Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS

J. Serb. Chem. Soc. *86* (4) 355–366 (2021)
JSCS–5426

Original scientific paper

Achillea clypeolata Sibth. & Sm. essential oil composition and QSRR model for predicting retention indices

MILICA AĆIMOVIĆ[1]*, LATO PEZO[2], MIRJANA CVETKOVIĆ[3], JOVANA STANKOVIĆ[3] and IVANA ČABARKAPA[4]

[1]*Institute of Field and Vegetable Crops Novi Sad, Maksima Gorkog 30, 21000 Novi Sad, Serbia,* [2]*University of Belgrade, Institute of General and Physical Chemistry, Studentski trg 10–12, 1000 Belgrade, Serbia,* [3]*University of Belgrade, Institute of Chemistry, Technology and Metallurgy, Njegoševa 12, 11000 Belgrade, Serbia and* [4]*University of Novi Sad, Institute of Food Technology, Bulevar cara Lazara 1, 21000 Novi Sad, Serbia*

(Received 24 May 2020, revised 17 January, accepted 5 February 2021)

Abstract: The aim of this study was the prediction model of retention indices of compounds from the aboveground parts of *Achillea clypeolata* Sibth. & Sm. essential oil, obtained by hydrodistillation and analysed by GC–MS. The quantitative structure–retention relationship analysis was applied in order to anticipate the retention time of the obtained compounds. The selection of the seven molecular descriptors was done by a genetic algorithm. The chosen descriptors were uncorrelated and were used to construct an artificial neural network. A total of 40 experimentally obtained retention indices was used to build this prediction model. The coefficient of determination for the training, testing and validation cycles were: 0.950, 0.825 and 1.000, respectively, indicating that this model could be used for prediction of retention indices for *A. clypeolata*, essential oil compounds.

Keywords: hydrodistillation; GC–MS; artificial neural networks.

INTRODUCTION

Achillea clypeolata Sibth. & Sm., yellow or moonshine yarrow, is a Balkan endemic species, spread across North and Central Greece, South Albania, North Macedonia, East Serbia, Bulgaria, Southeast Romania, West and European Turkey according to a study by Nedelcheva.[1] It is dominantly diploid ($2n$ = 18), a perennial species, shortly tomentose, silver–grey. Erect stem, simple and up to 60 cm long. The rhizome is well developed and woody. Leaves are pinnatisec, plane, and weakly glandular–punctate. Basal leaves are 8–20 cm long and 2–4.5 cm wide, petiolate, the lobes ovate, serrate to pinnatifid, with acute teeth. Cauline leaves all more or less distant, about twice as long as the internodes, while the

* Corresponding author. E-mail: acimovicbabicmilica@gmail.com
https://doi.org/10.2298/JSC200524008A

upper leaves are 1–2 cm, sessile. Corymbs with many capitulars, peduncles, 2 mm tomentose. Involucres 3–4 mm in diameter, bracts elliptical or lanceolate, 1.5 mm. Ligules 1 mm yellow. Peripheral flowers with rounded ligules 5-fold shorter than involucres. It blooms from June to July. Pollination is entomophylous and anemophylous, with dispersal of fruits and seeds in its habitats. However, in nature it hybridizes with *A. neilreichii*, *A. setacea* and *A. panonica*. This plant is heliophytic, termophyte that occupies dry, neutral soils in arid meadows. Nowadays, it spontaneously grows only on limestone in Serbia, and it is listed as a critically endangered species. According to Contreras–Medina and Luna–Vega,[2] the plant is economically important, as a decorative plant, and has been identified as important to the plant genetic fund.

A. clypeolata tastes bitter like mugwort. This species is rich in sesquiterpenes, diterpenes and phenolic compounds, as well as flavonoids.[3–6] Furthermore, the content of essential oil is low (0.05–0.1 %), and according to the essential oil composition, there are two chemotypes.[7] According to a study by Simić *et al.*,[8] one chemotype contains *E*-γ-bisabolene, 1,8-cineole, borneol and caryophyllene oxide. The other one contains 1,8-cineole and camphor as the dominant compounds as documented in thestudy by Chalcat *et al.*[9]

