Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 43 (2008) 714–740
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

QSPR modeling of octanol/water partition coefficient for vitamins by optimal descriptors calculated with SMILES

A.A. Toropov [a,*], A.P. Toropova [a], I. Raska Jr. [b]

[a] Uzbek Academy of Sciences, Institute of Geology and Geophysics, Khodzhibaev street 49, 100041 Tashkent, Uzbekistan
[b] 3rd Medical Department, 1st Faculty of Medicine, Charles University in Prague, U Nemocnice 1, 12808 Prague 2, Czech Republic

Received 6 March 2007; received in revised form 13 April 2007; accepted 24 May 2007
Available online 3 June 2007

Abstract

Simplified molecular input line entry system (SMILES) has been utilized in constructing quantitative structure–property relationships (QSPR) for octanol/water partition coefficient of vitamins and organic compounds of different classes by optimal descriptors. Statistical characteristics of the best model (vitamins) are the following: $n = 17$, $R^2 = 0.9841$, $s = 0.634$, $F = 931$ (training set); $n = 7$, $R^2 = 0.9928$, $s = 0.773$, $F = 690$ (test set). Using this approach for modeling octanol/water partition coefficient for a set of organic compounds gives a model that is statistically characterized by $n = 69$, $R^2 = 0.9872$, $s = 0.156$, $F = 5184$ (training set) and $n = 70$, $R^2 = 0.9841$, $s = 0.179$, $F = 4195$ (test set).
© 2007 Elsevier Masson SAS. All rights reserved.

Keywords: QSPR; SMILES; Octanol/water partition coefficient; Vitamins

1. Introduction

Quantitative structure–property/activity relationships (QSPR/QSAR) are tools for predicting numerical data on physicochemical property or biological activity of substances which have not been studied experimentally. The prediction is based on correlation between property/activity of interest and molecular descriptors [1]. Recently, the so-called optimal descriptors have been recruited for the QSPR/QSAR analysis [2]. Optimal descriptors can be calculated with molecular graphs [3,4] and also with simplified molecular input line entry system (SMILES) [5–7].

Octanol/water partition coefficient is an important characteristic from biochemical point of view. QSPR models of octanol/water partition coefficient of vitamins and non-vitamins based on optimal descriptors calculated with molecular graphs have been examined in Refs. [3,4]. The present study is aimed to compare the mentioned models with ones based on optimal descriptors calculated with the SMILES notation.

Numerical data on the octanol/water partition coefficient are taken from Ref. [3]. Vitamins under consideration, their Chemical Abstract Service (CAS) numbers and structures are represented in Table 1. The information on set of organic compounds (non-vitamins) is summarized in Table 2.

2. Method

Descriptors utilized in present study have been defined as the following:

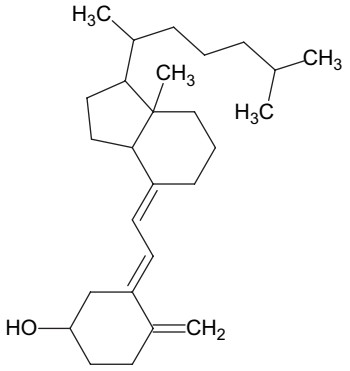
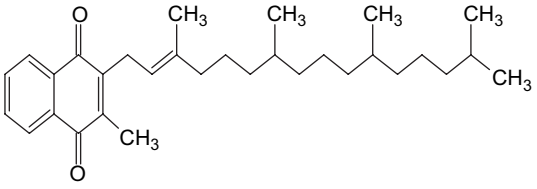
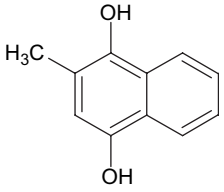
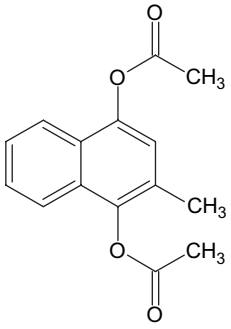
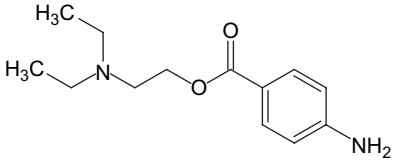
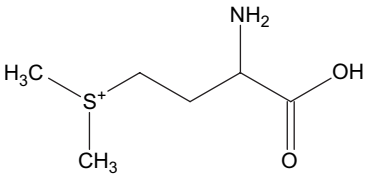
$$\mathrm{DCW(LimNI)} = \mathrm{CW(HC)} \prod_{k=1}^{N} \mathrm{CW(SF}_k\mathrm{)} \quad (1)$$

where SF_k is a symbol of SMILES, excepting the fragment of four characters in '[S+]' as well as two symbol images of Cl and Br; $CW(SF_k)$ is the correlation weights of the SF_k; the ')' has been replaced by '(', because these characters are the indicators of the same phenomena, i.e., molecular branching; N is the number of the SF_k in the given SMILES; HC is the

* Corresponding author.
E-mail address: aatoropov@yahoo.com (A.A. Toropov).

0223-5234/$ - see front matter © 2007 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2007.05.007

Table 1
Names, Chemical Abstract Service (CAS) numbers, and structures of vitamins examined in the study

Name	CAS No.	Structure
Vitamin D3	67-97-0	
Vitamin 2′,3′-*trans*-K1	84-80-0	
Vitamin K3H2	481-85-6	
Vitamin K4	573-20-6	
Vitamin H3	59-46-1	
Vitamin U	1115-84-0	

(*continued on next page*)

Table 1 (*continued*)

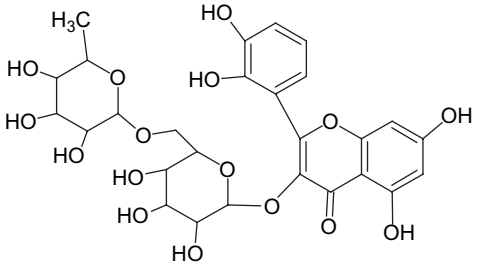
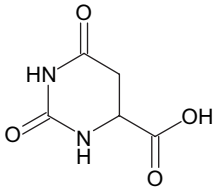
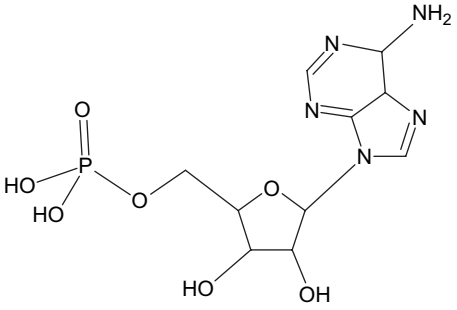
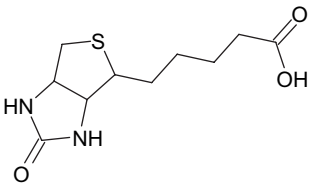
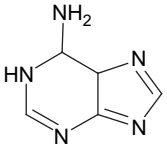
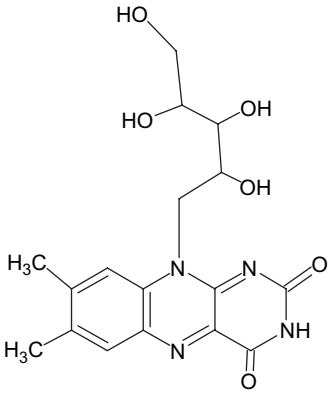
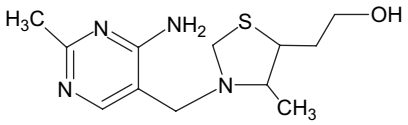
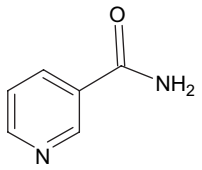
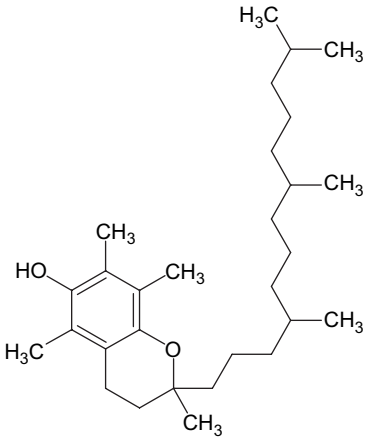
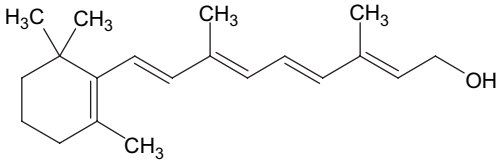
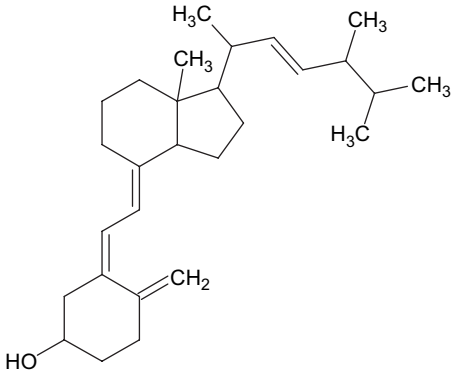
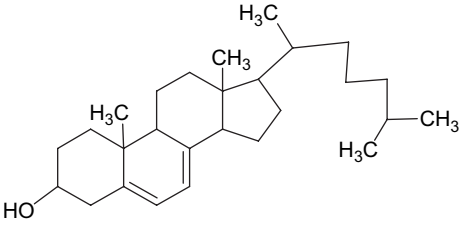
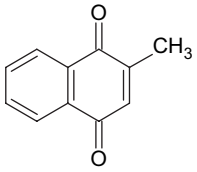
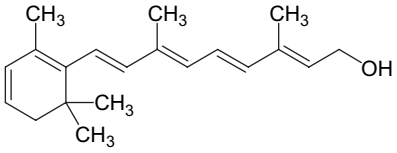
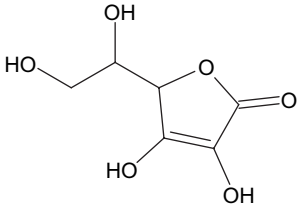
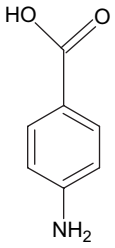
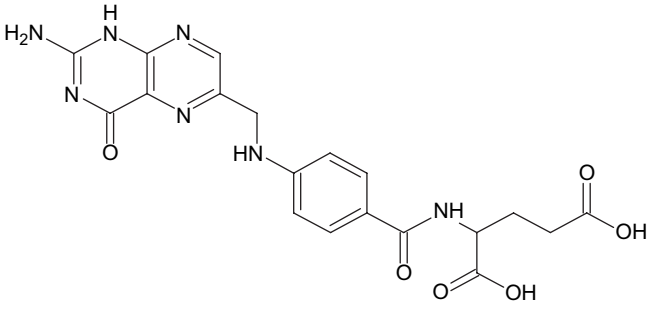
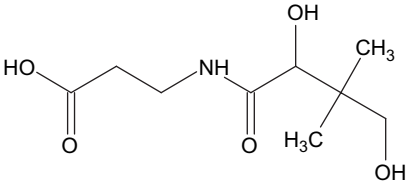
Name	CAS No.	Structure
Vitamin P	153-18-4	
Vitamin B13	65-86-1	
Vitamin B8	61-19-8	
Vitamin B7	58-85-5	
Vitamin B4	73-24-5	
Vitamin B2	83-88-5	

