Chemistry for Sustainable Development **17** (2009) 195–199

UDC 544.16

QSAR Modelling of Maximum Permissible Concentrations of Organic Substances in the Air of Working Area

A. A. TOROPOV[1], V. O. KUDYSHKIN[2], I. N. NURGALIEV[2], R. Z. AKHMEROV[1] and S. SH. RASHIDOVA[2]

[1]*Institute of Geology and Geophysics, Academy of Sciences of Republic of Uzbekistan, Ul. Khodzhibaeva 49, Tashkent 100041 (Uzbekistan)*

[2]*Institute of Chemistry and Physics of Polymers, Academy of Sciences of Republic Uzbekistan, Ul. A. Kadyri 76, Tashkent 100041 (Uzbekistan)*

E-mail: carbon@uzsci.net

(Received April 25, 2008; revised December 29, 2008)

Abstract

On the basis of the data of SMILES (simplified molecular input line entry system), optical descriptors were determined. It was established that they give a satisfactory forecast of the logarithms of maximum permissible concentrations of the substances that may be present in the air of the working area during production and processing of plastics.

Key words: QSAR, SMILES, descriptor, maximum permissible concentrations, processing of plastics

INTRODUCTION

During plastics processing, the air of the working area may contain hazardous substances, such as monomers, plasticizers, stabilizers, as well as the products of destruction of processed plastics. In this situation, the concentration of hazardous substances in the air of the working area should not exceed maximum permissible concentration values (MPC). The application of mathematical modelling methods to predict physicochemical and medical-biological properties of substances allows one to save resources substantially. These investigations were called QSPR/QSAR (Quantitative Structure–Property/Activity Relationships) [1–7]. We have already made an attempt to carry out modelling of MPC by means of optimization of correlation weights of local invariants of graphs that had been successfully used previously to model the activity of monomers in radical copolymerization [1, 7]. However, the application of calculations according to the algorithm similar to that described in [1, 7] did not give any positive results; no model could be constructed in that case.

Recently, the approach involving representation of the molecular structure through SMILES (Simplified Molecular Input Line Entry System) has become widespread [8]. In the present paper we report on the results of the use of descriptors calculated with SMILES for QSAR modelling of the logarithms of MPC of hazardous substances in the air of the working area during plastics processing [9].

EXPERIMENTAL

The molecular structure of the substances under consideration is shown in Table 1. The used SMILES were built up using the ChemSketch software [10].

Optimal descriptors were calculated using equation

$$DCW = PCW(SA_k) \quad (1)$$

where SA_k – is a SMILES attribute transferring indivisible information, that is, separate

TABLE 1

Molecular structures of substances used to build up the model

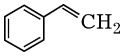
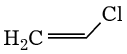
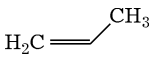
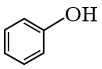
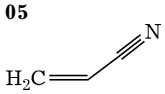
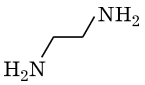
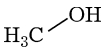
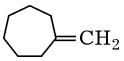
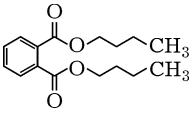
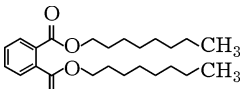
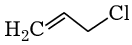
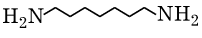
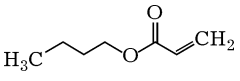
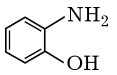
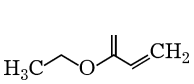
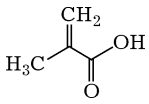
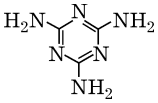
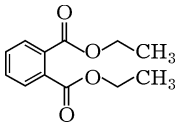
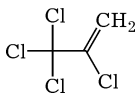
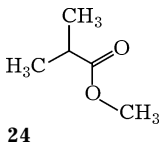
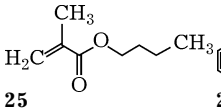
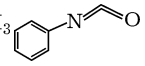
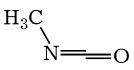
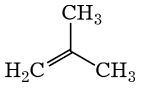
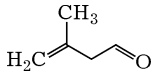
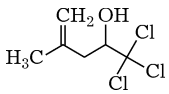
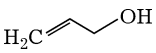
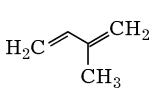
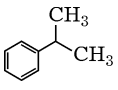
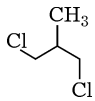
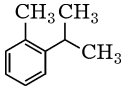
 **01**	 **02**	**03**
 **04**	 **05**	 **06**
 **07**	 **08**	 **09**
 **10**	 **11**	 **12**
 **13**	 **14**	 **15**
 **16**	 **17**	 **18**
 **19**	 **20**	 **21**
 **22**	 **23**	 **24**
 **25**	 **26**	 **27**
 **28**	 **29**	 **30**
 **31**	 **32**	 **33**
 **34**	 **35**	 **36**

TABLE 2

Statistical quality of models obtained in three attempts of optimization of correlation weights

Sample No.	R^2	s	F
1	0.6616/0.6456	0.553/0.613	37/24
2	0.6616/0.6468	0.553/0.612	37/24
3	0.6616/0.6527	0.554/0.607	37/24

Note. The first value stands for the learning sample (n = 2), the second for the reference sample (n = 15).