However, this plant is not investigated thoroughly, only its antioxidant and antimicrobial properties have been confirmed to the present day. Antioxidant activity of *A. clypeolata* leaf, flower, and root methanolic extract is determined by total reducing power assay and DPPH. It was reported that total reducing power ranged between 10.66 mg AAE/ml for root extract, to 11.90 mg AAE/ml for flower extract. *In vitro* DPPH tests showed similar antioxidant activities, and according to these results *A. clypeolata* can be used as a potential natural antioxidant source.[6] Investigations by Simić *et al.*[8] showed that *A. clypeolata* essential oil express antimicrobial activity against *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Staphylococcus aureus*. However, its wide application is recorded in traditional medicine of Bulgaria and Serbia. In her study, Nedelcheva,[1] stated that, in Bulgaria, it is used to treat: hemorrhoids, wounds, bleeding, gastro-intestinal atony, bed wetting, kidney inflammation, amenorrhoea, inflamed gums and liver diseases. However, Zlatković *et al.*[10] concluded that in Serbian traditional medicine it is mainly used as an antidiabetic drug.

In the study by Wolfender *et al.*,[11] quantitative structure retention relationship (QSRR) provides insight in relation between the chemical structure and the physicochemical or biological properties. A systematic study was presented in the paper by Héberger,[12] where the QSRR analysis in gas chromatography (GC) was presented for planar chromatography, column liquid and micellar liquid chromatography and affinity chromatography. Lately, numerous publications have been related to the QSRR analysis.[13,14] The chemical compound structure is explored by their mathematical models, presented by so–called molecular des-

criptors, which transfer the compound data through the symbolic representation of a molecule into a numerical value as reported by Khezeli *et al.*[15] Marrero-Ponce *et al.*[16] determined that the molecular descriptors should be chosen to avoid overfitting, in order to obtain statistically significant results, and to establish clear relationships between molecular structure and its descriptors. In a study by Micić *et al.*,[17] it was shown that GC–MS is a unique technique that yields a large number of the quantitatively comparable, reproducible data and exact retention time for large sets of compounds.

In a study by Tropsha and Golbraikh,[18] the numerical model that represents the relation between the molecular descriptors and the retention time can be established by numerous machine learning algorithms, or by using the artificial neural network (ANN), which is used in this study, and has already been proven to be an excellent tool according to the literature.[11,19]

The aim of this paper was to establish a new QSRR model for predicting the retention times of chemical compounds in *A. clypeolata* essential oil obtained by hydrodistillation and analyzed by GC–MS using the genetic algorithm (GA) variable selection method and the artificial neural network (ANN) model.

MATERIAL AND METHODS

Plant material

A. clypeolata was collected on 7th July 2018, on Mt. Rtanj. The plant species were in full flowering stage by this date. The plant aboveground parts were cut manually at the upper 15 cm of the plant, and the biomass was placed in an air-dryer until constant weight at 35 °C to avoid essential oil losses. Voucher specimens were confirmed and deposited at the Herbarium BUNS, the University of Novi Sad, Faculty of Sciences, Department of Biology and Ecology, under the acquisition number 2-1448.

Essential oil extraction

Air-dried aerial parts of *A. clypeolata* were submitted to hydrodistillation (Clevenger apparatus, 3 h). Then, the essential oil was dried over anhydrous sodium sulfate and stored in a dark glass vial at 4 °C for further analysis. Dried aerial parts of *A. clypeolata* were found to contain 0.04 % of pale–yellow oil.

Essential oil analysis

The essential oil was analyzed using an HP 5890 gas chromatograph coupled to an HP 5973 MSD and fitted with a capillary column HP–5MS (30 m×0.25 mm×0.25 μm film thickness). Analytical conditions were as follows: helium was used as carrier gas; inlet pressure was 25 kPa; linear velocity: 1 ml/min at 210 °C; injector temperature: 250 °C; injection mode: splitless. MS scan conditions were: source temperature, 200 °C; interface temperature, 250 °C; electron energy, 70 eV; mass scan range, 40–350 amu. Temperature program: 60 to 285 °C at a rate of 4.3 °C/min. The components were identified based on their linear retention index relative to C8–C32 *n*-alkanes, comparison with data reported in the literature (Wiley and NIST databases). Percentage (relative) of the identified compounds was computed from GC peak area.

