ACS ES&T Water

pubs.acs.org/estwater

Review

Application of Quantitative Structure–Property Relationship Predictive Models to Water Treatment: A Critical Review

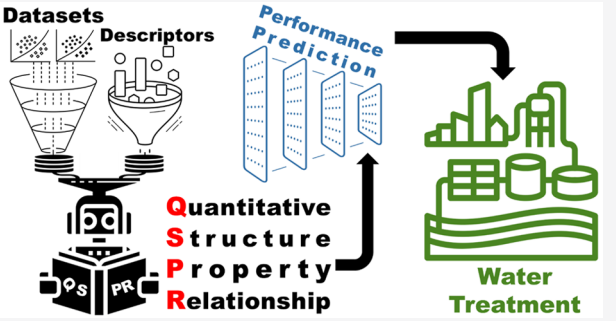
Dion Awfa,[‖] Mohamed Ateia,*[,‖] David Mendoza, and Chihiro Yoshimura

Cite This: https://dx.doi.org/10.1021/acsestwater.0c00206

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The large number of pollutants in water requires the application of various water treatment techniques. However, it is time-consuming, costly, and laborious to experimentally determine effective techniques for pollutant removal. As an alternative solution, quantitative structure–property relationship (QSPR) modeling has been applied to water treatments, including adsorption, membrane filtration, coagulation, ozonation, the Fenton reaction, photolysis, and photocatalysis. This work is a critical review of the application of QSPR models to water treatment. This modeling approach has proven to be useful for both significantly reducing the experimental load and predicting the treatment characteristics and performance, which are based on the chemical structures involved, the availability of molecular properties with minimal computational cost, and the applicability for regulatory purposes. Although current studies can serve as a basis for further model development, methods of testing the applicability of QSPR models under environmentally relevant conditions have not been explored. We also examine current priorities in ongoing research and the potential development of QSPR models for water treatment applications.



1. INTRODUCTION

Quantitative structure–property relationship (QSPR) [or quantitative structure–activity relationship (QSAR)] modeling was established by Corwin Hansch more than 50 years ago and was initially considered to be a branch of physical organic chemistry.[1] QSPR models are generally used to correlate the physical, chemical, and biological properties of a compound with its physicochemical characteristics. They were first applied in pharmaceutical research to predict the toxicity of chemicals and/or study certain properties of drugs.[2] QSPR models have since been developed and have evolved from simple series of regression methods to methods of analyzing very large data sets consisting of numerous molecules with diverse structures using several machine learning techniques. QSAR/QSPR has been applied with high accuracy to the prediction of various characteristics of chemical compounds, for example, the gas chromatographic retention time,[3] the toxicity,[4,5] and the behavior of electrochemical systems.[6] Therefore, QSPR modeling has been widely used in academic research, industrial applications, and government research worldwide.[7]

Numerous environmental contaminants have been reported in the literature (>85000 are registered by the U.S. Environmental Protection Agency, and 700–1000 new chemicals enter the market annually), and it is time-consuming, costly, and laborious to obtain experimental data on the treatment of such a large number of pollutants.[5] Thus, mathematical modeling, including machine learning, is a key approach to understanding, predicting, and managing the chemical properties and reactivities of legacy and emerging compounds. Consequently, researchers and practitioners in the field of water treatment have used QSPR modeling to investigate techniques such as adsorption, membrane filtration, photocatalysis, ozone treatment, reduction, and other advanced oxidation processes. This approach has proven to be useful and advantageous for both reducing the experimental work significantly and predicting treatment characteristics and performance, which are based on the chemical structures involved, the availability of easily obtained input parameters with minimal computational cost, and the applicability for regulatory purposes.[8,9] However, the application of QSPR modeling to water treatment is still emerging, and further training of environmental engineers and/or chemists is necessary to avoid model oversimplification. Therefore, better practices for developing and validating QSPR models and the extrapolation of model application outside the model domain are needed.[10]

Received: October 18, 2020
Revised: December 21, 2020
Accepted: January 5, 2021

To the best of our knowledge, no review article about the applications of QSPR modeling to water treatment technologies has been published. Several reviews have discussed the fundamentals and interpretation tools of QSPR/QSAR models.[11−13] Focused review papers have been published on the design and toxicity of drugs[1,14−16] and nanomaterial development and risk analysis.[17−20] Overall, these studies reveal that the use of QSPR models cannot completely replace laboratory investigations, but QSPR models can play an important role in supplementing and broadening our understanding of a system of interest and optimizing treatment system conditions. By contrast, a few review papers have discussed the applications of QSPR modeling in environmental studies. Most of these papers focused on the toxicity of chemical compounds and their environmental impacts. Delgado et al.[21] reviewed the potential application of QSPR models to rank various emerging contaminants and select relevant compounds as indicators to monitor drinking water treatment systems, including the parent compounds, metabolites, and transformation products of several pharmaceuticals, personal care products, and endocrine-disrupting compounds. Chen et al.[22] described the development and application of QSPR models to study the toxicity, formation, properties, and removal of disinfection byproducts in water. Villaverde et al.[23] reviewed pesticide risk assessment with a focus on the role of computational modeling using QSPR models. Here, we present the first critical review of recent developments and applications of QSPR modeling of various water treatment processes. A comprehensive understanding of the use of QSPR models for water treatment applications would support the prediction of treatment efficacy and the elucidation of treatment mechanisms.

2. FUNDAMENTALS OF QSPR MODELS

QSPR Principles. The underlying assumptions in QSPR modeling are based on the fact that the geometric, steric, and electronic properties of a molecule affect its intrinsic properties (e.g., biological activity, melting point, and absorption).[24] The models can interpolate the unknown properties of compounds in a certain group using either measured or calculated molecular parameters of the entire group and suitable mathematical and statistical methods.[25] For example, to minimize the gap between the pace of nanomaterial innovation and the development of nanospecific risk governance, regulators in the European Union and United States apply QSPR models to predict the risk or toxicity of newly developed nanomaterials using molecular descriptors.[26] The use of QSPR models is based on two main principles. (1) Under similar environmental conditions, compounds with similar structures show comparable behavior, and (2) variations in structure and composition among compounds are responsible for their behavioral differences.[22] The term "descriptors" refers to the predictor variables, which are also called quantum chemistry parameters, features, attributes, independent variables, or structural/compositional components. Descriptors are divided into two classes: those based on experimental measurements, such as the octanol−water partition coefficient (log K_{OW}), acid dissociation constant (pK_a), molar refractivity, and other physicochemical properties, and theoretical molecular descriptors that rely on symbolic representations of molecules to extract information, such as the molecular surface area or molecular interaction field. In addition, the terms "activities", "end points", and "dependent variables" may refer to reactivity, toxicity, bioactivity, or other response variables.[21]

Development of QSPR Models. The development of a QSPR model generally starts with the selection of molecular descriptors and the associated response variables. Molecular descriptor data can be computed or collected from experiments, the literature, and databases. Response variables can be determined experimentally or calculated using models from other studies. The data are collected and divided into two sets for running and validating the model. Then, a statistical tool is used to determine the appropriate model for the data series, and it is fitted to the curve using regression analysis, generalized linear model Bayesian inference, machine learning, etc. After validation is complete, the model can be used to predict the behavior of new molecules that belong to the group used in the model. Figure 1 shows a simplified flow diagram for a typical QSPR model.

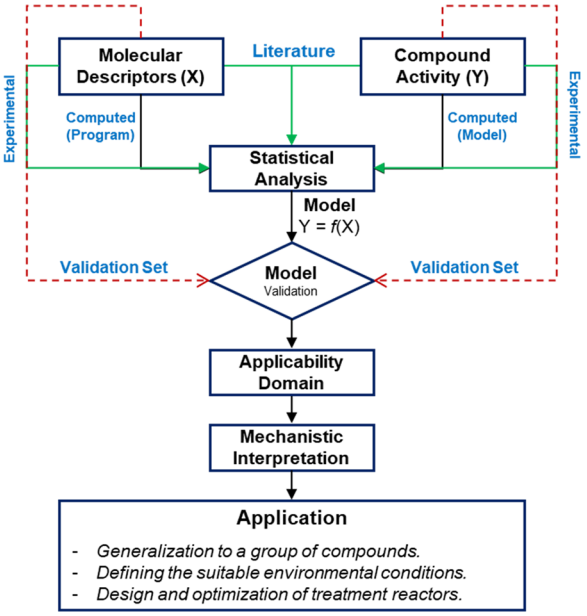


Figure 1. Simplified flow diagram of typical QSPR models of water treatment applications.

Although QSPR modeling has been in use for decades, some studies still fail to follow the guidelines of the Organization of Economic Co-operation and Development (OECD), specifically, the OECD Principles for the Validation of QSPR. Figure 2 shows a summary of good practices for each step of QSPR modeling based on the application of QSPR in previous studies. Dearden et al.[27] provided detailed descriptions of common mistakes in QSPR modeling. According to the OECD, the following five requirements must be met if a QSPR study is to be accepted: (1) a defined end point, (2) an unambiguous algorithm, (3) a defined domain of applicability, (4) appropriate measures of goodness of fit, robustness, and predictivity, and (5) a mechanistic interpretation, if possible.

Validation of QSPR Models. Model validation is an integrated step in which the predictivity of the developed QSPR model is verified. A model can be validated using both data splitting and statistical analysis, as summarized briefly in this section. Initially, the database is divided into (1) a training set, which is further split into a calibration subset and an internal validation subset, and (2) an external validation set,

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

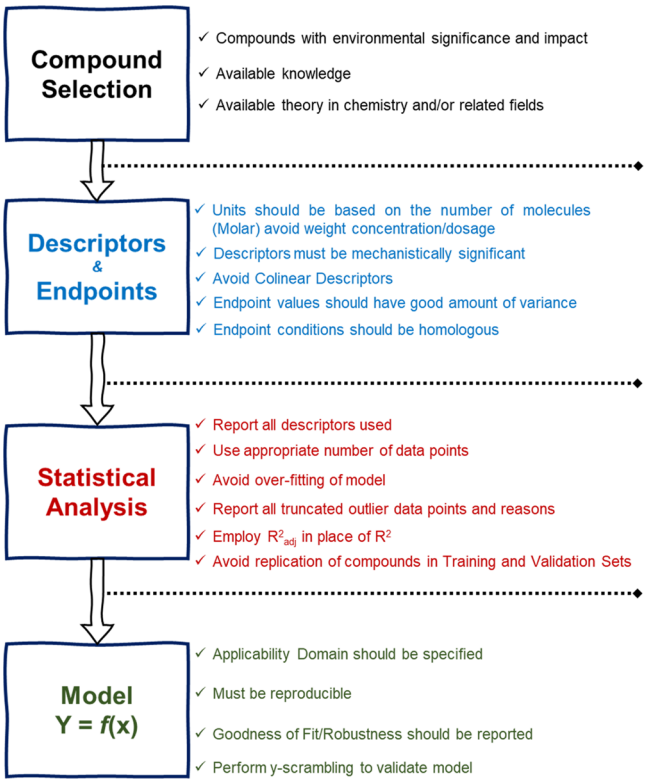


Figure 2. Proposed good practices for QSPR modeling.

which accounts for >30% of the data set and never includes data that were used to develop the model.[22] However, this conventional validation approach might result in a bias in the resultant optimal model because of the use of the same training set within the model, especially in multiple linear regression, partial least-squares, and principal component regression models.[28] As an alternative, Roy et al.[28] developed the double cross-validation technique, in which training sets with various compositions can be obtained by dividing the training set into n calibration and validation sets. In addition, other validation methods were recently introduced, such as response randomization (i.e., repetitive randomization of the response data of N compounds in the training set while the descriptor matrix is fixed) and bootstrapping (i.e., randomly selecting samples from the data set where the subsample of the data is repeatedly analyzed).[29] Thus, emerging validation techniques should be applied and compared with conventional methods.