Table 1 (*continued*)

Name	CAS No.	Structure
Vitamin B1	59-43-8	
Vitamin B3	98-92-0	
Vitamin E	59-02-9	
Vitamin A	68-26-8	
Vitamin D2	50-14-6	
Vitamin D3	434-16-2	

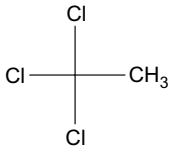
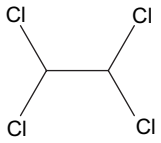
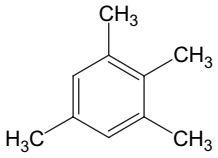
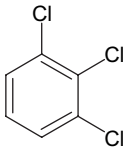
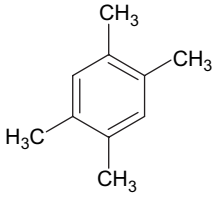
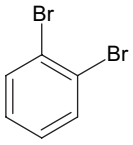
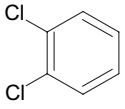
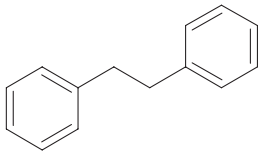
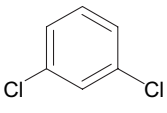
(*continued on next page*)

Table 1 (*continued*)

Name	CAS No.	Structure
Vitamin K0	58-27-5	
Vitamin A2	79-80-1	
Vitamin C	50-81-7	
Vitamin H’	150-13-0	
Vitamin B9	59-30-3	
Vitamin B5	79-83-4	

The CAS numbers for compounds under consideration were taken from US National Library of Medicine (http://chem.sis.nlm.nih.gov/chemidplus).

Table 2
Names, Chemical Abstract Service (CAS) numbers, and structures for organic compounds (non-vitamins) examined in the study

Name	CAS No.	Structure
1,1,1-Trichloroethane	71-55-6	
1,1,2,2-Tetrachloroethane	79-34-5	
1,2,3,5-Tetramethylbenzene	527-53-7	
1,2,3-Trichlorobenzene	87-61-6	
1,2,4,5-Tetramethylbenzene	95-93-2	
1,2-Dibromobenzene	583-53-9	
1,2-Dichlorobenzene	95-50-1	
1,2-Diphenylethane	103-29-7	
1,3-Dichlorobenzene	541-73-1	

(*continued on next page*)

Table 2 (*continued*)

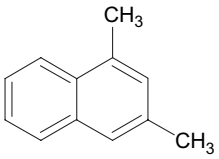
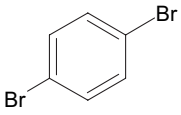
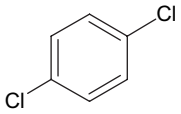
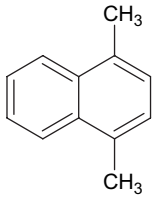
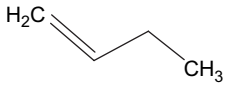
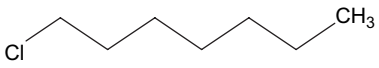
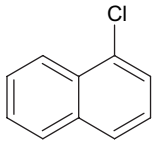
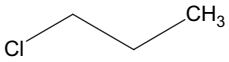
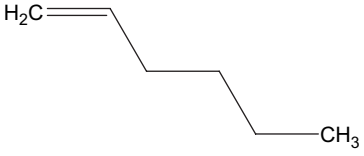
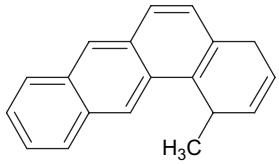
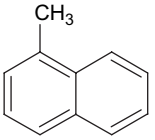
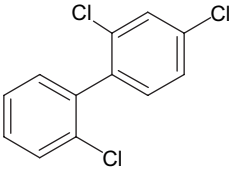
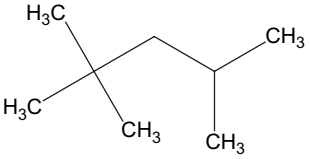
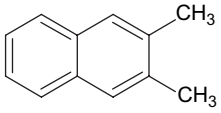
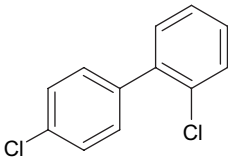
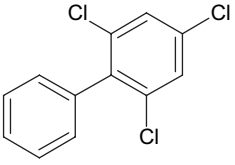
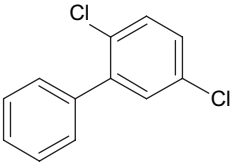
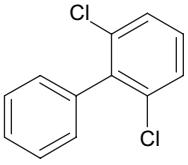
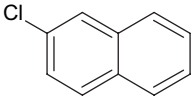
Name	CAS No.	Structure
1,3-Dimethylnaphthalene	575-41-7	
1,4-Dibromobenzene	106-37-6	
1,4-Dichlorobenzene	106-46-7	
1,4-Dimethylnaphthalene	571-58-4	
1-Butene	106-98-9	
1-Chloroheptane	629-06-1	
1-Chloronaphthalene	90-13-1	
1-Chloropropane	540-54-5	
1-Hexene	592-41-6	
1-Methylbenz(*a*)anthracene	2498-77-3	

Table 2 (*continued*)

Name	CAS No.	Structure
1-Methylnaphthalene	90-12-0	
2,2′,4-Trichlorobiphenyl	37680-66-3	
2,2,4-Trimethylpentane	540-84-1	
2,3-Dimethylnaphthalene	581-40-8	
2,4′-Dichlorobiphenyl	34883-43-7	
2,4,6-Trichlorobiphenyl	35693-92-6	
2,5-Dichlorobiphenyl	34883-39-1	
2,6-Dichlorobiphenyl	33146-45-1	
2-Chloronaphthalene	91-58-7	

(*continued on next page*)

Table 2 (*continued*)

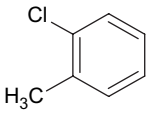
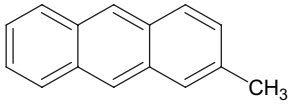
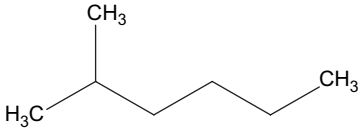
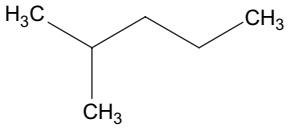
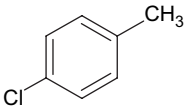
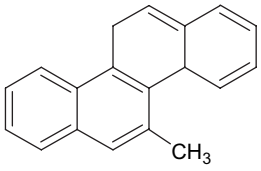
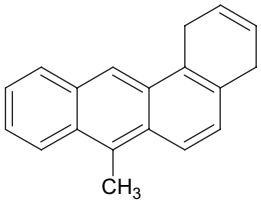
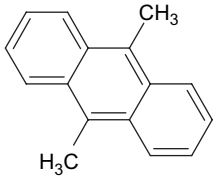
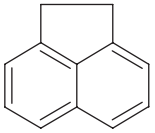
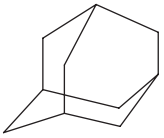
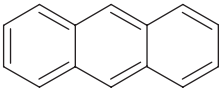
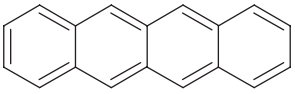
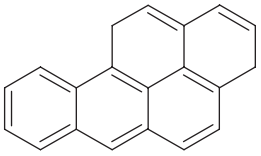
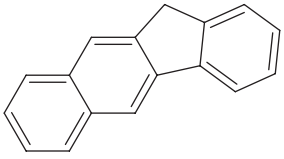
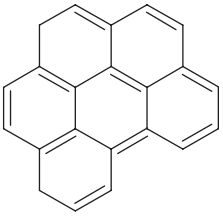
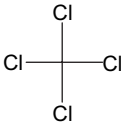
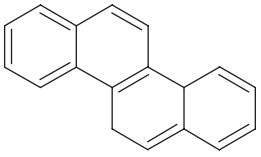
Name	CAS No.	Structure
2-Chlorotoluene	95-49-8	
2-Methylanthracene	613-12-7	
2-Methylhexane	591-76-4	
2-Methylpentane	107-83-5	
4-Chlorotoluene	106-43-4	
5-Methylchrysene	3697-24-3	
7-Methylbenz(*a*)anthracene	2541-69-7	
9,10-Dimethylanthracene	781-43-1	
Acenaphthene	83-32-9	

