ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

UDC 541.66:544.015.32
E. Pesterev[1],
L. Ognichenko[1], PhD,
A. Artemenko[1], PhD,
M. Kichova[1],
N. Muratov[2], PhD, Assoc. Prof.,
V. Kuz'min[1], DSc., Prof.
[1] O.V. Bogatsky Physico-chemical institute of the NAS of Ukraine, 86, Lustdorfska doroga, Odesa, Ukraine, 65080, E-mail: ogni@ukr.net
[2] Odesa Polytechnic National University, 1 Shevchenko Ave., Odesa, Ukraine, 65044

QUANTITATIVE STRUCTURE–PROPERTY RELATIONSHIP ANALYSIS OF CRITICAL PROPERTIES OF INDIVIDUAL ORGANIC COMPOUNDS

Є. Пестерев, Л. Огніченко, А. Артеменко, М. Кічьова, Н. Муратов, В. Кузьмін. **Кількісний аналіз зв'язку між структурою та критичними властивостями індивідуальних органічних сполук**. У роботі представлені розроблені та проаналізовані адекватні QSPR моделі щодо критичних властивостей (тиск, температура, об'єм) для бази даних, яка містить 399 органічних сполук таких класів, як насичені та ненасичені вуглеводні, ароматичні вуглеводні, гетероциклічні сполуки, спирти, прості та складні етери та їх похідні. В даній базі представлені карбон-, галоген-, оксиген-, нітроген та сульфуровмісні сполуки. На попередньому етапі роботи всі молекулярні структури сполук бази даних були змодельовані, стандартизовані, перевірені щодо зв'язності і унікальності. Розрахунок структурних дескрипторів проводився на 2D рівні моделювання молекулярної структури у рамках симплексного підходу. Для диференціації вершин симплексів використовувалися не тільки мітки, що відображали природу атомів, але й інші атомні характеристики, зокрема, ван-дер-ваальсові взаємодії із параметрів універсального силового поля, величини потенціалів інформаційних полів, які зважувалися за різними атомними характеристиками (частковий заряд, поляризуємість, електронегативність, ліпофільність). Для кожної молекули разом з симплексами (чотирьохатомними фрагментами) були розраховані менші за розміром фрагменти – двійки, трійки. Всього було розраховано 4939 2D структурних параметрів. Для встановлення зв'язку між структурними дескрипторами і величинами критичних властивостей використовувався метод випадкового лісу (Random Forest – RF). Побудовані 2D RF QSPR моделі мають високу якість апроксимації ($R^2 = 0{,}99$) і прогнозуючу здатність ($R^2_{oob} = 0{,}9…0{,}97$). В результаті фізико-хімічної інтерпретації побудованих моделей показано, що найбільший вплив на критичні властивості мають електростатичні чинники. Для сполук бази, для яких відсутня інформація щодо експериментальних значень досліджуваних властивостей, приведені результати прогнозу і оцінено область застосовності розроблених QSPR моделей.
Ключові слова: критичні властивості, симплексний підхід, інформаційне поле молекул, метод випадкового лісу, QSPR

E. Pesterev, L. Ognichenko, A. Artemenko, M. Kichova, N. Muratov, V. Kuz'min. **Quantitative structure-property relationship analysis of critical properties of individual organic compounds**. This work presents the development and analysis of robust QSPR models for critical properties (pressure P_C, temperature T_C, and volume V_C). The object of this study is a database consisting of 399 different organic compounds. These compounds include saturated and unsaturated hydrocarbons, aromatic hydrocarbons, heterocyclic compounds, alcohols, simple and complex ethers, and their derivatives. The database encompasses carbon-, halogen-, oxygen-, nitrogen-, and sulfur- containing compounds. The molecular structures of investigated compounds were previously modeled, standardized, and validated with respect to connectivity and uniqueness. Structural descriptors were calculated at the 2D molecular modeling level using the simplex approach. To differentiate simplex vertices, not only atom types but also other atomic characteristics were used, including van der Waals interactions (from the universal force field), and potentials of informational fields weighted by atomic properties such as partial charge, polarizability, electronegativity, and lipophilicity. In addition to simplexes, smaller fragments consisting of two and three atoms were also used for each compound of database. A total of 4,939 2D structural descriptors were calculated. The Random Forest (RF) method was applied to establish the relationships between the structural descriptors and the critical properties (P_C, T_C, V_C). The developed 2D RF QSPR models showed high approximation accuracy ($R^2 = 0.99$) and predictive ability ($R^2_{oob} = 0.90…0.97$). Physicochemical interpretation revealed that electrostatic factors have the greatest impact on critical properties. For compounds lacking experimental data on the studied properties, predictions were presented using the developed RF models, and the applicability domain of the developed QSPR models was evaluated.
Keywords: critical properties, simplex approach, informational field of molecules, random forest method, QSPR

Introduction

Knowledge of basic thermodynamic properties is essential for calculating various empirical and semi-empirical constants, particularly in equations of state that describe substance behavior under a wide range of conditions. The study of critical states is important for processes such as mixture separation and vortex formation, helping to elucidate the mechanisms of condensation, evaporation, diffusion, and related phenomena [1, 2]. This understanding is vital for developing new research methods and controlling such processes. As critical temperature, pressure, and volume are fundamental in studying phase behavior, fast and accurate models are required to predict these properties.

DOI: 10.15276/opu.1.71.2025.24

Analysis of recent publications and problem statement

Quantitative structure-property relationship (QSAR/QSPR) studies are widely used in solving chemical problems. However, most existing methods for predicting critical properties are applicable only to homogeneous classes of compounds and often require additional experimental data.

The earliest developed QSPR models were based on small, related compound sets, limiting their generalizability [3]. For example, the study [3] presents QSPR models for a set of 47 alkyl benzenes. Support vector regression (SVR) and machine learning technology based on statistical learning theory (SLT), integrated with topological indices was applied to the prediction of critical properties. The prediction accuracy of the model was discussed on the basis of the leave-one-out cross-validation. The drawbacks of this study include the homogeneity of the compounds in the dataset and the relatively small number of compounds [3]. A very recent study by Chinese colleagues is presented in reference [4]. This study presents a QSPR model that predicts the critical temperature of alkenes based on quantum-chemical and topological molecular descriptors. A dataset of 103 alkenes was divided into a training set (58 compounds) and a test set (45 compounds). Quantum-chemical parameters were obtained using DFT calculations with Gaussian-16, while topological indices were computed via AlvaDesc. Using stepwise multiple regression, a five-descriptor linear model was built, including parameters such as Ehomo and HATS3s. The model showed excellent statistical performance and strong predictive power on the external validation set, confirming its reliability for estimating alkene critical temperatures. Despite the very good statistical performance of the model, its drawbacks include the homogeneity of the compounds in the dataset and the complexity of calculating the quantum-chemical descriptors.

However, attempts to analyze datasets of diverse organic compounds have been undertaken only relatively recently. For example, the study [5] presents QSPR model for predicting the critical temperatures of 165 structurally diverse organic compounds. The model utilizes hybrid optimal descriptors derived from both molecular graphs and SMILES notation, processed through the CORAL software using the Monte Carlo method. A key focus of the work is robust external validation, addressing a common limitation in prior models on the same dataset, where either validation was omitted or showed weak predictive performance. Here, the dataset was split three times into training, calibration, and validation sets to assess consistency and reliability. The resulting models demonstrated strong statistical performance, with coefficient of determination (r^2) values of 0.98 for the training set, 0.95 for the calibration set, and 0.94 for the external validation set. Despite the relatively heterogeneous dataset and good statistical performance of the model, its limitations include the relatively small number of compounds and the low interpretability of the descriptors used.

Given the accumulation of experimental data of critical temperature, pressure, and volume on various compound classes, developing models for broader, more representative sets is timely and necessary.

The purpose of research

The aim of this work is to develop reliable QSPR models with high predictive ability for critical properties across a diverse database of organic compounds, utilizing a simplex-based representation of molecular structure.

Consequently, to accomplish the stated goal, it is necessary to perform the following tasks:

– model the molecular structures of compounds from the database, perform *data curation*;

– calculate the structural descriptors using the simplex approach;

– develop adequate RF-QSPR models for critical properties of the studied compounds;

– assess the applicability domain (AD) of each model;

– a perform the physicochemical interpretation of the developed models.

Materials and methods of research

The object of this study is a database consisting of 399 organic compounds of different classes (see Table 1), for which known values of critical properties [6] were available within the following ranges:

– Critical pressure (P_C): data available for 371 compounds; range: 9.6…85 atm;

– Critical temperature (T_C): data available for 395 compounds; range: 190.6…926 K;

– Critical volume (V_C): data available for 301 compounds; range: 99…1000 mL/mol.

The database includes carbon-, halogen-, oxygen-, nitrogen-, and sulfur-containing compounds. Among the studied entries are representatives of various classes of organic compounds, including satu-

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

rated and unsaturated hydrocarbons, aromatic hydrocarbons, heterocyclic compounds, alcohols, simple and complex ethers, and their derivatives.