There are also several methods of data splitting in the QSPR model, including (1) simple random sampling, (2) convenience sampling, (3) simple trial and error, (4) CADEX, (5) DUPLEX, (6) systematic sampling, and (7) stratified sampling.[30] Very few studies have compared the data splitting approaches in terms of bias and variance or examined their effects on QSPR model performance in water treatment applications. Thus, the relative benefits of QSPR model development for water treatment applications have not yet been fully assessed. The general advantages and disadvantages of each data splitting and operation method are summarized in Table 1. In addition to data splitting, statistical indicators are also an important part of a QSPR model and are employed to evaluate the performance (i.e., goodness of fit, R^2, and predictivity, Q^2) of developed QSPR models. Table 2 describes the most common statistical indicators for QSPR models.

3. QSPR MODEL APPLICATION TO WATER TREATMENT: AN OVERVIEW

In environmental engineering, we assess the impacts of chemicals on the environment and living organisms, for example, by exploring how similarly structured drugs have different effects on or produce different symptoms in the human body. In addition, we focus on pollutant control via water or wastewater treatment using various technologies. Overall, QSPR models have been successfully applied to predict the performance of several water treatment processes (Figure 3) and have shown high prediction accuracy (R^2 > 0.7,

Table 1. Data Splitting Methods

data splitting method	advantages	disadvantages	ref
simple random sampling (SRS)	easy to perform	splitting of data suffers from variance or bias, especially when the data are non-uniformly distributed	92
	efficiently implemented		
convenience sampling	efficient when dealing with time series	the presence of long-term trends within the data, or differences in the features and events observed during the different time intervals, can lead to poor model performance	93
	splitting data according to discrete time intervals		
simple trial and error	overcome the high variance of the SRS method by repeating the random sampling several times and then averaging the results	based on trial and error that cannot guarantee that it will find the best subsets	94
		time-consuming and high computational costs	
		vague theoretical background	
CADEX	draws samples based on distance, selecting points farthest from those already included in the sample, and ensures maximum coverage of the data	prohibits its use on large data sets due to the computational complexity	95
DUPLEX	improved version of CADEX	prohibits its use on large data sets due to the computational complexity	96
systematic sampling	deterministic approach designated for naturally ordered data sets	high variance and also bias of the model performance	97
	for the ordered data set, a random starting sample is chosen and then each Kh sample is taken	sensitivity to periodicities in data	
stratified sampling	partitions the data into H homogeneous groups of size Nh, and data are sampled from within each stratum	prohibits its use on large data sets due to the computational complexity	30
	ensures that adequate representation of input−output tuplets can be achieved		

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 2. Statistical Indicators Commonly Used in QSPR Models

abbreviation	definition	description
TSS	total sum of squares	same as the definition
SSR	sum of squares due to the regression model	same as the definition
MSR	mean square of regression	same as the definition
SSE	residual sum of squares	same as the definition
RMS	residual mean squares	same as the definition
SD	standard deviation	it shows the variation or dispersion level of the data from the average value; the lower it is, the closer it is to the mean
SE	standard error	it measures the accuracy of predictions
RMSE	root-mean-square error	it measures the accuracy of predictions
MBD	mean bias deviation	it reflects the deviation degree of measured and predicted values
AE	absolute error	it measures how close forecasts or predictions are to the eventual outcomes
AAE	average absolute error	it measures how close forecasts or predictions are to the eventual outcomes
PRESS	predictive residual sum of squares	it measures the accuracy of predictions in internal validation
r	Pearson r	it reflects the relationship between two variables
R, R^2	multicorrelation coefficient	it reflects the relationship between a dependent variable y and independent variables x; R^2 increases when more variables are included
R_{2adj}	adjusted R^2	it overcomes the problem from R^2, and it is more reliable
Q^2_{F1}, Q^2_{F2}, Q^2_{F3}, Q^2, Q^2_{LOO}, Q^2_{ext}	coefficient of external validation data	they define the actual predictive power of the model more precisely
CCC	concordance correlation coefficient	it measures the predictivity of model
F	Fisher ratio	it checks the significance of independent variables on the dependent variable; if F is higher, the equation is significant
t	t value	it checks the significance of an individual independent variable to the dependent variable; the higher the value, the more important the variable
SDR	standard residuals	a statistic for checking potential outliers; an observation with SDRi outside the range of ±2.5 or ±2.0 may be considered an outlier
h_j	leverage	a measure of importance of data in developing a model, 0 (not important) to 1 (very influential)
D	Cook's distance	a statistic for checking outliers, an observation with $D_i > 1$ may be considered an outlier

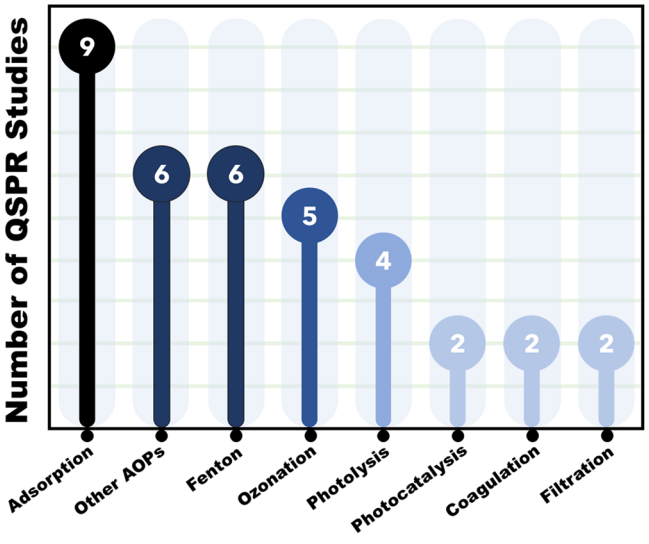


Figure 3. Water treatment methods for which QSPR models have been used for performance prediction. The numbers in parentheses represent the number of studies in the literature for each method.

and $Q^2 > 0.5$). Table 3 summarizes the studies that have employed QSPR modeling to predict the performance of water treatment processes over the past two decades. These studies meet the requirements for comparing compounds and structures and predicting emerging contaminant behavior (e.g., chemicals with limited access and/or with no available analytical standards). Note, however, that these QSAR models might have overemphasized or underestimated some aspects and/or ignored important features.[31] In addition, most previous models were developed on the basis of studies that were run under unrealistic conditions (i.e., ultrapure water matrices in batch mode) with concentrations of target contaminants ranging from micrograms per liter to milligrams per liter. Therefore, QSPR models of various water treatment technologies based on consensus modeling (i.e., calculating an average result for representative individual models that provide an equally reliable response prediction) under environmentally relevant conditions (e.g., in the presence of background organic matter and inorganic species) have yet to be developed to properly assess the applicability of QSPR modeling to real applications.

4. APPLICATION TO ADSORPTION

Activated carbon (AC) is the most frequently used adsorbent because of its large specific surface area and high porosity.[32] The organic contaminant adsorption mechanisms of AC involve hydrophobic interactions, $\pi-\pi$ interactions, electrostatic attraction, and hydrogen bonding.[33] The first QSPR adsorption models of AC and AC cloth used multiple linear regression analysis as well as artificial neural networks and were applied to more than 300 chemical compounds.[2] The models were very simple and were based only on the relationship between the molecular connectivity indices (MCIs) and the adsorption affinity of organic molecules on both adsorbents. Molecular connectivity modeling relates the MCIs to specific properties of a class of compounds. The MCI is a fundamental descriptor that can be computed from structural information and may be related to several physicochemical properties.[34]

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 3. QSPR Predictive Model Application in Water Treatments

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Photocatalysis				
17 metal oxides under solar irradiation	*Escherichia coli*	2.0−3.0 × 10^9 bacteria/mL	ultrapure water	vertical and adiabatic ionization potential energy of the highest occupied molecular orbital (E_{HOMO}) energy of the lowest unoccupied molecular orbital (E_{LUMO}) $E_{LUMO} - E_{HOMO}$ molecular hardness (η) absolute electronegativity average of the α and β E_{LUMO} larger of the E_{HOMO} of the α and β orbitals dipole moment (μ) mass molecular volume (V) density surface area metal cation radii enthalpy of formation of metal oxides and metal atoms enthalpy and Gibbs energy difference between the neutral/cation and neutral/anion heat capacities	linear regression	not available (NA)	dark condition F = 33.83 R^2 = 0.87 SD = 0.48 irradiation condition F = 20.51 R^2 = 0.81 SD = 0.63	89
UVA-TiO_2 UVA-activated carbon fiber/TiO_2 (ACF/TiO_2)	sulfamethoxazole sulfapyridine sulfadoxine sulfamidine sulfadimethoxine sulfamerazine sulfamethoxypyridazine sulfachloropyridazine sulfadiazine sulfafurazole	20	ultrapure water (pH 7.00)	dipole moment (μ) most positive partial charge on a hydrogen atom (qH$^+$) most negative partial charge on a carbon atom (qC$^-$) most negative partial charge on a nitrogen atom (qN$^-$) E_{LUMO} E_{HOMO} exact polarizability (Ep) molecular volume (V) Fukui indices	partial least squares (PLS)	NA	UV-TiO_2 R^2 = 0.92 F = 16.08 SE = 0.07 P = 0.01 $Q^2_{cumulative}$ = 0.51 UV-ACF/TiO_2 R^2 = 0.89 F = 20.69 SE = 0.06 P = 0.004 $Q^2_{cumulative}$ = 0.83	85
				Photolysis				
solar irradiation	naphthalene 1-methylnaphthalene 2-methylnaphthalene phenanthrene anthracene 9-methylanthracene	NA	ultrapure water	molecular weight (Mw) final heat of formation (HOF) total energy (TE) electronic energy (EE) core−core repulsion energy (CCR = TE − EE) average molecular polarizability (α)	PLS	NA	R = 0.96 $Q^2_{cumulative}$ = 0.84 P = 3 × 10^{-7} R^2_{adj} = 0.75	76

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Photocatalysis				
	9,10-dimethylanthracene			μ				
	pyrene			E_{LUMO}				
	fluoranthene			E_{HOMO}				
	chrysene			$E_{LUMO} - E_{HOMO}$				
	naphthacene			$(E_{LUMO} - E_{HOMO})^2$				
	benz[*a*]anthracene			$E_{LUMO} + E_{HOMO}$				
	benzo[*a*]pyrene diphenyl			qH^+				
	1-methylphenanthrene			qC^-				
	2-methylphenanthrene							
	3-methylphenanthrene							
	9-phenylanthracene							
	perylene							
	dibenz[*a*,*h*]anthracene							
artificial light irradiation	1,2,3,7-T_4CDD	NA	ultrapure water	Mw	PLS	NA	R = 0.99	75
	1,3,6,8-T_4CDD			HOF			$Q^2_{cumulative}$ = 0.92	
	1,2,3,4,7-P_5CDD			TE			P = 0.00002	
	1,2,3,4,7,8-H_6CDD			EE			R^2_{adj} = 0.89	
	1,2,3,4,6,7,8-H_7CDD			CCR				
	1,2,3,4,6,7,8,9-O_8CDD			α				
	2,3,7,8-T_4CDD			μ				
	2,7-D_2CDD			E_{LUMO}				
	1,2,3,4-T_4CDD			E_{HOMO}				
	1,2,3,7,8-P_5CDD			$E_{LUMO} - E_{HOMO}$				
	1,2,3,6,7,8-H_6CDD			$(E_{LUMO} - E_{HOMO})^2$				
	1,2,3,7,8,9-H_6CDD			$E_{LUMO} + E_{HOMO}$				
				largest positive atomic charge on a chlorine atom (q_{Cl})				
				largest negative atomic charge on the carbon atom that connects with the above chlorine atom (q_{Cl-C})				
				qH^+				
				qC^-				
UVC irradiation	haloacetic acid DBPs (10 types)	0.1−0.3	ultrapure water (pH 7.0)	5271 descriptors (5255 molecular descriptors and 16 quantum-chemical descriptors)	multiple linear regression (MLR)	simple random sampling (SRS)	R^2_{adj} = 0.86	77
	halonitromethane DBPs (6 types)						RMSE = 0.18	
	haloacetonitrile DBPs (7 types)						$Q^2_{cumulative}$ = 0.87	
	haloacetamide DBPs (8 types)							
	trihalomethane DBPs (9 types)							
UVA irradiation	Bismarck Brown Y	20−60	ultrapure water (pH 4.0, 6.0, and 9.0)	topological polar surface area (PSA)	stepwise regression	NA	pH 4.0	78
	Acid Red 73			α			R^2 = 0.78	

