Odor Threshold prediction by means of the Monte Carlo method

Andrey A. Toropov[1*], Alla P. Toropova[1], Luigi Cappellini[2], Emilio Benfenati[1], Enrico Davoli[2]

[1]*IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy, Department of Environmental Health Sciences, Laboratory of Environmental Chemistry and Toxicology*

[2]*IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy, Department of Environmental Health Sciences, Laboratory of Mass Spectrometry*

Published version of this paper could be find here https://doi.org/10.1016/j.cplett.2016.08.018

Ecotoxicology and Environmental Safety, Volume 133, November 2016, Pages 390-394

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.

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

*Corresponding author:
Andrey A. Toropov
Laboratory of Environmental Chemistry and Toxicology,
IRCCS - Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy
Tel: +39 02 3901 4595
Fax: +3902 3901 4735
E-mail: andrey.toropov@marionegri.it

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 was detected (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 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 should be utilized (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,b; 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:

$$pOT = -5.5431(\pm 0.0035) + 0.1418(\pm 0.0001) * DCW(1,22) \quad (3)$$

$$pOT = -5.4618(\pm 0.0036) + 0.1331(\pm 0.0001) * DCW(1,17) \quad (4)$$

$$pOT = -5.5541(\pm 0.0033) + 0.1527(\pm 0.0001) * DCW(1,28) \quad (5)$$

Table 1 contains the statistical characteristics of the models for *pOT* calculated with Eqs. 3-5

[Table 1 around here]

3.2. Mechanistic interpretation

Table 2 contains the list of molecular features, which are statistically stable promoters of increase or decrease of the *pOT*. These data selected according to the following principles: (i) molecular features extracted from SMILES with large prevalence in the training and calibration sets; (ii) molecular

features which have positive correlation weights for all three runs of the Monte Carlo optimization; and (iii) molecular features which have negative correlation weights for all three runs of the Monte Carlo optimization.

One can see, there are stable promoters of the *pOT* increase related to all distributions. Table 3 contains interpretation of SMILES attributes as a molecular features.

3.3. Domain of applicability

Building up a model can be accompanied by probabilistic characteristics of the influence of the parameters and/or conditions upon endpoint (Toropova et al., 2015).

An example of the probabilistic parameter is a statistical defect of an attribute. The defect can be calculated as the following:

$$defect(A_k) = \frac{P_T(A_k) - P_C(A_k)}{N_T(A_k) + N_C(A_k)} \quad (6)$$

where $P_T(A_k)$ and $P_C(A_k)$ are probabilities of attribute A_k in the training and the calibration sets, respectively; $N_T(A_k)$ and $N_C(A_k)$ are prevalence of the attribute A_k in the training and the calibration sets, respectively. It should be noted if $N_C(A_k) = 0$ then the $defect_k = 1$.

The criterion to define the domain of applicability is the following: SMILES falls into domain of applicability if defect (SMILES) calculated with Eq. 7 is less than average value of defects of SMILES over the training set (inequality 8).

$$defect(SMILES) = \sum defect(A_k) \quad (7)$$

$$defect(SMILES) < 2 \times \overline{defect(SMILES)} \quad (8)$$

The sum of *defect(SMILES)* is a measure of a distribution defect (DD):

$$DD = \sum_{Train\&Calib} defect(SMILES) \quad (9)$$

The suggested criteria (Eqs. 6-9) are calculated with data on the training and calibration sets without any data on the external validation set. Our hypothesis is: if the first distribution D1 is characterized by the distribution defect = DD1 and distribution D2 is characterized by the distribution defect = DD2, then the distribution D1 is better than distribution D2 if DD1 is smaller than DD2.

The comparison of models from the literature (Pal et al., 2014; Cliff et al., 2011; Xu et al., 2012; Hansen et al., 2016) shows that the CORAL software can be a satisfactory and useful tool to predict the odor thresholds for a large variety of organic compounds (Table 4).

Supplementary materials section contains screenshots of the models calculated with the CORAL software (http://www.insilico.eu/coral). Technical details related to total data set (i.e. three

distributions into the training, calibration, and validation sets examined in this work as well as experimental and calculated values of odor threshold) are available on the request.

4. Conclusions

The suggested models for odor threshold have good statistical quality. It must be stressed that odor measurements are complex, the odor sensitivity in population is lognormally distributed, and the database used is comprehensive of all data in literature. The additional checking up of the approach by means of comparison of three different distributions into the training, calibration, and validation sets confirms the reproduction of the statistical quality for the three random events: i.e. building up the QSAR model for random splits. The suggested models have mechanistic interpretation via groups of promoters of increase/decrease for *pOT*. Finally, the special criteria for domain of applicability and quality of distribution into the training and validation sets are defined for the approach. Thus, the approach gives models in accordance with OECD principles (OECD, 2007).