Table 1

The values of critical properties of the studied compounds

№	Compound	empirical formula	P_C, atm		T_C, K		V_C, mL/mol	
			Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1	methane	CH_4	45.4	60.5	190.6	389.1	99	130
2	tetrafluoromethane	CF_4	36.9	44.8	227.6	330.8	140	180
3	ethene	C_2H_4	49.7	55.9	282.4	356.5	129	140
4	chlorotrifluoromethane	$CClF_3$	38.7	40.5	302	359.4	180	179
5	1,1-difluoroethene	$C_2H_2F_2$	44	51.3	302.8	358.0	154	168
6	ethane	C_2H_6	48.2	54.3	305.4	404.9	148	164
7	tetrafluoroethene	C_2F_4	38.9	44.2	306.4	337.8	175	179
8	acetylene	C_2H_2	60.6	57.7	308.3	384.1	113	132
9	fluoromethane	CH_3F	58	55.3	317.8	350.0	124	130
10	fluoroethene	C_2H_3F	51.7	51.2	327.8	378.9	144	151
11	bromotrifluoromethane	$CBrF_3$	39.2	43.7	340.2	395.6	200	201
12	1,1,1-trifluoroethane	$C_2H_3F_3$	37.1	44.2	346.2	378.6	221	194
13	1-chloro-1,1,2,2,2-pentafluoroethane	C_2ClF_5	31.2	34.6	353.2	405.0	252	250
14	prop-1-ene	C_3H_6	45.6	51.2	365	434.8	181	186
15	chlorodifluoromethane	$CHClF_2$	49.1	48.3	369.2	376.0	165	177
16	propane	C_3H_8	41.9	44.9	369.8	443.9	203	209
17	fluoroethane	C_2H_5F	49.6	50.0	375.3	410.9	169	168
18	ethenone	C_2H_2O	64	60.5	380	447.8	145	138
19	dichlorodifluoromethane	CCl_2F_2	40.7	43.3	385	432.4	217	215
20	1,1-difluoroethane	$C_2H_4F_2$	44.4	46.7	386.6	368.2	181	202
21	propa-1,2-diene	C_3H_4	54	50.8	393	366.4	162	152
22	cyclopropane	C_3H_6	54.2	49.1	397.8	429.1	170	184
23	methoxymethane	C_2H_6O	53	52.8	400	464.4	178	180
24	prop-1-yne	C_3H_4	55.5	51.6	402.4	413.7	164	157
25	2-methylpropane	C_4H_{10}	36	38.6	408.1	445.9	263	264
26	1-chloro-1,1-difluoroethane	$C_2H_3ClF_2$	40.7	43.2	410.2	423.3	231	222
27	chloromethane	CH_3Cl	65.9	64.0	416.3	436.3	139	161
28	2-methylprop-1-ene	C_4H_8	39.5	41.4	417.9	441.6	239	234
29	1,1-dichloro-1,2,2,2-tetrafluoroethane	$C_2Cl_2F_4$	32.6	34.0	418.6	437.4	294	264
30	1,2-dichloro-1,1,2,2-tetrafluoroethane	$C_2Cl_2F_4$	32.2	34.2	418.9	434.7	293	272
31	but-1-ene	C_4H_8	39.7	42.3	419.6	457.4	240	233
32	buta-1,3-diene	C_4H_6	42.7	46.4	425	454.4	221	220
33	butane	C_4H_{10}	37.5	39.3	425.2	466.5	255	274
34	chloroethene	C_2H_3Cl	55.3	51.3	429.7	459.8	169	187
35	methanamine	CH_5N	73.6	61.5	430	442.5	140	144
36	trimethylamine	C_3H_9N	40.2	44.9	433.2	479.8	254	238
37	2,2-dimethylpropane	C_5H_{12}	31.6	34.8	433.8	508.7	303	304
38	methoxyethene	C_3H_6O	47	48.4	436	480.8	205	207
39	1,1,2-trichloro-1,2,2-trifluoroethane	$C_2Cl_3F_3$	33.7	35.8	437.2	497.3	304	290
40	dimethylamine	C_2H_7N	52.4	54.6	437.6	469.3	187	183
41	methoxyethane	C_3H_8O	43.4	44.4	437.8	504.5	221	224
42	but-1-yne	C_4H_6	46.5	49.0	440	467.6	220	218
43	buta-1,2-diene	C_4H_6	44.4	45.7	443.7	438.8	219	220
44	3-methylbut-1-ene	C_5H_{10}	34.7	37.3	450	486.3	300	292
45	dichloro(fluoro)methane	$CHCl_2F$	51	49.1	451.6	482.3	197	208
46	tetradecafluorohexane	C_6F_{14}	18.8	24.7	451.7	492.3	442	452

1	2	3	4	5	6	7	8	9
47	but-1-en-3-yne	C_4H_4	49	51.4	455	451.8	202	187
48	ethanamine	C_2H_7N	55.5	54.1	456	501.6	178	189
49	cyclobutane	C_4H_8	49.2	44.1	459.9	445.7	210	232
50	2-methylbutane	C_5H_{12}	33.4	33.8	460.4	495.4	306	315
51	chloroethane	C_2H_5Cl	52	54.8	460.4	486.0	199	211
52	acetaldehyde	C_2H_4O	55	57.3	461	450.9	154	157
53	pent-1-ene	C_5H_{10}	40	36.9	464.7	493.2	300	294
54	2-methylbut-1-ene	C_5H_{10}	34	37.2	465	480.2	294	295
55	ethoxyethane	$C_4H_{10}O$	35.9	38.8	466.7	524.3	280	279
56	oxirane	C_2H_4O	71	56.9	469	429.9	140	160
57	pentane	C_5H_{12}	33.3	35.1	469.6	498.2	304	312
58	2-methylbut-2-ene	C_5H_{10}	34	35.9	470	480.0	318	291
59	methanethiol	CH_4S	71.4	56.9	470	459.0	145	180
60	trichloro(fluoro)methane	CCl_3F	43.5	46.1	471.2	520.7	248	245
61	hexadecafluoroheptane	C_7F_{16}	16	23.5	474.8	530.1	664	505
62	ethoxyethene	C_4H_8O	40.2	41.4	475	507.3	260	271
63	ethenyl formate	$C_3H_4O_2$	57	51.6	475	514.4	210	201
64	propan-2-amine	C_3H_9N	50	47.6	476	499.2	229	243
65	penta-1,4-diene	C_5H_8	37.4	40.1	478	511.7	276	267
66	2-methyloxirane	C_3H_6O	48.6	49.1	482.2	484.2	186	213
67	2-methylbuta-1,3-diene	C_5H_8	38	39.7	484	479.2	276	272
68	2-chloropropane	C_3H_7Cl	46.6	45.1	485	514.1	230	247
69	methyl formate	$C_2H_4O_2$	59.2	54.4	487.2	479.1	172	181
70	but-2-yne	C_4H_6	50.2	44.6	488.6	422.2	221	213
71	2,2-dimethylbutane	C_6H_{14}	30.4	31.5	488.7	514.4	359	355
72	3,3-dimethylbut-1-ene	C_6H_{12}	32.1	32.5	490	510.9	340	342
73	furan	C_4H_4O	54.3	51.6	490.2	519.9	218	209
74	pent-1-yne	C_5H_8	40	40.6	493.4	485.2	278	265
75	(3E)-penta-1,3-diene	C_5H_8	39.4	40.0	496	480.8	275	280
76	3-methylbuta-1,2-diene	C_5H_8	40.6	37.7	496	479.3	267	277
77	propanal	C_3H_6O	47	49.8	496	499.4	223	210
78	diethylamine	$C_4H_{10}N$	36.6	40.5	496.6	523.6	301	285
79	propan-1-amine	C_3H_9N	46.8	49.3	497	522.9	233	241
80	2-methylpentane	C_6H_{14}	29.7	31.1	497.5	517.6	367	360
81	ethanethiol	C_2H_6S	54.2	51.1	499	504.4	207	216
82	2,3-dimethylbutane	C_6H_{14}	30.9	30.4	499.9	523.9	358	367
83	2-(propan-2-yloxy)propane	$C_6H_{14}O$	28.4	30.6	500	544.5	386	396
84	2,3-dimethylbut-1-ene	C_6H_{12}	32	32.4	501	515.3	343	343
85	penta-1,2-diene	C_5H_8	40.2	38.3	503	481.7	276	274
86	1-chloropropane	C_3H_7Cl	45.2	44.8	503	516.0	254	238
87	(methylsulfanyl)methane	C_2H_6S	54.6	53.3	503	461.7	201	213
88	bromoethane	C_2H_5Br	61.5	51.2	503.8	517.9	215	216
89	hex-1-ene	C_6H_{12}	31.3	35.0	504	528.6	350	339
90	3-methylpentane	C_6H_{14}	30.8	30.3	504.4	514.9	367	364
91	2-methylpropan-2-ol	$C_4H_{10}O$	39.2	41.2	506.2	522.8	275	266
92	methyl acetate	$C_3H_6O_2$	46.3	46.3	506.8	510.7	228	226
93	hexa-1,5-diene	C_6H_{10}	34	38.0	507	538.0	328	323
94	2-chloro-2-methylpropane	C_4H_9Cl	39	38.6	507	526.3	295	290
95	hexane	C_6H_{14}	29.3	32.0	507.4	513.9	370	364
96	acetyl chloride	C_2H_3ClO	58	52.3	508	502.0	204	204
97	propan-2-one	C_3H_6O	46.4	49.3	508.1	484.6	209	213
98	propan-2-ol	C_3H_8O	47	50.2	508.3	514.2	220	229
99	ethyl formate	$C_3H_6O_2$	46.8	47.5	508.4	515.5	229	227
100	dichloromethane	CH_2Cl_2	60	59.9	510	485.9	193	200
101	cyclopentane	C_5H_{10}	44.5	38.8	511.6	477.0	260	291
102	methanol	CH_4O	79.9	63.0	512.6	430.0	118	142
103	2-methylpropanal	C_4H_8O	41	42.6	513	523.7	274	255