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Photocatalysis				
	Chrysoidine			μ			F = 21	
	Gold Orange			$E_{LUMO} - E_{HOMO}$			pH 6.0	
	Orange IV			$(E_{LUMO} - E_{HOMO})^2$			R^2 = 0.51	
	Mordant Black 11a			$E_{LUMO} + E_{HOMO}$			F = 16	
	Mordant Black 17			qC^+/qC^-			pH 9.0	
	Acid Orange 7			qH^+/qH^-			R^2 = 0.88	
	Acid Orange 20			η			F = 36	
	calconcarboxylic acid			electronegativity				
	Brilliant Red 5SKHa			softness (s)				
	Fast Fuchsine G			electrophilicity index (ω)				
	Acid Chrome Blue K			vertical ionization potential (IP)				
	Amaranthe			vertical electron affinity (EA)				
	Bordeaux Red							
	Acid Red 26a							
	Acid Orange 10							
	Ponceau 4R							
				Ozonation				
ozone (O_3)	36 PPCPs	NA	ultrapure water spiked with Suwanee River NOM (pH 7.5)	>30 descriptors	MLR for each individual water matrix	SRS	MLR	56
			river waters (pH 7.0–8.0)		artificial neural network (ANN) for compiled data		R^2 = 0.86–0.92 Q^2 = 0.65–0.9	
ozone	40 organic micropollutants	NA	ultrapure water (pH 5–8)	36 descriptors	MLR	NA	R^2 = 0.83 Q^2_{LOO} = 0.78 Q^2_{BOOT} = 0.72	60
ozone	32 organic compounds	25–160	ultrapure water (pH 7.0)	μ E_{LUMO} E_{HOMO} $E_{LUMO} - E_{HOMO}$ qC^- qN^- bond order (BO) Fukui indices	MLR	NA	R^2 = 0.72 Q^2 = 0.60	62
ozone	33 organic compounds	25–160	ultrapure water (pH 4.0)	μ E_{LUMO} E_{HOMO} $E_{LUMO} - E_{HOMO}$ qC^- qC^+ qH^+ bond order (BO) Fukui indices	stepwise regression	NA	R^2 = 0.763 Q^2 = 0.62 SD = 0.72 F = 25.72	63

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Ozonation				
ozone	136 organic compounds	NA	ultrapure water (pH 7.0)	>3000 descriptors	MLR support vector machine (SVM)	SRS	MLR R^2 = 0.75, RMSE = 1.1 Q^2_{LOO} = 0.7 Q^2_{EXT} = 0.8 SVM R^2 = 0.85 RMSE = 0.8	98
				Fenton				
Fenton	benzoic acid caffeic acid *p*-coumaric acid *p*-hydroxybenzoic acid 3-(4-hydroxyphenyl)propionic acid *β*-resorcylic acid syringic acid 3,4,5-trimethoxybenzoic acid vanillic acid veratric acid	100	ultrapure water (pH 3.5)	Hammett constant (σ, σ^-, and σ^+)	linear regression	NA	R = 0.79 SD = 0.07	67
polyacrylonitrile fiber Fe complexes	28 synthetic dyes	0.05 mmol/L	ultrapure water (pH 6.0)	72 molecular descriptors	MLR	SRS	R^2 = 0.82 F = 18.58 P = 0.000 Q^2 = 0.68	73
Fenton	33 organic compounds	100	ultrapure water (pH 3.0)	surface area grid (SAG) molecular volume Fukui indices E_{LUMO} E_{HOMO} $E_{LUMO} - E_{HOMO}$ qC^- qH^+	MLR	SRS	R^2_{adj} = 0.80 Q^2_{LOO} = 0.72	99
3D electro-Fenton	pyridine quinoline indole isoquinoline	20	ultrapure water (pH 3, 5, 7, 8, 9, and 10)	E_{LUMO} E_{HOMO} $E_{LUMO} - E_{HOMO}$	NA	NA	R^2 = 0.99	100
Fenton	28 organic compounds	10−100	ultrapure water (pH 3.0)	>20 descriptors	MLR	SRS	R^2 = 0.82 SD = 0.22 RMSE = 0.05 Q^2 = 0.74	101
Fenton	18 organic compounds	100	ultrapure water (pH 3.0)	>20 descriptors	PLS	NA	R^2_{adj} = 0.72	71

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Fenton				
							P = 0.000 RMSE = 0.168 Q^2 = 0.62	
				Other AOPs				
OH*	78 aromatic compounds	NA	ultrapure water (neutral pH)	~1169 descriptors	MLR	systematic sampling	r^2 = 0.74 F = 38.01 Q^2 = 0.69	102
O_3	nondissociated phenol groups	NA	ultrapure water	k	linear regression	NA	r^2 = 0.79−0.97	103
O_3/ClO_2	dissociated phenol groups			Hammett constant (σ, σ^-, and σ^+)			SD = 0.19−0.99	
O_3/HOCl	aniline groups			Taft constants (σ^*)				
ClO_2	benzene derivative groups							
HOCl	olefin groups							
$HFeO_4^-$	amines and amine derivative groups							
	tertiary amine groups							
	primary and secondary amine groups							
	aromatic group							
supercritical water oxidation (SCWO)	amino nitrogen organics (10 types)	1 mM	ultrapure water	μ	MLR	stratified sampling	log k_{TN} model	104
	nitro nitrogen organics (8 types)			total energy of a molecule (E_{B3YLP})			R^2 = 0.73	
	azo organics (5 types)			E_{LUMO}			RMSE = 0.013	
	heterocyclic organics (8 types)			E_{HOMO}			SD = 0.113	
	mixed organics			$E_{LUMO} - E_{HOMO}$			Q^2 = 0.57	
				atomic charge densities			$T_{TN_{50}}$ model	
				Fukui indices			R^2 = 0.96 RMSE = 69.39 SD = 8.33 Q^2 = 0.93	
UV/free chlorine	3-methylbenzoic acid	0.6−1.3	ultrapure water (pH 7.2)	Hammett constant (σ)	linear regression	NA	R^2 = 0.82−0.98	105
	4-fluorobenzoic acid							
	2-chlorobenzoic acid							
	2-iodobenzoic acid							
	3-cyanobenzoic acid							
	3-nitrobenzoic acid							
UV/H_2O_2	fluoroquinolones (9 types)	20 μM	ultrapure water (pH 2, 3, 5, 7, 9, 11)	~20 descriptors	MLR	SRS	R^2_{adj} = 0.81	106
	sulfonamides (11 types)						RMSE = 0.2 Q^2_{LOO} = 0.78	

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Other AOPs				
oxygen (O_2) hydrogen peroxide (H_2O_2) ozone (O_3) hydroxyl radical (OH*)	30 organic compounds	100	ultrapure water (pH 3.0)	~30 descriptors	MLR	SRS	Q^2_{BOOT} = 0.77 R^2 = 87 RMSE = 0.000 SD = 0.005 F = 28.99	107
				Coagulation				
novel covalently bound inorganic−organic hybrid (CBHyC)	8 phosphorus compounds	0.2 and 5	ultrapure water (pH 5, 7, and 9)	μ total energy of a molecule (E_{B3YLP}) E_{LUMO} E_{HOMO} $E_{LUMO} - E_{HOMO}$ atomic charge densities of oxygen atom atomic charge densities of phosporus qH^+ qC^- molecular weight Fukui indices P═O and P−C bond lengths logarithm of the octanol−water partition coefficient (log P)	PCA and MLR	NA	R^2_{adj} = 0.97 RMSE = 0.04 F = 59.33	52
ferric oxyhydroxide	28 organic compounds	10−100	ultrapure water (pH 3.0)	>20 descriptors	MLR	SRS	R^2 = 0.79 SD = 0.18 RMSE = 0.032 Q^2 = 0.60	101
				Filtration				
polyamide NF membrane	acetaminophen	0.0065−0.065	ultrapure water and clean membrane	molecular weight	PLS and MLR	NA	MLR	45
	phenacetine caffeine metronidazole phenazone sulfamethoxazole naproxen ibuprofen carbamazepine atrazine estradiol estrone nonylphenol bisphenol A		ultrapure water and fouled membrane with alginate	solubility log K_{ow} log P dipole moment molar volume molecular length molecular width molecular depth equivalent width molecular weight cutoff (MWCO) pure water permeability (PWP) magnesium sulfate salt rejection (SR) charge of the membrane as a ζ potential (ZP) hydrophobicity as contact angle (CA) pressure (P)	MLR		R^2 = 0.75 F = 52.5 PLS and MLR R^2 = 0.75 F = 60.2	

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Filtration				
				cross flow velocity (v)				
				back diffusion mass transfer coefficient (k)				
				flux (J)				
				ratio of pure water permeation flux J_0 and back diffusion mass transfer coefficient (J_0/k)				
				recovery				
polyamide NF	67 non-ionic organic compounds	NA	ultrapure water (pH 6.3)	45 descriptors	PLS	SRS	PLS	48
					ANN		R^2 = 0.7	
					Q^2 = 0.7			
					ANN			
					R^2 = 0.89			
					Q^2 = 0.88			
				Adsorption				
activated carbon	368 organic compounds	NA	ultrapure water	molecular connectivity indexes (MCI)	artificial neural network (ANN)	SRS	MLR	108
					MLR		r^2 = 0.83	
					F = 1741			
					ANN			
					r^2 = 0.88			
					F = 2538			
activated carbon cloth (ACC)	55 organic compounds (mainly aromatic)	NA	ultrapure water	molecular connectivity indexes (MCI)	MLR	SRS	R^2_{adj} = 0.91	2
							SE = 0.121	
activated carbon	fenoprofen	0.002	ultrapure water (pH 4.0)	ratio of equilibrium concentration with initial concentration (Ce/Co)	MLR	NA	ultrapure water, fresh carbon	109
	clofibric acid		surface water (pH 8.0)	log P			R^2 = 0.58	
	ibuprofen		wastewater (pH 6.5)	charge (C)			Q^2_{LOO} = 0.47	
	ketoprofen			activated carbon concentration (Cc)			ultrapure water, preloaded carbon	
	diclofenac			pK_a			R^2 = 0.63	
	gemfibrozil			MW			Q^2_{LOO} = 0.5	
	bezafibrate						surface water, pre-loaded carbon	
	naproxen						R^2 = 0.78	
	phenazone						Q^2_{LOO} = 0.75	
	cyclophosphamide							
	aminopyrine							
	carbamazepine							
	pentoxifylline							
	terbutaline							
	propanolol							
	sotalol							
	salbutamol							
	pindolol							

