Chemosphere 92 (2013) 1029–1034

Contents lists available at SciVerse ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Development of a model for predicting reaction rate constants of organic chemicals with ozone at different temperatures

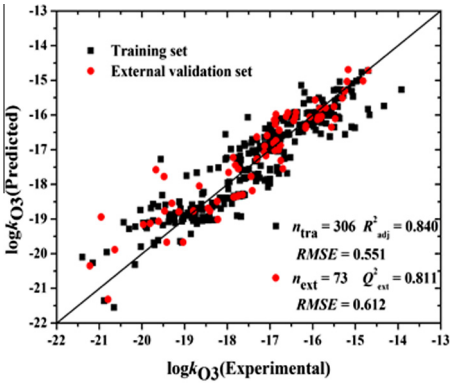
Xuehua Li *, Wenxing Zhao, Jing Li, Jingqiu Jiang, Jianji Chen, Jingwen Chen

Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), School of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian 116024, China

HIGHLIGHTS

- A PLS model was developed for reaction rate constants with tropospheric ozone (k_{O3}).
- The model can be used to predict k_{O3} over the temperature range 178–409 K.
- The functional groups cover C=C, C≡C, —OH, —CHO, —O—, >C=O, —COOH, —NH—, —NO_2, —X.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 19 December 2012
Received in revised form 6 March 2013
Accepted 20 March 2013
Available online 16 April 2013

Keywords:
Ozone
Reaction rate constant
Quantitative structure–activity relationship
Temperature dependence
Partial least-squares (PLSs)

ABSTRACT

To assess the persistence and fate of volatile organic compounds in the troposphere, the rate constants for the reaction with ozone (k_{O3}) are needed. As k_{O3} values are only available for hundreds of compounds, and experimental determination of k_{O3} is costly and time-consuming, it is of importance to develop predictive models on k_{O3}. In this study, a total of 379 logk_{O3} values at different temperatures were used to develop and validate a model for the prediction of k_{O3}, based on quantum chemical descriptors, Dragon descriptors and structural fragments. Molecular descriptors were screened by stepwise multiple linear regression, and the model was constructed by partial least-squares regression. The cross validation coefficient Q^2_{CUM} of the model is 0.836, and the external validation coefficient Q^2_{ext} is 0.811, indicating that the model has high robustness and good predictive performance. The most significant descriptor explaining logk_{O3} is the *BELm2* descriptor with connectivity information weighted atomic masses. k_{O3} increases with increasing *BELm2*, and decreases with increasing ionization potential. The applicability domain of the proposed model was visualized by the Williams plot. The developed model can be used to predict k_{O3} at different temperatures for a wide range of organic chemicals, including alkenes, cycloalkenes, haloalkenes, alkynes, oxygen-containing compounds, nitrogen-containing compounds (except primary amines) and aromatic compounds.

1. Introduction

Large numbers of synthetic volatile organic chemicals (VOCs) have been emitted into the troposphere. These organic chemicals are expected to be removed by chemical degradation processes such as photolysis and reactions with tropospheric oxidants (O_3, **.**OH and **.**NO_3 radicals) (Gramatica et al., 2003a; Ahn et al., 2006; Picquet-Varrault et al., 2010). Ozone is of fundamental importance for the chemical reactions in the troposphere. The reaction of VOCs with O_3 is a significant pathway determining their fate in the troposphere (Gramatica et al., 1999; Avzianova and Ariya, 2002;

* Corresponding author. Tel./fax: +86 411 8470 7844.
E-mail address: lixuehua@dlut.edu.cn (X. Li).

0045-6535/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chemosphere.2013.03.040

Fatemi, 2006). Therefore, the rate constants for the reaction with ozone (k_{O3}) are needed to assess the atmospheric fate and persistence of VOCs.

The kinetics of the reaction with O_3 has been investigated as a function of temperature (Winterhalter et al., 2009; Kim et al., 2011; Leather et al., 2011). k_{O3} shows a positive temperature dependence (Leather et al., 2010, 2011; Kim et al., 2011), and accurate k_{O3} values at different temperatures are required for the persistence assessment under atmospheric conditions (Khamaganov and Hites, 2001). However, experimental k_{O3} values are available only for hundreds of compounds (Atkinson and Carter, 1984; Grosjean et al., 1993; McGillen et al., 2008; Kim et al., 2011). In addition, the 'benign by design' concept requires information on compounds' degradability to be available at an early stage, even before synthesis (Ruecker and Kuemmerer, 2012). Quantitative structure–activity relationship (QSAR) models are successful in predicting reaction kinetic parameters only from molecular structural information. Since QSAR models are cost-effective and independent of authentic chemical standards, they are crucial for persistence assessment of existing and new chemicals.