QSRR analysis

The molecular structure dataset was presented in the form of .smi files (simplified molecular input line entry specification) which were used in the molecular descriptors calculation. The .smi files were collected from Pub Chem. The similar approach was noticed in a paper by Matyushin *et al.*,[20] where .smi notation of the molecule structure was used as an input for the model. In that study, a neural network was used for the estimation of gas chromatographic retention indices on non-polar stationary phases. In the study by Dong *et al.*,[21] the calculation of the specified molecular descriptors for each chemical compound was performed using molecular descriptor software PaDel as in the study by Yap.[22] After the calculation was completed, the data were randomly separated and independently chosen into training, testing and validation sets (60, 20 and 20 % of data, respectively), in order to determine the predictive artificial neural network model (ANN). A series of 100,000 randomly generated ANN topologies were tested, changing the number of hidden neurons (from 1 to 20) and initial values of weights and biases the training process. The optimization process was performed based on validation error minimization. ANN was developed to form a reliable model to predict the retention times from PaDel-calculated descriptors. The evaluation of the performances of the developed model was done by comparing the predicted and experimentally obtained retention times of the observed chemical compounds used for the model construction. The model overfitting was also checked. All calculations were performed by an eight-core personal computer and the PaDel database was used to calculate the molecular descriptors (which included 1D, 2D and 3D descriptors, Micić *et al.*[17]). The genetic algorithm (GA) must be applied in order to reduce the number of parameters (calculated by PaDel). This task was performed, using Heuristic Lab, to select the most relevant molecular descriptors for *RT* prediction. GA is a stochastic optimization method inspired by evolution theory.[23,24] The correlation between the descriptors was examined and collinear descriptors were detected using factor analysis. Statistical investigation of the data has been performed mainly by the Statistica 10 software.[25]

Artificial neural network (ANN)

A multi-layer perceptron model (MLP) covered input, hidden and output layer was used, considering that it is proven to be quite capable of approximating nonlinear functions.[26] Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm was used for ANN modelling. ANN results, including the weight values, depend on the initial assumptions of parameters necessary for ANN construction and fitting.[27,28] A series of various topologies were used, in which the number of hidden neurons varied from 10 to 20 and the training process of each network was run 100,000 times with random initial values of weights and biases. The optimization process was performed on the basis of validation error minimization. ANN calculations were performed with Statistica 10. Yoon's interpretation method was used to determine the relative influence of molecular descriptors on retention time.[29] This method was applied based on the weight coefficients of the developed ANN.

RESULTS AND DISCUSSION

Chemical profile of A. clypeolata essential oil

In the *A. clypeolata* essential oil 40 compounds were detected, that represented 99.3 % of total oil composition (Table I). Among them 3 not identified compounds (NI) compromised 1.0 %. As it can be seen, the most abundant compounds in *A. clypeolata* essential oil were 1,8-cineole (45.1 %) and camphor (18.2 %). Sixteen compounds had average relative abundance over 1.0 %. Mono-

terpene hydrocarbons (55.7 %) and their oxygenated derivatives (29.5 %) were dominant in the chemical composition.