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
				Adsorption				
	atenolol							
	metoprolol							
	clenbuterol							
	aminopyrine							
activated carbon	aromatic, no H-bond d/a groups	NA	ultrapure water (pH 5.3−8.0)	equilibrium adsorption capacity (qe)	MLR	NA	aromatic, no H-bond d/a groups	110
	aromatic, H-bond d/a groups			log *P*			R^2 = 0.84	
	aliphatic, no H-bond d/a groups			equilibrium concentration (Ce)			Q^2 = 0.8	
	aliphatic, H-bond d/a groups			dipole moment			aromatic, H-bond d/a groups	
				polarizability			R^2 = 0.81	
				pK_a			Q^2 = 0.76	
							aliphatic, no H-bond d/a groups	
							R^2 = 0.84	
							Q^2 = 0.81	
							aliphatic, H-bond d/a groups	
							R^2 = 0.61	
							Q^2 = 0.55	
activated carbon	3483 organic compounds	NA	ultrapure water	∼40 descriptors	PLS	DUPLEX and k-means clustering	DUPLEX	111
							R^2_{adj} = 0.82	
							$Q^2_{External}$ = 0.86	
							k-means clustering	
							R^2_{adj} = 0.83	
							$Q^2_{External}$ = 0.62	
activated carbon	13 pharmaceuticals	1−480	ultrapure water (pH 6.5)	log *P*	MLR	NA	R^2_{adj} = 0.74	35
				MW			Q^2_{LOO} = 0.55	
				polar surface area (PSA)			SE = 40.94	
				hydrogen bond acceptor (HBA)				
				hydrogen bond donor (HBD)				
				molar volume (MV)				
graphene oxide (GO)	44 aromatic organic compounds	NA	ultrapure water	zero-point vibrational energy (ZPE)	MLR	NA	R^2 = 0.93	112
				ZPE-corrected total molecular energy (*E*)			Q^2_{LOO} = 0.90	
				E_{HOMO}				
				E_{LUMO}				
				α				
				dipole moment (*d*)				
				electrophilicity index (ω)				
				absolute electronegativity				

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS ES&T Water XXXX, XXX, XXX−XXX

Table 3. continued

treatment	targeted contaminants	contaminant concentration (mg/L)	water matrices	QSPR model descriptors	model/algorithm design	data splitting method	statistical indicator value	ref
single-wall carbon nanotubes	40 organic compounds	NA	ultrapure water	Adsorption absolute hardness quantum mechanically computed enthalpy (H) Gibbs free energy (G) >30 descriptors	PLS	SRS	$R^2 > 0.9$ $Q^2_{LOO} > 0.86$	39
activated carbon β-cyclodextrin polymer	200 organic micropollutants	0.002	ultrapure water (pH 6.6)	3656 descriptors	elastic net stepwise regression all possible regression	SRS	best model $R^2_{adj} = 0.84$ $Q^2 = 0.70$	40

The main descriptors in this study were molecular size and flexibility, molecular volume, topology of unsaturated atoms and heteroatoms, and a critical dimension. More recently, Zhao et al.[35] developed QSPR models of the removal of cationic pharmaceuticals by AC. Six main descriptors were found to be the most influential parameters in the QSPR models: octanol−water partition coefficient (log P), molecular weight (MW), polar surface area (PSA), hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), and molar volume. The log P values were estimated only for the adsorption of neutral molecules, and they were found to significantly affect the adsorption affinity.[36] In addition, the molecular size, PSA, H bonding (HBA and HBD), and ionic strength also make important contributions to the adsorption affinity. The surface functional groups and associated hydrogen bonding on the surface of AC can interact with pharmaceuticals.

Because AC fails to remove many emerging contaminants and is prone to pore blockage by organic matter, numerous alternative adsorbents have been developed; in particular, carbon nanotubes (CNTs) and graphene have attracted considerable attention for environmental applications.[37,38] Ghosh et al.[39] developed a QSPR model to predict the adsorption of 40 organic compounds on single-walled CNTs. In this study, more than 30 selected descriptors were used as input data for the QSPR model, including the Moriguchi octanol−water partition coefficient (MLOGP2), summation of the X values of the vertices that are joined to three other nonhydrogen vertices in the connected molecular graph (ETA Shape Y), number of tertiary amines (nRNR2), number of H atoms attached to α-C (H-051), and presence or absence of C−O at a topological distance of 6 (B06 [C−O]). Each of these descriptors represents adsorption mechanisms such as hydrophobic interactions (e.g., MLOGP2), molecular shape/sieving capacity (ETA Shape Y), electrostatic interaction (H-051), and the formation of hydrogen bonds (B06 [C−O]). By contrast, the nRNR2 value indicates a detrimental effect on the adsorption of organic compounds due to the absence of free hydrogen atoms. In addition, another research group developed a QSPR model of carbonaceous materials, specifically, graphene. Adsorption on graphene oxide was found to be driven by the mean polarizability, which originates from quantum mechanical exchange interactions between electrons of parallel spin. The models of both CNTs and graphene suggested that it might be possible to use the QSPR model, in conjunction with an experimental approach, to clarify the mechanism. A QSPR model was also developed for β-cyclodextrin, a porous polymer adsorbent.[40] In this study, more than 3000 input descriptors were used to predict the β-cyclodextrin adsorption of 200 organic contaminants and its mechanism using three QSPR methods (elastic net, stepwise regression, and all possible regression). It was found that group C1 (electrostatic interaction descriptors), group C2 (hydrophobic descriptors), group C3 (molecular size, mass, and shape descriptors), group C4 (atoms, bonds, and functional group descriptors), and group C5 (molecular connectivity) affect the adsorption behavior. This study is the most comprehensive QSPR study to compare various QSPR models and algorithms and to predict the adsorption process with a high goodness of fit and predictivity. Note, however, that the adsorption of organic compounds is affected by the properties of the adsorbent (e.g., specific surface area, pore volume, pore size distribution, inner diameter, outer diameter, length, purity, and

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

surface charge), organic compound (e.g., molecular size, geometrical properties, hydrophobicity, aromaticity, aliphaticity, and polarity), and background solution (e.g., pH, ionic strength, and background organic matter). Current QSPR models of adsorption are limited to the application of a single adsorbent (i.e., they rely on the properties of that adsorbent) and do not consider the effects of the background solution. In addition to QSPR, other predictive models based on the linear solvation energy relationship (LSER) have also shown good applicability as predictive tools and provide mechanistic insight into intermolecular interactions between organic molecules (i.e., the adsorbate properties) and some carbon nanomaterials (i.e., the adsorbent properties).[41] In addition to these advances in predictive QSPR and LSER models of adsorption, the community must focus on the development of predictive models that include complex environments to address real-world applications.[42]

5. APPLICATION TO MEMBRANE FILTRATION

Among various types of membrane filtration, nanofiltration (NF) has attracted considerable attention for environmental remediation owing to its properties, which fall between those of ultrafiltration and reverse osmosis membrane filtration. NF has superior characteristics, including high versatility, a high-MW cutoff (100−5000 Da), membrane surface charge due to the dissociation of surface functional groups, and the ability to adsorb charged solutes.[43] In addition, NF membranes operate with no phase change (i.e., no chemical reaction) and typically have high rejections of multivalent inorganic salts and small organic molecules.[44] Yangali-Quintanilla et al.[45] developed a QSPR model to predict the removal of pharmaceuticals and endocrine disruptors by polyamide NF. The model was based on the quantification of rejection considering the physiochemical properties of the organic compound, membrane characteristics, and operating conditions. Two QSPR models, partial least-squares multiple linear regression (PLS-MLR) and multiple linear regression (MLR), were used to obtain the best model. It was found that PLS-MLR (R^2 = 0.75; F = 60.2) outperformed MLR (R^2 = 0.75; F = 52.5), and the main QSPR descriptors were the equivalent width, molecular depth, and molecular length (calculated using Molecular Modeling Pro); log P (calculated using Tox Web Software); and salt rejection (SR, membrane properties). The equivalent width, molecular depth, and molecular length have been recognized as the main chemical properties that cause rejection owing to size exclusion in the filtration process.[46] The effect of log P is reasonable, because the hydrophobicity of contaminants affects the membrane rejection by adsorption and subsequent partitioning mechanisms.[47] Finally, SR is related to the charge of the membrane and also to the cake-enhanced concentration polarization.[46] Shahmansouri and Bellona[48] also developed a QSPR model of NF of 67 organic micropollutants with 45 descriptors. They developed two types of models: QSPR models that predict rejection using the flux and employing molecular descriptors as independent variables and QSPR models that indirectly predict the rejection using molecular descriptors to predict first the fitting parameters of the Spiegler−Kedem model and then the rejection. Both models showed a high goodness of fit (R^2 > 0.8) and high predictivity (Q^2 > 0.7). It was found that the rejection of non-ionic organic compounds could be explained by the molecular depth, Wilke−Chang diffusion coefficient, and flux. By contrast, the results of the second model indicate that the permeability coefficient is a function of solute diffusivity and that the reflection coefficient can be predicted from the total solvent accessible surface area and diffusion coefficient. However, both QSPR model studies of the NF process can be applied to only the NF270 membrane. Therefore, a better understanding of how QSPR models change when the surface chemistry and morphology of the membrane are varied will be important for predicting the rejection of various organic and/or inorganic compounds, which have different levels of affinity for membranes.

6. APPLICATION TO COAGULATION

Coagulation is a physicochemical process that reduces the repulsive potential of the electrical double layer of colloids using various types of coagulants (e.g., ferric chloride, aluminum sulfate, polyaluminum chloride, and polyaluminum iron sulfate).[49,50] Research on coagulation mechanisms and approaches to improving the treatment performance is always focused on a target contaminant and the associated water quality standards. Cheng et al.[51] developed a QSPR model to predict the coagulation process using ferric oxyhydroxide as a coagulant. The most positive partial charge on a hydrogen atom ($q\mathrm{H}^+$) and the minimum and maximum positive partial charges on hydrogen atoms bonded to a carbon atom ($q\mathrm{CH}^+_{\mathrm{min}}$ and $q\mathrm{CH}^+_{\mathrm{max}}$, respectively) were found to be the main descriptors in the best QSPR model. The $q\mathrm{H}^+$ value indicates the active site of a molecule to be dehydrogenated, and $q\mathrm{CH}^+$ represents the ability of a hydrogen atom bonded to a carbon atom to be oxidized. Both of these descriptors are strongly associated with active coagulation sites. Another research group also developed a QSPR model of the removal of phosphorus compounds using a polymer coagulant.[52] This study examined the coagulation mechanism and predicted the behavior of a novel covalently bound inorganic−organic hybrid coagulant (CBHyC) in the removal of eight phosphorus compounds. Electrostatic attraction and hydrophobic interactions were found to be the main removal mechanisms of CBHyC. The atomic charges of the phosphorus atoms ($q\mathrm{P}$) and hydrogen atoms ($q\mathrm{H}^+$) and the energy gap between the energy of the highest occupied molecular orbital (E_{HOMO}) and the energy of the lowest unoccupied molecular orbital (E_{LUMO}) ($E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}}$) were the main descriptors in the QSPR model. Both studies indicated that the charges of the chemical compounds (e.g., $q\mathrm{H}^+$, $q\mathrm{CH}^+$, and $q\mathrm{P}$) are the essential descriptors in QSPR models of coagulation. However, the current QSPR model of coagulation is based on polymer and ferric coagulants. Thus, the development of QSPR models of other types or groups of coagulants is necessary to cover a wider range of QSPR model applications, which can lead to significant cost savings in optimization and parameter determination for contaminant removal by coagulation.