Software

The CORAL software utilized in this work is free available on the Internet (http://www.insilico.eu/coral).

Acknowledgments

Authors thank the EC project PeptiCAPS (Project reference: 686141) for financial support.

References

Abraham, M.H., Sanchez-Moreno, R., Cometto-Muñiz, J.E., Cain, W.S., 2012. An algorithm for 353 odor detection thresholds in humans. Chem. Senses 37, 207–218.

Abraham, M.H., Gola, J.M.R., Cometto-Muñiz, J.E., 2016. An assessment of air quality reflecting the chemosensory irritation impact of mixtures of volatile organic compounds. Environ. Int. 86, 84–91.

ASTM E679 2004. Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration SeriesMethod of Limits, ASTM E679-04, pp. 1–7. ASTM International,West Conshohocken, PA.

Bankowski, Z., Howard-Jones, N., 1986. International Guiding Principles for Biomedical Research Involving Animals [Committee of International Organizations of Medical Science (CIOMS), Geneva].

Benigni, R., Zito, R., 2004. The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. Mutat. Res. 566, 49–63.

Bianchi, G., Palmiotto, M., Davoli, E., 2013. Environmental odour pollution: a complex GC-MS, olfactometry and diffusion modeling approach to define air quality. Compr. Anal. Chem. 61, 455-470.

Blanes-Vidal, V., Baelum, J., Nadimi, E.S., Lofstrom, P., Christensen, L.P., 2014. Chronic exposure to odorous chemicals in residential areas and effects on human psychosocial health: dose-response relationships. Sci. Total. Environ. 490, 545-54.

Boelens, M.H., van Gemert, L.J., 1986. Physicochemical parameters related to organoleptic properties of flavour components. In: Birch, G.G., Lindley, M.G., editors. Developments in food flavours. London: Elsevier Science B.V. p. 23–49.

Brodzik, K., Faber, J., Lomankiewicz, D., Golda-Kopek, A., 2014. In-vehicle VOCs composition of unconditioned, newly produced cars. J. Environ. Sci. 26, 1052–1061.

Capelli, L., Sironi, S., Del Rosso, R., Bianchi, G., Davoli, E., 2012a. Olfactory and toxic impact of industrial odour emissions. Wat. Sci. Tech. 66, 1399-1406.

Capelli, L., Sironi, S., Del Rosso, R., Bianchi, G., Davoli, E., 2012b. Evaluating the dispersion of toxic odour emissions from complex sources. J. Environ. Sci. Health. A Tox. Hazard. Subst. Environ. Eng. 47, 1113-1122.

Cliff, M., Stanich, K., Trujillo, J.M., Toivonen, P., Forney, C.F., 2011. Determination and prediction of odor thresholds for odor active volatiles in a neutral apple juice matrix. J. Food Quality 34, 177–186.

CORAL software, 2016. http://www.insilico.eu/coral, Accessed April 29, 2016

EC regulation 1907/2006. Of the European Parliament and the council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.

EU Regulation, 1223/2009. Of The European Parliament And Of The Council of 30 November 2009 on cosmetic products. Official Journal of the European Union L 342/59-209 .

Hansen, M.J., Jonassen, K.E.N., Løkke, M.M., Adamsen, A.P.S., Feilberg, A., 2016. Multivariate prediction of odor from pig production based on in-situ measurement of odorants. Atmos. Environ. 135, 50-58.

Hoshika, Y., Imamura, T., Muto, G., Van Gemert, L.J., Don, J.A., Walpot, J.I., 1993. International comparison of odor threshold values of several odorants in Japan and in The Netherlands. Environ. Res. 61(1), 78-83.

Hobbs, P.J., Misselbrook, T.H., Dhanoa, M.S., Persaud, K.C., 2001. Development of a relationship between olfactory response and major odorants from organic wastes. J. Sci. Food Agric. 81, 188-193.

Le Berre, E., Beno, N., Ishii, A., Chabanet, C., Etievant, P., Thomas-Danguin, T., 2008. Just noticeable differences in component concentrations modify the odor quality of a blending mixture. Chem. Senses 33, 389–395.

Kumar, A., Singh, B.P., Punia, M., Singh, D., Kumar, K., Jain, V.K., 2014. Assessment of indoor air concentrations of VOCs and their associated health risks in the library of Jawaharlal Nehru University, New Delhi. Environ. Sci. Pollut. Res. Int. 21, 2240–2248.

