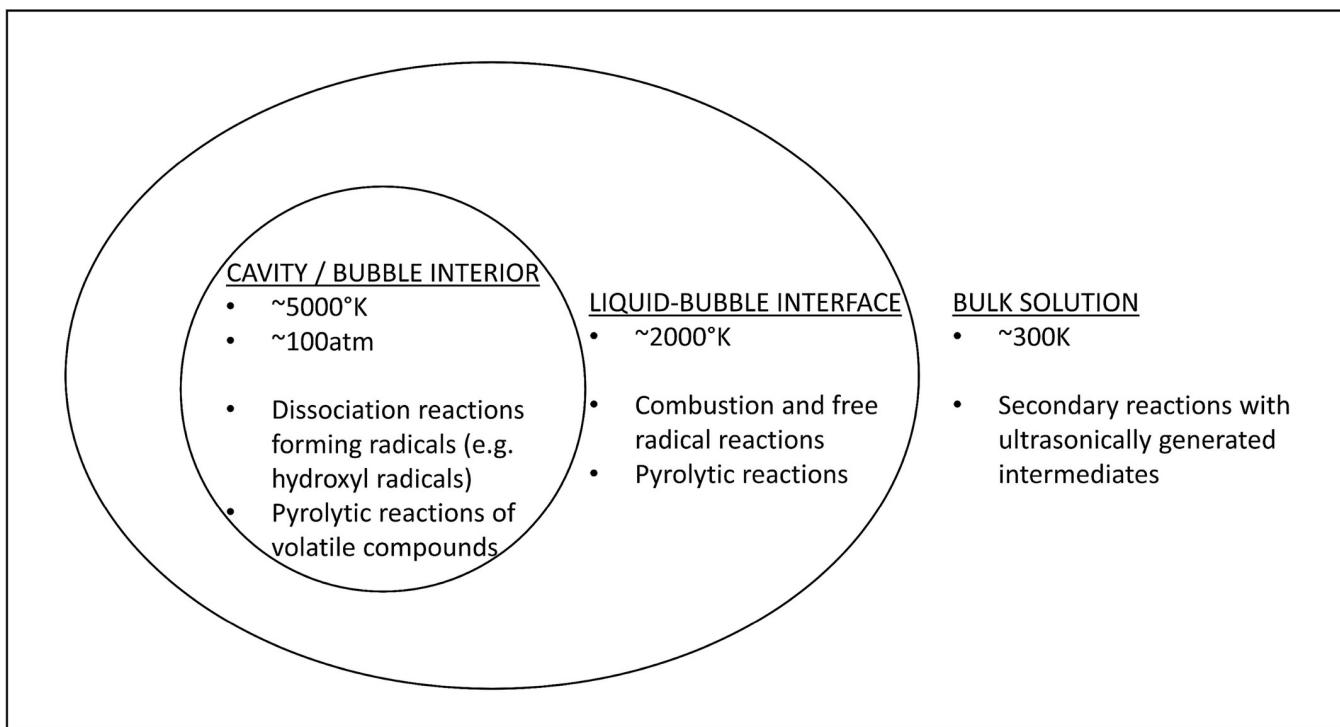


Fig 1. Three reaction zones in the cavitation process (following Adewuyi [18]).

<https://doi.org/10.1371/journal.pwat.0000082.g001>

Among challenges in method development, the large variability of the structure of organic micropollutants and therefore their chemical behaviour has to be considered. Previous studies of the sonolysis of organic micropollutants observed a very different behaviour of various micropollutants, showing for example a qualitative influence of the polarity on the degradability with sonolysis, stating that a higher polarity results in a slower degradation [12, 19, 20].

The calculation of such mathematical, predictive *in-silico* models via *quantitative structure-property relationship* (QSPR) (or *quantitative structure-activity relationship* (QSAR)) can be a useful tool to correlate various biological, physical or chemical properties of a molecule with its chemical structure, which therefore is translated into numerous numerical parameters, so-called molecular descriptors [21]. In general, QSPR/QSAR modelling studies usually utilize a set of chemicals with known experimental target endpoint and their calculated molecular descriptors (training set) to select relevant descriptors and to develop a correlation equation.

The use of QSPR modelling in regulatory and industrial purposes receives growing support. To ensure the quality and reliability of such models, the OECD (*Organisation of Economic Cooperation and Development*) defined five principles, which should be met in good practiced QSPR modelling. (1) a defined endpoint, (2) an unambiguous algorithm, (3) a defined domain of application, (4) appropriate measures of goodness-of-fit, robustness, and predictivity, and (5) a mechanistic interpretation, if possible [22, 23].

In addition to its predictive value, QSPR modelling can help to identify molecular properties and sites which are important for the degradation and therefore can contribute to a broader understanding of reaction pathways. This was shown in our previous study on the sonolytic degradation of 32 phenol derivates and the first predictive QSPR model on sonolysis [24]. By interpreting some selected model descriptors, a potential influence of the polarity and the occurrence of strong hydrogen bonds could be identified. Due to the limitations of the

small underlying dataset though, it could not be excluded that these two descriptors were only selected because of a dataset anomaly and the simultaneous occurrence of a stabilizing mesomeric effect. Therefore, in this study we increased the experimental dataset from the previous study to 92 organic micropollutants, that were experimentally investigated in a standardized laboratory setup with fixed parameters under the same reproducible conditions to ensure the needed homogeneity of the underlying dataset [25, 26]. The QSPR workflow from Glienke et al. [24] was executed again to ensure comparability as well as sufficient model quality, including multiple validation methods. The selected model descriptors were interpreted within the experimental background to connect the quantitative mathematical model with the underlying experimental reaction pathways.

Material and methods

Reagents and materials

All sources of chemicals, including CAS-numbers, molecular weight, structures, SMILES-codes, and purity, are described in Tables A and B in [S1 Text](#). All chemicals were used as received and possessed a purity > 90%. Reaction solutions were prepared using freshly filtered ultrapure water ($\sigma \leq 0.055 \mu\text{S}/\text{cm}$, TOC < 5 ppb; GenPure Pro, Fisher Scientific).

Experimental data

The sonolytic degradation experiments at 860 kHz performed in our previous study (Glienke et al. [24]) on 32 phenol derivates were extended by 60 more organic micropollutants, including bisphenol derivates, pharmaceuticals, pesticides and herbicides. The dataset ultimately contained mostly aromatic compounds, with a focus on phenol derivates and anilines, but also substituted benzenes, azabzenes, naphthalenes and (benzo)azoles. The same laboratory setup and parameters were used to ensure data homogeneity and comparability.