7. APPLICATION TO OZONATION

Ozone is a widely used oxidant for water and wastewater treatment owing to its high oxidation and disinfection potentials.[53] Because of their electrophilic character, ozone molecules react specifically with high-electron density sites of contaminants (e.g., unsaturated bonds and aromatics).[54] Moreover, ozone can cause $^{\bullet}$OH radicals to form in water, which are less specific and react quickly with a wide range of organic functional groups.[55] The first QSPR model of ozonation was developed by Sudhakaran et al.[56] Their model

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

focused on the degradation of pharmaceuticals, personal care products, and pesticides in bench-scale studies of natural water matrices (e.g., the Colorado River, the Passaic River, the Ohio River, and synthetic Suwannee water). The molecular descriptors that affected the model were found to be $E_{LUMO} - E_{HOMO}$, electron affinity (EA), mean oxidation number, number of ring atoms (#ring atoms), number of ring atoms in five- or six-membered rings (#in56), number of halogens (#X), and oxygen/carbon ratio (O/C). Contaminants with large $E_{LUMO} - E_{HOMO}$ values showed low chemical reactivity toward ozone.[56] The #ring atoms and #in56 descriptors represent the presence of carbon-based ring systems (i.e., high electron density), which can easily react with ozone.[57] EA indicates the affinity of a molecule for electrons, where contaminants with low EA enhance ozonation because ozone and the contaminants do not compete for electrons.[58] With regard to #X, a high value of #X will cause electron deficiency in organic micropollutants and decrease the ozonation capacity.[59] O/C indicates the oxidizability of the compound, and high O/C values indicate low reactivity with ozone.[56] In addition, a more advanced QSPR model based on an artificial neural network algorithm was also developed to compile the data from these four rivers. It showed that the importance of the descriptors is ranked as follows: $E_{LUMO} - E_{HOMO}$ > #ring atoms > EA > #X. However, these QSPR models can be applied to only ozone at pH 7.00−8.00.

In a different study by Sudhakaran and Amy,[60] another QSPR model of ozone was developed using a wider pH range (5.00−8.00) and different input descriptors [i.e., double bond equivalence (DBE), weakly polar component of solvent accessible surface area (WPSA), and ionization potential (IP)].[60] DBE indicates the number of rings, double bonds, or triple bonds present in a compound, which is positively correlated with ozonation efficiency. WPSA describes the surface area of the compound, which consists of halogens. Ozonation can be inhibited by the presence of halogens because they are electrophilic and withdraw electrons from the chemical system, which causes the reactive molecular sites to become electron-deficient.[59] Finally, the IP is the amount of energy required for an electron to be released from a neutral chemical system.[61] Thus, when the IP is low, more electrons are available to ozone, and the ozonation efficiency is higher. These promising results showed that the QSPR model can be applied to predict the ozonation process in a wide pH range.

In recent years, two studies by the same research group have investigated QSPR model application to elucidate the ozonation mechanisms of organic compounds at neutral and acidic pH using different input descriptors.[62,63] At neutral pH (pH 7.0), three descriptors were well-correlated with the ozone reaction rate constant (k): the Fukui indices for the minimum nucleophilic attack constant [$f(+)_n$], the most positive partial charge on a carbon atom (qC^+), and E_{HOMO}. When $f(+)_n$ is larger, it becomes harder for the C−H bonds of aliphatic hydrocarbons and N−H bonds of amines to be ruptured by ozone. By contrast, as E_{HOMO} increases, electron donation becomes easier, and hence, compounds react more easily with ozone. Finally, qC^+ represents non-uniform electric charge on the main chain, which indicates the ease or complexity of valence bond breakage in organic molecules. However, at acidic pH (4.0), the maximum nucleophilic attack [$f(+)_x$], maximum electrophilic attack [$f(0)_x$], and most negative partial charge on the carbon atom (qC^-) were the main QSPR descriptors for the ozonation process. As $f(+)_x$, which indicates the likelihood of a nucleophilic attack, increases, it becomes easier for the main chain carbon at the site to be attacked by ozone. By contrast, rupturing the C−H bonds of aliphatic hydrocarbons and N−H bonds of amines becomes more difficult with an increase in $f(0)_x$. In addition, qC^- is usually associated with electron withdrawal. Despite this advance in the application of the QSPR model to ozonation, most studies have focused on the parent contaminants, and limited information about the formation of degradation products is available.[64] Full mineralization of contaminants is rarely realized during ozonation; therefore, it is also important to develop a predictive QSPR model for the transformation products. Furthermore, theoretical thermodynamic computations of target compounds, oxidants, intermediates, and derivatives are still limited by the absence of a reliable and affordable model chemistry and software.[65]

8. APPLICATION TO FENTON REACTIONS

Fenton reactions rely on the reaction of hydrogen peroxide (H_2O_2) with iron ions (Fe^{2+}) to form reactive oxygen species (ROS, mainly hydroxyl radicals) that can oxidize water contaminants. The optimal conditions for the Fenton process are affected mainly by the operating pH, Fe^{2+} concentration, H_2O_2 concentration, and operating temperature.[66] Several studies have examined the application of a QSPR model to predict and elucidate the mechanism of the Fenton process. The first QSPR model of the Fenton process was developed using the Hammett correlation,[67] which is used to predict the effect of substituents (methyl, amino, nitro, bromo, iodo, chloro, fluoro, etc.) on the rate constant of the reaction between oxidants and organic compounds. The Hammett constant (σ) represents the effects of inductive and resonance effects on substituents.[68,69] The results indicated that this method can predict the Fenton reaction of phenolic acid with modest success (R was approximately 0.8, and the standard deviation was 0.07). However, there is no validation method for this model owing to the simplicity of the Hammett correlation. Temperature is one of the most important operating parameters in the Fenton process during the degradation of organic compounds.[66] Thermal activation could improve the oxidative ability of hydrogen peroxide by producing more hydroxyl radicals, which have stronger oxidative activity in the Fenton process.[70] A high concentration of hydroxyl radicals in the reaction solution could increase the number of effective collisions between the oxidant and organic compounds. Thus, a QSPR model of the Fenton oxidation of 18 organic compounds in a temperature range of 15−60 °C was recently developed.[71] The main descriptors in this model were the temperature of the reaction system, reciprocal of the reaction temperature, number of carbon atoms per molecule (#C), ratio of the numbers of oxygen and carbon atoms, maximum positive partial charge on a hydrogen atom bonded with a carbon atom (qCH^+), minimum negative partial charge on a carbon atom (qC^-), square of the gap energy [$(E_{LUMO} - E_{HOMO})^2$], most positive partial charge on a hydrogen atom (qH^+), bond order (BO), and Fukui indices. A high #C value indicates that a chemical is a complex compound, and it will be difficult for hydroxyl radicals to attack it. The descriptors qCH^+, qC^-, and qH^+ and the Fukui indices indicate the changes in the charge on each atom during oxidation by the Fenton process. In addition, $E_{LUMO} - E_{HOMO}$ indicates the ability to lose electrons.[72] Finally, the BO in compounds and the chemical bonds in a compound tend to be more stable if

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

the BO is larger.[71] However, this Fenton process and QSPR model can be applied only at pH 3.0.

Many efforts have been made to improve the Fenton process and expand its application to a wider pH range. One promising method is to combine iron precursors with a supporting material. Li et al.[73] developed a catalyst consisting of a polyacrylonitrile fiber Fe complex and used a QSPR model to predict and enhance the understanding of this catalyst for the Fenton reaction at pH 6.0. The main descriptors were MW divided by the number of sulfonate groups (MW/S), the number of azo linkages ($N_{N=N}$), the number of aromatic rings (N_{AR}), and the inorganic character divided by the organic character (I/O). Dyes with more sulfonate groups have a relatively low MW/S and higher decoloration values, which could result from enhanced adsorption between the polyacrylonitrile fiber Fe complex and dye molecules. In this study, $N_{N=N}$ was also an important factor in dye mineralization. Moreover, aromatic intermediates with higher N_{AR} and lower I/O values may be easily attracted to the polyacrylonitrile fiber Fe complex, resulting in rapid mineralization. However, using dyes to assess the performance only at pH 6.0 is not recommended because the dissociation capacity of dyes varies with pH, and this variation can affect the process. Therefore, further comprehensive studies in a wider pH range are needed to explore the superiority of newly developed Fenton-based processes and comprehensively assess the QSPR application.

9. APPLICATION TO PHOTOLYSIS AND PHOTOCATALYSIS

The objective of QSPR studies of photolysis and photocatalysis is to determine the reaction rate constant (k) or half-life ($t_{1/2}$) of target pollutants under exposure to photons or radicals. Photolysis is categorized as direct or indirect, where direct photolysis is a transformation process driven mainly by the irradiation of chromophores and indirect photolysis causes degradation by the reaction of the contaminant with a reactive species generated by photosensitizers, which can absorb radiation to reach an excited state.[74] QSPR modeling has been applied to predict photolytic degradation of various organic contaminants, such as polychlorinated dibenzo-*p*-dioxins,[75] polycyclic aromatic hydrocarbons,[76] halogenated disinfection byproducts,[77] and azo dyes[78] under ultraviolet (UV) and solar irradiation. Frontier molecular orbital energies (i.e., E_{HOMO}, E_{LUMO}, and $E_{HOMO} - E_{LUMO}$), which are directly related to the IP, are reportedly key indicators in photolysis and radical oxidation reactions; they are highly correlated with the chemical reactivity and can affect the degree of electron transfer reactions.[77] These findings are consistent with QSPR modeling results for photolysis, which also indicated that the main descriptors that always contribute to photolysis of different types of contaminants are E_{HOMO}, E_{LUMO}, and $E_{HOMO} - E_{LUMO}$ (Table 3). The existing QSPR predictive models of photolysis showed both a high goodness of fit ($R^2 > 0.8$) and high predictivity ($Q^2 > 0.8$) (Table 3). The most recent study of application of the QSPR model to photolysis indicated that pH affected the QSPR model development.[78] If the pH become higher or lower than the pK_a, the use of the IP as a descriptor for targeted contaminants can be affected owing to chemical protonation or deprotonation in solution.[79] Moreover, indirect photolysis in natural water could be enhanced owing to the presence of triplet natural organic matter and carbonate radicals.[80] Despite this enhancement, in most cases, background water matrices (i.e., the organic and inorganic constituents) can act as scavengers for the radicals and introduce the inner filter effect, which can change the photolytic performance.[81] Therefore, to explore QSPR model application, the effect of solution pH on protonated and deprotonated species of contaminants and background water matrices needs to be considered in quantum chemical calculations and in the development of models of photolysis.

The application of photolysis for the removal of organic contaminants is very limited because the energy supplied by light is not sufficient for the reaction to begin and/or the rates of photolysis are sometimes too low to be of interest.[82] Thus, semiconductor materials have been used as photocatalysts to accelerate the kinetics of this process and to enhance its performance in treating contaminants in water or wastewater. In photocatalysis, light with an energy that is greater than the band gap of the semiconductor excites an electron from the valence band to the conduction band and creates an electron−hole pair, which triggers a series of reactions resulting in the photocatalytic degradation of pollutants.[83] The electrons and holes on the surface of the semiconductor will participate in redox reactions that produce ROS such as hydroxyl radicals ($^{\bullet}OH$) and superoxide anion radicals ($^{\bullet}O_2^-$). These reactive species are very strong and can destroy the bonds of stable and unreactive organic molecules to form organic intermediates before reaching the total mineralization of contaminants.[84] Among semiconductor materials, TiO_2 shows the greatest potential for use in green chemistry technology. Huang et al.[85] developed a QSPR model application for the prediction of sulfonamide degradation using TiO_2 and its composites with activated carbon fiber (ACF) under UVA irradiation (365 nm). QSPR models of the TiO_2 system generally use three key types of descriptors: (1) E_{HOMO} and E_{LUMO}, which represent the oxidation reaction in photocatalysis, in which electrons are gained or lost, (2) the Fukui indices (functions that describe the electron density in a frontier orbital), which represent the interaction of contaminants with $^{\bullet}OH$, and (3) charge distribution parameters, which represent the reactivity of the chemical. However, for ACF-TiO_2, the adsorptive descriptors [apparent sorption rate constant (k_{ad})] and Fukui indices were the main descriptors. This difference can be explained using the Langmuir−Hinshelwood model, in which the apparent rate constant of photocatalysis (k_{app}) is the integrated parameter of the intrinsic photocatalysis reaction rate constant (k') and k_{ad}.[86] It is reasonably deduced that k' and k_{ad} are positively correlated and play important roles in the prediction of the rate constant in the QSPR model of ACF-TiO_2. In addition, it was found that the physicochemical properties of the photocatalysts (e.g., specific surface area, functionalized groups, and pore size and distribution), physicochemical properties of the targeted contaminants, and background water matrices (i.e., the presence of organic and inorganic species) can affect the entire photocatalytic process.[87] Therefore, the actual performance of QSPR modeling of photocatalysis remains unclear until relevant environmental conditions and various types of photocatalysts are considered in model development.