Table 2 (*continued*)

Name	CAS No.	Structure
Adamantine	281-23-2	
Anthracene	120-12-7	
Benz(*b*)anthracene	92-24-0	
Benzene	71-43-2	
Benz(*a*)pyrene	50-32-8	
Benzo(*b*)fluorine	30777-19-6	
Benzo(*ghi*)perylene	191-24-2	
Carbon tetrachloride	56-23-5	
Chrysene	218-01-9	

(*continued on next page*)

Table 2 (*continued*)

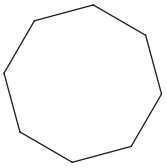
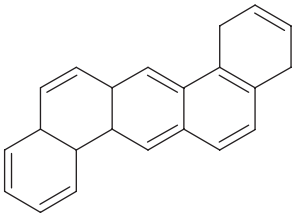
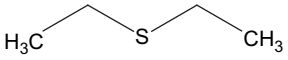
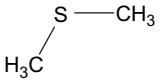
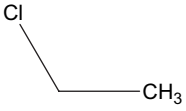
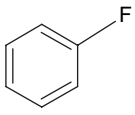
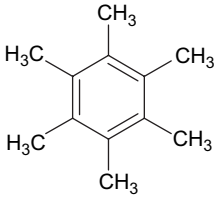
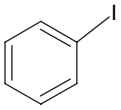
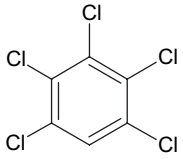
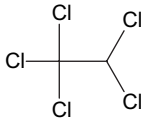
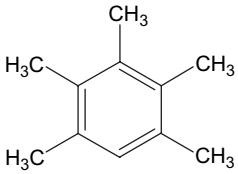
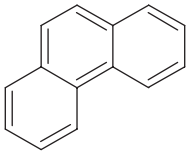
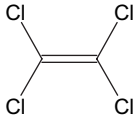
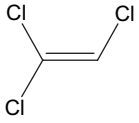
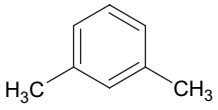
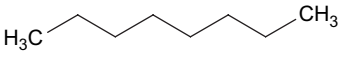
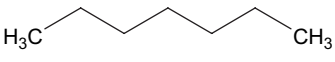
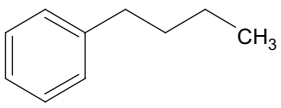
Name	CAS No.	Structure
Cyclohexane	110-82-7	
Cyclooctane	292-64-8	
Dibenz(*a*,*h*)anthracene	53-70-3	
Diethyl sulfide	352-93-2	
Dimethyl sulfide	75-18-3	
Ethylchloride	75-00-3	
Fluorobenzene	462-06-6	
Hexamethylbenzene	87-85-4	
Iodobenzene	591-50-4	

Table 2 (*continued*)

Name	CAS No.	Structure
Pentachlorobenzene	608-93-5	
Pentachloroethane	76-01-7	
Pentamethylbenzene	700-12-9	
Phenanthrene	85-01-8	
Tetrachloroethylene	127-18-4	
Trichloroethylene	79-01-6	
m-Xylene	108-38-3	
n-Octane	111-65-9	
n-Heptane	142-82-5	
n-Butylbenzene	104-51-8	

(*continued on next page*)

Table 2 (*continued*)

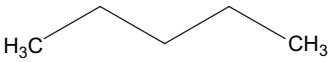
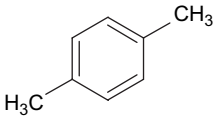
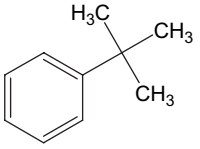
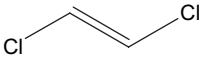
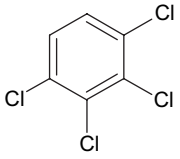
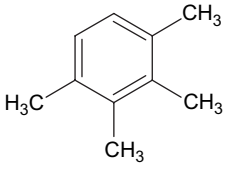
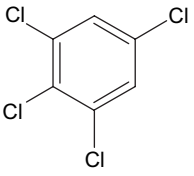
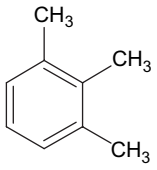
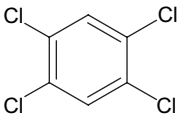
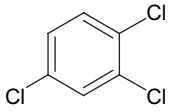
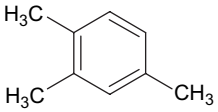
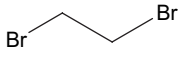
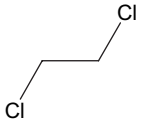
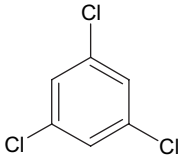
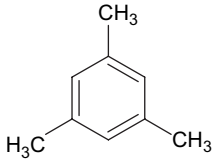
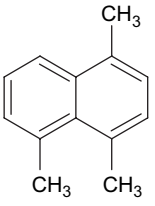
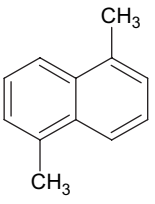
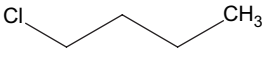
Name	CAS No.	Structure
n-Pentane	109-66-0	
p-Xylene	106-42-3	
tert-Butylbenzene	98-06-6	
trans-1,2-Dichloroethylene	156-60-5	
1,2,3,4-Tetrachlorobenzene	634-66-2	
1,2,3,4-Tetramethylbenzene	488-23-3	
1,2,3,5-Tetrachlorobenzene	634-90-2	
1,2,3-Trimethylbenzene	526-73-8	
1,2,4,5-Tetrachlorobenzene	95-94-3	

Table 2 (*continued*)

Name	CAS No.	Structure
1,2,4-Trichlorobenzene	120-82-1	
1,2,4-Trimethylbenzene	95-63-6	
1,2-Dibromoethane	106-93-4	
1,2-Dichloroethane	107-06-2	
1,3,5-Trichlorobenzene	108-70-3	
1,3,5-Trimethylbenzene	108-67-8	
1,4,5-Trimethylnaphthalene	2131-41-1	
1,5-Dimethylnaphthalene	571-61-9	
1-Chlorobutane	109-69-3	

(*continued on next page*)

Table 2 (*continued*)

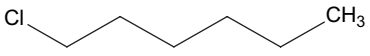
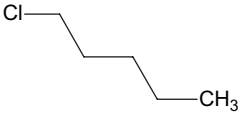
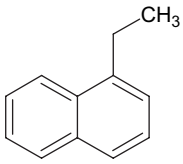
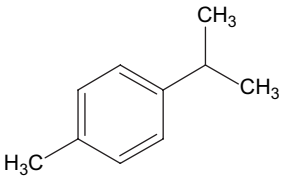
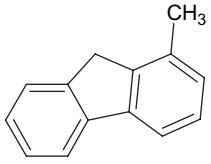
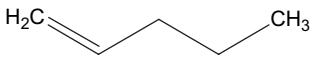
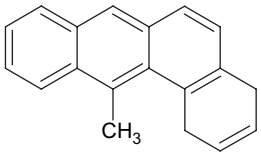
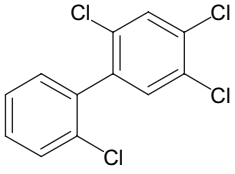
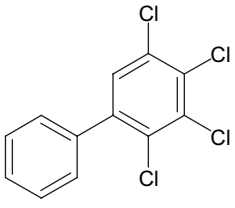
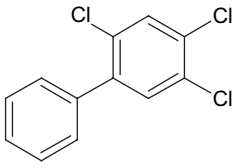
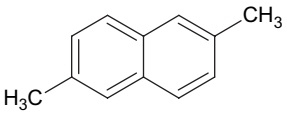
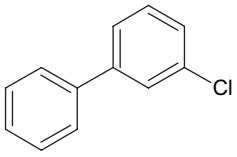
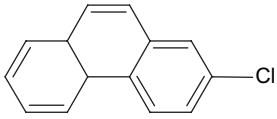
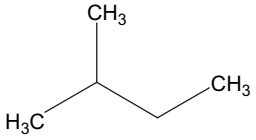
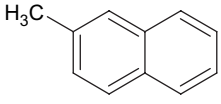
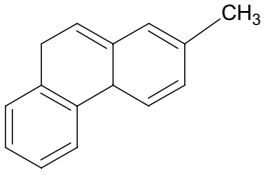
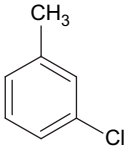
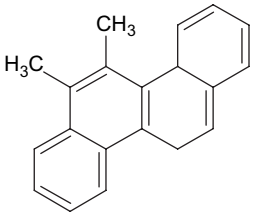
Name	CAS No.	Structure
1-Chlorohexane	544-10-5	
1-Chloropentane	543-59-9	
1-Ethylnaphthalene	1127-76-0	
1-Isopropyl-4-methylbenzene	99-87-6	
1-Methylfluorene	1730-37-6	
1-Pentene	109-67-1	
12-Methylbenz(*a*)anthracene	2422-79-9	
2,2′,4,5-Tetrachlorobiphenyl	70362-47-9	
2,3,4,5-Tetrachlorobiphenyl	33284-53-6	