The concentration of micropollutants were mainly analysed using a high-performance liquid chromatography (HPLC) (LC2000, Jasco), including a fluorescence detector (FP-2020Plus, Jasco), a multiwavelength detector (MD-2010Plus, Jasco), an autosampler (AS-2055Plus), a 100 μL injection loop and a RP C18 column (Dr. Maisch GmbH Kromasil 100 C18 10mm*4.6mm, 5 μm & 250mm*4.6mm, 5 μm) tempered at 40°C. The concentrations of organic dyes were measured using a spectral photometer (DR 3900, Hach Lange). All analytical methods are described in detail in Table C in [S1 Text](#).

Reaction rate constants were calculated following the pseudo-first-order kinetic equation ([Eq 1](#)) [27]. The average value of the kinetic constant obtained by a triple determination for each substance in 1/h was logarithmically transformed to serve as the modelling dataset.

$$\ln\left(\frac{c_0}{c_t}\right) = k \times t \quad (1)$$

With c_0 as start concentration, c_t as concentration of the analyte at time t and k as the rate constant.

QSPR modelling process

The executed QSPR modelling followed the process described in detail in Glienke et al. [24]. More information can be also found in the supplement material (Texts D–H in [S1 Text](#)). The PaDEL-descriptor software [28], version 2.21, was used to calculate all used model descriptors and fingerprints, which were then imported into QSARINS-software, version 2.2.4 [<http://www.qsar.it>] [29, 30].

Molecular descriptors

CDK (Chemistry Development Kit) fingerprints were calculated with a length of 1024 and a search depth of 8 based on all 92 molecules. They were reduced by pair-wise correlation >95% and constancy >90%. The dataset was then split into a representative training set and validation set in a ratio of 4:1 based on structural properties characterized by principal component analysis (PCA) of the CDK fingerprints.

Descriptor pool for modelling

The software PaDEL was used to calculate one- and bi-dimensional descriptors, PubChem fingerprints, substructure fingerprints and substructure fingerprint. PubChem fingerprints as well as substructure fingerprints have values of either 0 or 1, indicating the absence or the presence of the fingerprint within the molecule, respectively. It was dealt with redundant information and binary collinearity by filtering the descriptor pool for pair-wise correlation greater 95% and constancy greater 90%. The remaining descriptors (Text F in [S1 Text](#)) were normalized and imported to the QSARINS-software for further modelling. The splitting into training set and validation set obtained from chapter 2.3.1. was adopted. The distribution of the experimental endpoint and the structural domain of the descriptor pool were inspected for possible outliers and potential clusters within the dataset and examine the splitting.

QSPR modelling and validation

Multiple linear regression was used as the underlying mathematical approach ([Eq 2](#)). The algorithm thereby tried to minimize the sum of squares of the difference between experimental endpoint and its calculated value on basis of the training set.

$$\hat{Y} = b_0 + \sum_{j=1}^J (b_j \times X_j) = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_J X_J + u \quad (2)$$

All possible combinations of subsets of 2 descriptors were calculated before computing higher dimensional models using genetic algorithm (fitness function: Q^2_{loo} , population size: 400, generations per size: 100, mutation rate: 20%). The maximum number of descriptors was set to $n_{\text{feature,max}} = 13$, because the number of model descriptors should not exceed 1/5 of the number of molecules in the training set ($n_{\text{Tr}} = 75$). The variable significance level was set to ≤ 0.05 and the critical QUIK-value [[31](#)] to 0.050 to immediately dismiss all models with a high multicollinearity.

The best 10 models per size stored by the program were further analysed. Besides the internal LOO-cross validation conducted during the algorithm, an internal leave-many-out (LMO) cross validation, external validation, Y-scrambling, and Y-randomization were executed (Text H in [S1 Text](#)).

To obtain the final, most optimized model from the pool of calculated models, multi-criteria decision-making (MCDM) was used to select the best overall performing model from all stored models integrating all calculated statistical values.

For the overall best performing model, the applicability domain was defined by both a Williams and an Insubria plot and permuted and randomized response tests were applied to calculate the probability of chance correlation of the model descriptors.

Results and discussion

Experimentally derived k values and calculated descriptors

Sonolytic degradation experiments at 860 kHz were performed with 60 organic compounds in a standardized setup. The kinetic constants from a previous study for 32 phenol derivates were

supplemented [24] to obtain a total of 92 investigated compounds for the underlying dataset. Table D in [S1 Text](#) gives a full overview of all rate constants including standard deviation and their variation coefficient. The values of the experimental rate constants vary between 0.0035 min⁻¹ (Nicotinamide) and 0.2909 min⁻¹ (Tetrachlorocatechol).

The graphic display of the first two principal component analysis dimensions based on the 628 normalized descriptors used as the pool for the modelling process ([Fig 2A](#)) showed no structural outlier or hard clusters within the dataset. Additionally, the splitting based on CDK fingerprints resulted in an even distribution of validation molecules within the structural spectrum. The distribution of the rate constants ([Fig 2B](#)) however showed a high endpoint outlier of Tetrachlorocatechol relative to the rest of the dataset, even though it was not conspicuous in the PCA analysis. As the dataset for QSPR modelling purposes should be distributed normally at best, and Tetrachlorocatechol affected the modelling results negatively in pre-tests, this molecule was excluded from the dataset ([Fig 2C](#)).

QSPR model for k_{US}

Statistical quality of the QSPR model. The QSPR modelling procedure was carried out as described in chapter 2.3 based on the experimental values for the rate constant k_{US} . The best obtained model selected via MCDM is defined by the following equation for unstandardized coefficients:

$$\begin{aligned} \log k_{pred} = & -0.2851 + 0.4009 \times ALogP - 0.2053 \times AATS8m + 0.2490 \times ATSC8c - 0.2881 \\ & \times ATSC5e + 0.1036 \times MATS4p + 0.1036 \times VE1_Dzs + 0.1347 \times PubchemFP257 \\ & - 0.0912 \times PubchemFP365 + 0.1566 \times PubchemFP542 + 0.0821 \\ & \times PubchemFP688 - 0.1579 \times SubFP135 + 0.3104 \times SubFPC2 \end{aligned} \quad (3)$$

The coefficients, standardized coefficients, the confidence intervals (Co. int 95%) and the p-values of the model descriptors are listed in [Table 1](#). The model must not be considered suspect as the ratio of the confidence interval and the descriptor coefficient is below 1 and the respected p-values are below 0.05.

All binary correlation values of the 12 model descriptors ([Table E in S1 Text](#)) are well below the critical value of 0.7 [32]. With the additional executed QUIK test with a critical value of 0.05, severe multicollinearity between the model descriptors can be dismissed with high probability.

[Table 2](#) gives an overview of all statistical values calculated during the modelling process. The goodness of fit, model stability and predictability of the model seem very high based on good values for R^2 , Q^2_{lo} and R^2_{ext} , respectively. The good regression ability of the model, additionally shown in the regression plot in [Fig 3A](#), is also supported through a low value for the lack of fit (LOF) and simultaneously high values for the concordance correlation coefficients CCC_{tr} and CCC_{ext} . With good results of the Y-scrambling and Y-randomization tests ([Fig 3B and 3C](#)), and the calculations of the permuted and randomized endpoint randomization tests applied on the whole modelling process ([Fig C in S1 Text](#)), the possibility of mathematical chance correlation of the model descriptors can be dismissed.

