Chemosphere 154 (2016) 23–33

Contents lists available at ScienceDirect

Chemosphere

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

A QSAR for the prediction of rate constants for the reaction of VOCs with nitrate radicals

Michael Schindler

Bayer CropScience AG, Building 6692, Alfred Nobel Str. 50, D-40789 Monheim, Germany

HIGHLIGHTS

- A new QSAR predicting degradation of VOCs by NO_3 radicals was developed.
- Its extended applicability domain includes rate constants of (di)hydroxybenzenes.
- The QSAR uses a diversity-based work set ($n = 99$) representing $n = 258$ compounds.

ARTICLE INFO

Article history:
Received 18 February 2016
Received in revised form
16 March 2016
Accepted 20 March 2016

Handling Editor: Ralf Ebinghaus

Keywords:
QSAR
PLS
Nitrate degradation
Applicability domain
External validation
Diversity selection

ABSTRACT

A QSAR for the prediction of rate constants for the degradation of volatile organic compounds by nitrate radicals is developed using the Partial Least Squares technique. The QSAR is based on experimental data published in the literature for 260 compounds. They are modeled by a set of calculated descriptors from standard descriptor generation tools and from quantum chemistry.

Out of several diversity-based partitionings of the data set a diverse set of 99 compounds turned out to be the optimum choice with regard to simplicity and performance. The final QSAR model is characterized by $r^2 = 0.831$ (fit) and $q^2 = 0.823$ (prediction), and by an $r^2_{pred} = 0.862$ for the $n = 155$ external validation set. The QSAR needs 3 latent variables. The most important descriptors for the QSAR are the ionization potential, obtained from density functional theory, and the energy of the highest occupied molecular orbital, which are modulated by fingerprints indicating the presence of specific molecular fragments like functional groups or ring systems.

The applicability domain of the new QSAR was studied for some compound classes which are important for the crop protection industry, including (di)hydroxbenzenes and heterocyclic compounds.

1. Introduction

The lifetime of Volatile Organic Compounds (VOCs) in the troposphere depends on their degradation either by direct phototransformation or by indirect processes, such as their reaction with radicals. It is part of the process of approval of plant protection products to estimate the lifetime in air in order to assess the potential for long range transport (Regulation 1107/2009 EC). Therefore methodology is required to support the determination of the half-life of chemicals in the troposphere.

Degradation of organic chemicals in the troposphere can proceed via several mechanisms (Monks, 2005). Actually, the most important degradation route is via reaction with hydroxyl radicals, predominantly occurring during daytime. In comparison, the reaction with the (photochemically unstable) nitrate radicals occurs at night (Brown and Stutz, 2012; Khan et al., 2015).

The reaction rates of hydrocarbons with NO_3 radicals are approximately 30 times lower than with OH radicals, however, this may well be compensated by the increased concentration of NO_3 radicals at night-time: their concentration of 5.0×10^8 molecules cm^{-3} is two orders of magnitude larger than the OH radical concentration during the day-time (2.0×10^6 molecules cm^{-3}) (Ball, 2015). It has been found that NO_3 contributed 28% to the overall VOC initiated oxidation in the continental boundary layer over Europe (Geyer et al., 2001). Especially for phenols the reaction with NO_3 radical reactions appears to be an important degradation route (Atkinson et al., 1992; Carter et al., 1981; Thüner et al., 2004).

As the experimental derivation of reaction rates of VOCs in the troposphere is demanding, they are available only for a limited

---

E-mail address: michael.schindler@bayer.com.

http://dx.doi.org/10.1016/j.chemosphere.2016.03.096
0045-6535/

number of chemicals, mainly for the degradation by hydroxyl radicals. The number of experimentally obtained rate constants for NO_3 radical reactions is even less.

Therefore methods have been sought to use the experimental data available to predict rate constants for compounds whose rate constants have not been measured. Quantitative Structure-Activity Relations (QSARs) provide mathematical relationships between experimental observations and some molecular properties (descriptors) of the compounds under investigation. While it might be difficult to obtain the required experimental data, often sets of descriptors used to characterize the compounds are easily available. After having established a QSAR, it can be used to predict the rates for a set of new substances, thus avoiding potentially huge efforts of new experiments.

Based on the pioneering work of Atkinson (see Atkinson (1997) for a review), the half-life of a chemical in air by the main degradation routes, namely by reaction with OH radicals and ozone, can be predicted by QSARs such as those of the Atmospheric Oxidation Program for Microsoft Windows (AOPWIN) (Kwok and Atkinson, 1995), developed and maintained by the Syracuse Research Corporation (SRC) and the US EPA.

Only relatively few publications deal with predicting rate constants for the degradation of VOCs by nitrate radicals (see Mamy et al. (2015) for a comprehensive review). As of today there exists no equally accepted QSAR for the reaction with NO_3 radicals like AOPWIN for hydroxyl radicals.

Different theoretical approaches have been used to develop QSARs for special compound classes like alkylnaphthalenes (Long and Niu, 2007), broad ranges of VOCs (Kerdouci et al., 2010; Yu et al., 2014) or QSPRs (Quantitative Structure Property Relations) describing the tropospheric night-time degradation of 166 VOCs by reaction with nitrate radicals (Papa and Gramatica, 2008). The latter model was developed following the OECD principles for regulatory acceptance of QSARs (OECD, 2008).

In general, the main limitation of a QSAR arises from the fact that its applicability domain (AD) is restricted the types and number of compounds and by the property space spanned by the descriptors utilized when building the QSAR. This might become a source of error when uncritically using QSARs available in the literature. Frequently a check of the AD is impossible due to missing information. As an example, the AD of AOPWIN is not known as "The complete training sets for AOPWIN's estimation methodology are not available" (AOPWIN Help).

The objective of this work was to establish a new QSAR model for the prediction of rate constants $k(NO_3)$ for the degradation of VOCs by NO_3 radicals with an extended AD, as the ADs of the QSARs published previously do not cover some of the compound classes like substituted phenols which are of interest to the crop protection industry.

2. Material and methods

According to official guidance (OECD, 2007; EFSA, 2013; JRC, 2010; JRC, 2011; NAFTA, 2012), necessary prerequisites for developing a QSAR are:

- A set of compounds characterized by unambiguously defined molecular structures
- A consistent set of experimentally determined data,
- Descriptor sets describing the properties of the full set of compounds
- Software that is able to perform the steps from structure generation via descriptor calculation to QSAR construction and usage.

2.1. Sets of experimental data

The starting point for the new model was the QSPR data set ($n = 166$) used by Papa and Gramatica (2008),[1] augmented by IUPAC data (Atkinson et al., 2006). Experimental data from additional studies on (di)hydroxybenzenes (Olariu et al., 2004; Thüner et al., 2004), terpenes (Suh et al., 2001; Harrison and Ham, 2009, 2010; Ham, 2013), cyclo-alkenes (Stewart et al., 2013), cyclic ethers (Kind et al., 1996; Gai et al., 2011), acrylates (Wang et al., 2010), aldehydes (Martínez et al., 2002), pyrethroids (Wang et al., 2013) and effective 1st order rate constants for fenvalerate, phenothrin and resmethrin aerosols (Wang et al., 2013), although they lead to atmospheric lifetimes which differ greatly from those one would obtain by ignoring the complicated processes involving aerosols and assuming simple 1st order kinetics. This led to a set of 260 compounds which are listed in Appendix B.

Two data sets available in the literature had to be discarded: (1) alkylnaphthalenes were modeled by Long and Niu (2007) referring to experimental data from Phousongphouang and Arey (2003). These authors derived rate constants for reactions of order three, which are out of scope as we consider first-order reaction rates only; (2) rate constants for aromatic alcohols in solution (Barzaghi and Herrmann, 2004) do not correlate with those obtained in the gas phase. They, too, had to be excluded.

All rate constants were converted to their negative logarithms, pk(NO3). This is suggested by the linear free energy relationship (LFER) between the equilibrium constant K of a chemical reaction and its Free Energy ΔG, where K is defined as a quotient of rate constants $k_{\rightarrow}$ and $k_{\leftarrow}$ for the forward and backward reactions, and R and T are the gas constant and the absolute temperature, respectively

$$\Delta G = -RT\ln K = -RT\ln \frac{k_{\rightarrow}}{k_{\leftarrow}}$$

The vast majority of the $pk(NO_3)$s is in the range of 10–16, a set of small molecules (e.g., methane, ethane, methyl- and ethyl-formate, ethyl-acetate and dichloromethane) exceeds this range by up to two orders of magnitude.

2.2. Choice of descriptors

2.2.1. Tropospheric NO_3 chemistry

The initial choice of appropriate descriptors and the final physical interpretation of a QSAR is facilitated by information on the chemistry of nitrate radical reactions in the troposphere (Brown and Stutz, 2012; Khan et al., 2015; Mamy et al., 2015).

Nitrate radicals are formed by the oxidation of NO_2 by ozone

$$NO_2 + O_3 \rightarrow NO_3 + O_2$$

and by the equilibrium reaction involving dinitrogen pentoxide and a partner M

$$N_2O_5 + M \rightleftharpoons NO_3 + NO_2 + M$$

Degradation of VOCs is initiated by abstraction reactions that are observed for alkanes and aldehydes

$$NO_3 + RH \rightarrow HNO_3 + R,$$

---

[1] Two entries were removed from the data set: capronaldehyde (identical to hexanal), and Ocimene, (as *cis*- and *trans*-Ocimene are already present).