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

1	2	3	4	5	6	7	8	9
104	3-chloroprop-1-ene	C_3H_5Cl	47	50.0	514	511.0	234	222
105	2-methylpropan-1-amine	$C_4H_{10}N$	42	40.1	516	523.6	284	293
106	ethanol	C_2H_6O	63	54.4	516.2	510.7	167	179
107	2-methylpent-2-ene	C_6H_{12}	32.4	31.3	518	510.5	351	348
108	2,4-dimethylpentane	C_7H_{16}	27	28.4	519.7	535.1	418	408
109	2,2-dimethylpentane	C_7H_{16}	27.4	28.3	520.4	534.1	416	406
110	2-chlorobutane	C_4H_9Cl	39	39.1	520.6	544.6	305	294
111	1,1-dichloroethane	$C_2H_4Cl_2$	50	48.1	523	518.6	240	235
112	ethyl acetate	$C_4H_8O_2$	37.8	41.0	523.3	546.7	286	283
113	2,3-dimethylbut-2-ene	C_6H_{12}	33.2	31.3	524	508.2	351	338
114	butanal	C_4H_8O	40	42.7	524	533.1	278	270
115	butan-1-amine	$C_4H_{10}N$	41	42.6	524	538.9	288	295
116	ethenyl acetate	$C_4H_6O_2$	43	44.5	525	528.7	265	265
117	iodomethane	CH_3I	65	69.7	528	510.2	190	227
118	2-methylhexane	C_7H_{16}	27	28.5	530.3	538.5	421	412
119	methyl propanoate	$C_4H_8O_2$	39.5	39.2	530.6	547.5	282	284
120	1-ethoxybutane	$C_6H_{14}O$	30	31.2	531	562.6	390	402
121	2,2,3-trimethylbutane	C_7H_{16}	29.2	28.2	531.1	542.3	398	417
122	methylcyclopentane	C_6H_{12}	37.4	36.0	532.7	543.5	319	328
123	2,3,3-trimethylbut-1-ene	C_7H_{14}	28.6	30.7	533	517.8	400	387
124	triethylamine	$C_6H_{15}N$	30	31.9	535	532.1	390	391
125	3-methylhexane	C_7H_{16}	27.8	28.7	535.2	540.2	404	418
126	butan-2-one	C_4H_8O	41	39.9	535.6	522.7	267	272
127	1,2-dimethoxyethane	$C_4H_{10}O_2$	38.2	40.3	536	552.2	271	297
128	butan-2-ol	$C_4H_{10}O$	41.4	41.7	536	546.3	268	274
129	methyl prop-2-enoate	$C_4H_6O_2$	42	43.1	536	523.0	265	252
130	prop-2-enenitrile	C_3H_3N	45	50.8	536	449.6	210	166
131	3,3-dimethylpentane	C_7H_{16}	29.1	28.4	536.3	532.9	414	405
132	trichloromethane	$CHCl_3$	54	50.6	536.4	518.3	239	231
133	propan-1-ol	C_3H_8O	51	47.9	536.7	555.9	218.5	236
134	hept-1-ene	C_7H_{14}	28	32.2	537.2	554.3	440	387
135	2,3-dimethylpentane	C_7H_{16}	28.7	27.9	537.3	534.8	393	417
136	propyl formate	$C_4H_8O_2$	40.1	42.2	538	552.1	285	276
137	heptane	C_7H_{16}	27	28.1	540.2	548.8	432	419
138	oxolane	C_4H_8O	51.2	48.4	540.2	513.4	224	255
139	3-ethylpentane	C_7H_{16}	28.5	28.1	540.6	538.2	416	412
140	methyl 2-methylpropanoate	$C_5H_{10}O_2$	33.9	35.0	540.8	554.1	339	340
141	1-chlorobutane	C_4H_9Cl	36.4	40.2	542	534.2	312	298
142	2,2,4-trimethylpentane	C_8H_{18}	25.3	26.3	543.9	555.4	468	443
143	2-methylbutan-2-ol	$C_5H_{12}O$	39	36.7	545	558.1	319	328
144	prop-2-ynal	C_3H_6O	56.4	50.0	545	535.4	203	211
145	1,1-dimethylcyclopentane	C_7H_{14}	34	30.4	547	528.0	360	389
146	2-methylpropan-1-ol	$C_4H_{10}O$	42.4	42.0	547.7	560.6	273	273
147	acetonitrile	C_2H_3N	47.7	59.6	548	415.6	173	146
148	2,2-dimethylpropan-1-ol	$C_5H_{12}O$	39	37.0	549	569.4	319	329
149	propyl acetate	$C_5H_{10}O_2$	32.9	36.2	549.4	554.1	345	339
150	2,2-dimethylhexane	C_8H_{18}	25	25.9	549.8	562.3	478	458
151	2,5-dimethylhexane	C_8H_{18}	24.5	25.3	550	574.4	482	466
152	dipropylamine	$C_6H_{15}N$	31	31.6	550	562.0	407	407
153	2-methylpropyl formate	$C_5H_{10}O_2$	38.3	35.1	551	555.4	350	328
154	ethyl prop-2-enoate	$C_5H_8O_2$	37	39.7	552	550.6	320	299
155	ethyl 2-methylpropanoate	$C_6H_{12}O_2$	30	30.7	553	566.8	410	402
156	3-methylbutan-2-one	$C_5H_{10}O$	38	36.3	553.4	545.2	310	314
157	cyclohexane	C_6H_{12}	40.2	36.7	553.4	553.4	308	332
158	2,4-dimethylhexane	C_8H_{18}	25.2	25.6	553.5	560.7	472	470
159	pentanal	$C_5H_{10}O$	35	38.4	554	562.3	333	311
160	methyl butanoate	$C_5H_{10}O_2$	34.3	34.8	554.4	550.6	340	339