To define the applicability domain of the model, a Williams plot and a Insubria graph are used ([Fig 4](#)). As the Williams plot uses standardized residuals, experimental endpoint values are needed to see if a molecule lies within the applicability domain of the model. As the Insubria plot uses only the predicted endpoint, no experimental data is needed for external compounds. It can be seen that 1,2,4-Benzenetricarboxylic acid is the only structural outlier. Additionally for the Williams plot, Nicotinamide and 4,4'-Diaminodiphenylsulfone have standardized residuals higher than the critical value of 2.5, so their experimental rate constant

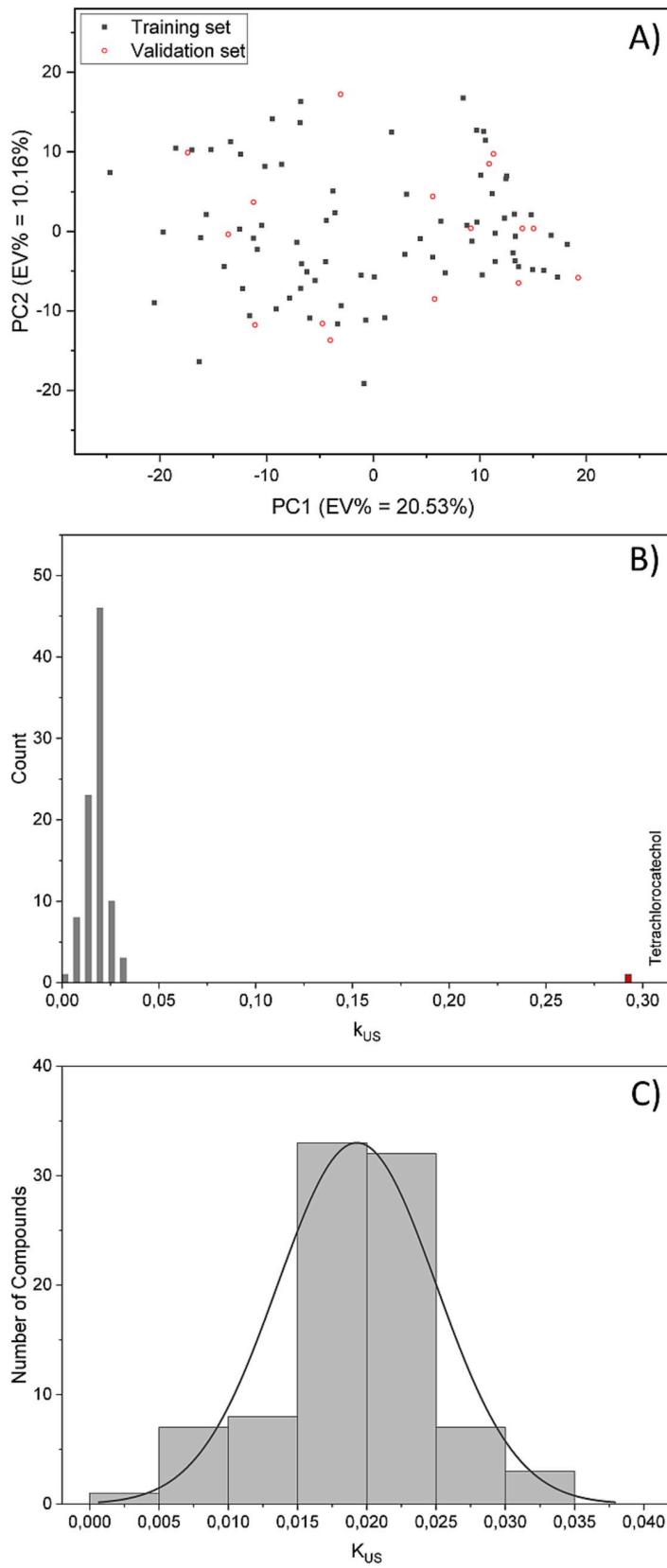


Fig 2. A) PCA analysis of the descriptor pool, B) Response distribution of all 92 molecules, C) Response distribution without Tetrachlorocatechol.

<https://doi.org/10.1371/journal.pwat.0000082.g002>

might have to be seen with caution. For Nicotinamide, this might be due to the very low value of the rate constant, almost at 0., leading to a higher deviation of the results within the triple determination.

Descriptor interpretation

Twelve descriptors were selected via genetic algorithm for the final model with the best overall statistical performance. When compared to the descriptors selected in the best ten models, the relevance of their structural information for the degradation of micropollutants with sonolysis at 860 kHz seems to be very high, as eight of these variables are present in all of the best ten models (Fig 5).

An overview of the model descriptors with a description of their mathematical background, their class and their qualitative influence on the rate constant is given in Table 1. Even though a complete mechanistical interpretation of the sonolytic degradation of micropollutants based on this QSPR model is restricted due to limitations of dataset size and descriptor interpretability, some interesting aspects can be identified.

The first selected descriptor relevant for describing the rate constant in sonolytic degradation is ALogP, calculated with an atomic approach, which considers the contribution of each atom of a molecule to its overall logP value [33, 34]. Within this QSPR model, the ALogP contributes positively to the calculation of the rate constant. More nonpolar molecules therefore seem to be degraded faster. The same influence was previously calculated in our QSPR model based on the sonolytic degradation of phenol derivates [24]. In that study however, larger polarity was equal to the occurrence of a stabilizing negative mesomeric effect due to limitations of the underlying dataset. Therefore, it could not definitely be sure which influence was

Table 1. Standardized coefficients, standardized coefficient, confidence intervals (Co. Int. 95%), p-values, background and class of the model descriptors and intercept.