TABLE 3

Numerical values of correlation weights for SA_k

SA_k	Sample 1	Sample 2	Sample 3
#	1.0000000	1.0000000	1.0000000
(	1.0013852	1.0017522	1.0019360
/	1.0171287	1.0217534	1.0243480
1	0.9961685	0.9951212	0.9944593
=	1.0012630	1.0016034	1.0017922
C	0.9999307	0.9999132	0.9999121
Cl	0.9961539	0.9951343	0.9945848
N	0.9982923	0.9978310	0.9975361
O	0.9971900	0.9964399	0.9960152
\	0.9937723	0.9921033	0.9911403
c	1.0013831	1.0017624	1.0019974
n	1.0000000	1.0000000	1.0000000

TABLE 4

Example of SMILES calculation for structure **01** (styrene)

SA_k	$CW(SA_k)$ in sample 1
C	0.9999307
=	1.0012630
C	0.9999307
c	1.0013831
1	0.9961685
c	1.0013831
c	1.0013831
c	1.0013831
c	1.0013831
c	1.0013831
1	0.9961685

Note. SMILES = "C = Cc1ccccc1", DCW = 1.0017401.

TABLE 5

Experimental and calculated values of log (MPC) for hazardous substances in the air of working area (MPC in mg/m^3)

Structure No.	SMILES	DCW	log $(MPC)_{exp}$	$log(MPC)_{calc}$	Difference
		Learning sample			
01	C=Cc1ccccc1	1.0017401	1.477	0.946	0.531
03	C=C	1.0011243	2.000	0.808	1.192
05	C=O	0.9983803	−0.301	0.197	−0.498
07	Oc1ccccc1	0.9978035	−0.523	0.069	−0.592
09	c1ccccc1	1.0006152	1.176	0.695	0.481
10	NCCN	0.9964494	0.301	−0.233	0.534
12	C=C1CCCCCC1	0.9930544	−1.000	−0.989	−0.011
14	O=C(OCCCCCCCCC)c1ccccc1C(=O)OCCCCCCCCC	0.9961811	0.000	−0.293	0.293
16	NCCCCCCCCN	0.9961043	−1.000	−0.310	−0.690
18	Nc1ccccc1O	0.9960995	0.000	−0.311	0.311
20	CC(=C)C(O)=O	1.0021587	1.000	1.039	−0.039
22	O=C(OCC)c1ccccc1C(=O)OCC	0.9970095	−0.301	−0.108	−0.193
23	C=C(Cl)C(Cl)(Cl)Cl	0.9939641	−1.000	−0.786	−0.214
24	CC(C)C(=O)OC	1.0008253	1.000	0.742	0.258
25	C=C(C)C(=O)OCCCC	1.0018811	1.477	0.977	0.500
27	C\N=C=O	0.9916506	−1.301	−1.301	0.000
29	CC(=C)CC=O	1.0021348	−0.301	1.033	−1.334
30	ClC(Cl)(Cl)C(O)CC(=C)C	0.9975491	0.301	0.012	0.289
32	Cl/C(Cl)=C(\Cl)Cl	1.0019851	1.000	1.000	−0.000
34	CC(C)c1ccccc1	1.0031807	0.699	1.266	−0.567
36	Cc1ccccc1C(C)C	1.0031112	1.000	1.251	−0.251
		Reference sample			
02	C=CCl	0.9972738	0.699	−0.049	0.748
04	CC=C	1.0010549	2.000	0.793	1.207
06	CC=O	0.9983111	0.699	0.182	0.517
08	C=CC#N	0.9993454	−0.301	0.412	−0.713
11	CO	0.9971210	0.699	−0.083	0.782
13	O=C(OCCCC)c1ccccc1C(=O)OCCCC	0.9967333	−0.301	−0.170	−0.131
15	C=CCCl	0.9972048	0.301	−0.065	0.366
17	C=CC(=O)OCCCC	0.9991804	1.000	0.375	0.625
19	C=CC(=O)OCC	0.9993189	0.699	0.406	0.293
21	Nc1nc(N)nc(N)n1	0.9968829	−0.301	−0.136	−0.165
26	O=C=N\c1ccccc1	0.9923294	−0.301	−1.150	0.849
28	CC(C)=C	1.0037606	2.000	1.396	0.604
31	C=CCO	0.9982420	0.301	0.166	0.135
33	CC(=C)C=C	1.0049587	1.602	1.662	−0.060
35	CC(CCl)CCl	0.9947980	−0.301	−0.601	0.300

symbol (for example, "c", "C", "N", "O", "=" *etc.*) or two symbols (for example, "Cl") to transfer the information about chemical element (Cl), $CW(SA_k)$ is the correlation weight for SA_k.