1	2	3	4	5	6	7	8	9
161	tetrachloromethane	CCl_4	45	44.5	556.4	529.6	276	275
162	(ethylsulfanyl)ethane	$C_4H_{10}S$	39.1	37.5	557	540.6	318	329
163	2-methylheptane	C_8H_{18}	24.5	25.3	559.6	566.2	488	472
164	fluorobenzene	C_6H_5F	44.9	45.3	560.1	609.2	271	264
165	cyclohexene	C_6H_{10}	42.9	37.5	560.4	511.1	292	311
166	2-methylpropyl acetate	$C_6H_{12}O_2$	30	30.9	561	572.1	414	395
167	pentan-3-one	$C_5H_{10}O_2$	36.9	37.7	561	552.9	336	309
168	1,2-dichloroethane	$C_2H_4Cl_2$	53	50.5	561	520.9	220	246
169	4-methylheptane	C_8H_{18}	25.1	25.7	561.7	571.8	476	469
170	3,3-dimethylhexane	C_8H_{18}	26.2	26.5	562	564.9	443	466
171	benzene	C_6H_6	48.3	47.1	562.1	593.2	259	254
172	butan-1-ol	$C_4H_{10}O$	43.6	42.2	562.9	563.5	274	278
173	2,3-dimethylhexane	C_8H_{18}	25.9	25.8	563.4	563.6	468	460
174	2,2,3-trimethylpentane	C_8H_{18}	26.9	27.2	563.4	569.8	436	454
175	3-methylheptane	C_8H_{18}	25.1	25.6	563.6	573.8	464	476
176	pentan-2-one	$C_5H_{10}O$	38.4	37.0	564	553.3	301	323
177	propanenitrile	C_3H_5N	41.3	48.5	564.4	504.0	230	198
178	3-ethylhexane	C_8H_{18}	25.7	25.8	565.4	567.8	455	466
179	ethyl butanoate	$C_6H_{12}O_2$	31	31.0	566	569.7	395	403
180	2,3,4-trimethylpentane	C_8H_{18}	26.9	26.5	566.3	555.2	461	444
181	oct-1-ene	C_8H_{16}	25.9	28.3	566.6	588.3	464	454
182	3-ethyl-2-methylpentane	C_8H_{18}	26.7	26.3	567	564.9	433	456
183	2,2,5-trimethylhexane	C_9H_{20}	23	24.3	568	585.2	519	537
184	pyrrolidine	C_4H_9N	55.4	49.7	568.6	524.6	249	246
185	octane	C_8H_{18}	24.5	25.4	568.8	576.6	492	487
186	3,4-dimethylhexane	C_8H_{18}	26.6	27.2	568.8	568.7	466	443
187	acetyl acetate	$C_4H_6O_3$	46.2	42.3	569	545.3	290	281
188	ethylcyclopentane	C_7H_{14}	33.5	31.1	569.5	560.2	375	400
189	4-methylpentan-2-one	$C_6H_{12}O$	32.3	31.4	571	556.4	371	354
190	2-methylbutan-1-ol	$C_5H_{12}O$	38	38.1	571	573.7	322	324
191	1,1,2-trichloroethene	C_2HCl_3	48.5	44.6	571	554.7	256	266
192	methylcyclohexane	C_7H_{14}	34.3	31.7	572.1	544.6	368	406
193	2,3,3-trimethylpentane	C_8H_{18}	27.8	27.0	573.5	568.5	455	445
194	3-ethyl-3-methylpentane	C_8H_{18}	27.7	26.7	576.5	577.0	455	452
195	1,2-dichloropropane	$C_3H_6Cl_2$	44	45.0	577	587.5	226	283
196	butyl acetate	$C_6H_{12}O_2$	31	30.6	579	560.6	400	395
197	thiophene	C_4H_4S	56.2	46.2	579.4	561.8	219	243
198	3-methylbutan-1-ol	$C_5H_{12}O$	38	37.9	579.5	564.7	329	324
199	1-butoxybutane	$C_8H_{18}O$	25	25.6	580	609.4	500	509
200	butanenitrile	C_4H_7N	37.4	44.6	582.2	532.7	285	245
201	but-3-enenitrile	C_4H_5N	39	46.2	585	521.2	265	224
202	pentan-1-ol	$C_5H_{12}O$	38	38.1	586	583.4	326	323
203	1,4-dioxane	$C_4H_8O_2$	51.4	46.3	587	561.2	238	264
204	nitromethane	CH_3NO_2	62.3	58.5	588	470.3	173	175
205	cycloheptane	C_7H_{14}	36.7	34.1	589	556.1	390	377
206	1,1-dimethylcyclohexane	C_8H_{16}	29.3	28.6	591	588.2	416	439
207	toluene	C_7H_8	40.6	42.0	591.7	635.1	316	305
208	non-1-ene	C_9H_{18}	23.1	26.3	592	601.4	580	504
209	ethane-1,2-diamine	$C_2H_8N_2$	62	51.6	593	545.2	206	222
210	piperidine	$C_5H_{11}N$	47	41.3	594	565.8	289	309
211	acetic acid	$C_2H_4O_2$	57.1	57.6	594.4	518.5	171	188
212	nonane	C_9H_{20}	22.8	24.3	594.6	604.6	548	542
213	dibutylamine	$C_8H_{19}N$	25	25.5	596	598.0	517	517
214	1,1,2-trichloroethane	$C_2H_3Cl_3$	41	46.2	602	554.9	294	262
215	propylcyclopentane	C_8H_{16}	29.6	28.4	603	590.2	425	461
216	ethylcyclohexane	C_8H_{16}	29.9	29.3	609	611.0	450	449
217	2-methylpropanoic acid	$C_4H_8O_2$	40	44.8	609	590.0	292	273

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

1	2	3	4	5	6	7	8	9
218	hexan-1-ol	$C_6H_{14}O$	40	32.1	610	597.4	381	395
219	propanoic acid	$C_3H_6O_2$	53	49.4	612	574.7	230	233
220	2-ethylhexan-1-ol	$C_8H_{18}O$	27.2	28.2	613	611.1	494	486
221	2-aminoethan-1-ol	C_2H_7NO	44	66.4	614	595.4	196	211
222	dec-1-ene	$C_{10}H_{20}$	21.8	26.6	615	620.0	650	560
223	prop-2-enoic acid	$C_3H_4O_2$	56	52.1	615	539.9	210	203
224	1,4-xylene	C_8H_{10}	34.7	34.7	616.2	643.4	379	381
225	1,3-xylene	C_8H_{10}	35	34.9	617	648.1	376	368
226	ethylbenzene	C_8H_{10}	35.6	35.4	617.1	652.3	374	375
227	decane	$C_{10}H_{22}$	20.8	22.2	617.6	645.0	603	641
228	morpholine	C_4H_9NO	54	49.1	618	580.4	253	274
229	tetrachloroethene	C_2Cl_4	44	42.9	620	558.1	290	288
230	pyridine	C_5H_5N	55.6	49.5	620	603.9	254	231
231	cyclohexanol	$C_6H_{12}O$	37	37.0	625	581.8	327	366
232	propane-1,2-diol	$C_3H_8O_2$	60	53.1	625	593.1	237	251
233	cyclopentanone	C_5H_8O	53	44.6	626	551.6	268	260
234	butanoic acid	$C_4H_8O_2$	52	43.2	628	593.8	292	276
235	cyclohexanone	$C_6H_{10}O$	38	40.3	629	574.1	312	326
236	1,2-xylene	C_8H_{10}	36.8	34.8	630.2	640.7	369	365
237	propan-2-ylbenzene	C_9H_{12}	31.7	31.9	631	657.9	428	443
238	chlorobenzene	C_6H_5Cl	44.6	43.1	632.4	668.5	308	317
239	heptan-1-ol	$C_7H_{16}O$	30	30.9	633	594.7	435	458
240	octan-2-ol	$C_8H_{18}O$	27	27.1	637	605.0	494	503
241	1-ethyl-3-methylbenzene	C_9H_{12}	28	30.7	637	652.8	490	442
242	1,3,5-trimethylbenzene	C_9H_{12}	30.9	32.3	637.3	661.6	433	434
243	propylbenzene	C_9H_{12}	31.6	31.4	638.3	653.8	440	458
244	undecane	$C_{11}H_{24}$	19.4	19.5	638.8	657.1	660	662
245	1-ethyl-4-methylbenzene	C_9H_{12}	29	29.6	640	648.2	470	458
246	ethane-1,2-diol	$C_2H_6O_2$	76	55.4	645	571.0	186	200
247	4-methylpyridine	C_6H_7N	44	43.4	646	609.5	311	287
248	1,2,4-trimethylbenzene	C_9H_{12}	31.9	32.1	649.1	640.9	430	443
249	(2-methylpropyl)benzene	$C_{10}H_{14}$	31	29.0	650	658.8	480	480
250	1-ethyl-2-methylbenzene	C_9H_{12}	30	30.7	651	643.5	460	448
251	pentanoic acid	$C_5H_{10}O_2$	38	38.2	651	588.6	340	334
252	1,2,3-trichloropropane	$C_3H_5Cl_3$	39	43.8	651	585.4	348	301
253	prop-1-en-2-ylbenzene	C_9H_{10}	33.6	34.7	654	650.4	397	400
254	1-(hexyloxy)hexane	$C_{12}H_{26}O$	18	17.6	657	668.0	720	752
255	tridecane	$C_{13}H_{28}$	17	17.2	657.8	680.3	780	759
256	1,4-diethylbenzene	$C_{10}H_{14}$	27.7	26.5	657.9	666.4	480	546
257	octan-1-ol	$C_8H_{18}O$	34	28.0	658	625.8	490	500
258	propane-1,3-diol	$C_3H_8O_2$	59	58.2	658	613.9	241	243
259	dodecane	$C_{12}H_{26}$	18	18.3	658.3	666.5	713	692
260	butylbenzene	$C_{10}H_{14}$	28.5	29.6	660.5	658.3	497	475
261	1,2,3-trimethylbenzene	C_9H_{12}	34.1	30.8	664.5	642.1	430	435
262	bromobenzene	C_6H_5Br	44.6	43.6	670	676.4	324	327
263	1,2,4,5-tetramethylbenzene	$C_{10}H_{14}$	29	30.5	675	669.5	480	471
264	phenylmethanol	C_7H_8O	46	47.7	677	702.6	334	324
265	dodecan-1-ol	$C_{12}H_{26}O$	19	18.8	679	682.2	718	719
266	2-(2-hydroxyethoxy)ethan-1-ol	$C_4H_{10}O_3$	46	47.7	681	608.9	316	317
267	1,3-dichlorobenzene	$C_6H_4Cl_2$	38	42.6	684	670.7	359	344
268	1,4-dichlorobenzene	$C_6H_4Cl_2$	39	42.3	685	645.3	372	331
269	methyl benzoate	$C_8H_8O_2$	36	37.9	692	691.0	396	393
270	tetradecane	$C_{14}H_{30}$	16	15.6	694	691.7	830	786
271	2-methylaniline	C_7H_9N	37	45.5	694	691.7	343	331
272	phenol	C_6H_6O	60.5	47.7	694.2	647.7	229	270
273	ethyl benzoate	$C_9H_{10}O_2$	32	34.7	697	698.4	451	452
274	1,2-dichlorobenzene	$C_6H_4Cl_2$	40.5	41.7	697.3	664.4	360	340