Nei, M., Niimura, Y., Nozawa, M., 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat. Rev. Genet. 9(12), 951-63.

Niimura, Y., 2012. Olfactory receptor multigene family in vertebrates: from the viewpoint of evolutionary genomics. Curr. Genomics 13, 103-114.

OECD, 2007. OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69. Guidance Document on the Validation of (Quantitative) Structure–Activity Relationship [(Q)SAR] Models. http://search.oecd.org/officialdocuments

285 Orzi, V., Scaglia, B., Lonati, S., Riva, C., Boccasile, G., Alborali, G.L., Adani, F., 2015. The role of
286 biological processes in reducing both odor impact and pathogen content during mesophilic
287 anaerobic digestion. Sci. Total Environ. 526, 116-126.
288 Pal, P., Mitra, I., Roy, K., 2014. QSPR modeling of odor threshold of aliphatic alcohols using
289 Extended Topochemical Atom (ETA) indices. Croat. Chem. Acta 87 (1), 29–37.
290 Palmiotto, M., Fattore, E., Paiano, V., Celeste, G., Colombo, A., Davoli, E., 2014. Influence of a
291 municipal solid waste landfill in the surrounding environment: Toxicological risk and odor
292 nuisance effects. Environ. Int. 68, 16–24.
293 Polster, J., Schieberle, P., 2015. Structure−Odor correlations in homologous series of alkanethiols
294 and attempts to predict odor thresholds by 3D-QSAR studies. J. Agric. Food Chem. 63 (5),
295 1419–1432.
296 Rosenkran, H.S., Cunningham, A.R., 2003. Environmental odors and health hazards. Sci. Total
297 Environ. 313, 15–24.
298 Sell, C.S., Pybus, D.H., 2006. Chemoreception. In: The Chemistry of Fragrances: From Perfumer to
299 Consumer. 2nd Edition. RCS Publisher. X001-X002. DOI:10.1039/9781847552044-FX001
300 Takeoka, G., Buttery, R., Turnbaugh, J., Benson, M., 1995. Odor thresholds of various branched
301 esters. Lebensm. Wiss. Technol. 28, 153–156.
302 Takeoka, G., Buttery, R.G., Ling, L., 1996. Odour thresholds of various branched and straight chain
303 acetates. Lebensm.Wiss. Technol. 29, 677–680.
304 Teranishi, R., Buttery, R., Schamp, N., 1987. The significance of low threshold odor compounds in
305 aroma research. In: Flavour Science and Technology (Martens, M., Dalen, G.A., Russwurm,
306 H., eds.) pp. 515–527. Wiley, NewYork, NY.
307 Toropov, A.A., Toropova, A.P., 2014. Optimal descriptor as a translator of eclectic data into endpoint
308 prediction: Mutagenicity of fullerene as a mathematical function of conditions. Chemosphere
309 104, 262-264.
310 Toropov, A.A., Toropova, A.P., 2015a. Quasi-QSAR for mutagenic potential of multi-walled carbon-
311 nanotubes. Chemosphere 124, 40-46.
312 Toropov, A.A., Toropova, A.P., 2015b. Quasi-SMILES and nano-QFAR: United model for
313 mutagenicity of fullerene and MWCNT under different conditions. Chemosphere 139, 18-22.
314 Toropova, A.P., Toropov, A.A., 2013. Optimal descriptor as a translator of eclectic information into
315 the prediction of membrane damage by means of various TiO2 nanoparticles. Chemosphere
316 93, 2650-2655.
317 Toropova, A.P., Toropov, A.A., 2014. CORAL software: Prediction of carcinogenicity of drugs by
318 means of the Monte Carlo method. Eur. J. Pharm. Sci. 52, 21-25.

Toropova, A.P., Toropov, A.A., Benfenati, E., Leszczynska, D., Leszczynski, J., 2015. QSAR model as a random event: A case of rat toxicity. Bioorg. Med. Chem. 23 (6), 1223-1230.

Valerio Jr, L.G., 2011. *In silico* toxicology models and databases as FDA critical path initiative toolkits. Hum. Genomics. 5(3), 200–207.

van Gemert, L. J., 1999. Compilations of odour threshold values in air and water Boelens Aroma Chemical Information Service (BACIS). Groen GB Huizen. The Netherlands.

Wu, C., Liu, J., Yan, L., Chen, H., Shao, H., Meng, T., 2015. Assessment of odor activity value coefficient and odor contribution based on binary interaction effects in waste disposal plant. Atmos. Environ. 103, 231–237.

Xu, X., Luan, F., Liu, H., Gao, Y., 2012. Prediction of the odor thresholds of oxygen and nitrogen containing heterocycles using the Quantitative Structure-Property Relationship approach. Asian J. Chem. 24(9), 3842-3848.