TABLE I. Quantitative profile of *A. clypeolata* essential oil

No.	Compound	Formula	RI^a, min	RI^b, min	Content, %
1	Tricyclene	$C_{10}H_{16}$	927	921	0.1
2	*α*-Thujene	$C_{10}H_{16}$	929	928	0.1
3	*α*-Pinene	$C_{10}H_{16}$	936	932	1.1
4	Camphene	$C_{10}H_{16}$	950	946	2.3
5	Thuja-2,4(10)-diene	$C_{10}H_{14}$	955	953	0.1
6	Sabinene	$C_{10}H_{16}$	975	969	0.3
7	*β*-Pinene	$C_{10}H_{16}$	979	974	1.4
8	dehydro-1,8-Cineole	$C_{10}H_{16}O$	993	988	0.1
9	*α*-Terpinene	$C_{10}H_{16}$	1016	1014	0.6
10	*p*-Cymene	$C_{10}H_{14}$	1022	1020	3.1
11	1,8-Cineole	$C_{10}H_{18}O$	1028	1026	45.1
12	*γ*-Terpinene	$C_{10}H_{16}$	1053	1054	1.2
13	*p*-Mentha-2,4(8)-diene	$C_{10}H_{16}$	1081	1083	0.2
14	Linalool	$C_{10}H_{18}O$	1092	1095	0.4
15	*Z*-*p*-Menth-2-en-1-ol	$C_{10}H_{18}O$	1114	1118	0.2
16	*α*-Campholenal	$C_{10}H_{16}O$	1119	1122	0.2
17	Camphor	$C_{10}H_{16}O$	1136	1141	18.2
18	*Z*-Chrysanthenol	$C_{10}H_{16}O$	1156	1160	0.3
19	Borneol	$C_{10}H_{18}O$	1159	1165	2.7
20	*δ*-Terpineol	$C_{10}H_{18}O$	1161	1162	0.6
21	Terpinen-4-ol	$C_{10}H_{18}O$	1172	1176	2.8
22	*α*-Terpineol	$C_{10}H_{18}O$	1186	1190	2.4
23	Myrtenal	$C_{10}H_{14}O$	1192	1195	0.3
24	Thymol	$C_{10}H_{14}O$	1290	1289	0.9
25	Carvacrol	$C_{10}H_{14}O$	1300	1298	0.5
26	*E*-Caryophyllene	$C_{15}H_{24}$	1417	1408	0.5
27	*allo*-Aromadendrene	$C_{15}H_{24}$	1458	1458	1.3
28	Germacrene D	$C_{15}H_{24}$	1479	1484	1.2
29	NI[c]-1		1484	–	0.2
30	*γ*-Cadinene	$C_{15}H_{24}$	1512	1513	1.2
31	*δ*-Cadinene	$C_{15}H_{24}$	1521	1522	0.2
32	NI[c]-2		1571	–	0.4
33	Spathulenol	$C_{15}H_{24}O$	1573	1577	0.4
34	Caryophyllene oxide	$C_{15}H_{24}O$	1578	1582	3.2
35	*β*-Oplopenone	$C_{15}H_{24}O$	1603	1607	0.2
36	Muurola-4,10(14)-dien-1-*β*-ol	$C_{15}H_{24}O$	1622	1630	0.3
37	Caryophylla-4(12),8(13)-dien-5-*α*-ol	$C_{15}H_{24}O$	1630	1627	0.7
38	*epi*-*α*-Cadinol	$C_{15}H_{26}O$	1635	1638	2.4
39	*α*-Cadinol	$C_{15}H_{26}O$	1648	1652	1.5
40	NI[c]-3		1680	-	0.4
Monoterpene hydrocarbons					55.7
Oxygenated monoterpenes					29.5
Sesquiterpene hydrocarbons					4.4

TABLE I. Continued

Compound	Formula	*RI*[a], min	*RI*[b], min	Content, %
Oxygenated sesquiterpenes				8.7
NI[c]				1.0
Total identified				99.3

[a]Retention index experimental; [b]retention index from the NIST web book database; [c]not identified compound with mass spectrum –*m*/*z* (intensity): **NI–1**: 41.05 (17.0), 43.00 (9.0), 57.00 (100.0), 80.95 (10.0), 85.00 (50.0), 91.00 (9.0), 118.95 (35.0), 121.10 (8.0), 133.85 (11.0), 236.15 (12.0); **NI–2**: 41.05 (33.0), 57.00 (69.0), 66.95 (29.0), 69.05 (38.0), 79.05 (41.0), 91.00 (42.0), 105.00 (28.0), 134.00 (46.0), 135.00 (100.0), 150.00 (44.0); **NI–3**: 41.05 (34.0), 66.95 (40.0), 79.00 (72.0), 81.00 (45.0), 90.95 (80.0), 93.00 (42.0), 104.95 (49.0), 107.00 (48.0), 109.00 (100.0), 159.05 (61.0)

The prevailing compounds in the oils of *A. clypeolata* from Mt. Rtanj (43°46′34″ N, 21°53′36″ E), collected during July 1996, were 1,8-cineole (38.6 %) and camphor (19.9 %) in the study by Chalcat *et al.*[9]

It can be concluded that this slight variation in the chemical composition could be caused by the climate conditions during the year, collection time and exposition. Furthermore, aerial parts of *A. clypeolata* from the Mt. Rudina (43°41′35″ N, 21°55′18″ E) had significantly different composition: *E*-*γ*-bisabolene (17.9 %), 1,8-cineole (16.0 %), borneol (11.9 %) and caryophyllene oxide (11.5 %) as reported by Simić *et al.*[8] However, this diversity could be a phenomenon of endemism within genus according to Radulović *et al.*[30]