1	2	3	4	5	6	7	8	9
275	2-methylphenol	C_7H_8O	49.4	44.9	697.6	699.8	282	328
276	aniline	C_6H_7N	52.4	50.5	699	657.0	270	287
277	decan-1-ol	$C_{10}H_{22}O$	22	26.4	700	658.3	600	628
278	1-phenylethan-1-one	C_8H_8O	38	37.6	701	669.1	376	363
279	3-methylphenol	C_7H_8O	45	46.7	705.8	693.7	310	309
280	pentadecane	$C_{15}H_{32}$	15	14.3	707	711.1	880	842
281	3-methylaniline	C_7H_9N	41	42.1	709	681.2	343	341
282	N-butylaniline	$C_{10}H_{15}N$	28	30.7	721	685.9	518	506
283	iodobenzene	C_6H_5I	44.6	42.3	721	662.8	351	359
284	butyl benzoate	$C_{11}H_{14}O_2$	26	29.0	723	709.2	561	524
285	propane-1,2,3-triol	$C_3H_8O_3$	66	55.8	726	623.0	255	256
286	heptadecane	$C_{17}H_{36}$	13	12.4	733	738.8	1000	844
287	naphthalene	$C_{10}H_8$	40	38.7	748.4	707.1	410	349
288	benzoic acid	$C_7H_6O_2$	45	48.4	752	691.1	341	329
289	2-methylnaphthalene	$C_{11}H_{10}$	34.6	34.8	761	761.1	462	455
290	1-methylnaphthalene	$C_{11}H_{10}$	35.2	34.4	772	732.7	445	455
291	phenylbenzene	$C_{12}H_{10}$	38	35.4	789	763.6	502	456
292	1,3-dihydro-2-benzofuran-1,3-dione	$C_8H_4O_3$	47	44.4	810	665.1	368	333
293	1,2-diphenylbenzene	$C_{18}H_{14}$	38.5	33.6	891	869.5	769	725
294	1,3-diphenylbenzene	$C_{18}H_{14}$	34.6	34.3	924.8	877.7	784	732
295	1,4-diphenylbenzene	$C_{18}H_{14}$	32.8	35.3	926	881.9	779	727
296	(2Z)-4-methylpent-2-ene	C_6H_{12}	30	32.4	–	510.2 +[1]	360	348
297	hexafluoroethane	C_2F_6	–	35.3 +	292.8	374.0	224	221
298	ethyl propanoate	$C_5H_{10}O_2$	–	34.8 +	546	551.3	345	337
299	(2Z)-pent-2-ene	C_5H_{10}	–	36.7 +	–	484.1 +	300	302
300	(3Z)-hex-3-ene	C_6H_{12}	–	32.8 +	–	524.5 +	350	352
301	(2Z)-hex-2-ene	C_6H_{12}	–	32.7 +	–	520.9 +	351	352
302	hexadecylcyclopentane	$C_{21}H_{42}$	9.6	12.2	791	727.4	–	*739* –
303	pentadecylcyclopentane	$C_{20}H_{40}$	10.1	11.1	780	766.3	–	*849* –
304	nonadecane	$C_{19}H_{40}$	11	11.2	756	762.7	–	*909* –
305	icosane	$C_{20}H_{42}$	11	11.2	767	761.1	–	*780* –
306	tetradecylcyclopentane	$C_{19}H_{38}$	11.1	11.8	772	758.4	–	*912* –
307	octadec-1-ene	$C_{18}H_{36}$	11.2	13.2	739	734.4	–	932 +
308	tridecylcyclopentane	$C_{18}H_{36}$	11.9	12.0	761	750.0	–	930 +
309	octadecane	$C_{18}H_{38}$	11.9	11.9	745	750.3	–	929 +
310	icosan-1-ol	$C_{20}H_{42}O$	12	15.3	770	715.4	–	*717* –
311	dodecylcyclopentane	$C_{17}H_{34}$	12.8	12.8	750	739.7	–	916 +
312	hexadec-1-ene	$C_{16}H_{32}$	13.2	14.5	717	717.2	–	868 +
313	decylcyclohexane	$C_{16}H_{32}$	13.4	13.5	750	732.4	–	*897* -
314	hexadecane	$C_{16}H_{34}$	14	13.6	717	730.7	–	*892* -
315	heptadecan-1-ol	$C_{17}H_{36}O$	14	13.1	736	741.6	–	*909* -
316	octadecan-1-ol	$C_{18}H_{38}O$	14	12.6	747	744.9	–	*887* -
317	pentadec-1-ene	$C_{15}H_{30}$	14.4	15.6	704	694.9	–	821 +
318	decylcyclopentane	$C_{15}H_{30}$	15	14.5	723.8	711.4	–	*838* –
319	tetradec-1-ene	$C_{14}H_{28}$	15.4	17.1	689	681.0	–	769 +
320	nonylcyclopentane	$C_{14}H_{28}$	16.3	15.6	710.5	692.4	–	*821* –
321	tridec-1-ene	$C_{13}H_{26}$	16.8	18.5	674	663.9	–	754 +
322	octylcyclopentane	$C_{13}H_{26}$	17.7	17.8	694	678.6	–	*737* –
323	tributylamine	$C_{12}H_{27}N$	18	19.6	643	664.9	–	*696* –
324	dodec-1-ene	$C_{12}H_{24}$	18.3	19.6	657	655.9	–	*711* –
325	heptylcyclopentane	$C_{12}H_{24}$	19.2	18.0	679	673.4	–	722 +
326	undec-1-ene	$C_{11}H_{22}$	19.7	22.0	637	656.9	–	645 +
327	hexylcyclopentane	$C_{11}H_{22}$	21.1	20.5	660.1	664.3	–	664 +
328	2,2,5,5-tetramethylhexane	$C_{10}H_{22}$	21.6	21.9	581.5	614.0	–	*610* –