The application of photocatalysis for disinfection has also attracted much attention. In photocatalytic disinfection, microorganisms are inactivated or killed when ROS attack and break the cell wall and denature proteins, producing various end products.[88] Pathakoti et al.[89] developed a QSPR model to predict the photocatalytic disinfection of *Escherichia coli* using 17 semiconductors/metal oxides under dark and light conditions. Under dark conditions, the toxicity toward *E.*

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

coli is a process that is completely controlled by electronegativity, in which the ionic potential is the main contributor to the toxicity. As the ionic potential decreases, the level of screening of a cation in an oxide by surrounding anions decreases, resulting in a higher toxicity. Under light irradiation, electronegativity (E_{HOMO} and E_{LUMO}) and molar heat capacity are the main QSPR descriptors. The band gap ($E_{HOMO} - E_{LUMO}$) is related to the photoreactive ability of a metal oxide to produce radicals, whereas the molar heat is associated with the binding reaction on the surface of the metal oxide. Metal oxides with a sufficient band gap (i.e., one that is equal to or lower than the energy provided by light irradiation) and strong binding interaction will exhibit high photocatalytic disinfection activity.[90] In this regard, it was found that the physical separation of microorganisms by adsorption can play a significant role in disinfection by carbonaceous composites with TiO_2.[91] Thus, the development of QSPR models of composite materials should be pursued to cover a wider range of photocatalytic disinfection processes.

10. CONCLUSIONS AND FUTURE OUTLOOK

Many contaminants that significantly affect water environments have been reported in the literature. The application of the QSPR model together with laboratory experiments could play an important role in extending our understanding of water treatment processes. Because many parameters are involved, additional studies are required to determine the optimal model conditions for improving QSPR model prediction of water treatment processes. This review leads us to suggest several recommendations.

- Many QSPR models have been developed to predict water treatment processes and clarify their mechanisms. However, many of these models do not follow the OECD principles for the validation of QSPR models. Moreover, there is still no agreement regarding the use of data splitting to validate QSPR models. Therefore, we proposed good practices for QSPR modeling as simple guidelines for developing new QSPR models (Figure 1).
- Researchers have typically used a single QSPR model/algorithm; the use of a combination of QSPR models and algorithms to enhance the performance of QSPR models has rarely been reported. Moreover, most investigations have used high contaminant concentrations (micrograms per liter to milligrams per liter). Because QSPR models can support/predict and enhance the understanding of water treatment processes compared to laboratory experiments alone, the performance of a combined QSPR model/algorithm for other treatments under more realistic concentrations (up to micrograms per liter) is a high priority for future work.
- The type and number of input descriptors can affect the performance of a QSPR model. Further research is needed to understand and identify the optimal type and number of input descriptors and find the best input data for QSPR models under different environmental conditions and for various water treatment technologies.
- Practical applications of QSPR models of water treatment are limited because available studies have been conducted under simplified laboratory-scale conditions (i.e., in ultrapure water) and were based on batch experimental systems (i.e., at laboratory scale). These conditions are inadequate for evaluating and predicting water treatments in real treatment systems. Therefore, data from pilot- or full-scale experiments must be collected, and more advanced QSPR models should be developed to evaluate the reliability of QSPR models for the prediction of real treatment systems and the elucidation of their mechanisms.

This review examined the current state of knowledge and the limitations of the application of QSPR models for prediction of water treatment processes. We hope that it will promote the use of this approach and result in further development of QSPR applications in water treatment.

AUTHOR INFORMATION

Corresponding Author

Mohamed Ateia − *Department of Chemistry, Northwestern University, Evanston, Illinois 60208, United States;* orcid.org/0000-0002-3524-5513; Email: ateia@northwestern.edu

Authors

Dion Awfa − *Water and Wastewater Engineering Research Group, Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Bandung 40132, Indonesia*

David Mendoza − *Department of Civil and Environmental Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan*

Chihiro Yoshimura − *Department of Civil and Environmental Engineering, Tokyo Institute of Technology, Tokyo 152-8552, Japan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsestwater.0c00206

Author Contributions

‖D.A. and M.A. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.A. is grateful for financial support by the Collaborative Water-Energy Research Center (CoWERC), supported by the Binational Industrial Research and Development Foundation under Energy Center Grant EC-15. This research was also funded by the Japan Society for the Promotion of Science (JSPS KAKENHI, Grant 18H01566).

REFERENCES

(1) Kubinyi, H. QSAR and 3D QSAR in Drug Design Part 2: Applications and Problems. *Drug Discovery Today* **1997**, *2*, 538−546.

(2) Brasquet, C.; Bourges, B.; Le Cloirec, P. Quantitative Structure-Property Relationship (QSPR) for the Adsorption of Organic Compounds onto Activated Carbon Cloth: Comparison between Multiple Linear Regression and Neural Network. *Environ. Sci. Technol.* **1999**, *33* (23), 4226−4231.

(3) Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Qualitative Structure-Property Relationships Treatment. *Anal. Chem.* **1994**, *66* (11), 1799−1807.

(4) Farahani, S. R.; Sohrabi, M. R.; Ghasemi, J. B. A Detailed Structural Study of Cytotoxicity e Ff Ect of Ionic Liquids on the Leukemia Rat Cell Line IPC-81 by Three Dimensional Quantitative Structure Toxicity Relationship. *Ecotoxicol. Environ. Saf.* **2018**, *158*, 256−265.

(5) Yu, X. Quantitative Structure-Toxicity Relationships of Organic Chemicals against Pseudokirchneriella Subcapitata. *Aquat. Toxicol.* **2020**, *224*, 105496.

(6) Touhami, I.; Messadi, D. QSER Study for Half Wave Potential of Some PAHs. *Energy Procedia* **2019**, *157* (2018), 522−532.

(7) Roy, K.; Kar, S.; Das, R. N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*; Elsevier, 2015.
(8) Huang, Y.; Li, T.; Zheng, S.; Fan, L.; Su, L.; Zhao, Y.; Xie, H.; Li, C. QSAR Modeling for the Ozonation of Diverse Organic Compounds in Water. *Sci. Total Environ.* **2020**, *715*, 136816.
(9) Roy, J.; Ghosh, S.; Ojha, P. K.; Roy, K. Environmental Science Nano Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs). *Environ. Sci.: Nano* **2019**, *6*, 224−247.
(10) Gramatica, P. Principles of QSAR Modeling. *Int. J. Quant. Struct.-Prop. Relat.* **2020**, *5* (3), 61−97.
(11) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'Min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57* (12), 4977−5010.
(12) Yousefinejad, S.; Hemmateenejad, B. Chemometrics Tools in QSAR/QSPR Studies: A Historical Perspective. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 177−204.
(13) Polishchuk, P. Interpretation of Quantitative Structure - Activity Relationship Models: Past, Present, and Future. *J. Chem. Inf. Model.* **2017**, *57*, 2618.
(14) Danishuddin; Khan, A. U. Descriptors and Their Selection Methods in QSAR Analysis: Paradigm for Drug Design. *Drug Discovery Today* **2016**, *21* (8), 1291−1302.
(15) Lill, M. A. Multi-Dimensional QSAR in Drug Discovery. *Drug Discovery Today* **2007**, *12* (23−24), 1013−1017.
(16) Verma, R. P.; Hansch, C. Investigation of DNA-Binding Properties of Organic Molecules Using Quantitative Structure-Activity Relationship (QSAR) Models. *J. Pharm. Sci.* **2008**, *97* (1), 88−110.
(17) Ambure, P.; Aher, R. B.; Gajewicz, A.; Puzyn, T.; Roy, K. Chemometrics and Intelligent Laboratory Systems " NanoBRIDGES " Software: Open Access Tools to Perform QSAR and Nano-QSAR Modeling. *Chemom. Intell. Lab. Syst.* **2015**, *147*, 1−13.
(18) Choi, J.; Trinh, T. X.; Yoon, T.; Kim, J.; Byun, H. Chemosphere Quasi-QSAR for Predicting the Cell Viability of Human Lung and Skin Cells Exposed to Different Metal Oxide Nanomaterials. *Chemosphere* **2019**, *217*, 243−249.
(19) Burello, E. NanoImpact Review of (Q) SAR Models for Regulatory Assessment of Nanomaterials Risks. *NanoImpact* **2017**, *8* (July), 48−58.
(20) Yang, Y.; Lin, T.; Weng, X. L.; Darr, J. A.; Wang, X. Z. Data Flow Modeling, Data Mining and QSAR in High-Throughput Discovery of Functional Nanomaterials. *Comput. Chem. Eng.* **2011**, *35* (4), 671−678.
(21) Delgado, L. F.; Charles, P.; Glucina, K.; Morlay, C. QSAR-like Models: A Potential Tool for the Selection of PhACs and EDCs for Monitoring Purposes in Drinking Water Treatment Systems e A Review. *Water Res.* **2012**, *46* (19), 6196−6209.
(22) Chen, B.; Zhang, T.; Bond, T.; Gan, Y. Development of Quantitative Structure Activity Relationship (QSAR) Model for Disinfection Byproduct (DBP) Research: A Review of Methods and Resources. *J. Hazard. Mater.* **2015**, *299*, 260−279.
(23) Villaverde, J. J.; Sevilla-Morán, B.; López-Goti, C.; Alonso-Prados, J. L.; Sandín-España, P. Considerations of Nano-QSAR/QSPR Models for Nanopesticide Risk Assessment within the European Legislative Framework. *Sci. Total Environ.* **2018**, *634*, 1530−1539.
(24) Schultz, T. W.; Cronin, M. T. D.; Walker, J. D.; Aptula, A. O. Quantitative Structure - Activity Relationships (QSARs) in Toxicology: A Historical Perspective. *J. Mol. Struct.: THEOCHEM* **2003**, *622*, 1−22.
(25) Villaverde, J. J.; Sevilla-Morán, B.; López-Goti, C.; Alonso-Prados, J. L.; Sandín-España, P. Considerations of Nano-QSAR/QSPR Models for Nanopesticide Risk Assessment within the European Legislative Framework. *Sci. Total Environ.* **2018**, *634*, 1530−1539.
(26) Soeteman-Hernández, L. G.; Bekker, C.; Groenewold, M.; Jantunen, P.; Mech, A.; Rasmussen, K.; Sintes, J. R.; Sips, A. J. A. M.; Noorlander, C. W. NanoImpact Perspective on How Regulators Can Keep Pace with Innovation: Outcomes of a European Regulatory Preparedness Workshop on Nanomaterials and Nano- Enabled Products. *NanoImpact* **2019**, *14*, 100166.
(27) Dearden, J. C.; Cronin, M. T. D.; Kaiser, K. L. E. How Not to Develop a Quantitative Structure-Activity or Structure-Property Relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20* (3−4), 241−266.
(28) Roy, K.; Ambure, P. The "Double Cross-Validation" Software Tool for MLR QSAR Model Development. *Chemom. Intell. Lab. Syst.* **2016**, *159*, 108−126.
(29) Roy, K. On Some Aspects of Validation of Predictive Quantitative Structure-Activity Relationship Models. *Expert Opin. Drug Discovery* **2007**, *2* (12), 1567−1577.
(30) May, R. J.; Maier, H. R.; Dandy, G. C. Data Splitting for Artificial Neural Networks Using SOM-Based Stratified Sampling. *Neural Networks* **2010**, *23* (2), 283−294.
(31) Gramatica, P.; Pilutti, P.; Papa, E. Validated QSAR Prediction of OH Tropospheric Degradation of VOCs: Splitting into Training-Test Sets and Consensus Modeling. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1794−1802.
(32) Rodriguez-Narvaez, O. M.; Peralta-Hernandez, J. M.; Goonetilleke, A.; Bandala, E. R. Treatment Technologies for Emerging Contaminants in Water: A Review. *Chem. Eng. J.* **2017**, *323*, 361−380.
(33) Ahmed, M. B.; Zhou, J. L.; Ngo, H. H.; Johir, Md. A. H.; Sun, L.; Asadullah, M.; Belhaj, D. Sorption of Hydrophobic Organic Contaminants on Functionalized Biochar: Protagonist Role of π - π Electron-Donor-Acceptor Interactions and Hydrogen Bonds. *J. Hazard. Mater.* **2018**, *360*, 270−278.
(34) Nirmalakhandan, N. N.; Speece, R. E. QSAR Model for Predicting Henry's Constant. *Environ. Sci. Technol.* **1988**, *22* (11), 1349−1357.
(35) Zhao, Y.; Choi, J. W.; Bediako, J. K.; Song, M. H.; Lin, S.; Cho, C. W.; Yun, Y. S. Adsorptive Interaction of Cationic Pharmaceuticals on Activated Charcoal: Experimental Determination and QSAR Modelling. *J. Hazard. Mater.* **2018**, *360*, 529−535.
(36) Chen, W. Y.; Liu, Z. C.; Lin, P. H.; Fang, C. I.; Yamamoto, S. The Hydrophobic Interactions of the Ion-Exchanger Resin Ligands with Proteins at High Salt Concentrations by Adsorption Isotherms and Isothermal Titration Calorimetry. *Sep. Purif. Technol.* **2007**, *54* (2), 212−219.
(37) Ersan, G.; Kaya, Y.; Apul, O. G.; Karanfil, T. Adsorption of Organic Contaminants by Graphene Nanosheets, Carbon Nanotubes and Granular Activated Carbons under Natural Organic Matter Preloading Conditions. *Sci. Total Environ.* **2016**, *565*, 811−817.
(38) Ersan, G.; Apul, O. G.; Perreault, F.; Karanfil, T. Adsorption of Organic Contaminants by Graphene Nanosheets: A Review. *Water Res.* **2017**, *126*, 385−398.
(39) Ghosh, S.; Ojha, P. K.; Roy, K. Exploring QSPR Modeling for Adsorption of Hazardous Synthetic Organic Chemicals (SOCs) by SWCNTs. *Chemosphere* **2019**, *228*, 545−555.
(40) Ling, Y.; Klemes, M. J.; Steinschneider, S.; Dichtel, W. R.; Helbling, D. E. QSARs to Predict Adsorption Affinity of Organic Micropollutants for Activated Carbon and B-Cyclodextrin Polymer Adsorbents. *Water Res.* **2019**, *154*, 217−226.
(41) Apul, O. G.; Wang, Q.; Shao, T.; Rieck, J. R.; Karanfil, T. Predictive Model Development for Adsorption of Aromatic Contaminants by Multi-Walled Carbon Nanotubes. *Environ. Sci. Technol.* **2013**, *47* (5), 2295−2303.
(42) Apul, O. G.; Perreault, F.; Ersan, G.; Karanfil, T. Linear Solvation Energy Relationship Development for Adsorption of Synthetic Organic Compounds by Carbon Nanomaterials: An Overview of the Last Decade. *Environ. Sci. Water Res. Technol.* **2020**, *6*, 2949.