QSRR model validation

Prior to the GA calculation, the factor analysis was performed in order to eliminate the descriptors with equal or almost equal values for the examined molecules. Only one of the inter–correlated descriptors remained in the GA calculation. As a result of this preliminary consideration, about 400 descriptors remained for GA calculation. GA was used to select the most appropriate molecular descriptors for *RI* prediction, and the selection of the most relevant descriptors was done using the evolution simulation.[31,32] The number of elements on each chromosome (*i.e.*, observed compounds) was equal to the number of the molecular descriptors obtained in the PaDel base. The number of the elements was kept relatively low to maintain a small subset of descriptors according to a study by Todeschini and Consonni.[33] As a result, the probability of generating zero for an element was set at least 60 % greater than the probability of generating the value of one. The operators used were crossover and mutation. The probability of application of these operators was varied linearly with generation renewal (0.5 % for mutation and 90 % for crossover). A population size of 100 individuals was chosen for GA, and evolution was allowed for over 50 generations. The predicted retention indices and molecular descriptors were presented in Fig. 1 and also in Ttable S-I of the Supplementary material to this paper, confirming the adequate

prediction capabilities of the constructed ANN, by showing the relationship between the predicted and experimental retention values.

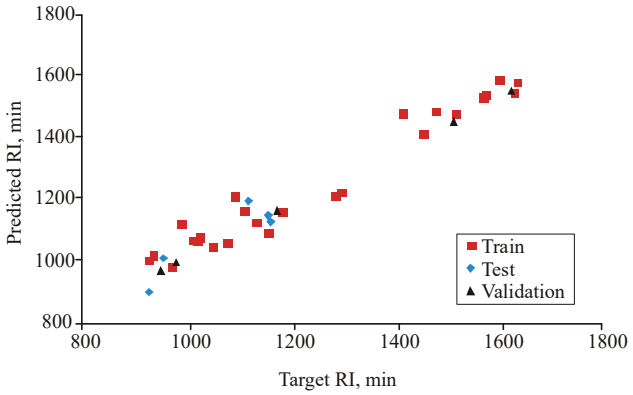

Fig. 1. Comparison of experimentally obtained *RI*s with ANN predicted values.

The principal component analyses (Fig. S-1 of the Supplementary material) showed the groups of chemical compounds in the first factor plane. These groups were mostly different according to molecular descriptors: VP-0, IC0 and ETA Epsilon 3.

The evolution of the generations was stopped when 90 % of the generations took the same fitness.

The ANN results, including the weight coefficients, depend on the initial presumptions of parameters which are vital for ANN development and fitting. Likewise, the number of neurons in the hidden layer can alter the result of the ANN model.

As a result, the seven most significant molecular descriptors selected by GA were: 2D autocorrelation descriptors (AATS0v – average Broto–Moreau autocorrelation – lag 0 / weighted by van der Waals volumes, and AATSC4c – average centered Broto–Moreau autocorrelation – lag 4 / weighted by charges)[33], 2D Barysz matrix descriptor (VR2 Dzi – normalized Randic–like eigenvector-based index from Barysz matrix / weighted by first ionization potential);[32] Chi path descriptors (VP-0 – valence path, order 0;[34] extended topochemical atom descriptor (ETA Epsilon 3, which shows a measure of electronegative atom count);[34] information content descriptors (IC0 – information content index (neighborhood symmetry of 0-order and BIC2 – bond information content index (neighborhood symmetry of 2-order).[33]

These descriptors encode different aspects of the molecular structure and were applied to develop a QSRR model. Table II represents the correlation matrix among these descriptors (none of the correlations were statistically significant).

TABLE II. The correlation coefficient matrix for the selected descriptors by GA

	AATSC4c	VR2Dzi	VP-0	ETAEpsilon3	IC0	BIC2
AATS0v	0.059	–0.015	–0.149	0.124	–0.034	0.143
AATSC4c		0.209	0.151	0.059	0.046	0.175
VR2Dzi			0.166	0.200	–0.131	0.177
VP-0				0.151	–0.012	0.151
ETAEpsilon3					–0.232	–0.297
IC0						0.022

The calibration and predictive capability of a QSRR model should be tested through model validation. The most widely used squared correlation coefficient (r^2) can provide a reliable indication of the fitness of the model, thus, it was employed to validate the calibration capability of a QSRR model.

Artificial neural network (ANN)

In order to explore the nonlinear relationship between *RI*s and the selected descriptors, ANN technique was used to build models. The ability to generalize the model was evaluated by an external test set. The statistical results of the MLP 7-8-1 network is shown in Table III.