Statistical characteristics of the models obtained in the three attempts of optimization of correlation weights with the correlation coefficient between DCW descriptor and log (MPC) as the target function [5, 6] are represented in Table 2. The numerical values of correlation weights are listed in Table 3. An example of the calculation of descriptor on the basis of correlation weights obtained in the first attempt of optimization is shown in Table 4. The data on the logarithms of MPC taken from [9] and calculated using the equation obtained with the help of the least squares procedure are presented in Table 5:

$$\log (\mathrm{MPC}) = -222.1501(\pm 7.2337) + 22.7081(\pm 7.2602)\mathrm{DCW} \quad (2)$$

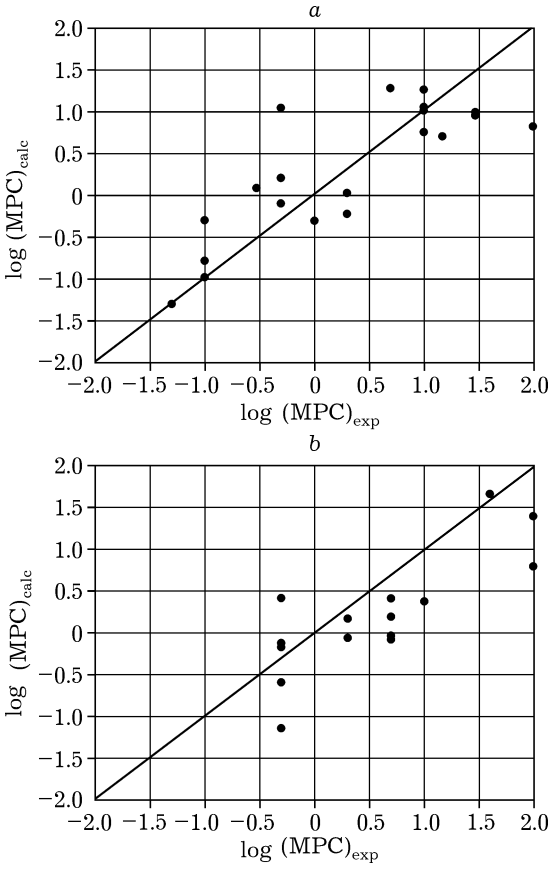


Fig. 1. Correlation between experimental and calculated values of log (MPC) for the structures of the learning (*a*) and reference (*b*) samples.

The numbers of structures in Table 5 correspond to their numbers in Table 1. Model No. 2 was obtained in the first attempt of optimization of correlation weights (see Table 3).

The model calculated using equation (2) for the learning sample and the reference sample, respectively, is presented in Fig. 2. All the three models obtained through optimization using Monte Carlo method (random search [5, 6] have close correlation coefficients (R) and standard deviation (s) both for the learning and reference samples.

CONCLUSIONS

Thus, the possibility to predict MPC of hazardous substances in the air of the working area with the help of optimal descriptors calculated using the SMILES system was demonstrated. The choice of the calculation equation for descriptor may be either additive [11] or multiplicative (1). Multiplicative form is preferable for nonlinear dependencies.

No data on QSAR modelling were found in literature; but the comparative analysis of the models based on the described technique, with the models built up by means of multivariate analysis using descriptors obtained with molecular graphs or the data of quantum-chemical calculations, provides evidence that SMILES serves as a representation of molecular structure convenient for QSAR analysis [3–6, 11]. Broadening of the available databases for physicochemical properties and biological activity [12, 13] stimulates further development of descriptors calculated using SMILES. The considered structures (see Table 1) can be described only in the form of one SMILES code because they are not complicated and so admit unambiguous SMILES representation.

REFERENCES

1 A. A. Toropov, V. O. Kudyshkin, N. L. Voropaeva *et al.*, *Zh. Str. Khim.*, 45, 6 (2004) 994.
2 E. A. Castro, A. P. Toropova, A. A. Toropov and R. Z. Akhmerov, *World J. Chem.*, 1, 1 (2006) 06.
3 D. Vidal, M. Thormann, *J. Chem. Inf. Model*, 45 (2005) 386.
4 D. Vidal, M. Thormann, *Ibid.*, 46 (2006) 836.
5 A. A. Toropov, E. Benfenati, *Current Drug Discovery Technol.*, 4 (2007) 77.

6 A. A. Toropov, A. P. Toropova, D. V. Mukhamedzhanova, I. Gutman, *Ind. J. Chem., Sec. A*, 44 (2005) 1545.
7 A. A. Toropov, V. O. Kudyshkin, N. L. Voropaeva *et al.*, *Vysokomol. Soyed.*, 43, 4 (2001) 760.
8 URL: http://www.daylight.com
9 Predelno-Dopustimye Kontsentratsii (PDK) Vrednykh Veshchestv v Vozdukhe Rabochey Zony. Gigiyenicheskiye Normativy. GN 2.2.5.686–98. Utverzhdeny Glavnym Gosudarstvennym Sanitarnym Vrachom RF, Postanovleniye No. ot 04.02.98.
10 URL: http://www.acdlabs.com/
11 A. A. Toropov, E. Benfenati, *Bioorg. Med. Chem.*, 16 (2008) 4801.
12 http://chem.sis.nlm.nih.gov/chemidplus/
13 http://webbook.nist.gov/chemistry/