Table 2 (*continued*)

Name	CAS No.	Structure
2,4,5-Trichlorobiphenyl	15862-07-4	
2,6-Dimethylnaphthalene	581-42-0	
2-Chlorobiphenyl	2051-60-7	
2-Chlorophenanthrene	No CAS[a]	
2-Methylbutane	78-78-4	
2-Methylnaphthalene	91-57-6	
2-Methylphenanthrene	2531-84-2	
3-Chlorotoluene	108-41-8	
5,6-Dimethylchrysene	3697-27-6	

(*continued on next page*)

Table 2 (*continued*)

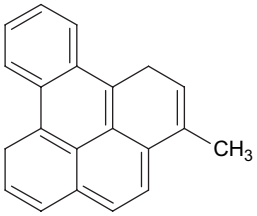
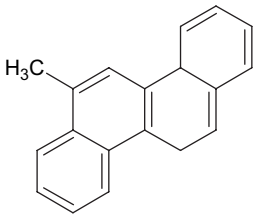
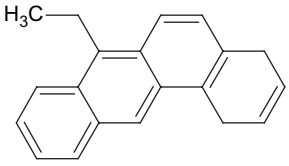
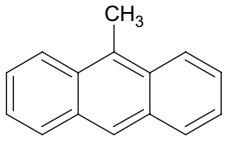
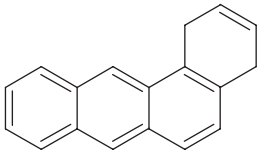
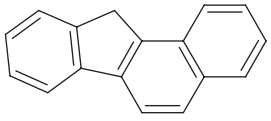
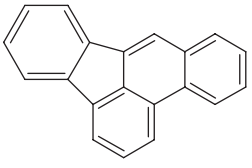
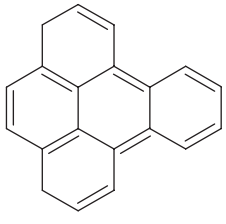
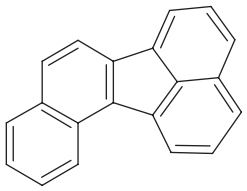
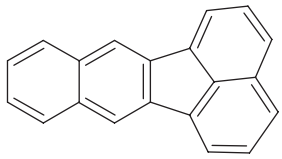
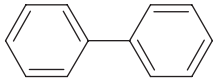
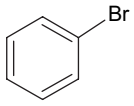
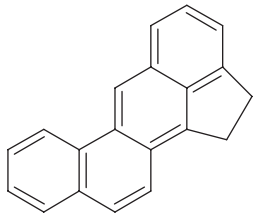
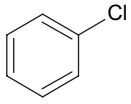
Name	CAS No.	Structure
Methylbenzo(*e*)pyrene	41699-04-1	
6-Methylchrysene	1705-85-7	
7-Ethylbenz(*a*)anthracene	3697-30-1	
9-Methylanthracene	779-02-2	
Benz(*a*)anthracene	56-55-3	
Benzo(*a*)fluorine	238-84-6	
Benzo(*b*)fluoranthene	205-99-2	
Benzo(*e*)pyrene	192-97-2	

Table 2 (*continued*)

Name	CAS No.	Structure
Benzo(*j*)fluoranthene	205-82-3	
Benzo(*k*)fluoranthene	207-08-9	
Biphenyl	92-52-4	
Bromobenzene	108-86-1	
Cholanthrene	479-23-2	
Chlorobenzene	108-90-7	
Cycloheptane	291-64-5	
Cyclohexene	110-83-8	
Cyclopentane	287-92-3	
Cyclopentene	142-29-0	

(*continued on next page*)

Table 2 (*continued*)

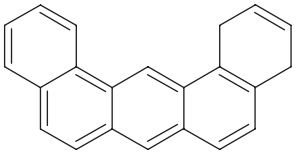
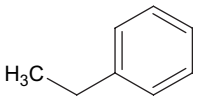
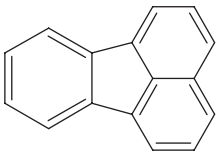
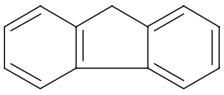
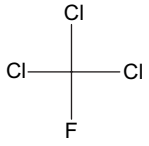
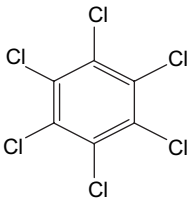
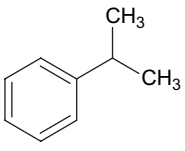
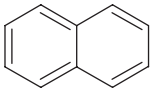
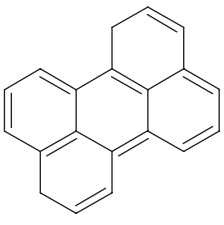
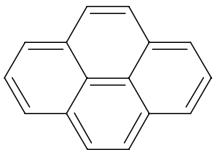
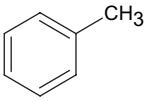
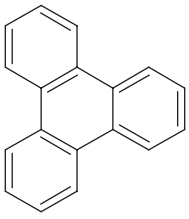
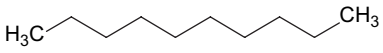
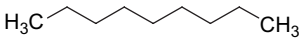
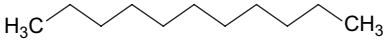
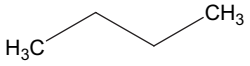
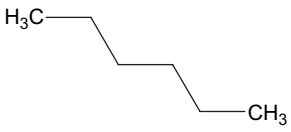
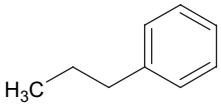
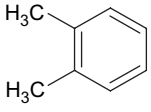
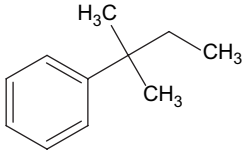
Name	CAS No.	Structure
Dibenz(*a*,*j*)anthracene	224-41-9	
Ethylbenzene	100-41-4	
Fluoranthene	206-44-0	
Fluorene	86-73-7	
Fluorotrichloromethane	75-69-4	
Hexachlorobenzene	118-74-1	
Isopropylbenzene	98-82-8	
Naphthalene	91-20-3	
Perylene	198-55-0	

Table 2 (*continued*)

Name	CAS No.	Structure
Pyrene	129-00-0	
Toluene	108-88-3	
Triphenylene	217-59-4	
n-Decane	124-18-5	
n-Nonane	111-84-2	
n-Undecane	1120-21-4	
n-Butane	106-97-8	
n-Hexane	110-54-3	
n-Propylbenzene	103-65-1	
o-Xylene	95-47-6	
tert-Amylbenzene	2049-95-8	

The CAS numbers for compounds under consideration were taken from US National Library of Medicine (http://chem.sis.nlm.nih.gov/chemidplus).

[a] CAS number for 2-chlorophenanthrene is not available at http://chem.sis.nlm.nih.gov/chemidplus.

Table 3
Construction of the (0,1) hypercode for SMILES notation

Number	Character	Comments
1	'%'	Indicator of the global SMILES attribute
2	'0'	Fluorine atom (F) is absent
	'1'	Fluorine atom (F) is present
3	'0'	Chlorine atom (Cl) is absent
	'1'	Chlorine atom (Cl) is present
4	'0'	Bromine atom (Br) is absent
	'1'	Bromine atom (Br) is present
5	'0'	Cycle indicated by 1 is absent
	'1'	Cycle indicated by 1 is present
6	'0'	Cycle indicated by 2 is absent
	'1'	Cycle indicated by 2 is present
7	'0'	Cycle indicated by 3 is absent
	'1'	Cycle indicated by 3 is present

For instance, the %010100 means that in given molecular system there are chlorine atom(s) and cycle(s) indicated by '1' and there are no fluorine atom, bromine atom, and naphthalene-like and anthracene-like cycles.

hypercode of the SMILES, the HC encoded the total molecular situation by means of the (0,1) sequence. In Table 3 construction of the HC is demonstrated. The HC is aimed to express cooperative features of molecular structure (presence and absence of fluorine, chlorine, and bromine as well as number and quality cycles, e.g., benzene-like, naphthalene-like, etc.). The CW(HC) is the correlation weight of the given HC.

LimNI is the limitation for number of presence of the SF_k or the HC in SMILES of the training set: if the number of given SF_k is less than the LimNI then the CW(SF_k) is equal to 1. In other words, this SF_k is blocked and influence of the SF_k for the model is absent.