Variable	Coefficient	Standardized coefficient	Standard error	Co. Int. 95%	p-value	Background of the Descriptor	Class
Intercept	-0.2851		0.0735	0.1469	0.0002		
ALogP	0.4009	0.5030	0.0468	0.0935	0.0000	Ghose-Crippen LogKow	2D
AATS8m	-0.2053	-0.2151	0.0688	0.1376	0.0039	average centered Broto-Moreau autocorrelation of lag 8 / weighted by mass	2D
ATSC8c	0.2490	0.2086	0.0744	0.1487	0.0013	Centered Broto-Moreau autocorrelation of lag 8 / weighted by charges	2D
ATSC5e	-0.2881	-0.2585	0.0611	0.1222	0.0000	Centered Broto-Moreau autocorrelation of lag 5 / weighted by Sanderson electronegativities	2D
MATS4p	0.1036	0.1369	0.0503	0.1005	0.0429	Moran autocorrelation of lag 4 / weighted by polarizabilities	2D
VE1_Dzs	0.1036	0.1597	0.0364	0.0727	0.0057	Coefficient sum of the last eigenvector from Barysz matrix / weighted by I-state	2D
PubchemFP257	0.1347	0.4037	0.0209	0.0419	0.0000	≥ 2 aromatic rings	FP
PubchemFP365	-0.0912	-0.2173	0.0240	0.0480	0.0003	C(~H)(~N)	FP
PubchemFP542	0.1566	0.4559	0.0203	0.0406	0.0000	O-C:C-[#1]	FP
PubchemFP688	0.0821	0.2493	0.0181	0.0362	0.0000	C-C-C-C-C-C-C	FP
SubFP135	-0.1579	-0.3139	0.0287	0.0574	0.0000	Vinylogous carbonyl or carboxyl derivative	FP
SubFPC2	0.3104	0.3196	0.0557	0.1113	0.0000	Secondary carbon	FP

<https://doi.org/10.1371/journal.pwat.0000082.t001>

Table 2. Calculated statistical parameters of the final QSPR model.

$N_{tr} = 75$, $N_{ext} = 17$, MCDM = 0.820
Fitting criteria:
$R^2 = 0.8651$, $R^2_{adj} = 0.8390$, $R^2 - R^2_{adj} = 0.0261$, $LOF = 0.0078$, $K_{xx} = 0.2818$, $\Delta K = 0.0502$, $RSME_{tr} = 0.0600$, $MAE_{tr} = 0.0472$, $RSS_{tr} = 0.2703$, $CCC_{tr} = 0.9277$, $s = 0.0660$, $F = 33.1346$
Internal validation criteria:
$Q^2_{loo} = 0.8010$, $R^2 - Q^2_{loo} = 0.0641$, $RMSE_{cv} = 0.0729$, $MAE_{cv} = 0.0570$, $PRESS_{cv} = 0.3988$, $CCC_{cv} = 0.8946$, $Q^2_{LMO} = 0.7791$, $R^2_{Yscr} = 0.1627$, $Q^2_{Yscr} = -0.2416$, $RMSE_{AverageYscr} = 0.1494$, $R^2_{Yrnd} = 0.1629$, $Q^2_{Yrnd} = -0.2456$
External validation criteria:
$RMSE_{ext} = 0.0735$, $MAE_{ext} = 0.0646$, $PRESS_{ext} = 0.0865$, $R^2_{ext} = 0.7836$, $Q^2_{F1} = 0.7697$, $Q^2_{F2} = 0.7603$, $Q^2_{F3} = 0.7977$, $CCC_{ext} = 0.8838$, $r_m^2 = 0.7459$, $\Delta r_m^2 = 0.0114$
Permuted Endpoint Randomization Test:
Average $R^2_{max} = 0.4037$, standard deviation $R^2 = 0.1185$, Probability of chance correlation = 0.0000
Average $Q^2_{max} = 0.2342$, standard deviation $Q^2 = 0.1078$, Probability of chance correlation = 0.0000
Randomized Endpoint Randomization Test:
Average $R^2_{max} = 0.3779$, standard deviation $R^2 = 0.1100$, Probability of chance correlation = 0.0000
Average $Q^2_{max} = 0.2663$, standard deviation $Q^2 = 0.1002$, Probability of chance correlation = 0.0000

<https://doi.org/10.1371/journal.pwat.0000082.t002>

the true reason behind the selection of this descriptor during the genetic algorithm. However, with the larger dataset in this study including more complex aromatic compounds, the presence of a substituent with negative mesomeric effect does not always goes along with low values for ALogP compared to the entire dataset (Table E in [S1 Text](#)). An example for that is 4-(4-hydroxyphenoxy)phenol, which is one of the more polar compounds of the dataset and simultaneously contains two hydroxy- and one ether-group, which both possess a positive mesomeric effect. Therefore, it seems that the selection of ALogP for the model is due to the actual influence of molecular polarity on the reactivity in sonolysis. This corresponds with qualitative experimental observations, where more nonpolar molecules could be degraded faster with sonolysis [19, 20]. A possible explanation for the increased reactivity of nonpolar compounds in sonolysis could be given by looking at potential degradation pathways. Hydrophilic molecules mostly react with ROS in the bulk solution, whereas hydrophobic thus nonpolar, non-volatile compounds undergo degradation in the bubble-liquid interface via thermal and/or radical pathways [35]. As the ROS concentration is higher in the interface, the degradation of hydrophobic compounds therefore tends to be faster. As ALogP has the largest coefficient in the model equation ([Table 1](#)), the high importance of the polarity for the degradation of organic micropollutants with high-frequency ultrasound could be quantified with this QSPR model.

Additionally, based on the fingerprint PubChem257, the occurrence of two or more aromatic rings within a molecule increases its rate constant, regardless of their actual count. This means that the fingerprint is either 0 for molecules with less than two aromatic rings, or 1 for molecules with two or more. Within the underlying dataset, an important distinction has to be made, as the bisphenol derivates have two separate aromatic rings, whereas some other molecules possess a system of fused aromatic rings, named polycyclic aromatic hydrocarbons (PAHs). The former can undergo electrophilic substitution of ROS such as hydroxyl radicals same as phenol to form polyhydroxyphenols, which can further degrade [36]. Therefore, if bisphenol derivates are compared to phenol derivates, more aromatic rings can act as additional reactive sites for substitution reactions, leading to an overall faster first degradation step. This reason might be a bit different for PAHs though because the fusion of aromatic rings leads to a change of bonds characteristics within the molecule. The C-C bonds in isolated aromatic rings

