



# Predicting the reaction rate constants of micropollutants with hydroxyl radicals in water using QSPR modeling<sup>☆</sup>



Xiaohui Jin<sup>a,\*</sup>, Sigrid Peldszus<sup>b</sup>, Peter M. Huck<sup>b</sup>

<sup>a</sup> Walkerton Clean Water Centre, Walkerton, Ontario N0G 2V0, Canada

<sup>b</sup> NSERC Chair in Water Treatment, Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

## HIGHLIGHTS

- QSPR model predicts hydroxyl radical rate constants of micropollutants.
- Multiple linear regression with forward selection method was used.
- Model was externally validated and the applicability domain was defined.
- Model determines susceptibility of contaminants to AOPs without experimentation.

## ARTICLE INFO

### Article history:

Received 24 December 2014

Received in revised form 7 May 2015

### Keywords:

Molecular descriptors  
Hydroxyl radical  
Reaction rate constant  
Micropollutants  
Applicability domain

## ABSTRACT

Quantitative structure–property relationship (QSPR) models which predict hydroxyl radical rate constants ( $k_{OH}$ ) for a wide range of emerging micropollutants are a cost effective approach to assess the susceptibility of these contaminants to advanced oxidation processes (AOPs). A QSPR model for the prediction of  $k_{OH}$  of emerging micropollutants from their physico-chemical properties was developed with special attention to model validation, applicability domain and mechanistic interpretation. In this study, 118 emerging micropollutants including those experimentally determined by the author and data collected from the literature, were randomly divided into the training set ( $n = 89$ ) and validation set ( $n = 29$ ). 951 DRAGON molecular descriptors were calculated for model development. The QSPR model was calibrated by applying forward multiple linear regression to the training set. As a result, 7 DRAGON descriptors were found to be important in predicting the  $k_{OH}$  values which related to the electronegativity, polarizability, and double bonds, etc. of the compounds. With outliers identified and removed, the final model fits the training set very well and shows good robustness and internal predictivity. The model was then externally validated with the validation set showing good predictive power. The applicability domain of the model was also assessed using the Williams plot approach. Overall, the developed QSPR model provides a valuable tool for an initial assessment of the susceptibility of micropollutants to AOPs.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Micropollutants such as endocrine disrupting chemicals (EDCs) and pharmaceuticals and personal care products (PPCPs) create unique challenges to water treatment because of the number of compounds detected and the diversity and complexity of their physico-chemical properties. These compounds are present at

low  $\text{ng L}^{-1}$  levels in source water samples and only a small number of them have been investigated to date (Padhye et al., 2012). Recent studies have summarized the presence of many trace-level EDCs and PPCPs in finished drinking water, distribution systems and in tap water (Delgado et al., 2012; Fan et al., 2013). Although their health effects are mostly unknown, the removal of these micropollutants from drinking water has been of concern to water utilities and environmental agencies due to the precautionary principle. Advanced treatment technologies such as advanced oxidation processes (AOPs), reverse osmosis, and nanofiltration have demonstrated effectiveness in removing micropollutants (Ivančev-Tumbas, 2014).

<sup>☆</sup> Note: The majority of the study was completed at the University of Waterloo. The first author is now affiliated with the Walkerton Clean Water Centre.

\* Corresponding author at: Walkerton Clean Water Centre, 20 Ontario Road, Walkerton, Ontario N0G 2V0, Canada.

E-mail address: [xjin@wcwc.ca](mailto:xjin@wcwc.ca) (X. Jin).

AOPs such as  $O_3/H_2O_2$ ,  $UV/H_2O_2$ , and  $UV/TiO_2$  produce a highly reactive oxidant, the hydroxyl radical, which reacts rapidly with most organic micropollutants and leads to their degradation (Huber et al., 2003). To investigate the removal efficiency of various organic micropollutants during AOPs in natural waters, it is necessary to obtain the reaction rate constants of these micropollutants for their reaction with hydroxyl radicals ( $k_{OH}$ ). Rate constants are needed when predicting the extent to which the original contaminants are eliminated from water, and they are therefore important for designing and optimizing treatment processes. Although kinetic data are available for a large number of chemicals for their reactions with hydroxyl radicals (Buxton et al., 1988), there is still a data gap for emerging micropollutants such as EDCs and PPCPs since it is time consuming and costly to establish these values experimentally. Thus, it is highly desirable and more cost-effective to develop a reliable model to predict the rate constants of numerous micropollutants.