Numerical values of the CW(SF_k) and CW(HC) are calculated using Monte Carlo optimization with correlation coefficient in role of target function [3,5–7]. Having numerical data on CW(SF_k) and CW(HC) which produce maximum of the correlation coefficient between octanol/water partition coefficient and the DCW(LimNI) one can calculate one variable model of view:

$$\log P = C_0 + C_1 \times \mathrm{DCW(LimNI)} \quad (2)$$

Predictive ability of Eq. (2) should be validated with an external test set. Splitting into training and test sets used in the study has been done randomly, but according to the following conditions: first, majority of the SF_k should be in the training set; and second, interval of the octanol/water partition coefficient for training and test sets should be similar.

The aim of the LimNI is to select rational list of the SF_k. Apparently, if the training set contains only one SF_k its correlation weight is hardly informative. Ideal situation is, if all SF_k take place in training set many times and in different SMILES. However, it is a possible situation when some of the SF_k are rare in the training set. Under such circumstances one can obtain a model that is good for training set but that is poor for external test set (overtraining). However, rational selection of the informative SF_k and HC is possible. One can study some sequences of the LimNI values and detect the best value of the LimNI. Of course the selection is dependent on composition of compounds under consideration and splitting into the training and test sets.

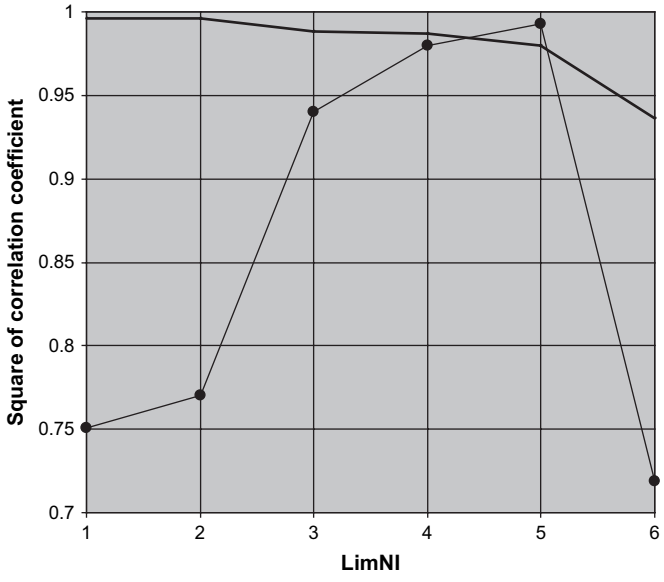


Fig. 1. Diagram of average values of correlation coefficient for the training and test sets (vitamins). The LimNI values varied from 1 to 6. Test set is marked by circles.

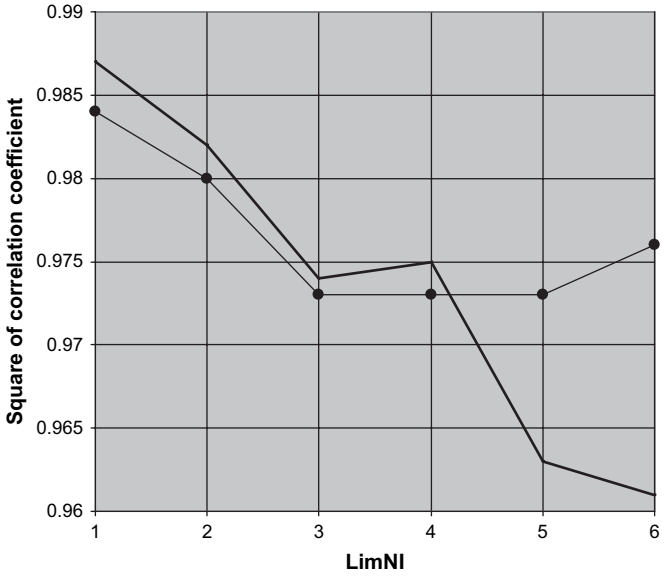


Fig. 2. Diagram of average values of correlation coefficient for the training and test sets (non-vitamins). The LimNI values varied from 1 to 6. Test set is marked by circles.

Table 4
Statistical characteristics of models obtained with the optimal descriptors calculated with SMILES notation

Probe	Training set, $n = 17$			Test set, $n = 7$		
	R^2	s	F	R^2	s	F
Panel 1[a]						
1	0.9841	0.634	931	0.9928	0.773	690
2	0.9832	0.654	876	0.9898	0.816	484
3	0.9776	0.754	654	0.9936	0.718	782
Probe	Training set, $n = 69$			Test set, $n = 70$		
	R^2	s	F	R^2	s	F
Panel 1[b]						
1	0.9872	0.156	5184	0.9841	0.179	4195
2	0.9875	0.154	5313	0.9815	0.195	3616
3	0.9876	0.153	5329	0.9835	0.183	4059

[a] Statistical characteristics of the DCW(5) based models of octanol/water partition coefficient for vitamins obtained in three probes of the Monte Carlo optimization.
[b] Statistical characteristics of the DCW(1) based models of octanol/water partition coefficient for organic compounds (non-vitamins) obtained in three probes of the Monte Carlo optimization.

Work sets of vitamins and organic compounds of different classes (non-vitamins) have been examined in the present study. SMILES notations for this study have been builded using ACD/ChemSketch software (http://www.acdlabs.com/).

3. Results and discussion

Diagrams of LimNI values from 1 to 6 versus correlation coefficient values on the training and test sets are presented in Figs. 1 and 2, for vitamins and organic compounds (which are not vitamins), respectively.

One can see from Fig. 1 that the best statistical quality of the model for test set of vitamins takes place if the LimNI is equal to 5. Also, one can see that best statistical quality of the model for test set of organic compounds (non-vitamins) takes place in case of LimNI = 1.

Three probes of the optimization with the DCW(5) for vitamins have shown that the statistical quality of the model

Table 5
Correlation weights for calculation the DCW(5) in modeling of octanol/water partition coefficient of vitamins obtained in three probes of the Monte Carlo optimization

SMILES attributes	CW(SA_k) and CW(CH) in probe 1	CW(SA_k) and CW(CH) in probe 1	CW(SA_k) and CW(CH) in probe 1	Number of the SA_k or HC in training set	Number of the SA_k or HC in test set
SA_k					
(	1.0147938	1.0203469	1.0097334	142	52
/	0.9972771	0.9823258	0.9946518	12	6
1	1.0533693	1.0540981	1.0296327	32	12
2	0.9663424	1.0252735	0.9921373	24	6
3	0.9877296	0.9809385	0.9950202	10	4
4	1.0	1.0	1.0	2	2
5	1.0	1.0	1.0	2	0
=	1.0211489	1.0327756	1.0205812	41	22
C	1.0271229	1.0213236	1.0194922	188	96
N	0.9343767	0.9452214	0.9681511	29	2
O	0.9049274	0.9110210	0.9430364	60	18
P	1.0	1.0	1.0	1	0
S	1.0	1.0	1.0	2	0
\	0.9755711	0.9720624	0.9785514	19	13
c	1.0334095	1.0267283	1.0211745	75	12
n	0.6737481	0.7161091	0.8078579	5	0
[S+]	1.0	1.0	1.0	1	0
CH					
%000000	1.0	1.0	1.0	1	1
%000100	1.0	1.0	1.0	4	3
%000110	1.0740580	0.9661000	1.0175784	7	1
%000111	1.0472086	0.9841456	0.9897869	5	2

Table 6
Correlation weights for calculating the DCW(1) in modeling of octanol/water partition coefficient of organic compounds (non-vitamins) obtained in three probes of the Monte Carlo optimization

SMILES attributes	CW(SA_k) and CW(CH) in probe 1	CW(SA_k) and CW(CH) in probe 1	CW(SA_k) and CW(CH) in probe 1	Number of the SA_k or HC in training set	Number of the SA_k or HC in test set
SA_k					
(	0.9995870	0.9997079	0.9994184	108	90
/	0.9926701	0.9953003	0.9957427	20	19
1	1.0026932	0.9743704	0.9784140	98	114
2	0.9694709	0.9857619	0.9798661	54	66
3	1.0113754	0.9913780	0.9949645	30	44
4	0.9762072	0.9773554	0.9776874	18	34
5	0.9810292	0.9822500	0.9825093	6	16
6	0.9913791	0.9912505	0.9908931	2	0
=	0.9877003	0.9847608	0.9841193	27	26
C	1.0437231	1.0406275	1.0406904	217	218
Br	1.0474591	1.0496393	1.0372826	4	3
Cl	1.0488146	1.0451557	1.0456801	58	47
F	1.0249216	1.0369138	0.9901785	1	1
I	1.0692090	1.0646279	1.0640478	1	0
S	0.9369619	0.9416182	0.9414262	2	0
\	0.9846079	0.9892496	0.9892186	20	22
C	1.0324902	1.0301223	1.0301152	428	556
HC					
%000000	1.0809405	1.0336775	1.0286452	10	7
%000100	1.0402384	1.0552063	1.0419716	12	15
%000110	1.0471497	1.0321126	1.0309267	5	7
%000111	0.9828912	1.0115677	1.0028036	15	21
%001000	1.0	1.0	1.0	0	1
%001100	1.0513386	1.0545302	1.0659787	2	1
%010000	1.0183750	0.9773115	0.9716974	10	4
%010100	1.0390667	1.0546975	1.0401921	7	8
%010110	1.0450687	1.0308914	1.0288885	7	4
%010111	1.0	1.0	1.0	0	1
%100100	1.0011817	1.0055463	1.0392843	1	0
%110000	1.0	1.0	1.0	0	1

is reproducing satisfactorily. Three probes of the DCW(1) optimization for the set of organic compounds have shown that reproducing also takes place in this case. The statistical characteristics are summarized in Table 4.