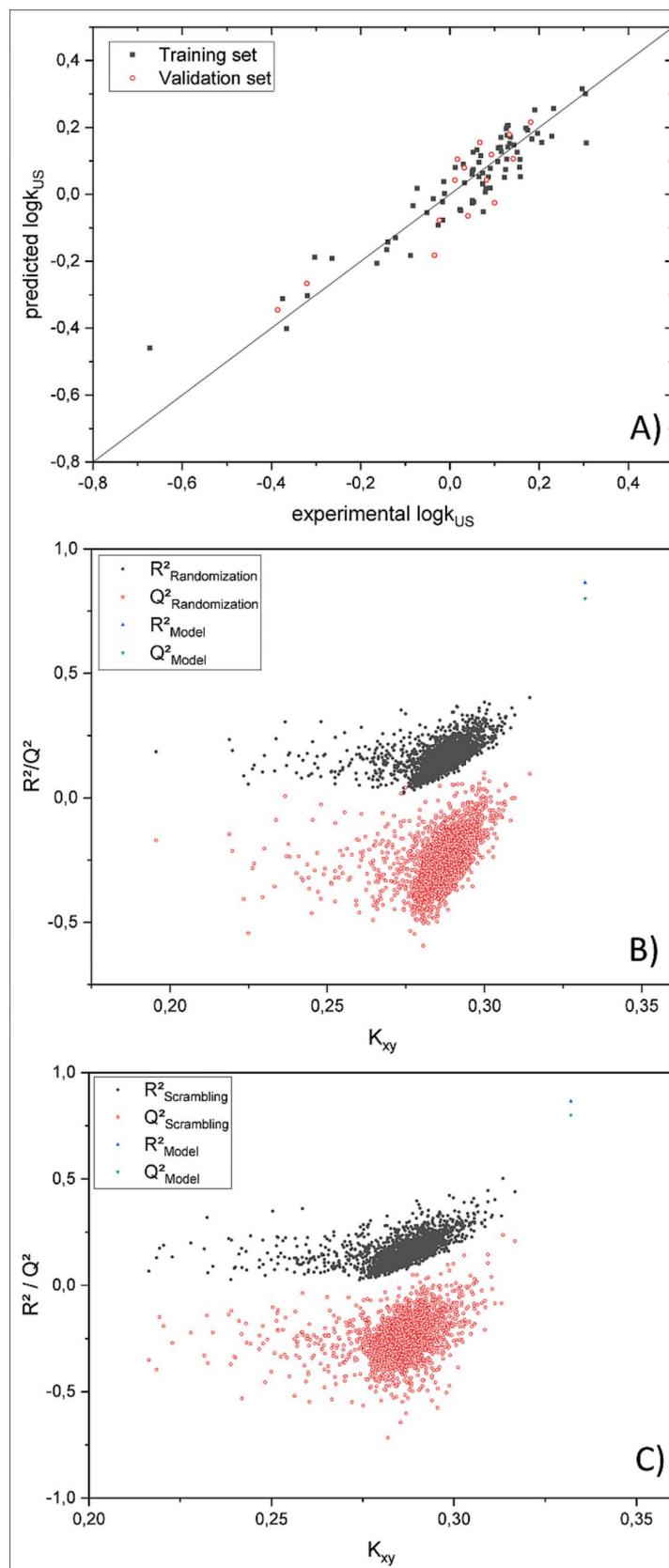


Fig 3. A) Regression plot of predicted vs experimental endpoint values, B) Y-randomization, C) Y-scrambling.

<https://doi.org/10.1371/journal.pwat.0000082.g003>

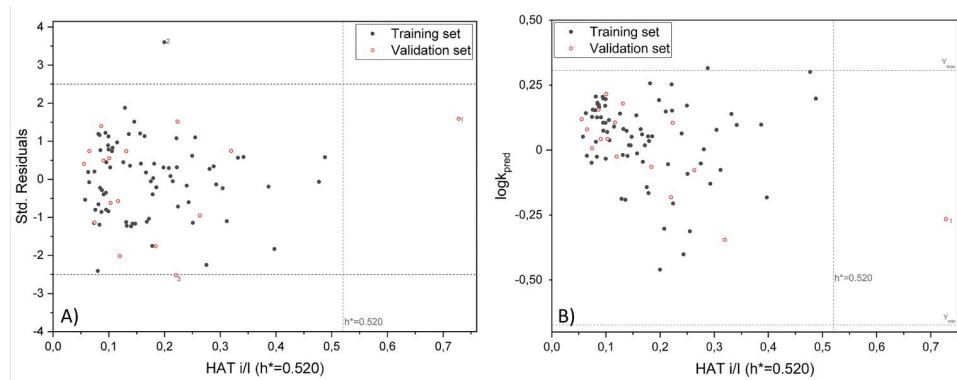


Fig 4. A) Williams plot with (1) 1,2,4-Benzenetricarboxylic acid, (2) Nicotinamide, (3) 4,4'-Diaminodiphenylsulfone, B) Insubria plot with (1) 1,2,4-Benzenetricarboxylic acid.

<https://doi.org/10.1371/journal.pwat.0000082.g004>

are all equal in length with properties between single and double bonds, whereas the bonds in fused aromatic rings tend to differ, as some C-C bonds possess more single properties, while some have more of a double bond character [37]. The latter can serve as a reactive site for electrophilic addition of ROS other than substitution reactions onto aromatic systems. As within addition reactions, only π -bonds have to be broken, this reaction tends to be faster than substitution reactions. Therefore, different mechanistic pathways might be predominant for PAHs compared to phenol derivates, leading to a faster degradation with sonolysis.

PubchemFP365 is a fingerprint for the substructure C(~H)(~N), regardless of the bond order or count. It has a negative influence on the rate constant, indicating its occurrence is

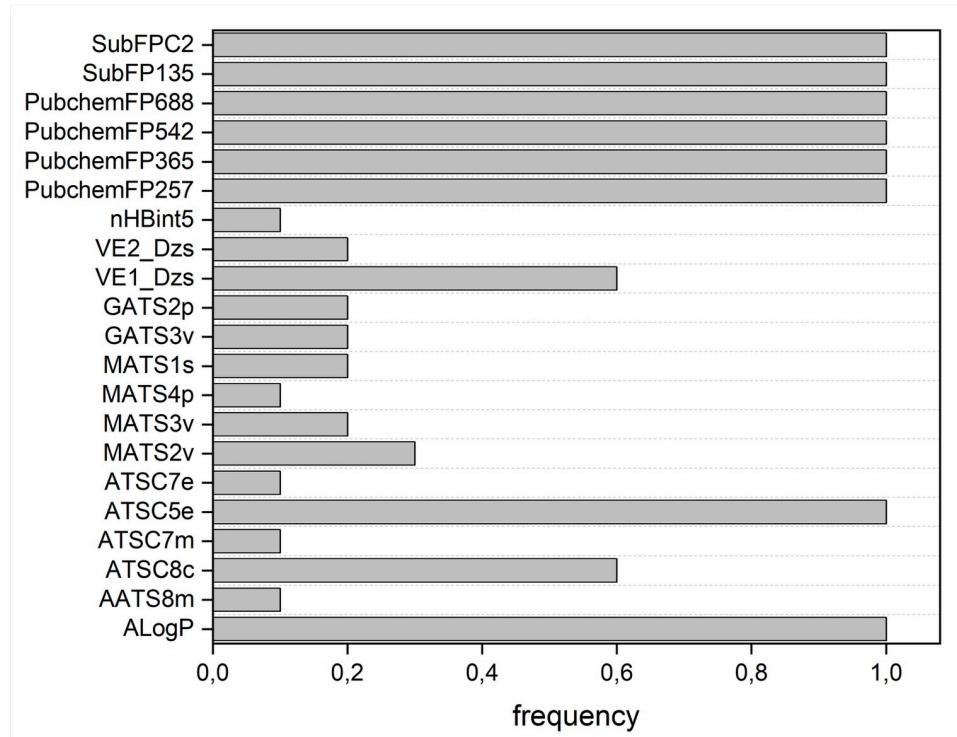


Fig 5. Descriptor frequencies for the 10 best models.