Yan, L., Liu, J., Fang, D., 2015. Use of a Modified Vector Model for Odor Intensity Prediction of Odorant Mixtures. Sensors 15, 5697-5709.

Zahn, J.A., DiSpirito, A.A., Do, Y.S., Brooks, B.E., Cooper, E.E., Hatfield, J.L., 2001.Correlation of human olfactory responses to airborne concentrations of malodorous volatile organic compounds emitted from swine effluent. J. Environ. Qual. 30, 624-634.

Table 1

The statistical characteristics of QSAR models for odor threshold calculated with Eqs. 3-5

Split	n_{train}	r^2_{train}	q^2_{train}	s_{train}	n_{calib}	r^2_{calib}	s_{calib}	n_{valid}	r^2_{valid}	s_{valid}
1, Eq.3	523	0.647	0.645	1.18	191	0.610	1.03	192	0.686	1.06
2, Eq.4	479	0.600	0.597	1.26	214	0.667	1.00	213	0.587	1.08
3, Eq.5	523	0.644	0.642	1.15	191	0.618	1.16	192	0.664	1.01

Table 2

Promoters of increase and decrease for *pOT*

split	Attribute Ak	CW(Ak) in run 1	CW(Ak) in run 2	CW(Ak) in run 3	Frequency in training set	Frequency in calibration set	Defect(Ak)
	Promoters of *pOT* increase						
1	HALO00000000	3.58826	4.04993	3.72319	466	171	0.0000
2	HALO00000000	2.47709	2.71629	2.71376	425	190	0.0000
3	HALO00000000	2.64024	2.09750	2.39838	463	173	0.0000
1	C...C.....	0.51254	0.48620	0.26053	358	147	0.0002
2	C...C.....	0.36085	0.44673	0.15952	328	154	0.0001
3	C...C.....	0.05871	0.10788	0.12833	365	139	0.0001
1	NOSP01000000	5.32536	5.22168	5.37597	298	123	0.0002
2	NOSP01000000	5.16421	5.37228	5.15435	277	130	0.0001
3	NOSP01000000	4.29786	4.00130	4.16129	292	124	0.0002
1	++++O---B2==	1.33859	1.32759	1.12530	282	108	0.0001
2	++++O---B2==	0.52102	0.65489	0.50825	260	111	0.0001
3	++++O---B2==	1.73585	1.60045	1.50177	272	111	0.0002
	Promoters of *pOT* decrease						
1	(.....	-1.32289	-1.28659	-1.27594	369	128	0.0001
2	(.....	-0.97827	-1.12017	-1.23431	353	140	0.0002
3	(.....	-1.38277	-1.28759	-1.19665	373	133	0.0000
1	O.....	-2.14641	-2.06157	-2.14781	351	133	0.0001
2	O.....	-1.51134	-1.61069	-1.61012	321	144	0.0000
3	O.....	-1.89988	-1.83508	-1.81488	346	137	0.0001
1	1.....	-1.43485	-1.32913	-1.76430	247	65	0.0004
2	1.....	-0.69631	-0.79625	-0.70094	231	70	0.0005
3	1.....	-1.90238	-1.83671	-1.95808	247	65	0.0004
1	++++N---B2==	-3.45368	-3.42215	-3.44855	50	7	0.0010
2	++++N---B2==	-1.90281	-1.90455	-1.90284	40	9	0.0008
3	++++N---B2==	-3.29875	-3.30467	-3.30140	43	13	0.0003

Table 3

Interpretation of the promoter of increase and decrease for *pOT*

Attribute A_k	**Comment**
Promoter of *pOT* increase	
HALO00000000	Absence of F, Cl, Br
C...C.....	Presence of carbon – carbon bonds (sp^3)
NOSP01000000	Presence of oxygen together with absence of nitrogen, sulfur and phosphorus
++++O---B2==	Presence of oxygen and double bonds
Promoter of *pOT* decrease	
(.....	Branching in molecular skeleton
O.....	Presence of oxygen
1.....	Presence of rings
++++N---B2==	Presence of nitrogen and double bond

Table 4

Statistical characteristics of QSAR models for odor threshold from the literature

Reference	**Number of compounds in training set**	**Determination coefficient for training set**	**Number of compounds in validation set**	**Determination coefficient for validation set**
Cliff et al., 2011	10	0.728	-	-
Xu et al., 2012	40	0.8012-0.8767	10	0.648-0.7746
Pal et al., 2014	42	0.809	11	0.813
Hansen et al., 2016	94	0.74	21	0.77