Correlation weights for calculation of the DCW(5) obtained in three probes of the Monte Carlo optimization (for vitamins) are listed in Table 5. The model based on results of the first optimization is as follows:

$$\log P = -6.8588(\pm 0.1096) + 7.1888(\pm 0.0599) \times \mathrm{DCW}(5)$$
$$n = 17,\ R^2 = 0.9841,\ s = 0.634,\ F = 931$$
$$n = 7,\ R^2 = 0.9928,\ s = 0.773,\ F = 690 \quad (3)$$

Correlation weights for calculation of the DCW(1) obtained in three probes of the Monte Carlo optimization (for non-vitamins) are listed in Table 6. The model based on results of the first optimization is as follows:

$$\log P = -9.3926(\pm 0.0195) + 9.3428(\pm 0.0131) \times \mathrm{DCW}(1)$$
$$n = 69,\ R^2 = 0.9872,\ s = 0.156,\ F = 5184$$
$$n = 70,\ R^2 = 0.9841,\ s = 0.179,\ F = 4195 \quad (4)$$

Table 7
Example of calculation of the DCW(5) for vitamin D3 (CAS 67-97-0) with correlation weights obtained in first optimization probe, DCW(5) = 2.3609496 SMILES = {C=C3CCC(O)C/C3=C\C=C1/CCCC2(C)C(CCC12)C(C)CCCC(C)C}

SA_k and HC	CW(SA_k) and CW(CH)
C	1.0271229
=	1.0211489
C	1.0271229
3	0.9877296
C	1.0271229
C	1.0271229
C	1.0271229
(	1.0147938
O	0.9049274
(	1.0147938
C	1.0271229
/	0.9972771
C	1.0271229
3	0.9877296
=	1.0211489
C	1.0271229
\	0.9755711
C	1.0271229
=	1.0211489

Table 7 (*continued*)

SA_k and HC	$CW(SA_k)$ and CW(CH)
C	1.0271229
1	1.0533693
/	0.9972771
C	1.0271229
C	1.0271229
C	1.0271229
C	1.0271229
2	0.9663424
(	1.0147938
C	1.0271229
(	1.0147938
C	1.0271229
(	1.0147938
C	1.0271229
C	1.0271229
C	1.0271229
1	1.0533693
2	0.9663424
(	1.0147938
C	1.0271229
(	1.0147938
C	1.0271229
(	1.0147938
C	1.0271229
C	1.0271229
C	1.0271229
C	1.0271229
(	1.0147938
C	1.0271229
(	1.0147938
C	1.0271229
%000111	1.0472086

Example of calculation of the DCW(5) with $CW(SF_k)$ and CW(HC) obtained in the first probe of optimization is demonstrated in Table 7.

Experimental and calculated values of octanol/water partition coefficient for vitamins with Eq. (3) as well as splitting into training and test sets are demonstrated in Table 8. Graphically the model is demonstrated in Fig. 3.

Experimental and calculated values of octanol/water partition coefficient for non-vitamins with Eq. (4) as well as splitting into training and test sets are demonstrated in Table 9. Graphically the model is demonstrated in Fig. 4.

Statistical characteristics of the model for octanol/water partition coefficient of vitamins described in Ref. [3] are the following: $n = 18$, $R^2 = 0.9804$, $s = 0.738$, $F = 800$ (training set) and $n = 6$, $R^2 = 0.9497$, $s = 0.910$, $F = 75$. Statistical characteristics of the model based on molecular graph (for case of non-vitamins) are $n = 69$, $R^2 = 0.9953$, $s = 0.096$, $F = 14136$ (training set) and $n = 70$, $R^2 = 0.9930$, $s = 0.119$, $F = 9631$ (test set). In other words the model calculated with Eq. (3) is better than that described in Ref. [3] and statistical characteristics of the model calculated with Eq. (4) are worse than the statistical characteristics of the model described in Ref. [3]. However, the difference of these models is not dramatic. Thus SMILES is a reasonable alternative of molecular graph for modeling octanol/water partition coefficient of vitamins and organic compounds of different classes by optimal descriptors.

Table 8
Splitting into training and test sets; experimental and calculated values of octanol/water partition coefficient for vitamins

CAS No.	SMILES notation	DCW(5)	Expr	Calc
	Training set			
67-97-0	C=C3CCC(O)C/C3=C\C=C1/CCCC2(C)C(CCC12)C(C)CCCC(C)C	2.3609496	10.24	10.11359
84-80-0	CC(C)CCCC(C)CCCC(C)CCCC(\C)=C\C\C2=C(/C)C(=O)c1ccccc1C2=O	2.6022886	11.71	11.84853
481-85-6	Oc1cc(C)c(O)c2ccccc12	1.3789122	2.76	3.05392
573-20-6	CC(=O)Oc1cc(C)c(OC(C)=O)c2ccccc12	1.3897700	3.59	3.13198
59-46-1	O=C(OCCN(CC)CC)c1ccc(N)cc1	1.2995111	2.14	2.48313
1115-84-0	C[S+](C)CCC(N)C(O)=O	1.0019351	0.28	0.34391
153-18-4	Oc1cccc(c1O)C=4Oc5cc(O)cc(O)c5C(=O)C=4OC3OC(COC2OC(C)C(O)C(O)C2O)C(O)C(O)C3O	0.6776114	−2.02	−1.98759
65-86-1	O=C(O)C1CC(=O)NC(=O)N1	0.8635871	−0.83	−0.65065
61-19-8	OP(O)(=O)OCC3OC(N2\C=N/C1/C2=N\C=N/C1N)C(O)C3O	0.5652061	−1.68	−2.79565
58-85-5	O=C1NC2C(CCCCC(O)=O)SCC2N1	1.0405178	0.39	0.62127
73-24-5	NC2N/C=N\C1=N\C=N/C12	0.9132504	−0.09	−0.29363
83-88-5	OCC(O)C(O)C(O)CN\2c3cc(C)c(C)cc3/N=C1\C/2=N/C(=O)NC1=O	0.8868278	−1.46	−0.48357
59-43-8	Cc2nc(N)c(CN1CSC(CCO)C1C)cn2	0.6157999	−3.93	−2.43194
98-92-0	NC(=O)c1cccnc1	0.8046813	−0.37	−1.07411
59-02-9	CC(C)CCCC(C)CCCC(C)CCCC1(C)CCc2c(O1)c(C)c(C)c(O)c2C	2.5982015	12.18	11.81915
68-26-8	CC1(C)CCCC(\C)=C1\C=C\C(\C)=C\C=C\C(\C)=C\CO	1.7569609	5.68	5.77164
59-30-3	OC(=O)C(CCC(O)=O)NC(=O)c1ccc(cc1)NCc2cnc3NC(/N)=N\C(O)c3n2	0.4407125	−2.81	−3.69061
	Test set			
50-14-6	C=C3CCC(O)C\C3=C\C=C1/CCCC2(C)C(CCC12)C(C)/C=C/C(C)C(C)C	2.4810102	10.44	10.97669
434-16-2	CC(C)CCCC(C)C1CCC2C\3=C\C=C4\CC(O)CCC4(C)C/3CCC12C	2.1426149	8.65	8.54403
58-27-5	O=C2\C=C(\C)C(=O)c1ccccc12	1.3637084	2.20	2.94463
79-80-1	CC=1/C=C\CC(C)(C)C=1/C=CC(\C)=C\C=C\C(\C)=C\CO	1.9144941	7.41	6.90412
50-81-7	OC=1C(OC(=O)C=1O)C(O)CO	0.8319783	−1.85	−0.87787
150-13-0	O=C(O)c1ccc(N)cc1	1.1501860	0.83	1.40966
79-83-4	CC(C)(CO)C(O)C(=O)NCCC(O)=O	0.8712608	−1.69	−0.59548

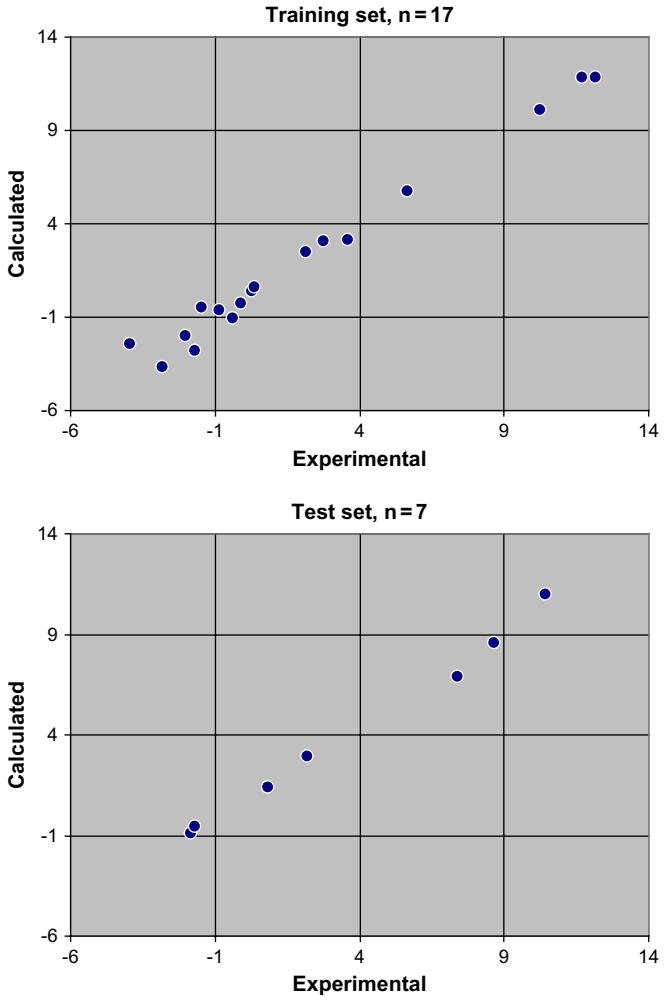


Fig. 3. Plot of experimental versus calculated values of the octanol/water partition coefficient for vitamins: probe 1 of the DCW(5) optimization.