<https://doi.org/10.1371/journal.pwat.0000082.g005>

hindering for the sonolytic degradation of organic micropollutants. Within the dataset, the occurrence of that fingerprint mostly represents the presence of nitrogen as a heteroatom within a ring like pyridine, of a substituted nitro group or an amino group. Nitrogen bound to carbon has an electron-withdrawing or negative inductive effect, respectively, leaving less electron density to the immediate intermolecular environment. The aromatic ring is generally deactivated toward electrophilic aromatic substitution with negative inductive effects by polarizing the σ -bond system [37].

The occurrence of PubchemFP542, which represents the substructure O-C:C-[#1], where an oxygen atom is substituted to an aromatic ring with a single bond, increases the degradation speed. Within the underlying dataset, this is true for aromatic compounds with hydroxy-substituents or ether groups. As -OH and -OR groups have a positive mesomeric effect, the electron density in the aromatic ring increases and is stabilized. The aromatic ring therefore has an increased nucleophilic character, likely making it more accessible for the reaction with electrophilic ROS.

The fingerprint PubchemFP688, representing the substructure C-C-C-C-C-C-C, regardless of its count, also increases the rate constant within the model equation. In the dataset, the descriptor has a value of 1 for all compounds containing more than one fused aromatic ring or an aromatic system with an alkyl- or carboxyl-substituent. As discussed before, the presence of fused rings accelerates the degradation due to the developed double bond characteristics for some of the C-C bonds. Alkyl groups substituted onto aromatic rings on the other hand possess a positive isomeric effect by increasing the electron density in the aromatic ring near their substituent region due to their hybrid orbitals [38]. This in turn increases the reactivity towards electrophilic substitution reactions in *ortho* or *para* position [39].

SubFP135 is a fingerprint for the substructure of vinyllogous carbonyl or carboxyl derivative ([#6X3] (= [OX1])[#6X3] =, [#6X3][#7,#8,#16,F,Cl,Br,I]), regardless of its count. The substructure represents the presence of a carbonyl or carboxyl group connected by a double or aromatic bond to another heteroatom (O, N, S, halogens) (Fig 6). In the model equation it has a negative influence on the rate constant. On the one hand, this is because the carbonyl group next to an aromatic bond can influence the electron density of the aromatic system by its negative mesomeric effect. The decreased electron density within the aromatic ring will lower reactivity towards the electrophilic ROS, resulting in slower degradation. On the other hand, the connection of a carbonyl/carboxyl-group with another functional group over a double bond leads to a charge distribution, generally lowering the overall reactivity of that structural side (Fig 6). Within this descriptor, the influence of the negative mesomeric effect, which previously could not be distinguished from other possible influences [24], is still represented in this model, indicating its relevance for the sonolytic degradation of aromatic compounds.

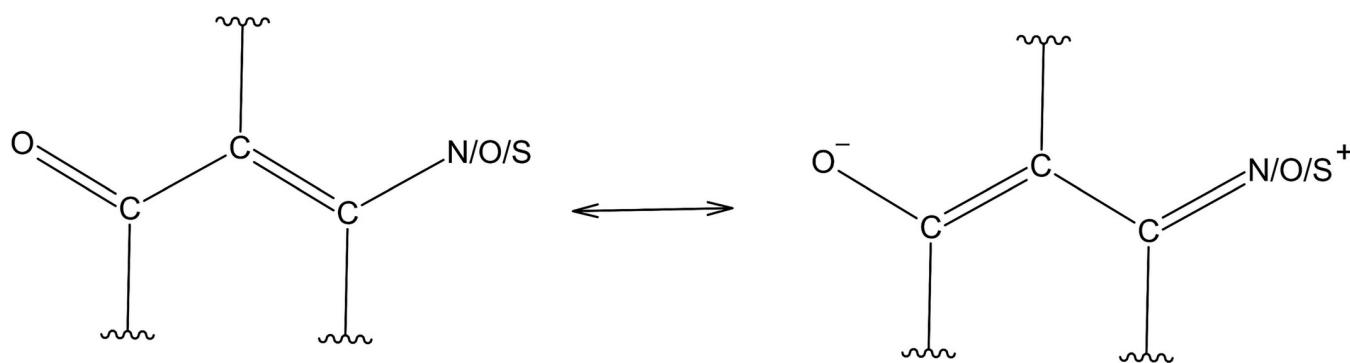


Fig 6. Charge distribution within the substructure of the fingerprint SubFP135.

<https://doi.org/10.1371/journal.pwat.0000082.g006>

SubFPC2 is a fingerprint count, which represents the number of secondary carbon atoms within a molecule. With higher values for that descriptor, the rate constant of a substance increases. For once, a high number of secondary carbons, which are mostly consecutive in the underlying dataset, could increase the hydrophobicity of a molecule, resulting in additional degradation reactions in the bubble-liquid interface or in the bubble. There, long alkyl-chains can undergo alkane pyrolysis through a homolytic break of a C-C bond [40]. This would generally increase the overall degradation speed of the target molecule.

Qualitative model interpretation

After the individual descriptor interpretation, a qualitative interpretation of the model as a whole can be done to show that the model can not only be used for quantitative calculation of the endpoint but also for describing some trends within the dataset. Serving as an example, 4-butylbenzene-1,3-diol and 4-aminobenzenesulfonamide can be compared. Fig 7 shows the chemical structure, the rate constant, and the normalized descriptor values of these two compounds.

As seen in Fig 7, 4-butylbenzene-1,3-diol degrades 3 times faster than 4-aminobenzenesulfonamide. The much higher reactivity can be explained by selected model descriptors and therefore influences. First, 4-butylbenzene-1,3-diol is more nonpolar, which is favourable for the sonolytic degradation. Additionally, 4-butylbenzene-1,3-diol possesses secondary carbon atoms, expressed by the fingerprint PubChemFP688 and the fingerprint count SubFPC2. The

	4- butylbenzene-1,3-diol	4-aminobenzenesulfonamide
<chem>Oc1ccc(cc1)CCCC</chem>	<chem>N#Cc1ccc(cc1)S(=O)(=O)N#N</chem>	
rate constant k_{US} [min $^{-1}$]	0.0285	0.0091
ALogP	0.2088	0.1390
AATS8m	0.0828	0.0053
ATSC8c	0.5057	0.5755
ATSC5e	0.6743	0.6654
MATS4p	0.4885	0.3749
VE1_Dzs	0.6565	0.4704
PubchemFP257	0	0
PubchemFP365	0	0
PubchemFP542	1	0
PubchemFP688	1	0
SubFP135	0	0
SubFPC2	0.6	0

Fig 7. Comparison of 4-butylbenzene-1,3-diol and 4-aminobenzenesulfonamide with chemical structure, endpoint values and normalized model descriptors.