TABLE III. ANN model summary (performance and errors), for training, testing and validation cycles; performance term represent the coefficients of determination, while error terms indicate a lack of data for the ANN model

Net. name	Performance			Error			Train.	Error	Hidden	Output
	Train.	Test.	Valid.	Train.	Test.	Valid.	algor.	funct.	activat.	activat.
MLP 7-8-1	0.950	0.825	1.000	1656.639	1236.916	531.5842	BFGS 8	SOS	Exponential	Tanh

The predicted *RI*s presented in Table S-I confirm the good quality of the constructed ANN, by showing the relationship between the predicted and experimental retention values. The obtained results reveal the reliability of the ANN models for predicting the *RI*s of compounds in *A. clypeolata* essential oil determined by GC–MS. There are two groups of compounds according to the carbon chain length, first monoterpenes (C_{10}) and the second group sesquiterpenes (C_{15}). The two compounds (*cis* and *trans* isomers) are allocated from other monoterpenes. The retention indices of monoterpene compounds were in the range of 927 and 1300, while in the case of sesquiterpenes indices ranged between 1417 and 1648. Furthermore, thymol and carvacrol are the structures containing both a phenyl and a hydroxy group, suggesting both benzene and 1-butanol as model compounds according to a study by Roon *et al.*[36]

Molecular descriptors

Separation of compounds in GC and their retention indices are linked to affinity towards mobile and stationary phases. Affinity and solubility of the separ-

ated molecules directly depend on their chemical structure and physicochemical properties, which could be expressed by molecular descriptors. We have utilized seven molecular descriptors for predictions of RI in the obtained ANN model.

2D average Moreau–Broto are spatial autocorrelation descriptors,[37] and could be weighted with charges (AATSC4c) or by van der Waals volumes (AATS0v). According to Hollas,[38] these descriptors explain the molecular structure and physicochemical features of atoms. 2D autocorrelation descriptors are obtained by interatomic distances obtained within the geometry matrix which is determined by the set of atomic characteristics.[33] García-Domenech *et al.*[39] determined that Chi path descriptors belong to the group of connectivity indices which show the numerical possibilities of two identical molecules encountering each other and are obtained from the bond accessibilities. The Chi path index descriptor, used in this work, was the average simple path order 1 (VP–0). 2D Barysz distance matrix is a weighted distance matrix accounting simultaneously for the presence of heteroatoms and multiple bonds in the molecule, and VR2 Dzi was used in the study by Todeschini and Consonni.[33] Information content descriptors were calculated as information content of molecules, based on the calculation of equivalence classes from the molecular graph. The indices of neighbourhood symmetry also take into account neighbour degree and edge multiplicity. The used Information content descriptors were IC0 and BIC2. Extended topochemical atom descriptor ETA Epsilon 3 was used in research by Roy and Ghosh,[35] as a measure of electronegative atom count. The influence of seven most important input variables, identified using genetic algorithm on *RI* was studied in this section. VP0 was the most influential parameter with approximately relative importance of 72.1 %, while the influence of IC0 and AATS0v were 6.6 and 5.1%, respectively. ETA Epsilon 3 and VR2 Dzi were influential at levels 4.9 and 3.9 %, respectively. The influence of BIC2 and AATSC4c were 3.8 and 3.7 %.

CONCLUSION

A. clypeolata essential oil obtained by hydrodistillation was analyzed by GC–MS. Analysis showed that the most abundant compounds were *p*-cymene (24.4 %), limonene (13.5 %) and linalool (8.3 %). The QSRR model for estimating retention times of essential oil compounds was developed for a series of 40 compounds employing the ANN modelling approach. The results demonstrated that the ANN model was adequate in predicting retention times of the *A. clypeolata* essential oil compounds. A suitable model with high statistical quality and low prediction errors was derived. The following seven molecular descriptors were suggested by genetic algorithm: two 2D average Moreau–Broto descriptors (AATSC4c and AATS0v), Chi path descriptor VP-0, 2D Barysz distance matrix

descriptor VR2 Dzi, two Information content descriptors (IC0 and BIC2), and Extended topochemical atom descriptor ETA Epsilon 3.

Selected molecular descriptors were not auto correlated which was suggested by correlation coefficient matrix; thus, descriptors were suitable for QSRR analysis.

The results demonstrated that the ANN model was adequate in predicting the RIs of the compounds in *A. clypeolata* essential oil obtained by hydrodistillation and analyzed by GC–MS. The coefficient of determination for training cycle was 0.950, which is a good indication that this model could be used as a fast mathematical tool for prediction of retention time values for compounds in *A. clypeolata* essential oil obtained by GC–MS analysis due to low prediction error and moderately high r^2.The suitable model with high statistical quality and low prediction errors was derived, and it could be further used for estimation RI of newly detected compounds.