Quantitative structure–property relationships (QSPR) have been widely used as a modeling tool to develop relationships between the properties of chemicals and their structural characteristics (Eriksson et al., 2003). The application of QSPR in water treatment studies is relatively recent, and only a few QSPR models have been proposed for ozonation, chlorination, AOPs, membrane filtration, and activated carbon adsorption (Delgado et al., 2012). To date, a small number of QSPR models have been published to predict  $k_{OH}$  values of organic compounds in the aqueous phase using several different approaches. The fixed-descriptors approach, a Hammett-type linear free energy relationship was employed for  $k_{OH}$  modeling (Peres et al., 2010; Zimbron and Reardon, 2005; Lee and von Gunten, 2012), in which the substituent constant  $\sigma$  was used to model the effect of functional groups of substituted aromatics. However, Hammett-type relationships are only applicable to substituted aromatics with known substituent constants and cannot be applied to other non-aromatic compounds. The selected-descriptors approach is often used when significant descriptors are unknown for developing empirical models, in which significant descriptors are selected from a pool of candidate descriptors using statistical methods (Kusić et al., 2009; Toropov et al., 2012; Sudhakaran and Amy, 2013). However, advanced statistical skills are needed for the selection of descriptors (e.g., genetic algorithm, principal component analysis). In addition, the final set of descriptors is selected based on the principal of best-fit to the training set compounds, sometimes making the model interpretation difficult. The group contribution approach has also been used to predict  $k_{OH}$  values for compounds with a wide range of functional groups (Monod and Doussin, 2008; Minakata et al., 2009). However, availability of data for all possible functional groups and the assumption of additivity of rate constants limit the use of the group contribution method (Minakata et al., 2009). In addition, the models developed thus far are not necessarily applicable to the emerging micropollutants with their diverse structures since many of these existing models use only a few if any emerging micropollutants such as PPCPs or EDCs in model development.

With a special focus on the emerging micropollutants, the objective of this study was therefore to develop a robust QSPR model for predicting  $k_{OH}$  values of a wide range of micropollutants. A large number of micropollutants with diverse structures including many EDCs and PPCPs were selected for model development and their rate constants collected from the literature. This data set was then split into a training set and a validation set. The training set was used to calibrate the model which was then externally validated using the validation set. In addition, the applicability domain of the model was defined by a leverage approach. Using the applicability domain it can be determined whether the model is applicable to a new, unknown compound. This overall approach

ensured that the developed models were applicable to micropollutants with diverse structures and a wide range of  $k_{OH}$  values. The developed model gives an indication of the susceptibility of micropollutants to react with hydroxyl radicals by providing their  $k_{OH}$  values. These rate constants are needed to predict removals in natural waters and the developed model will therefore be helpful in assessing the efficiency of AOP technologies with respect to the degradation of micropollutants.

## 2. Materials and methods

### 2.1. Data set

A total of 118 emerging micropollutants were used for developing the QSPR models in this study, in which  $k_{OH}$  values of 22 micropollutants were determined experimentally in a previous study using competition kinetics (Jin et al., 2012), and the  $k_{OH}$  values of the other 96 micropollutants were collected from the literature. Micropollutants included in this study were very heterogeneous in structure and included a number of chemical classes (e.g. phenols, polycyclic aromatic hydrocarbons, alkanes, halogenated aromatic compounds, organophosphorus compounds, etc.) thus covering a wide spectrum of physico-chemical properties. The micropollutants included in this study and their  $k_{OH}$  values are provided in Table 1. The  $k_{OH}$  values range from  $5.4 \times 10^7$  ( $M^{-1} s^{-1}$ ) to  $1.7 \times 10^{10}$  ( $M^{-1} s^{-1}$ ). The total compound set was divided into the training set and validation set through the property sampling method (Leonard and Roy, 2006). This was accomplished by sorting the micropollutants according to their descending  $k_{OH}$  values, then taking one compound out of every four compounds. Compounds taken out were used as the validation set, and the remaining compounds were used as the training set. As a result, about 25% of the total data set was used for the validation set ( $n = 89$  for the training set, and  $n = 29$  for the validation set).

A large number of molecular descriptors were calculated using DRAGON software (Milano Chemometrics and QSAR Research Group, University of Milano-Bicocca, Milan, Italy), and these descriptors were then used as independent variables for modeling. The chemical name or registration number was used to search the SMILES code of the chemical structure from the ChemIDplus Advanced online database (United States National Library of Medicine). The SMILES code of the chemical structure was then used as input for the software DRAGON to generate the molecular descriptors. As a result, 951 descriptors including constitutional descriptors, topological descriptors, connectivity indices, information indices, 2D autocorrelations, eigenvalue-based indices, 3D MoRSE descriptors, WHIM descriptors, molecular properties, functional group counts, and atom-centered fragments, etc. were calculated. A list of the DRAGON descriptors is available (Todeschini et al., 2005). Most of these descriptors are reviewed in a textbook by Todeschini and Consonni (2000). The correctness of the SMILES code was then validated by comparison of the molecular weights reported in the databases with those calculated by the software. To minimize the redundant information, descriptors with constant values among micropollutants ( $n = 142$ , mostly functional group counts descriptors) were removed, and descriptors found to be pairwise correlated by greater than 95% were excluded ( $n = 110$ ). Hence, 699 descriptors were used for QSPR modeling.

### 2.2. QSPR modeling

Multiple linear regression (MLR) was used in this study to identify a linear relationship between  $k_{OH}$  and the above described set of molecular descriptors. MLR is among the most widely used

Download English Version:

<https://daneshyari.com/en/article/6307401>

Download Persian Version:

<https://daneshyari.com/article/6307401>

[Daneshyari.com](https://daneshyari.com)