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

(43) Oatley-Radcliffe, D. L.; Walters, M.; Ainscough, T. J.; Williams, P. M.; Mohammad, A. W.; Hilal, N. Nanofiltration Membranes and Processes: A Review of Research Trends over the Past Decade. *J. Water Process Eng.* **2017**, *19* (July), 164−171.
(44) Köhler, S. J.; Lavonen, E.; Keucken, A.; Schmitt-Kopplin, P.; Spanjer, T.; Persson, K. Upgrading Coagulation with Hollow-Fibre Nanofiltration for Improved Organic Matter Removal during Surface Water Treatment. *Water Res.* **2016**, *89*, 232−240.
(45) Yangali-Quintanilla, V.; Sadmani, A.; McConville, M.; Kennedy, M.; Amy, G. A QSAR Model for Predicting Rejection of Emerging Contaminants (Pharmaceuticals, Endocrine Disruptors) by Nanofiltration Membranes. *Water Res.* **2010**, *44* (2), 373−384.
(46) Chang, E. E.; Liang, C. H.; Huang, C. P.; Chiang, P. C. A Simplified Method for Elucidating the Effect of Size Exclusion on Nanofiltration Membranes. *Sep. Purif. Technol.* **2012**, *85*, 1−7.
(47) Koo, C. H.; Mohammad, A. W.; Suja', F.; Meor Talib, M. Z. Review of the Effect of Selected Physicochemical Factors on Membrane Fouling Propensity Based on Fouling Indices. *Desalination* **2012**, *287*, 167−177.
(48) Shahmansouri, A.; Bellona, C. Application of Quantitative Structure-Property Relationships (QSPRs) to Predict the Rejection of Organic Solutes by Nanofiltration. *Sep. Purif. Technol.* **2013**, *118*, 627−638.
(49) Jiang, J. Q. The Role of Coagulation in Water Treatment. *Curr. Opin. Chem. Eng.* **2015**, *8*, 36−44.
(50) Sillanpää, M.; Ncibi, M. C.; Matilainen, A.; Vepsäläinen, M. Removal of Natural Organic Matter in Drinking Water Treatment by Coagulation: A Comprehensive Review. *Chemosphere* **2018**, *190*, 54−71.
(51) Cheng, Z.; Yang, B.; Chen, Q.; Ji, W.; Shen, Z. Characteristics and Difference of Oxidation and Coagulation Mechanisms for the Removal of Organic Compounds by Quantum Parameter Analysis. *Chem. Eng. J.* **2018**, *332*, 351−360.
(52) Chu, Y. B.; Li, M.; Liu, J. W.; Xu, W.; Cheng, S. H.; Zhao, H. Z. Molecular Insights into the Mechanism and the Efficiency-Structure Relationship of Phosphorus Removal by Coagulation. *Water Res.* **2018**, *147*, 195−203.
(53) Gomes, J.; Costa, R.; Quinta-Ferreira, R. M.; Martins, R. C. Application of Ozonation for Pharmaceuticals and Personal Care Products Removal from Water. *Sci. Total Environ.* **2017**, *586*, 265−283.
(54) Umar, M.; Roddick, F.; Fan, L.; Aziz, H. A. Application of Ozone for the Removal of Bisphenol A from Water and Wastewater - A Review. *Chemosphere* **2013**, *90* (8), 2197−2207.
(55) Nöthe, T.; Fahlenkamp, H.; von Sonntag, C. Ozonation of Wastewater: Rate of Ozone Consumption and Hydroxyl Radical Yield. *Environ. Sci. Technol.* **2009**, *43* (15), 5990−5995.
(56) Sudhakaran, S.; Calvin, J.; Amy, G. L. QSAR Models for the Removal of Organic Micropollutants in Four Different River Water Matrices. *Chemosphere* **2012**, *87* (2), 144−150.
(57) von Gunten, U. Ozonation of Drinking Water: Part I. Oxidation Kinetics and Product Formation. *Water Res.* **2003**, *37*, 1443−1467.
(58) Gutsev, G. L.; Boldyrev, A. I. Theoretical Estimation of the Maximal Value of the First, Second, and Higher Electron Affinity of Chemical Compounds. *J. Phys. Chem.* **1990**, *94* (6), 2256−2259.
(59) Lei, H.; Snyder, S. A. 3D QSPR Models for the Removal of Trace Organic Contaminants by Ozone and Free Chlorine. *Water Res.* **2007**, *41* (18), 4051−4060.
(60) Sudhakaran, S.; Amy, G. L. QSAR Models for Oxidation of Organic Micropollutants in Water Based on Ozone and Hydroxyl Radical Rate Constants and Their Chemical Classification. *Water Res.* **2013**, *47* (3), 1111−1122.
(61) Güsten, H. Predicting the Abiotic Degradability of Organic Pollutants in the Troposphere. *Chemosphere* **1999**, *38* (6), 1361−1370.
(62) Zhu, H.; Shen, Z.; Tang, Q.; Ji, W.; Jia, L. Degradation Mechanism Study of Organic Pollutants in Ozonation Process by QSAR Analysis. *Chem. Eng. J.* **2014**, *255*, 431−436.
(63) Zhu, H.; Guo, W.; Shen, Z.; Tang, Q.; Ji, W.; Jia, L. QSAR Models for Degradation of Organic Pollutants in Ozonation Process under Acidic Condition. *Chemosphere* **2015**, *119*, 65−71.
(64) Lee, Y.; Von Gunten, U. Advances in Predicting Organic Contaminant Abatement during Ozonation of Municipal Wastewater Effluent: Reaction Kinetics, Transformation Products, and Changes of Biological Effects. *Environ. Sci. Water Res. Technol.* **2016**, *2* (3), 421−442.
(65) Tentscher, P. R.; Lee, M.; Von Gunten, U. Micropollutant Oxidation Studied by Quantum Chemical Computations: Methodology and Applications to Thermodynamics, Kinetics, and Reaction Mechanisms. *Acc. Chem. Res.* **2019**, *52* (3), 605−614.
(66) Babuponnusami, A.; Muthukumar, K. A Review on Fenton and Improvements to the Fenton Process for Wastewater Treatment. *J. Environ. Chem. Eng.* **2014**, *2* (1), 557−572.
(67) Peres, J. A.; Domínguez, J. R.; Beltran-Heredia, J. Reaction of Phenolic Acids with Fenton-Generated Hydroxyl Radicals: Hammett Correlation. *Desalination* **2010**, *252* (1−3), 167−171.
(68) Mukherjee, R. N.; Rajan, O. A.; Chakravorty, A. Electron Transfer in Groups of Iron, Cobalt, and Copper Triazene 1-Oxides: Hammett Correlation, Ligand Redistribution, and Crystal Field Effects. *Inorg. Chem.* **1982**, *21* (2), 785−790.
(69) Jovanovic, S. V.; Tosic, M.; Simic, M. G. Use of the Hammett Correlation and Σ+ for Calculation of One-Electron Redox Potentials of Antioxidants. *J. Phys. Chem.* **1991**, *95* (26), 10824−10827.
(70) Lee, C.; Yoon, J. Temperature Dependence of Hydroxyl Radical Formation in the Hv/Fe 3+/H2O2 and Fe3+/H2O 2 Systems. *Chemosphere* **2004**, *56* (10), 923−934.
(71) Cheng, Z.; Yang, B.; Chen, Q.; Shen, Z.; Yuan, T. Quantitative Relationships between Molecular Parameters and Reaction Rate of Organic Chemicals in Fenton Process in Temperature Range of 15.8 °C-60 °C. *Chem. Eng. J.* **2018**, *350*, 534−540.
(72) Xiong, L.; Tang, J.; Li, Y.; Li, L. Phototoxic Risk Assessment on Benzophenone UV Filters: In Vitro Assessment and a Theoretical Model. *Toxicol. In Vitro* **2019**, *60* (5), 180−186.
(73) Li, B.; Dong, Y.; Ding, Z. Heterogeneous Fenton Degradation of Azo Dyes Catalyzed by Modified Polyacrylonitrile Fiber Fe Complexes: QSPR (Quantitative Structure Peorperty Relationship) Study. *J. Environ. Sci. (Beijing, China)* **2013**, *25* (7), 1469−1476.
(74) Wang, L.; Zhang, Q.; Chen, B.; Bu, Y.; Chen, Y.; Ma, J.; Rosario-Ortiz, F. L. Photolysis and Photocatalysis of Haloacetic Acids in Water: A Review of Kinetics, Influencing Factors, Products, Pathways, and Mechanisms. *J. Hazard. Mater.* **2020**, *391*, 122143.
(75) Chen, J.; Quan, X.; Peijnenburg, W. J. G. M.; Yang, F. Quantitative Structure-Property Relationships (QSPRs) on Direct Photolysis Quantum Yields of PCDDs. *Chemosphere* **2001**, *43* (2), 235−241.
(76) Chen, J.; Peijnenburg, W.J.G.M; Quan, X.; Chen, S.; Martens, D; Schramm, K.-W; Kettrup, A Is It Possible to Develop a QSPR Model for Direct Photolysis Half-Lives of PAHs under Irradiation of Sunlight? *Environ. Pollut.* **2001**, *114* (1), 137−143.
(77) Zhang, Y.; Xiao, Y.; Zhang, Y.; Lim, T. T. UV Direct Photolysis of Halogenated Disinfection Byproducts: Experimental Study and QSAR Modeling. *Chemosphere* **2019**, *235*, 719−725.
(78) Zhang, G.; Zhang, S. Quantitative Structure-Activity Relationship in the Photodegradation of Azo Dyes. *J. Environ. Sci. (Beijing, China)* **2020**, *90*, 41−50.
(79) Chowdhury, P.; Sarathy, S. R.; Das, S.; Li, J.; Ray, A. K.; Ray, M. B. Direct UV Photolysis of Pharmaceutical Compounds: Determination of PH-Dependent Quantum Yield and Full-Scale Performance. *Chem. Eng. J.* **2020**, *380*, 122460.
(80) Arnold, W. A. One Electron Oxidation Potential as a Predictor of Rate Constants of N-Containing Compounds with Carbonate Radical and Triplet Excited State Organic Matter. *Environ. Sci. Process. Impacts* **2014**, *16* (4), 832−838.
(81) Koumaki, E.; Mamais, D.; Noutsopoulos, C.; Nika, M. C.; Bletsou, A. A.; Thomaidis, N. S.; Eftaxias, A.; Stratogianni, G. Degradation of Emerging Contaminants from Water under Natural Sunlight: The Effect of Season, PH, Humic Acids and Nitrate and