<https://doi.org/10.1371/journal.pwat.0000082.g007>

butyl-group not only increases the electron density in the aromatic ring, which will increase the reactivity towards electrophilic substitution of a ROS, but also additional degradation mechanisms through pyrolysis of the alkyl group can take place, which further increases the degradability of the compound. The presence of two hydroxyl-groups substituted to the aromatic system (expressed with PubChemFP542) further increases the reactivity of the aromatic system due to the positive mesomeric effect of that functional group. All things considered, the higher rate constant of 4-butylbenzene-1,3-diol compared to for example 4-aminobenzenesulfonamide could not only be quantified by the calculated QSPR model, but also the structural influences responsible for the higher degradability are visible qualitatively.

Overall, even though a complete mechanistical interpretation of the sonolytic degradation of the investigated molecules only based on the calculated QSPR model is limited due to limitations of the interpretability of some model descriptors, a lot of interesting aspects could be determined and connected to previous experimental studies and the theoretical knowledge of sonolytic degradation pathways.

Conclusions

In this study, QSPR modelling was used to quantify the structural influence of organic micropollutants on the degradability with high-frequency sonolysis. The experimental data was obtained under standardized conditions with fixed test parameter to ensure the homogeneous quality of the rate constants. The modelling process, executed with the software QSARINS, included multiple validation techniques and all five OECD principles for applicable QSPR models were respected.

The overall best performing model was selected using a multi-criteria decision-making tool based on all calculated statistical parameters. It consists of 12 model descriptors and shows good regression abilities as well as robustness and predictability ($R^2 = 0.8651$, $CCC_{tr} = 0.9277$, $Q^2_{loo} = 0.8010$, $R^2_{ext} = 0.7836$, $CCC_{ext} = 0.8838$, $Q^2_{F1} = 0.7697$). The results of Y-scrambling and -randomization as well as permuted and randomized response modelling allows for the exclusion of chance correlation.

The interpretation of selected model descriptors resulted in insights for high-frequency sonolysis. The following structural influences could be quantified with the conducted QSPR modelling based on the underlying structural spectrum of mainly aromatic compounds:

1. ALogP as a measure of the molecular polarity increases the rate constant of a molecule. This indicates that more nonpolar compounds are degraded faster.
2. The occurrence of more than one aromatic ring in a molecule also increases the degradability. For molecules with two or more non fused rings, this is probably due to additional reactive sites for an electrophilic attack of ROS. For compounds with fused aromatic rings, the change of bond characteristics of aromatic bonds to more of single/double bond properties makes electrophilic addition reactions possible, which enhances the reactivity towards reactive species.
3. A nitrogen as a heteroatom within a ring or as a substituent onto an aromatic system decreases the rate constant, presumably due to its electron-drawing properties and negative inductive effect.
4. Substituted hydroxy- or ether-groups onto aromatic systems enlarge the reactivity of the molecule towards sonolytic degradation due to a positive mesomeric effect.
5. Substituted alkyl-groups increase the rate constant of a molecule due to a positive isomeric effect.

6. Substituted vinylogous carbonyl and carboxyl derivates decrease the degradability due to negative mesomeric effects and charge distribution.
7. A larger number of secondary carbon atoms enhances the degradability due to a related lower polarity and additional alkane pyrolysis.

Supporting information

S1 Text.

(DOCX)

Acknowledgments

We express our gratitude to Prof. Paolo Gramatica, University of Insurbia, Varese, Italy for providing access to QSARINS software and her work in QSAR-modelling. We also thank Prof. Dr. Emma Schymanski, University of Luxembourg, Luxembourg, Dr. Thomas Bocklitz, Leibniz Institute of Photonic Technology (IPHT) and Friedrich-Schiller-University, Jena, Germany, and Prof. Dr. Christoph Steinbeck, Friedrich-Schiller-University, Jena, Germany, for their support. We acknowledge support by the German Research Foundation Projekt-Nr. 512648189 and the Open Access Publication Fund of the Thueringer Universitaets- und Landesbibliothek Jena.

Author Contributions

Conceptualization: Judith Glienke.

Data curation: Judith Glienke.

Formal analysis: Judith Glienke.

Funding acquisition: Michael Stelter, Patrick Braeutigam.

Investigation: Judith Glienke.

Methodology: Judith Glienke.

Project administration: Judith Glienke.

Resources: Michael Stelter, Patrick Braeutigam.

Software: Judith Glienke.

Supervision: Patrick Braeutigam.

Validation: Judith Glienke.

Visualization: Judith Glienke.

Writing – original draft: Judith Glienke.

Writing – review & editing: Judith Glienke, Patrick Braeutigam.

References

1. Arihilam NH, Arihilam EC. Impact and control of anthropogenic pollution on the ecosystem—A review. *J Biosci Biotechnol Discov.* 2019; 4(3):54–9.
2. Muter O, Bartkevics V. Advanced analytical techniques based on high-resolution mass spectrometry for the detection of micropollutants and their toxicity in aquatic environments. *Curr Opin Environ Sci Heal.* 2020; 18:1–6.