Several QSAR models have been developed for k_{O3}. Two multiple linear regression (MLR) models were developed to predict the reactivity of heterogeneous chemicals using six molecular descriptors (Pompe and Veber, 2001; Gramatica et al., 2003b). The model of Gramatica et al. has a good predictive ability in both internal and external validation. Fatemi (2006) developed a nonlinear QSAR model to predict k_{O3} by using artificial neural networks (ANN) and six molecular descriptors. Later on, the projection pursuit regression (PPR) model with seven molecular descriptors was developed and validated (Ren et al., 2007). Two nonlinear models (Fatemi, 2006; Ren et al., 2007) for diverse compounds have a sufficient goodness-of-fit and great predictive ability, but the algorithms are not transparent (Cronin and Schultz, 2001). In addition, two models are reported to predict k_{O3} for aliphatic compounds. For aliphatic alkenes and dialkenes, based on a summation of the inductive and steric effects, a prediction model of the gas-phase ozonolysis was developed by McGillen et al. (2008). The model has good predictive ability for aliphatic olefins. For aliphatic compounds, a support vector machine (SVM) model containing three descriptors also has great predictive power (Yu et al., 2012). Virtually, all of the previous QSAR models for k_{O3} have been limited to room-temperature predictions. Furthermore, some models have a narrow applicability domain and are only applicable for specific chemical classes.

In this study, using partial least-squares (PLSs) regression, a new QSAR model with wide applicability domain and temperature dependence was developed for the prediction of k_{O3}. The proposed model, based on 379 $\log k_{O3}$ values for 166 organic compounds at different temperatures, was evaluated by an internal and external validation procedure. The leverage approach was used to characterize the applicability domain of the model.

2. Materials and methods

2.1. Data set

Experimental k_{O3} values (Table S1, reported in $cm^3\ s^{-1}$molecule^{-1}) were collected from the literature (Atkinson and Carter, 1984; Grosjean et al., 1993; Grosjean and Grosjean, 1994, 1995, 1996; Grosjean et al., 1996; Atkinson, 1997; McGillen et al., 2008; Leather et al., 2011) and EPI Suite™ (Version 4.0) (http://www.epa.gov/oppt/exposure). The dataset comprises 379 $\log k_{O3}$ values for 166 organic compounds (including alkenes, cycloalkenes, haloalkenes, alkynes, oxygen-containing compounds, nitrogen-containing compounds and aromatic compounds) at different temperatures (178–409 K). The data were randomly divided into a training set and a validation set with a ratio of 4:1. The number of compounds, statistics of the $\log k_{O3}$ values and temperature range are listed in Table 1. Fig. 1 shows the normal distribution of $\log k_{O3}$ values with mean of −17.24 and standard deviation (SD) of 1.44 for the total dataset.

2.2. Calculation of molecular structural descriptors

Three classes of molecular structural descriptors were considered in the model development, including 14 quantum chemical descriptors, 18 molecular fragments and 18 categories of Dragon descriptors (Todeschini and Consonni, 2000). All the descriptors are listed in Table S2. The quantum chemical descriptors were calculated by the PM6 method in MOPAC 2009 (Fatemi, 2006; Puzyn et al., 2008; Wang et al., 2009). The following keywords were used for optimization: EF, GNORM = 0.100, MMOK, GEO-OK, PM6, MULLIK, GRAPH, ESR, HYPERFINE and POLAR. The Dragon descriptors were calculated by the DRAGON software (Version 2.0) (http://www.disat.unimid.it/chm/moldes.html). The descriptor values of all the compounds are listed in Table S3. In addition, $1/T$ was included as a predictor variable on the basis of relationship of $\log k_{O3}$ and temperature (T) in the Arrhenius equation.

2.3. Development of the model

MLR and PLS analysis methods were used to select optimal descriptors and build QSAR models. The following procedures were followed:

(1) Stepwise MLR analysis was employed to select the significant descriptors by using the SPSS software (SPSS 13.0). The MLR model was obtained with the maximum adjusted determination coefficient (R^2_{adj}) (OECD, 2007) and with each descriptor having the variable inflation factor (VIF) < 10 (Norusis, 1997). R^2_{adj} is defined as:

$$R^2_{adj} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p-1)}{\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)} \quad (1)$$

where $\hat{y}_i$ and y_i are the predicted value and observed value for the ith compound, respectively; $\bar{y}$ is the response mean; n is the number of the objects and p is the number of descriptors. The selected descriptors in the MLR model were used in the next step.

(2) PLS regression analysis was used to eliminate redundant descriptors manually and build the optimal model by using the Simca software (Simca 6.0, Umetri AB & Erisoft AB). The conditions of the computation were: cross validation rounds = 7, maximum iteration = 200, missing data tolerance = 50% and significance level limit = 0.05. The optimal model was obtained with the maximum R^2_{adj} and cumulative cross-validation coefficient (Q^2_{CUM}) that is the cumulative variance of the dependent variable that can be explained by the extracted PLS components (Ding et al., 2006; Wang et al., 2009). Q^2_{CUM} describes the robustness of a PLS model. Generally, $Q^2_{CUM} > 0.5$ can be regarded as good and $Q^2_{CUM} > 0.9$ as excellent (Eriksson et al., 2003).

2.4. Model validation

The developed model was evaluated by external validation. The external predictive squared correlation coefficients (Q^2_{ext-1} (Schüürmann et al., 2008) and Q^2_{ext-2} (Consonni et al., 2010)) were used along with the root-mean-square error ($RMSE$) to evaluate the predictive capability of the model. The parameters are calculated as: