Ecotoxicology and Environmental Safety 133 (2016) 390–394

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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

Odor threshold prediction by means of the Monte Carlo method

Andrey A. Toropov [a,*], Alla P. Toropova [a], Luigi Cappellini [b], Emilio Benfenati [a], Enrico Davoli [b]

[a] IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Department of Environmental Health Sciences, Laboratory of Environmental Chemistry and Toxicology, Milan, Italy
[b] IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Department of Environmental Health Sciences, Laboratory of Mass Spectrometry, Milan, Italy

ARTICLE INFO

Article history:
Received 28 June 2016
Received in revised form
25 July 2016
Accepted 27 July 2016
Available online 6 August 2016

Keywords:
QSPR/QSAR
Odor threshold
Monte Carlo method
Optimal descriptor
CORAL software

ABSTRACT

A large set of organic compounds (n=906) has been used as a basis to build up a model for the odor threshold (mg/m^3). The statistical characteristics of the best model are the following: n=523, r^2=0.647, RMSE=1.18 (training set); n=191, r^2=0.610, RMSE=1.03, (calibration set); and n=192, r^2=0.686, RMSE=1.06 (validation set). A mechanistic interpretation of the model is presented as the lists of statistical promoters of the increase and decrease in the odor threshold.

1. Introduction

The sense of smell and taste are generally referred as the chemical senses, as they give information about the chemistry of the environment. From an evolutionary point of view, chemoreception is generally thought to be one of the most primitive senses that was developed (Nei et al., 2008; Niimura, 2012). The potential of chemicals to impact the human olfactory system and to cause apparent health effects has been described (see for example Rosenkran and Cunningham, 2003; Orzi et al., 2015).

Coding and processing steps of odor receptor activation system, are in the olfactory bulb, one of the limbic brain structures where part of the autonomic processes are regulated. Although the sense of smell is a conscious perception, sensory information from the outer environment are brought by sensory neurons centrally, where they constantly and unconsciously modulate also the activity of the motor neurons of the autonomic nervous system. Our response to odors from the external environment is both voluntary and involuntary and environments quality can be highly degraded by odor pollution events, causing annoyance and difficulties in public health management. Measuring odor is fundamental to predict environmental effects on the population, and perception, exposure-response relationships, play a fundamental role for annoyance in the long-time exposure pathways, making it enable for risk assessment in this field.

Odor intensity measurements are difficult due to the fact that presently no analytical instrument is available to measure objectively odor intensity. Sensorial analysis, even performed by trained assessors, are problematic as there are no fixed reference points and no easy comparison (Sell and Pybus, 2006) and odor thresholds (the lowest concentration of an odorant that is perceivable by human nose) are usually measured (Hoshika et al., 1993) to describe the odor potency of a specific molecule.

The odor threshold is an important characteristic of a compound from the ecologic and biochemical viewpoints too (Hobbs et al., 2001; Zahn et al., 2001; Yan et al., 2015; Hansen et al., 2016). Volatile organic compounds are air pollutants that account for a substantial proportion of total pollutant concentrations (Kumar et al., 2014; Brodzik et al., 2014; Yan et al., 2015). Most of these compounds can easily cause pungent sensations even on very low concentration (Wu et al., 2015; Abraham et al., 2012, 2016) and their environmental presence are complex to measure (Bianchi et al., 2013) and difficult to describe (Capelli et al., 2012a).

Because of the interaction between these odorants, mixtures of many negligible odor pollutants can generate a stronger odor impact (Le Berre et al., 2008). The odor pollution induced by these substances lowers the quality of life (Palmiotto et al., 2014; Yan et al., 2015), causing potential threats to human health both from a toxicological point of view (Capelli et al., 2012b) and from a stress-related, psychosocial effect (Blanes-Vidal, 2014).

A substance with low odor threshold in the environment can be associated to a number of odors. Thus, the ambiguous nature of

* Corresponding author.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

http://dx.doi.org/10.1016/j.ecoenv.2016.07.039
0147-6513/

odorant receptors along with various characteristics of olfactory data has stimulated search for information about threshold data for odor of various compounds which have applications in the field of bioscience, food chemistry and environmental pollution (Pal et al., 2014).

Fundamental research has well shown the effect of structural changes of odorant compounds to their odor threshold values (Takeoka et al., 1995, 1996). For instance, pentyl acetate has the lowest odor threshold of all the straight chain acetates; however, the addition of a methyl group in the 1-position caused a 20-fold increase in the odor threshold. Systematic studies to define relationships between structure and odor thresholds have been carried out on homologous series and isomers of substance groups (for example see Boelens and van Gemert, 1986). Estimation of such effects is a complex task: to determine which compounds contribute to the overall apple aroma, long-term analytical methods must be used (ASTM, E679 2004). In addition, this estimation involves measurement of the compound's odor threshold in water (Teranishi et al., 1987).

The odor threshold value is defined by many factors including the temperature, pressure, presence / absence of other substances and so on. Extraction of all these data is a time-consuming technical problem. Consequently, development of computational models for the odor threshold is an attractive alternative. Usually, such models are based on quantitative structure – property/activity relationships (QSPRs/QSARs).

The QSAR models can also assist in the detection of potential odorant components from large databases, which reduces the need for time-consuming synthesis and testing a large number of compounds. Before any substance (pharmaceutical, cosmetic, chemical, etc.) can be brought into the European market, its safety to human health and the environment must be evaluated. The QSPR/QSAR paradigm is now supported by the Registration, Evaluation, and Authorization of Chemicals (EC regulation, 1907/2006; EU Regulation, 1223/2009), a legislative initiative of the European Commission and the Organization for Economic Cooperation and Development (OECD, 2007).

The QSPR/QSAR models are used by the Food and Drug Administration (FDA) (Benigni and Zito, 2004; Valerio Jr, 2011) for minimizing the rate of false negatives and false positives saving incalculable costs for manufacturers. The Council for International Organizations of Medical Sciences (CIOMS) (Bankowski and Howard-Jones, 1986) also recommends the QSPR/QSAR methods before animal experiments for the advancement of biological knowledge.

The predictive model for the odor threshold can find applications in many practical aspects: (i) perfume manufacture; (ii) medicine; (iii) chemical technology; (iv) drug discovery and (v) regulation. There are attempts to build up QSPR/QSAR for the odor threshold (Xu et al., 2012; Pal et al., 2014; Polster and Schieberle, 2015). The aim of present work was building up a predictive model for the odor threshold (mg/m^3) using the Monte Carlo method available via the CORAL software (CORAL, 2016).

2. Method

2.1. Data

The numerical data on the odor threshold (OT) were taken from a large database available (van Gemert, 1999). Data distribution does not appear to be unimodal, symmetric. For several odorants, several OT were available with large differences from average, median and mode values. For example for hydrogen sulfide values of 0.23, 0.0042 and 0.012 mg/m^3 (average, median and mode respectively) are obtained indicating not only large OT ranges, but also asymmetrical distributions. In the present study the average OT values have been used, a classical approach for OT description (see for example Chemoreception: In Ref. Sell and Pybus, 2006).

Three splits, into training, calibration, and validation sets were examined.

2.2. Optimal descriptors

Optimal descriptors (Toropov and Toropova, 2014, 2015a, 2015b; Toropova and Toropov, 2013, 2014) which are involved to build up the QSAR model for the *pOT*, (minus decimal logarithm of odor threshold) are the following:

$$DCW(T^*, N^*) = \sum CW(S_k) + \sum CW(SS_k) + \sum CW(SSS_k) + CW(BOND) + CW(NOSP) + CW(HALO) + CW(PAIR) \quad (1)$$

In Eq. (1), the s_k, ss_k, and sss_k are combinations of one, two, and three "SMILES atoms". The "SMILES atom" is a fragment of the SMILES notation, which contains one symbol or two symbols, which cannot be examined separately (e.g. 'Cl', 'Br', etc.). The *CW(s_k), CW(ss_k),* and *CW(sss_k)* are correlation weights of the above-mentioned "SMILES atoms". The correlation weights are coefficients, which are used to calculate the descriptor. The numerical data for the correlation weights are obtained by the Monte Carlo method optimization procedure, which gives maximum correlation coefficient between endpoint and the optimal descriptor. The BOND, NOSP, HALO, and PAIR are global attributes of SMILES which reflects the presence of various kinds of chemical bonds (BOND); the presence of nitrogen, oxygen, sulfur, and phosphorus (NOSP); the presence of halogens, i.e. fluorine, chlorine, bromine, and iodine (HALO); and presence of various combinations of SMILES atoms. The *CW(BOND), CW(NOSP),* and *CW(HALO),* and *CW(PAIR)* are correlation weights of the global attributes of SMILES. The detailed description of the above- listed local (s_k, ss_k, and sss_k) and global (BOND, NOSP, HALO, and PAIR) attributes of SMILES is available in the literature (Toropova et al., 2015) as well as at web site of the CORAL software (CORAL, 2016).

The T is the threshold, i.e. a coefficient used to classify SMILES attributes into two classes (i) rare or noise; and (ii) active. The rare attributes are blocked (their correlation weights are fixed zero). The coefficient can be 1, 2, …, M. The T* is threshold which gives preferable statistical quality of the model for the calibration set. The N is the number of epochs of the Monte Carlo optimization. The N* is the number which gives preferable statistical quality for the calibration set. The T* and N* are calculated according to scheme suggested in works (Toropova et al., 2015). Having the numerical data for the correlation weights, one can calculate the *DCW(T*,N*)* for the training set and define regression parameters C_0 and C_1 for the following model

$$pOT = C_0 + C_1 \times DCW(T^*, N^*) \quad (2)$$

The predictive potential of the model calculated with Eq. (2) should be checked up with external validation set (Toropova et al., 2015).

3. Results and discussion

3.1. QSAR models

The Monte Carlo optimization with T* and N* which are selected according to scheme suggested in work (Toropova et al., 2015) gives the following models:

Please cite this article as: Toropov, A.A., et al., Odor threshold prediction by means of the Monte Carlo method. Ecotoxicol. Environ. Saf. (2016), http://dx.doi.org/10.1016/j.ecoenv.2016.07.039