Table 9
Splitting into training and test sets; experimental and calculated values of octanol/water partition coefficient for organic compounds (non-vitamins)

CAS No.	SMILES notation	DCW(1)	Expr	Calc
	Training set			
71-55-6	CC(Cl)(Cl)Cl	1.2777820	2.481	2.54546
79-34-5	ClC(Cl)C(Cl)Cl	1.3401565	2.644	3.12821
527-53-7	Cc1cc(C)c(C)c(C)c1	1.4998545	4.738	4.62024
87-61-6	Clc1cccc(Cl)c1Cl	1.4589229	4.281	4.23783
95-93-2	Cc1cc(C)c(C)cc1C	1.5010943	4.738	4.63182
583-53-9	Brc1cccc1Br	1.4049742	3.588	3.73379
95-50-1	Clc1cccc1Cl	1.3921707	3.568	3.61417
103-29-7	C(Cc1ccccc1)c2ccccc2	1.5807279	4.888	5.37582
541-73-1	Clc1cccc(Cl)c1	1.3910208	3.568	3.60343
575-41-7	Cc2cc(C)cc1ccccc12	1.4828092	4.614	4.46099
106-37-6	Brc1ccc(Br)cc1	1.4038137	3.868	3.72295
106-46-7	Clc1ccc(Cl)cc1	1.3910208	3.568	3.60343
571-58-4	Cc1ccc(C)c2ccccc12	1.4828092	4.614	4.46099
106-98-9	C=CCC	1.2669754	2.266	2.44450
629-06-1	CCCCCCCCl	1.4411311	4.110	4.07160
90-13-1	Clc2cccc1ccccc12	1.4259632	4.029	3.92989
540-54-5	CCCCl	1.2144016	1.994	1.95331
592-41-6	CCCCC=C	1.3801897	3.324	3.50224

Table 9 (*continued*)

CAS No.	SMILES notation	DCW(1)	Expr	Calc
2498-77-3	CC2/C=C\Cc1ccc3cc4ccccc4cc3c12	1.6937238	6.313	6.43152
90-12-0	Cc2cccc1ccccc12	1.4218665	3.965	3.89161
37680-66-3	Clc2cc(Cl)ccc2c1ccccc1Cl	1.6707780	6.169	6.21715
540-84-1	CC(C)CC(C)(C)C	1.5184749	4.536	4.79421
581-40-8	Cc1cc2ccccc2cc1C	1.4840349	4.614	4.47244
34883-43-7	Clc2ccccc2c1ccc(Cl)cc1	1.5930156	5.456	5.49063
35693-92-6	Clc2cc(Cl)cc(Cl)c2c1ccccc1	1.6693981	6.169	6.20425
34883-39-1	Clc2ccc(Cl)cc2c1ccccc1	1.5930156	5.456	5.49063
33146-45-1	Clc2cccc(Cl)c2c1ccccc1	1.5930156	5.456	5.49063
91-58-7	Clc1ccc2ccccc2c1	1.4259632	4.029	3.92989
95-49-8	Cc1ccccc1Cl	1.3854123	3.504	3.55103
613-12-7	Cc1ccc2cc3ccccc3cc2c1	1.5514004	5.139	5.10182
591-76-4	CC(C)CCCC	1.4572700	4.267	4.22238
107-83-5	CC(C)CCC	1.3962228	3.738	3.65203
106-43-4	Cc1ccc(Cl)cc1	1.3842681	3.504	3.54034
3697-24-3	Cc2cc4ccccc4c1C\C=C3\C=C/C=C\C3c12	1.6726967	6.313	6.23507
2541-69-7	Cc3c4ccc1C/C=C\Cc1c4cc2ccccc23	1.6937238	6.313	6.43152
781-43-1	Cc2c1ccccc1c(C)c3ccccc23	1.6178951	5.788	5.72307
83-32-9	c1cc2cccc3CCc(c1)c23	1.4236613	4.070	3.90838
281-23-2	C1C3CC2CC(CC1C2)C3	1.4562297	3.982	4.21266
120-12-7	c1ccc2cc3ccccc3cc2c1	1.4864099	4.490	4.49463
92-24-0	c1cccc2cc3cc4ccccc4cc3cc12	1.6097791	5.664	5.64724
71-43-2	c1ccccc1	1.2670229	2.142	2.44494
50-32-8	C1\C=C2\C=C/Cc3ccc4cc5ccccc5c1c4c23	1.6545516	6.124	6.06554
30777-19-6	c1cccc2Cc3cc4ccccc4cc3c12	1.5760854	5.399	5.33245
191-24-2	c3ccc2/C=C\C1=C\Cc5ccc6C\C=C/c4c6c5c1c2c34	1.7101095	6.584	6.58461
56-23-5	ClC(Cl)(Cl)Cl	1.2840153	2.875	2.60370
218-01-9	C\3=C\C=C/C4=C/Cc2c(ccc1ccccc12)C/34	1.6025798	5.664	5.57998
110-82-7	C1CCCCC1	1.3520126	3.354	3.23898
292-64-8	C1CCCCCCC1	1.4728257	4.472	4.36772
53-70-3	C\1=C4/C=5C/C=C\CC=5\C=C/C4=C/C3C/1/C=C\C2/C=C\C=C/C23	1.7435861	6.838	6.89738
352-93-2	CCSCC	1.2018906	1.900	1.83642
75-18-3	CSC	1.1033018	0.842	0.91533
75-00-3	CCCl	1.1635285	1.465	1.47801
462-06-6	Fc1ccccc1	1.2498420	2.285	2.28442
87-85-4	Cc1c(C)c(C)c(C)c(C)c1C	1.6325289	6.036	5.85979
591-50-4	Ic1ccccc1	1.3547123	3.265	3.26421
608-93-5	Clc1cc(Cl)c(Cl)c(Cl)c1Cl	1.6021830	5.707	5.57628
76-01-7	ClC(Cl)C(Cl)(Cl)Cl	1.4044148	3.627	3.72857
700-12-9	Cc1cc(C)c(C)c(C)c1C	1.5654328	5.387	5.23293
85-01-8	c2cc3ccc1ccccc1c3cc2	1.4864099	4.490	4.49463
127-18-4	Cl/C(Cl)=C(/Cl)Cl	1.3043395	3.020	2.79358
79-01-6	Cl\C=C(/Cl)Cl	1.2345513	2.267	2.14157
108-38-3	Cc1cccc(C)c1	1.3791015	3.440	3.49207
111-65-9	CCCCCCCC	1.5222436	4.926	4.82942
142-82-5	CCCCCCC	1.4584746	4.397	4.23364
104-51-8	CCCCc1ccccc1	1.5035770	4.738	4.65502
109-66-0	CCCCC	1.3388387	3.339	3.11590
106-42-3	Cc1ccc(C)cc1	1.3791015	3.440	3.49207
98-06-6	CC(C)(C)c1ccccc1	1.5010943	4.118	4.63182
156-60-5	Cl\C=C\Cl	1.1684971	1.514	1.52443
	Test set			
634-66-2	Clc1ccc(Cl)c(Cl)c1Cl	1.5288759	4.994	4.89138
488-23-3	Cc1ccc(C)c(C)c1C	1.5010943	4.738	4.63182
634-90-2	Clc1cc(Cl)c(Cl)c(Cl)c1	1.5276132	4.994	4.87958

Table 9 (*continued*)