Acknowledgments. The Authors gratefully acknowledge Adonis Sokobanja Ltd. for their administrative and technical support. We would especially like to thank Nebojša Stanojević for his help during collection of *A. clypeolata at* Mt. Rtanj. The Authors are grateful to Milica Rat, PhD, Research Associate and Curator Assistant in Herbarium BUNS, Novi Sad, Serbia, for identifying and archiving this plant species.

SUPPLEMENTARY MATERIAL

Predicted retention indices and molecular descriptors of essential oil compounds in *A. clypeolata* obtained by hydrodistillation and PCA ordination of molecular descriptors are available electronically from the Journal web site, http://www.shd.org.rs/JSCS/, or from the corresponding author on request.

И З В О Д

ХЕМИЈСКИ САСТАВ ЕТАРСКОГ УЉА *Achillea clypeolata* Sibth. & Sm. И QSRR МОДЕЛ ЗА ПРЕДВИЂАЊЕ РЕТЕНЦИОНОГ ВРЕМЕНА

МИЛИЦА АЋИМОВИЋ[1], ЛАТО ПЕЗО[2], МИРЈАНА ЦВЕТКОВИЋ[3], ЈОВАНА СТАНКОВИЋ[3] и ИВАНА ЧАБАРКАПА[4]

[1]Институт за ратарство и повртарство Нови Сад, Максима Горког 30, 21000 Нови Сад, [2]Универзитет у Београду, Институт за општу и физичку хемију, Студентски трг 10–12, 1000 Београд, [3]Универзитет у Београду, Институт за хемију, технологију и металургију, Његошева 12, 11000 Београд и [4]Универзитет у Новом Саду, Институт за храну Технологија, Булевар цара Лазара 1, 21000 Нови Сад

Циљ ове студије био је израда модела за предвиђања ретенционог времена хемијских једињења из есенцијалног уља надземних делова биљке *Achillea clypeolata* Sibth. & Sm., добијеног хидродистилацијом и анализираног GC–MS техником. Квантитативна анализа хемијске структуре и предвиђања ретенционог времена (quantitative structure––retention relationship – QSRR) је примењена да би се предвидело време задржавања хемијских једињења добијених коришћењем GC–MS анализе. Избор седам молекулских дескриптора извршен је коришћењем факторске анализе и генетског алгоритма. Примећено је да изабрани дескриптори нису били у међусобној корелацији, па су коришћени као улазни подаци при изградњи вештачке неуронске мреже. У изградњи модела предвиђања ретенционих времена коришћено је укупно 40 експериментално добијених ретенционих времена. Коефицијент детерминације током циклуса припреме, тестирања и

валидације достигао је вредности 0,950; 0,825 и 1,000, редом, што указује на то да се овај модел може користити за предвиђање ретенционих времена хемијских једињења добијених из есенцијалног уља *A. clypeolata*.

(Примљено 24. маја 2020, ревидирано 17. јануара, прихваћено 5. фебруара 2021)