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX

Identification of Photodegradation by-Products. *Chemosphere* **2015**, *138*, 675−681.
(82) Kanakaraju, D.; Motti, C. A.; Glass, B. D.; Oelgemöller, M. Solar Photolysis versus TiO2-Mediated Solar Photocatalysis: A Kinetic Study of the Degradation of Naproxen and Diclofenac in Various Water Matrices. *Environ. Sci. Pollut. Res.* **2016**, *23*, 17437.
(83) Awfa, D.; Ateia, M.; Fujii, M.; Johnson, M. S.; Yoshimura, C. Photodegradation of Pharmaceuticals and Personal Care Products in Water Treatment Using Carbonaceous-TiO2 Composites: A Critical Review of Recent Literature. *Water Res.* **2018**, *142*, 26.
(84) Matos, J.; Laine, J.; Herrmann, J. M. Association of Activated Carbons of Different Origins with Titania in the Photocatalytic Purification of Water. *Carbon* **1999**, *37* (11), 1870−1872.
(85) Huang, X.; Feng, Y.; Hu, C.; Xiao, X.; Yu, D.; Zou, X. Mechanistic QSAR Models for Interpreting Degradation Rates of Sulfonamides in UV-Photocatalysis Systems. *Chemosphere* **2015**, *138*, 183−189.
(86) Ateia, M.; Alalm, M. G.; Awfa, D.; Johnson, M. S.; Yoshimura, C. Modeling the Degradation and Disinfection of Water Pollutants by Photocatalysts and Composites: A Critical Review. *Sci. Total Environ.* **2020**, *698*, 134197.
(87) Awfa, D.; Ateia, M.; Fujii, M.; Yoshimura, C. Photocatalytic Degradation of Organic Micropollutants: Inhibition Mechanisms by Different Fractions of Natural Organic Matter. *Water Res.* **2020**, *174*, 115643.
(88) Dalrymple, O. K.; Stefanakos, E.; Trotz, M. A.; Goswami, D. Y. A Review of the Mechanisms and Modeling of Photocatalytic Disinfection. *Appl. Catal., B* **2010**, *98* (1−2), 27−38.
(89) Pathakoti, K.; Huang, M. J.; Watts, J. D.; He, X.; Hwang, H. M. Using Experimental Data of Escherichia Coli to Develop a QSAR Model for Predicting the Photo-Induced Cytotoxicity of Metal Oxide Nanoparticles. *J. Photochem. Photobiol., B* **2014**, *130*, 234−240.
(90) Laxma Reddy, P. V.; Kavitha, B.; Kumar Reddy, P. A.; Kim, K. H. TiO2-Based Photocatalytic Disinfection of Microbes in Aqueous Media: A Review. *Environ. Res.* **2017**, *154*, 296−303.
(91) Shimizu, Y.; Ateia, M.; Wang, M.; Awfa, D.; Yoshimura, C. Disinfection Mechanism of E. Coli by CNT-TiO2 Composites: Photocatalytic Inactivation vs. Physical Separation. *Chemosphere* **2019**, *235*, 1041−1049.
(92) Tourassi, G. D.; Frederick, E. D.; Markey, M. K.; Floyd, C. E. Application of the Mutual Information Criterion for Feature Selection in Computer-Aided Diagnosis. *Med. Phys.* **2001**, *28* (12), 2394−2402.
(93) Bowden, G. J.; Maier, H. R.; Dandy, G. C. Optimal Division of Data for Neural Network Models in Water Resources Applications. *Water Resour. Res.* **2002**, *38* (2), 2-1−2-11.
(94) Bowden, G. J.; Dandy, G. C.; Maier, H. R. Input Determination for Neural Network Models in Water Resources Applications. Part 1 - Background and Methodology. *J. Hydrol.* **2005**, *301* (1−4), 75−92.
(95) Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11* (1), 137−148.
(96) Snee, R. D. Validation of Regression Models: Methods and Examples. *Technometrics* **1977**, *19* (4), 415.
(97) Zhang, G. P.; Berardi, V. L. Time Series Forecasting with Neural Network Ensembles: An Application for Exchange Rate Prediction. *J. Oper. Res. Soc.* **2001**, *52* (6), 652−664.
(98) Huang, Y.; Li, T.; Zheng, S.; Fan, L.; Su, L.; Zhao, Y.; Xie, H.-B.; Li, C. QSAR Modeling for the Ozonation of Diverse Organic Compounds in Water. *Sci. Total Environ.* **2020**, *715*, 136816.
(99) Jia, L.; Shen, Z.; Guo, W.; Zhang, Y.; Zhu, H.; Ji, W.; Fan, M. QSAR Models for Oxidative Degradation of Organic Pollutants in the Fenton Process. *J. Taiwan Inst. Chem. Eng.* **2015**, *46*, 140−147.
(100) Hou, B.; Ren, B.; Deng, R.; Zhu, G.; Wang, Z.; Li, Z. Three-Dimensional Electro-Fenton Oxidation of N-Heterocyclic Compounds with a Novel Catalytic Particle Electrode: High Activity, Wide PH Range and Catalytic Mechanism. *RSC Adv.* **2017**, *7* (25), 15455−15462.
(101) Cheng, Z.; Yang, B.; Chen, Q.; Ji, W.; Shen, Z. Characteristics and Difference of Oxidation and Coagulation Mechanisms for the Removal of Organic Compounds by Quantum Parameter Analysis. *Chem. Eng. J.* **2018**, *332*, 351−360.
(102) Kušić, H.; Rasulev, B.; Leszczynska, D.; Leszczynski, J.; Koprivanac, N. Prediction of Rate Constants for Radical Degradation of Aromatic Pollutants in Water Matrix: A QSAR Study. *Chemosphere* **2009**, *75* (8), 1128−1134.
(103) Lee, Y.; von Gunten, U. Quantitative Structure-Activity Relationships (QSARs) for the Transformation of Organic Micropollutants during Oxidative Water Treatment. *Water Res.* **2012**, *46* (19), 6177−6195.
(104) Cheng, Z.; Yang, B.; Chen, Q.; Gao, X.; Tan, Y.; Yuan, T.; Shen, Z. Quantitative-Structure-Activity-Relationship (QSAR) Models for the Reaction Rate and Temperature of Nitrogenous Organic Compounds in Supercritical Water Oxidation (SCWO). *Chem. Eng. J.* **2018**, *354*, 12−20.
(105) Zhou, S.; Zhang, W.; Sun, J.; Zhu, S.; Li, K.; Meng, X.; Luo, J.; Shi, Z.; Zhou, D.; Crittenden, J. C. Oxidation Mechanisms of the UV/Free Chlorine Process: Kinetic Modeling and Quantitative Structure Activity Relationships. *Environ. Sci. Technol.* **2019**, *53* (8), 4335−4345.
(106) Luo, X.; Wei, X.; Chen, J.; Xie, Q.; Yang, X.; Peijnenburg, W. J. G. M. Rate Constants of Hydroxyl Radicals Reaction with Different Dissociation Species of Fluoroquinolones and Sulfonamides: Combined Experimental and QSAR Studies. *Water Res.* **2019**, *166*, 115083.
(107) Cheng, Z.; Chen, Q.; Pontius, F. W.; Gao, X.; Tan, Y.; Ma, Y.; Shen, Z. Two New Predictors Combined with Quantum Chemical Parameters for the Selection of Oxidants and Degradation of Organic Contaminants: A QSAR Modeling Study. *Chemosphere* **2020**, *240*, 124928.
(108) Brasquet, C.; Le Cloirec, P. QSAR for Organics Adsorption onto Activated Carbon in Water: What about the Use of Neural Networks? *Water Res.* **1999**, *33* (17), 3603−3608.
(109) De Ridder, D. J.; McConville, M.; Verliefde, A. R. D.; Van Der Aa, L. T. J.; Heijman, S. G. J.; Verberk, J. Q. J. C.; Rietveld, L. C.; Van Dijk, J. C. Development of a Predictive Model to Determine Micropollutant Removal Using Granular Activated Carbon. *Drinking Water Eng. Sci.* **2009**, *2* (2), 57−62.
(110) de Ridder, D. J.; Villacorte, L.; Verliefde, A. R. D.; Verberk, J. Q. J. C.; Heijman, S. G. J.; Amy, G. L.; van Dijk, J. C. Modeling Equilibrium Adsorption of Organic Micropollutants onto Activated Carbon. *Water Res.* **2010**, *44* (10), 3077−3086.
(111) Ray, S.; Roy, K. Modeling Adsorption of Organic Compounds on Activated Carbon Using ETA Indices. *Chem. Eng. Sci.* **2013**, *104*, 427−438.
(112) Lata, S.; Vikas. Externally Predictive Quantum-Mechanical Models for the Adsorption of Aromatic Organic Compounds by Graphene-Oxide Nanomaterials. *SAR QSAR Environ. Res.* **2019**, *30* (12), 847−863.

https://dx.doi.org/10.1021/acsestwater.0c00206
ACS EST Water XXXX, XXX, XXX−XXX