3. Barbosa MO, Moreira NFF, Ribeiro AR, Pereira MFR, Silva AMT. Occurrence and removal of organic micropollutants: An overview of the watch list of EU Decision 2015/495. *Water Res.* 2016; 94:257–79. <https://doi.org/10.1016/j.watres.2016.02.047> PMID: 26967909
4. Westall F, Brack A. The Importance of Water for Life. *Space Sci Rev.* 2018; 214(2):50.
5. Popkin BM, D'Anci KE, Rosenberg IH. Water, hydration, and health. *Nutr Rev.* 2010; 68(8):439–58. <https://doi.org/10.1111/j.1753-4887.2010.00304.x> PMID: 20646222
6. Schwarzenbach RP, Escher BI, Fenner K, Hofstetter TB, Johnson CA, von Gunten U, et al. The Challenge of Micropollutants in Aquatic Systems. *Science.* 2006 25; 313(5790):1072–7. <https://doi.org/10.1126/science.1127291> PMID: 16931750
7. Saleh IA, Zouari N, Al-Ghouti MA. Removal of pesticides from water and wastewater: Chemical, physical and biological treatment approaches. *Environ Technol Innov.* 2020; 19:101026.
8. Rasheed T, Bilal M, Nabeel F, Adeel M, Iqbal HMN. Environmentally-related contaminants of high concern: Potential sources and analytical modalities for detection, quantification, and treatment. *Environ Int.* 2019; 122:52–66. <https://doi.org/10.1016/j.envint.2018.11.038> PMID: 30503315
9. Jung C, Son A, Her N, Zoh K-D, Cho J, Yoon Y. Removal of endocrine disrupting compounds, pharmaceuticals, and personal care products in water using carbon nanotubes: A review. *J Ind Eng Chem.* 2015; 27:1–11.
10. Benotti MJ, Trenholm RA, Vanderford BJ, Holady JC, Stanford BD, Snyder SA. Pharmaceuticals and Endocrine Disrupting Compounds in U.S. Drinking Water. *Environ Sci Technol.* 2009; 43(3):597–603. <https://doi.org/10.1021/es801845a> PMID: 19244989
11. Kanakaraju D, Glass BD, Oelgemöller M. Advanced oxidation process-mediated removal of pharmaceuticals from water: A review. *J Environ Manage.* 2018; 219:189–207. <https://doi.org/10.1016/j.jenvman.2018.04.103> PMID: 29747102
12. de Andrade FV, Augusti R, de Lima GM. Ultrasound for the remediation of contaminated waters with persistent organic pollutants: A short review. *Ultrason Sonochem.* 2021; 78:105719. <https://doi.org/10.1016/j.ultsonch.2021.105719> PMID: 34450413
13. Park J-S, Her N, Oh J, Yoon Y. Sonocatalytic degradation of bisphenol A and 17 α -ethinyl estradiol in the presence of stainless steel wire mesh catalyst in aqueous solution. *Sep Purif Technol.* 2011; 78(2):228–36.
14. Joseph CG, Li Puma G, Bono A, Krishnaiah D. Sonophotocatalysis in advanced oxidation process: A short review. *Ultrason Sonochem.* 2009; 16(5):583–9. <https://doi.org/10.1016/j.ultsonch.2009.02.002> PMID: 19282232
15. Titchou FE, Zazou H, Afanga H, El Gaayda J, Ait Akbour R, Nidheesh PV, et al. Removal of organic pollutants from wastewater by advanced oxidation processes and its combination with membrane processes. *Chem Eng Process—Process Intensif.* 2021; 169:108631.
16. Riesz P, Kondo T. Free radical formation induced by ultrasound and its biological implications. *Free Radic Biol Med.* 1992; 13(3):247–70. [https://doi.org/10.1016/0891-5849\(92\)90021-8](https://doi.org/10.1016/0891-5849(92)90021-8) PMID: 1324205
17. Suslick KS, Hammerton DA, Cline RE. Sonochemical hot spot. *J Am Chem Soc.* 1986; 108(18):5641–2.
18. Adewuyi YG. Sonochemistry: Environmental Science and Engineering Applications. *Ind Eng Chem Res.* 2001; 40(22):4681–715.
19. Xu LJ, Chu W, Graham N. Sonophotolytic degradation of phthalate acid esters in water and wastewater: Influence of compound properties and degradation mechanisms. *J Hazard Mater.* 2015; 288:43–50. <https://doi.org/10.1016/j.jhazmat.2015.02.023> PMID: 25682516
20. Wu Z, Ondruschka B. Roles of Hydrophobicity and Volatility of Organic Substrates on Sonolytic Kinetics in Aqueous Solutions. *J Phys Chem A.* 2005; 109(29):6521–6. <https://doi.org/10.1021/jp051768e> PMID: 16833997
21. Tropsha A, Gramatica P, Gombar V. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb Sci.* 2003; 22(1):69–77.
22. Organisation for Economic Co-operation and Development. GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE)STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SAR] MODELS. 2014 Sept 03 [Cited 2022 Oct 12]. Available from: <https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm>
23. Gramatica P. On the Development and Validation of QSAR Models. In: *Methods in molecular biology.* 2013. pp. 499–526. https://doi.org/10.1007/978-1-62703-059-5_21 PMID: 23086855
24. Glienke J, Schillberg W, Stelter M, Braeutigam P. Prediction of degradability of micropollutants by sonolysis in water with QSPR—a case study on phenol derivates. *Ultrason Sonochem.* 2022; 82:105867. <https://doi.org/10.1016/j.ultsonch.2021.105867> PMID: 34920352

25. Dearden JC, Cronin MTD, Kaiser KLE. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR QSAR Environ Res.* 2009; 20(3–4):241–66. <https://doi.org/10.1080/10629360902949567> PMID: 19544191
26. Gramatica P. Principles of QSAR Modeling. *Int J Quant Struct Relationships.* 2020; 5(3):61–97.
27. Vieira WT, de Farias MB, Spaolonzini MP, da Silva MGC, Vieira MGA. Latest advanced oxidative processes applied for the removal of endocrine disruptors from aqueous media—A critical report. *J Environ Chem Eng.* 2021; 9(4):105748.
28. Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J Comput Chem.* 2011; 32(7):1466–74. <https://doi.org/10.1002/jcc.21707> PMID: 21425294
29. Gramatica P, Chirico N, Papa E, Cassani S, Kovarich S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J Comput Chem.* 2013; 34(24):2121–32.
30. Gramatica P, Cassani S, Chirico N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J Comput Chem.* 2014; 35(13):1036–44. <https://doi.org/10.1002/jcc.23576> PMID: 24599647
31. Todeschini R, Consonni V, Maiocchi A. The K correlation index: theory development and its application in chemometrics. *Chemom Intell Lab Syst.* 1999; 46(1):13–29.
32. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Cop).* 2013; 36(1):27–46.
33. Ghose AK, Crippen GM. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J Comput Chem.* 1986; 7(4):565–77.
34. Ghose AK, Pritchett A, Crippen GM. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions. *J Comput Chem.* 1988; 9(1):80–90.
35. Serna-Galvis EA, Porras J, Torres-Palma RA. A critical review on the sonochemical degradation of organic pollutants in urine, seawater, and mineral water. *Ultrason Sonochem.* 2022; 82:105861. <https://doi.org/10.1016/j.ultsonch.2021.105861> PMID: 34902815
36. Inoue M, Masuda Y, Okada F, Sakurai A, Takahashi I, Sakakibara M. Degradation of bisphenol A using sonochemical reactions. *Water Res.* 2008; 42(6–7):1379–86. <https://doi.org/10.1016/j.watres.2007.10.006> PMID: 17976685
37. Roberts JD, Caserio MC. Basic Principles of Organic Chemistry. Second Edi. Menlo Park, CA: W. A. Benjamin, Inc; 1977.
38. Sebastian JF. The electronic effects of Alkyl groups. *J Chem Educ.* 1971; 48(2):97–8.
39. Cohn H, Hughes ED, Jones M-H, Peeling MG. Effects of Alkyl Groups in Electrophilic Additions and Substitutions. *Nature.* 1952; 169(4294):291–291.
40. Suslick KS, Gawienowski JJ, Schubert PF, Wang HH. Alkane sonochemistry. *J Phys Chem.* 1983; 87(13):2299–301.