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

1	2	3	4	5	6	7	8	9
329	3,3,5-trimethylheptane	$C_{10}H_{22}$	22.9	23.1	609.6	623.5	–	605 +
330	1,1,2,2,3,3,4,4,5,5,6-undecafluoro-6-(trifluoromethyl)cyclohexane	C_7F_{14}	23	27.5	486.8	534.2	–	*485 –*
331	2,2,4-trimethylhexane	C_9H_{20}	23.4	25.0	573.7	586.7	–	520 +
332	dodecafluorocyclohexane	C_6F_{12}	24	28.8	457.2	492.9	–	*382 –*
333	2,2,4,4-tetramethylpentane	C_9H_{20}	24.5	26.0	574.7	564.5	–	482 +
334	2,2,3-trimethylhexane	C_9H_{20}	24.6	25.8	588	577.4	–	514 +
335	2,2,3,3-tetramethylhexane	$C_{10}H_{22}$	24.8	23.8	623.1	592.2	–	*578 –*
336	2,2,3,4-tetramethylpentane	C_9H_{20}	25.7	26.2	592.7	580.5	–	488 +
337	tert-butylcyclohexane	$C_{10}H_{20}$	26.3	27.3	659	634.7	–	*574 –*
338	butan-2-ylcyclohexane	$C_{10}H_{20}$	26.4	29.3	669	657.6	–	597 +
339	3,3-diethylpentane	C_9H_{20}	26.4	24.4	610	605.5	–	515 +
340	2,3,3,4-tetramethylpentane	C_9H_{20}	26.8	25.9	607.6	590.9	–	500 +
341	2,2,3,3-tetramethylpentane	C_9H_{20}	27	25.5	607.6	586.9	–	505 +
342	(2E)-oct-2-ene	C_8H_{16}	27.3	27.3	580	584.7	–	462 +
343	propylcyclohexane	C_9H_{18}	27.7	29.2	639	649.5	–	544 +
344	1-methyl-4-(propan-2-yl)benzene	$C_{10}H_{14}$	27.9	28.2	653	667.5	–	495 +
345	1,1,3-trimethylcyclopentane	C_8H_{16}	27.9	28.4	569.5	586.4	–	432 +
346	propan-2-ylcyclohexane	C_9H_{18}	28	26.7	640	601.7	–	526 +
347	1-methyl-2-(propan-2-yl)benzene	$C_{10}H_{14}$	28.6	29.1	670	660.5	–	482 +
348	1-methyl-3-(propan-2-yl)benzene	$C_{10}H_{14}$	29	27.7	666	669.9	–	468 +
349	1,1,2-trimethylcyclopentane	C_8H_{16}	29	28.1	579.5	592.3	–	424 +
350	butan-2-ylbenzene	$C_{10}H_{14}$	29.1	29.9	664	652.7	–	481 +
351	tert-butylbenzene	$C_{10}H_{14}$	29.3	29.3	660	653.0	–	463 +
352	benzylbenzene	$C_{13}H_{12}$	29.4	32.0	767	771.9	–	509 +
353	1-ethyl-1-methylcyclopentane	C_8H_{16}	29.5	28.4	592	584.5	–	428 +
354	propan-2-ylcyclopentane	C_8H_{16}	29.6	28.2	601	582.0	–	426 +
355	(2-methylpropyl)cyclohexane	$C_{10}H_{20}$	30.8	26.9	659	647.5	–	587 +
356	phenoxybenzene	$C_{12}H_{10}O$	31	36.2	766	779.6	–	491 +
357	butylcyclohexane	$C_{10}H_{20}$	31.1	25.3	667	652.9	–	601 +
358	decanenitrile	$C_{10}H_{19}N$	32.1	25.6	622	623.6	–	*565 -*
359	1-ethoxypropane	$C_5H_{12}O$	32.1	34.5	500.6	537.2	–	335 +
360	trifluoroacetic acid	$C_2HF_3O_2$	32.2	45.2	491.3	456.2	–	203 +
361	hexafluorobenzene	C_6F_6	32.6	39.7	516.7	544.2	–	271 +
362	3,5-dimethylphenol	$C_8H_{10}O$	33.8	36.8	715.6	705.3	–	384 +
363	1,2,3,4-tetrahydronaphthalene	$C_{10}H_{12}$	34.7	32.8	719	635.8	–	455 +
364	N,N-dimethylaniline	$C_8H_{11}N$	35.8	34.9	687	660.7	–	420 +
365	dimethyl oxalate	$C_4H_6O_4$	39.3	44.7	628	566.5	–	289 +
366	ethenylbenzene	C_8H_8	39.4	41.4	647	638.6	–	341 +
367	anisole	C_7H_8O	41.2	43.4	641	686.0	–	329 +
368	benzonitrile	C_7H_5N	41.6	48.2	699.4	647.3	–	285 +
369	(methylsulfanyl)ethane	C_3H_8S	42	44.4	533	518.0	–	259 +
370	benzaldehyde	C_7H_6O	46	46.4	695	680.8	–	295 +
371	4-methylphenol	C_7H_8O	50.8	45.3	704.6	694.6	–	321 +
372	prop-2-enal	C_3H_4O	51	51.6	506	485.8	–	193 +
373	N-methylaniline	C_7H_9N	51.3	40.5	701	678.0	–	351 +
374	formaldehyde	CH_2O	65	60.7	408	373.2	–	124 +
375	dibromomethane	CH_2Br_2	71	59.3	583	532.8	–	234 +
376	bromomethane	CH_3Br	85	63.8	464	474.5	–	184 +
377	anthracene	$C_{14}H_{10}$	–	34.6 +	883	842.3	–	*493 -*

1	2	3	4	5	6	7	8	9
378	phenanthrene	$C_{14}H_{10}$	–	34.6 +	878	812.6	–	494 -
379	1,1,2,2-tetrachloro-1,2-difluoroethane	$C_2Cl_4F_2$	–	37.8 +	551	537.2	–	323 +
380	dimethoxymethane	$C_3H_8O_2$	–	41.5 +	497	501.2	–	263 +
381	(ethyldisulfanyl)ethane	$C_4H_{10}S_2$	–	38.3 +	642	559.1	–	352 +
382	1H-pyrrole	C_4H_5N	–	53.2 +	640	526.9	–	217 +
383	cyclopentene	C_5H_8	–	44.8 +	506	483.3	–	257 +
384	propyl propanoate	$C_6H_{12}O_2$	–	31.3 +	578	569.7	–	400 +
385	2,3-dimethylpyridine	C_7H_9N	–	38.3 +	655.4	656.7	–	344 +
386	2,5-dimethylpyridine	C_7H_9N	–	38.4 +	644.2	655.2	–	359 +
387	3,4-dimethylpyridine	C_7H_9N	–	38.9 +	683.8	652.3	–	344 +
388	3,5-dimethylpyridine	C_7H_9N	–	39.7 +	667.2	658.1	–	340 +
389	4-methylaniline	C_7H_9N	–	42.1 +	667	699.7	–	352 +
390	2,3-dimethylphenol	$C_8H_{10}O$	–	36.5 +	722.8	701.2	–	385 +
391	2,4-dimethylphenol	$C_8H_{10}O$	–	35.8 +	707.6	719.8	–	396 +
392	2,5-dimethylphenol	$C_8H_{10}O$	–	36.2 +	723	702.5	–	391 +
393	2,6-dimethylphenol	$C_8H_{10}O$	–	35.9 +	701	712.0	–	381 +
394	2-ethylphenol	$C_8H_{10}O$	–	36.2 +	703	702.3	–	397 +
395	3,4-dimethylphenol	$C_8H_{10}O$	–	35.7 +	729.8	708.6	–	396 +
396	3-ethylphenol	$C_8H_{10}O$	–	36.2 +	716.4	696.9	–	396 +
397	4-ethylphenol	$C_8H_{10}O$	–	36.0 +	716.4	699.5	–	407 +
398	Ethoxybenzene	$C_8H_{10}O$	–	34.8 +	647	687.7	–	403 +
399	formic acid	CH_2O_2	–	61.3 +	580	406.8	–	141 +

Note: for compounds with unknown experimental values of critical properties, their predicted values are given and it is also noted whether the compound is within the scope of the developed model: «+» – yes, «–» – no

At the previous stage of this work, all molecular structures of the compounds from the aforementioned database were modeled using the ChemAxon software suite [7] (specifically, the MarvinSketch.exe program – a graphical editor for chemical compounds; Standardizer.exe program – for standardization of molecular structures). Then, *data curation* was performed with HiT QSAR software [8], which included searching for erroneous structures, verifying connectivity and uniqueness, and filtering out duplicates.

To calculate structural descriptors, the simplex representation of molecular structure method (SiRMS) [8] was used, which represents a molecule as a system of different simplexes – tetraatomic molecular fragments with fixed composition and structure. The total number (N) of all possible simplexes in an n-atomic molecule is given by:

$$N = \frac{n!}{(n-4)! \cdot 4!}. \quad (1)$$

An important feature of the simplex approach is that when identifying the vertices of simplexes, not only labels reflecting the nature of atoms (a complex of atomic properties) are used, but also various weighting parameters (properties) of atoms.

In this work, to differentiate the vertices of simplexes, in addition to the nature of atoms, an informational description [9] was applied – that is, potentials of informational fields (Inf. potential), which allow accounting for the mutual influence of atoms. These informational fields were weighted by atomic characteristics such as partial charge (Chg), refraction (Rf), electronegativity (En), and lipophilicity (Lip).

The calculation of structural parameters was performed at the 2D level of molecular structure representation, where only the topology of the molecule is considered. In this representation, the molecule is modeled as a molecular graph, with each vertex acting as a source of information and information transmitted along the graph edges.

The number of simplexes of a certain type serves as a structural descriptor. The total number of descriptor types can be quite large, depending on the structural diversity of the compounds under study. Besides simplexes, smaller fragments consisting of two or three atoms can also be used.

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

The range of values of atomic characteristics, which are real numbers, is preliminarily divided into discrete groups. An atom's label is determined based on the group it belongs to. When calculating simplex descriptors, the following interval ranges for atomic property values were used:

– Inf.potential (Rf): $A < 2.41 \le B < 3.56 \le C < 4.71 \le D < 5.86 \le E < 7.00 \le F < 8.15 \le G$;

– Inf.potential (En): $A < 3.59 \le B < 4.50 \le C < 5.42 \le D < 6.33 \le E < 7.24 \le F < 8.16 \le G$;

– Inf.potential (Chg): $A < -3.56 \le B < -2.27 \le C < -0.99 \le D < 0.30 \le E < 1.58 \le F < 2.87 \le G$;

– Inf.potential (Lip): $A < -4.98 \le B < -2.67 \le C < -0.36 \le D < 1.96 \le E < 4.27 \le F < 6.58 \le G$.