CAS No.	SMILES notation	DCW(1)	Expr	Calc
526-73-8	Cc1cccc(C)c1C	1.4394001	4.089	4.05543
95-94-3	Clc1cc(Cl)c(Cl)cc1Cl	1.5288759	4.738	4.89138
120-82-1	Clc1cc(Cl)c(Cl)cc1	1.4577180	4.281	4.22657
95-63-6	Cc1cc(C)c(C)cc1	1.4382112	4.089	4.04432
106-93-4	BrCCBr	1.1952114	1.738	1.77402
107-06-2	ClCCCl	1.2203257	1.458	2.00866
108-70-3	Clc1cc(Cl)cc(Cl)c1	1.4577180	4.281	4.22657
108-67-8	Cc1cc(C)cc(C)c1	1.4382112	4.089	4.04432
2131-41-1	Cc1ccc(C)c2cccc(C)c12	1.5463640	5.263	5.05477
571-61-9	Cc2cccc1c2cccc1C	1.4840349	4.614	4.47244
109-69-3	CCCCCl	1.2674990	2.523	2.44939
544-10-5	CCCCCCCl	1.3807600	3.581	3.50756
543-59-9	CCCCCCl	1.3229180	3.052	2.96716
1127-76-0	CCc2cccc1ccccc12	1.4840349	4.494	4.47244
99-87-6	Cc1ccc(cc1)C(C)C	1.5010943	4.368	4.63182
1730-37-6	Cc2cccc1c3ccccc3Cc12	1.5189286	4.874	4.79845
109-67-1	C=CCCC	1.3223715	2.795	2.96205
2422-79-9	Cc3c2c(ccc1C/C=C\Cc12)cc4ccccc34	1.6923249	6.313	6.41845
70362-47-9	Clc2cc(c1ccccc1Cl)c(Cl)cc2Cl	1.7508891	6.882	6.96561
33284-53-6	Clc1c(cc(Cl)c(Cl)c1Cl)c2ccccc2	1.7494430	6.882	6.95210
15862-07-4	Clc2cc(c1ccccc1)c(Cl)cc2Cl	1.6693981	6.169	6.20425
581-42-0	Cc1ccc2cc(C)ccc2c1	1.4828092	4.614	4.46099
2051-60-7	Clc1cc(ccc1)c2ccccc2	1.5188725	4.743	4.79792
No CAS[a]	Clc1cc2/C=C\C3/C=C\C=C/C3c2cc1	1.5803166	5.203	5.37198
78-78-4	CC(C)CC	1.3377329	3.209	3.10557
91-57-6	Cc1ccc2ccccc2c1	1.4218665	3.965	3.89161
2531-84-2	CC=1\C=C/C2c3ccccc3C\C=C2\C=1	1.5255088	5.139	4.85992
108-41-8	Cc1cc(Cl)ccc1	1.3842681	3.504	3.54034
3697-27-6	Cc2c4ccccc4c1C\C=C3\C=C/C=C\C3c1c2C	1.7458322	6.962	6.91836
41699-04-1	C\C2=C\Cc3c1ccccc1c5C/C=C\c4ccc2c3c45	1.7003133	6.773	6.49309
1705-85-7	Cc2cc1C4/C=C\C=C/C4=C/Cc1c3ccccc23	1.7002017	6.313	6.49204
3697-30-1	CCc3c4ccc1C/C=C\Cc1c4cc2ccccc23	1.7677787	6.842	7.12340
779-02-2	Cc2c3ccccc3cc1ccccc12	1.5514004	5.139	5.10182
56-55-3	c2cc1C/C=C\Cc1c3cc4ccccc4cc23	1.6227712	5.664	5.76863
238-84-6	c1ccc2c3Cc4ccccc4c3ccc2c1	1.5760854	5.399	5.33245
205-99-2	c1cccc5c1cc4c2c5cccc2c3ccccc34	1.6515890	6.124	6.03787
192-97-2	c2cc1C\C=C/c3c1c5c2C\C=C/c5c4ccccc34	1.6424240	6.124	5.95224
205-82-3	c1cc5ccccc5c4c1c2cccc3cccc4c23	1.6515890	6.124	6.03787
207-08-9	c3c5c1cccc2cccc(c12)c5cc4ccccc34	1.6502249	6.124	6.02512
92-52-4	c1cc(ccc1)c2ccccc2	1.4510638	4.030	4.16440
108-86-1	Brc1ccccc1	1.3413165	3.005	3.13905
479-23-2	c2c1c5ccccc5ccc1c4CCc3cccc2c34	1.6877213	6.418	6.37544
108-90-7	Clc1ccccc1	1.3273754	2.855	3.00880
291-64-5	C1CCCCCC1	1.4111268	3.913	3.79128
110-83-8	C1\C=C/CCC1	1.3051916	2.810	2.80154
287-92-3	C1CCCC1	1.2953748	2.795	2.70983
142-29-0	C1\C=C/CC1	1.2505152	2.251	2.29071
224-41-9	c2cc1C/C=C\Cc1c3cc4c5ccccc5ccc4cc23	1.7748630	6.838	7.18959
100-41-4	CCc1ccccc1	1.3802415	3.320	3.50272
206-44-0	c1cccc4c1c2cccc3cccc4c23	1.5100608	4.950	4.71560
86-73-7	c1cccc2Cc3ccccc3c12	1.4552984	4.225	4.20396
75-69-4	ClC(Cl)(Cl)F	1.2321240	2.435	2.11889
118-74-1	Clc1c(Cl)c(Cl)c(Cl)c(Cl)c1Cl	1.6790050	6.420	6.29401
98-82-8	CC(C)c1ccccc1	1.4394001	3.719	4.05543
91-20-3	c1cccc2ccccc12	1.3623024	3.316	3.33512
198-55-0	c2ccc1C\C=C/c3c1c2c5C\C=C/c4cccc3c45	1.6424240	6.124	5.95224
129-00-0	c1ccc2ccc3cccc4ccc1c2c34	1.5100608	4.950	4.71560
108-88-3	Cc1ccccc1	1.3224211	2.791	2.96252
217-59-4	c3cc4c1ccccc1c2ccccc2c4cc3	1.6097791	5.664	5.64724
124-18-5	CCCCCCCCCC	1.6582681	5.984	6.10027
111-84-2	CCCCCCCCC	1.5888008	5.455	5.45125
1120-21-4	CCCCCCCCCCC	1.7307727	6.513	6.77766
106-97-8	CCCC	1.2827528	2.810	2.59190
110-54-3	CCCCCC	1.3973769	3.868	3.66281
103-65-1	CCCc1ccccc1	1.4405899	3.849	4.06654
95-47-6	Cc1ccccc1C	1.3802415	3.440	3.50272
2049-95-8	CCC(C)(C)c1ccccc1	1.5667268	4.647	5.24502

[a] 2-Chlorophenanthrene.

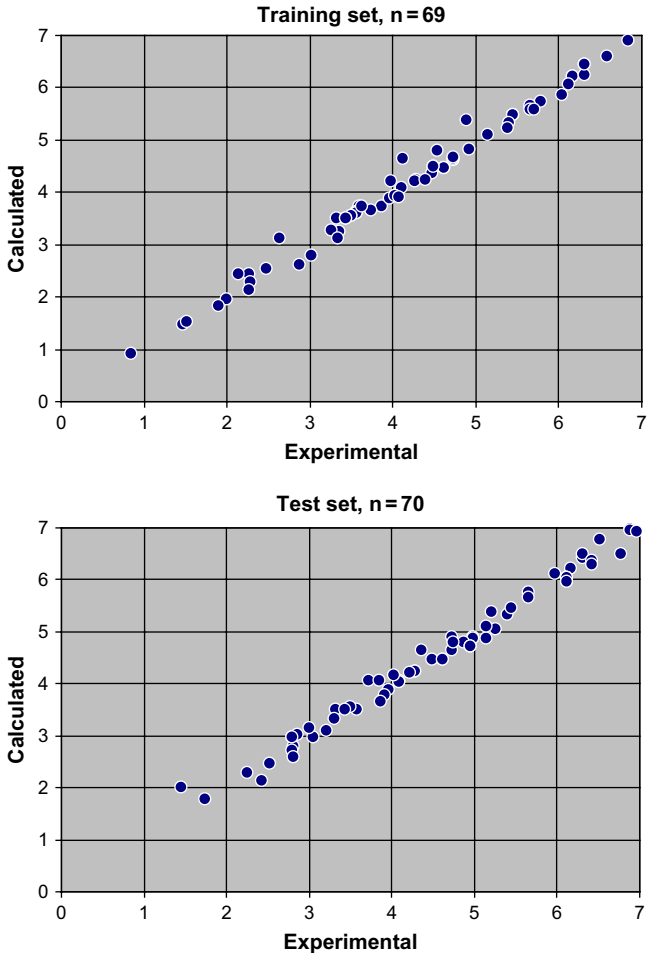


Fig. 4. Plot of experimental versus calculated values of the octanol/water partition coefficient for organic compounds (non-vitamins): probe 1 of the DCW(1) optimization.

4. Conclusions

SMILES based optimal descriptors give a reasonably good model for octanol/water partition coefficient of vitamins and organic compounds of different classes (non-vitamins).

References

[1] Qian-Nan Hu, Yi-Zeng Liang, Kai-Tai Fang, J. Data Sci. 1 (2003) 361–389.
[2] P.R. Duchowicz, E.A. Castro, A.A. Toropov, E. Benfenati, Top. Heterocycl. Chem. 3 (2006) 1–38.
[3] I. Raska Jr., A. Toropov, Eur. J. Med. Chem. 41 (2006) 1271–1278.
[4] I. Raska, A. Toropov, I. Gutman, D. Zavrsnik, S. Spirtovic, in: I. Gutman (Ed.), Mathematical Methods in Chemistry, Prijepolje Museum, Prijepolje, 2006, pp. 197–208.
[5] A.A. Toropov, A.P. Toropova, D.V. Mukhamedzhanova, I. Gutman, Indian J. Chem. 44A (2005) 1545–1552.
[6] A. Toropov, K. Nesmerak, I. Raska Jr., K. Waisser, K. Palat, Comput. Biol. Chem. 30 (2006) 434–437.
[7] A.A. Toropov, E. Benfenati, Comput. Biol. Chem. 31 (2007) 57–60.