The vertices of the simplexes were also differentiated based on the characteristics of van der Waals (VDW) interactions using parameters from the universal force field [10]. The Lennard-Jones van der Waals parameters were developed within the conceptual framework of the exponential-6 form:

$$E_{\text{vdw}} = A \cdot e^{-Bx} - \frac{C_6}{x^6}. \tag{2}$$

The given equation can be rearranged into a mathematically equivalent form containing a well depth term D_{ij} (kcal/mol), a distance term x_{ij} (Å), and a shape parameter ζ [10]:

$$E_{vdw} = \left[D_{ij} \left(\frac{6}{\zeta - 6} \right) e^{\zeta} \right] e^{-\zeta (x/x_{ij})} - \frac{D_{ij} \left(\frac{\zeta}{\zeta - 6} \right) x_{ij}^6}{x^6}. \tag{3}$$

A comparison of like terms in equation 2 and 3 leads to equation 4 for the distance:

$$x_i = \frac{\zeta}{B_i}, \tag{4}$$

equation 5 for the well depth:

$$D_i = \frac{C_{6ii} \left(\frac{\zeta - 6}{\zeta} \right)}{x_i^6}, \tag{5}$$

and equation 6 for the repulsive preexponential term:

$$A_i = D_i \left(\frac{6}{\zeta - 6} \right) e^{\zeta}. \tag{6}$$

Hence, the coefficient of VDW attraction was calculated using the formula:

$$C_{\text{attr}} = D_i \left(\frac{\zeta}{\zeta - 6} \right) x_i^6, \tag{7}$$

and the coefficient of VDW repulsion C_{rep} by equation 6.

The entire ranges of parameter values of van der Waals interactions were divided into the following intervals:

– VDW attraction: $A < 50 \le B < 100 \le C < 250 \le D < 400 \le E < 650 \le F < 2000 \le G$,

– VDW repulsion: $A < 20000 \le B < 32000 \le C < 50000 \le D < 100000 \le E$.

In addition, all atoms corresponding to the vertices of the simplexes were divided by labels into three groups: *D* – donors of potential H-bond, *A* – acceptors of potential H-bond, and *I* – indifferent.

The following integral parameters were also used as descriptors: molecular weight, electronic refraction, total electronegativity, and lipophilicity.

To establish the relationship between structural descriptors and critical property values (P_C, T_C, V_C), the Random Forest (RF) method [11] was applied.

Regarding the assessment of the reliability of QSAR models, their validation using an external test set (consisting of compounds that are not included in the training set) is of great importance. QSAR models used to predict the properties of yet unstudied molecules must exhibit sufficiently adequate statistical characteristics [12].

Various methods of forming a "test set" are used to assess predictive ability, during which some molecules are excluded from the model development process, and the remaining molecules form a new training set. This model is then used to predict the properties of the test set compounds.

A Random Forest model is an ensemble of individual decision trees, with a new training set for each tree being formed from the full set of compounds to construct each tree. Compounds that were not included in the training set fall into the so-called "out-of-bag" set and are used to evaluate the model's predictive ability. In the Random Forest method, the size of the control set (out of bag – oob) is approximately 33% of the number of objects in the training set.

For regression models, the most well-known indicator of approximation ability is the coefficient of determination (R^2), which is the ratio of the «explained variance» of a property to the total variance of that property in the set:

$$R^2 = \frac{SS - RSS}{SS}, \tag{8}$$

$$RSS = \sum_{i=1}^{N} (y_i^{\text{pred}} - y_i^{\text{obs}})^2, \tag{9}$$

$$SS = \sum_{i=1}^{N} (y_i^{\text{obs}} - \bar{y})^2, \tag{10}$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i^{\text{obs}}, \tag{11}$$

where: y_i^{pred} and y_i^{obs} – the value of property calculated using the developed regression model and the observed property value for the *i*-th object of the training set, respectively; N – the number of objects in the training set; $\bar{y}$ – the average value of a property for a set of N compounds.

An important statistical characteristic is the root mean square error (RMSE), which is calculated by the formula:

$$RMSE = \sqrt{\frac{RSS}{N}}. \tag{12}$$

It is clear that RMSE characterizes the descriptive ability of the regression model with respect to the compounds included in the training set.

For the "out-of-bag" set in RF models, the formulas for calculating the coefficient of determination (R^2_{oob}) and the root mean square error ($RMSE_{\text{oob}}$) are identical to those given above, except that instead of the observed value y_i^{obs}, the predicted value y_i^{pred} for the i-th object in the control set is used.

For regression models, an important statistical metric is the mean absolute error (MAE) – an indicator of the average magnitude of errors in a set of predictions, regardless of their direction. It is calculated as the average absolute difference between the predicted and observed values of a property and is used to evaluate the performance of a regression model:

$$MAE = \frac{1}{N \sum_{i=1}^{N} | y_i^{\text{pred}} - y_i^{\text{obs}} |}. \tag{13}$$

Research results

For all compounds in the studied database, a total of 4.939 2D structural descriptors were calculated. Prior to model development, the initial descriptor sets were refined by removing constant parameters and retaining only one representative from each group of mutually correlated descriptors (correlation threshold of 0.95). Adequate RF-QSPR models were then constructed for the critical properties of the organic compounds in the dataset (see Table 2). The applicability domain (AD) of each model was assessed for all compounds, in accordance with the method described in [13].

Table 2

Statistical characteristics of the developed RF-QSPR models

Property	R^2	R^2_{oob}	$RMSE_{\text{ws}}$	$RMSE_{\text{oob}}$
P_C	0.99	0.90	1.6	4.2
T_C	0.99	0.90	13.4	36.3
V_C	0.99	0.97	8.9	23.6

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

The predicted values of the critical properties of the studied compounds, as obtained from the developed models, are presented in Table 1. For compounds with no available experimental data, it is indicated whether the compound falls within the applicability domain (AD) of the corresponding model [13]. For critical pressure and temperature, all predicted values fall within the AD. However, for critical volume, some compounds lie outside the AD.

The approximation quality of the models is notably high. The absolute approximation errors fall within the following ranges:

– For critical temperature: 0.02…69.7 K;
– For critical volume: 0.02…64 mL/mol;
– For critical pressure: 0.002…8.4 atm.

For the training set, the mean absolute errors (MAE) are:

– MAE(V_C) = 5.6 mL/mol;
– MAE(P_C) = 1.01 atm;
– MAE(T_C) = 9.4 K.

For the out-of-bag (oob) set, the MAE values are:

– MAE(V_C) = 15.2 mL/mol;
– MAE(P_C) = 2.7 atm;
– MAE(T_C) = 25.7 K.

In accordance with the SiRMS approach, physicochemical interpretation can be performed – that is, the relative contributions of various physicochemical factors to the target properties can be determined. To achieve this, the summed contributions of the model parameters are analyzed separately for each differentiation group. In the case of the Random Forest model, these contributions are estimated using the original method described in [13]. Due to the nonlinear nature of the model, the contributions are unsigned; that is, they do not indicate the direction of influence of a given parameter.

Fig. 1 presents factor diagrams for the developed models, illustrating the relative contributions of specific physicochemical factors to the critical properties of organic compounds.

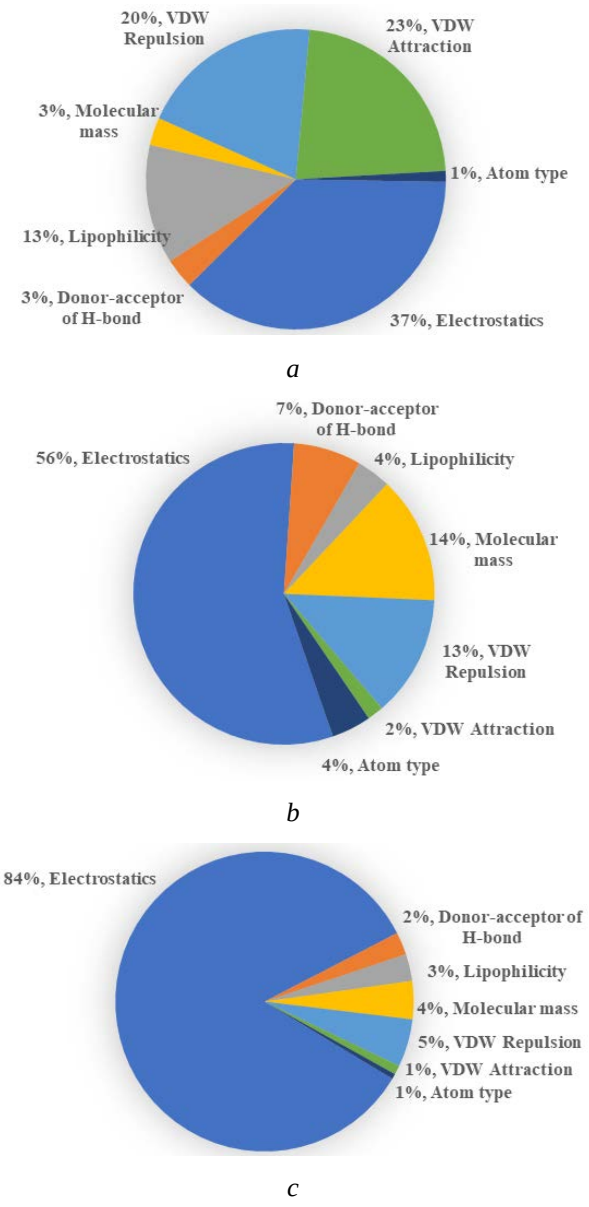

Fig. 1. Relative contribution of physicochemical factors to the values of critical properties within the framework of the developed RF models for: critical pressure (*a*), critical temperature (*b*), critical volume (*c*)

The diagram shows that in all cases, electrostatic factors exert the strongest influence on the critical properties. In the case of critical pressure, van der Waals interactions are also significant – specifically, van der Waals repulsion (23%) and attraction (20%). This confirms that critical point parameters are determined largely by the energy of intermolecular interactions.

As shown in Table 1, experimental values for all three critical properties are available for compounds 1…295. Based on this, a subset comprising the first 295 compounds was created, and corresponding RF QSPR models were developed. The predictive performance of these models was evaluated using test sets composed of those compounds among entries 296…399 for which experimental data were available. The following coefficients of determination (R^2_{test}) were obtained for the test sets: 0.90 for critical temperature, 0.98 for critical volume, and 0.92 for critical pressure. These results confirm the high predictive accuracy of the developed models.

Conclusions

Thus, based on the 2D representation of molecular structure (where only molecular topology is considered) robust QSPR models were developed to describe the fundamental critical properties of

compounds from various classes. These models exhibit high approximation accuracy and strong predictive capability. For compounds lacking experimental data on the studied properties, predictions were made and presented using the developed RF models. Furthermore, analysis of these models enabled the assessment of the impact of different atomic characteristics on the target properties. It was found that critical point parameters are primarily determined by electrostatic factors.

Література

1. Білий О.В. Фізична хімія: навч. посіб. Київ : ЦУЛ, 2003. 364 с.
2. Vapor–Liquid Critical Properties of Elements and Compounds. Review of Recent Data for Hydrocarbons and Non-hydrocarbons / Ambrose D., Tsonopoulos C., Nikitin E., Morton D., Marsh K. *Chem. Eng. Data*. 2015. Vol. 60, No 12. P. 3444–3482. DOI: https://doi.org/10.1021/acs.jced.5b00571.
3. Yang S., Lu W., Chen N., Hu Q. Support vector regression based QSPR for the prediction of some physicochemical properties of alkyl benzenes. *J. Mol. Struct*. 2005. Vol. 719. P. 119. DOI: https://doi.org/10.1016/j.theochem.2004.10.060.
4. Quantitative Structure-Property Relationship for Critical Temperature of Alkenes with Quantum-Chemical and Topological Indices / Huoyu R., Zhiqiang Z., Guofang J., Zhanggao L., Zhenzhen X. *Russ. J. Phys. Chem*. 2022. Vol. 96, No 11. P. 2329–2334. DOI: https://doi.org/10.1134/S0036024422110267.
5. Bouhedjar K., Nacereddine A. K., Ghorab H., Djerourou A. QSPR Modeling for Critical Temperatures of Organic Compounds Using Hybrid Optimal Descriptors. *International Journal of Quantitative Structure-Property Relationships (IJQSPR).* 2019. Vol. 4, No 4. P. 15–26. DOI: https://doi.org/10.4018/IJQSPR.2019100102.
6. Reid R.C., Prausnitz J. M., Sherwood T.K. The properties of gases and liquids. McGraw-Hill, 1977. 710 p.
7. MarvinSketch, Standardizer. (1998-2020) Chemaxon Ltd. URL: http://www.chemaxon.com.
8. Simplex Representation of molecular structure as universal QSAR/QSPR tool / Kuz'min V., Artemenko A., Ognichenko L. et al. *Struct. Chem.* 2021. Vol. 32, No 4. P. 1365–1392. DOI: https://doi.org/10.1007/s11224-021-01793-z.
9. Ognichenko L.N., Kuz'min V.E., Artemenko A.G. New structural descriptors of molecules on the basis of symbiosis of the informational field model and simplex representation of molecular structure. *QSAR&Comb.Sci*. 2009. Vol 28, No 9. P. 939–945.
10. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations / Rappe A.K., Casewit C.J., Colwell K.S. et al. *J. Am. Chem. Soc*. 1992. Vol. 114, No. 25. P. 10024–10035. DOI: https://doi.org/10.1021/ja00051a040.
11. Breiman L. Random Forest. *Machine Learning*. 2001. Vol. 45. P. 5–32. DOI: https://doi.org/10.1023/A:1010933404324.
12. Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals. *OECD Papers*. 2006. Vol. 6. P. 79–157. DOI: https://doi.org/10.1787/oecd_papers-v6-art37-en.
13. Interpretation of QSAR models based on Random Forest method / Kuz'min V.E, Polishchuk P.G., Artemenko A.G. et al. *Mol. Inf.* 2011. Vol. 30, No 6–7. P. 593–603. DOI: https://doi.org/10.1002/minf.201000173.

References

1. Bilyi, O.V. (2003). Physical chemistry. Kyev: CUL.
2. Ambrose, D., Tsonopoulos, C., Nikitin, E., Morton, D. & Marsh, K. (2015). Vapor–Liquid Critical Properties of Elements and Compounds. Review of Recent Data for Hydrocarbons and Non-hydrocarbons. *Chem. Eng. Data*, 60(12), 3444–3482. DOI: https://doi.org/10.1021/acs.jced.5b00571.
3. Yang, S., Lu, W., Chen, N., & Hu, Q. (2005). Support vector regression based QSPR for the prediction of some physicochemical properties of alkyl benzenes. *J. Mol. Struct*., 719, 119. DOI: https://doi.org/10.1016/j.theochem.2004.10.060.
4. Huoyu, R., Zhiqiang, Z., Guofang, J., Zhanggao, L., & Zhenzhen, X. (2022). Quantitative Structure-Property Relationship for Critical Temperature of Alkenes with Quantum-Chemical and Topological Indices. *Russ. J. Phys. Chem.*, 96(11), 2329–2334. DOI: https://doi.org/10.1134/S0036024422110267.
5. Bouhedjar, K., Nacereddine, A. K., Ghorab, H., & Djerourou, A. (2019). QSPR Modeling for Critical Temperatures of Organic Compounds Using Hybrid Optimal Descriptors. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, *4*(4), 15–26. DOI: https://doi.org/10.4018/IJQSPR.2019100102.
6. Reid, R. C., Prausnitz, J. M., & Sherwood, T. K. (1977). *The properties of gases and liquids*. McGraw-Hill.

ISSN 2076-2429 (print)
ISSN 2223-3814 (online)

7. Chemaxon Ltd. (1998–2020). *MarvinSketch, Standardizer*. Retrieved from http://www.chemaxon.com.
8. Kuz'min, V., Artemenko, A., & Ognichenko, L. et al. (2021). Simplex Representation of molecular structure as universal QSAR/QSPR tool. *Struct. Chem*., *32*(4), 1365–1392. DOI: https://doi.org/10.1007/s11224-021-01793-z.
9. Ognichenko, L.N., Kuz'min, V.E., & Artemenko, A.G. (2009). New structural descriptors of molecules on the basis of symbiosis of the informational field model and simplex representation of molecular structure. *QSAR&Comb.Sci*., 28(9), 939–945.
10. Rappe, A.K., Casewit, C.J., & Colwell, K.S. et al. (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc*., *114*(25), 10024–10035. DOI: https://doi.org/10.1021/ja00051a040.
11. Breiman, L. (2001). Random Forest. *Machine Learning*, *45*, 5–32. DOI: https://doi.org/10.1023/A:1010933404324.
12. Organisation for Economic Co-operation and Development. (2006). *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals*. *OECD Papers*, *6*, 79–157. DOI: https://doi.org/10.1787/oecd_papers-v6-art37-en.
13. Kuz'min, V.E, Polishchuk, P.G., & Artemenko, A.G. et al. (2011). Interpretation of QSAR models based on Random Forest method. *Mol. Inf.*, *30*(6-7), 593–603. DOI: https://doi.org/10.1002/minf.201000173.

Пестерев Єгор Костянтинович; Egor Pesterev, ORCID: https://orcid.org/0009-0002-7941-3571
Огніченко Людмила Миколаївна; Liudmyla Ognichenko, ORCID: https://orcid.org/0000-0002-3344-6119
Артеменко Анатолій Григорович; Anatoliy Artemenko, ORCID: https://orcid.org/0000-0002-9592-1384
Кічьова Марія Євгенівна; Mariia Kichova, ORCID: https://orcid.org/0000-0002-6294-3261
Муратов Наіль Ніязович; Nail Muratov, ORCID: https://orcid.org/0000-0001-8417-657X
Кузьмін Віктор Євгенович; Victor Kuz'min, ORCID: https://orcid.org/0000-0002-2753-0453

Received May 04, 2025
Accepted June 12, 2025