REFERENCES

1. A. Nedelcheva, *J. Appl. Pharm. Sci.* **2** (2012) 165 (https://doi.org/10.7324/JAPS.2012.2828)
2. R. Contreras-Medina, I. Luna-Vega, in *Global Advances in Biogeography*, L. Stevens, Ed., InTech, Rijeka, 2012, pp. 61–70 (https://doi.org/10.5772/34315)
3. I. Aljančić, S. Macura, N. Juranić, S. Anđelković, N. Ranđelović, S. Milosavljević, *Phytochemistry* **43** (1996) 169 (https://doi.org/10.1016/0031-9422(96)00271-3)
4. M. Todorova, M. Krasteva, M. Markova, E. Tsankova, R. Taskova, D. Peev, *Phytochemistry* **49** (1998) 2371 (https://doi.org/10.1016/S0031-9422(98)00201-5)
5. I. Werner, P. Mucaji, A. Presser, S. Glasl, *Z. Naturforsch*. **62** (2007) 267 (https://doi.org/10.1002/chin.200725175)
6. J. Cvetkovic, M. Dimitrijevic, M. Ilic, S. Simonovic, V. Stankov-Jovanovic, V. Mitic, G. Stojanovic, in *Proceedings of Congress of Chemists and Chemical Engineers of Bosnia and Herzegovina with international participation* (2014), Sarajevo, Bosnia and Hercegovina, *Bulletin of the Chemists and Technologists of Bosnia and Herzegovina*, 2014, p. 99
7. M. Aćimović, M. Zorić, V. D. Zheljazkov, L. Pezo, I. Čabarkapa, J. Stanković Jeremić, M. Cvetković, *Molecules* **25** (2020) 5482 (https://doi.org/10.3390/molecules25225482)
8. N. Simić, R. Palić, V. Ranđelović, *Flavour Fragr. J*. **20** (2005) 127 (https://doi.org/10.1002/ffj.1391)
9. J. C. Chalcat, S. D. Petrovic, Z. A. Maksimović, M. S. Gorunović, *J. Essent. Oil Res.* **15** (2005) 549 (https://doi.org/10.1080/10412905.2005.9698991)
10. B. Zlatković, S. Bogosavljević, A. Radivojević, M. Pavlović,. *J. Ethnopharmacol*. **151** (2014) 704 (https://doi.org/10.1016/j.jep.2013.11.037)
11. J. L. Wolfender, G. Martia, A. Thomas, S. Bertranda, *J. Chromatogr., A* **1382** (2015) 136 (https://doi.org/10.1016/j.chroma.2014.10.091)
12. K. Héberger, *J. Chromatogr., A* **1158** (2007) 273 (https://doi.org/10.1016/j.chroma.2007.03.108)
13. R. Kaliszan, *Chem. Rev.* **107** (2007) 3212 (https://doi.org/10.1021/cr068412z)
14. L. Wu, P. Gong, Y. Wu, K. Liao, H. Shen, Q. Qi, H. Liu, G. Wang, H. Hao, *J. Chromatogr., A* **1303** (2013) 39 (https://doi.org/10.1016/j.chroma.2013.06.041)
15. T. Khezeli, A. Daneshfar, R. Sahraei, *Talanta* **150** (2016) 577 (https://doi.org/10.1016/j.talanta.2015.12.077)
16. Y. Marrero-Ponce, S. J. Barigye, M. E. Jorge-Rodriguez, T. Tran-Thi-Thu, *Chem. Pap*. **72** (2017) 57 (https://doi.org/10.1007/s11696-017-0257-x)
17. D. Micić, S. Ostojić, L. Pezo, S. Blagojević, B. Pavlić, Z. Zeković, S. Đurović, *Ind. Crops Prod.* **138** (2019) (https://doi.org/10.1016/j.indcrop.2019.06.001)
18. A. Tropsha, A. Golbraikh, *Curr. Pharm. Des.* **13** (2007) 3494 (https://doi.org/10.2174/138161207782794257)
19. B. Zisi, I. Sampsonidis, S. Fasoula, K. Papachristos, M. Witting, H.G. Gika, P. Nikitas, A. Pappa-Louisi *Metabolites* **7** (2017) 7 (https://doi.org/10.3390/metabo7010007)
20. D. D. Matyushin, A. Y.Sholokhova, A. K. Buryak, *J. Chromatogr., A* **1607** (2019) 460395 (https://doi.org/10.1016/j.chroma.2019.460395)

21. J. Dong, D. S. Cao, H. Y. Miao, S. Liu, B. C. Deng, Y. H. Yun, N. N. Wang, A. P. Lu, W. B. Zeng, A. F. Chen, *J. Cheminformatics* **7** (2015) 60 (https://doi.org/10.1186/s13321-015-0109-z)
22. C. W. Yap, *J. Comput. Chem*. **32**,(2011) 1446 (https://doi.org/10.1002/jcc.21707)
23. D. E. Goldberg, *Genetic algorithms in search, optimisation and machine learning*, Addison-Wesley, Longman, Boston, MA, 1989 (ISBN:0201157675) (https://doi.org/10.5860/choice.27-0936)
24. A. Tropsha, *Mol. Inform.* **29** (2010) 476 (https://doi.org/10.1002/minf.201000061)
25. Statistica 10 software (StatSoft, Inc. Statistica, ver. 10, data analysis software system) (http://www.statsoft.com/, last accessed 10 January 2019)
26. R. Aalizadeh, N. S. Thomaidis, A. A. Bletsou, P. Gago-Ferrero, *J. Chem. Inf. Model.* **56** (2016) 1384 (https://doi.org/10.1021/acs.jcim.5b00752)
27. Q. Xu, C. Wei, R. Liu, S. Gu, J. Xu, *Chemom. Intell. Lab. Syst.* **146** (2015) 313 (https://doi.org/10.1016/j.chemolab.2015.06.001)
28. P. Kojic, R. Omorjan, *Chem. Eng. Res. Des.* **125** (2018) 398 (https://doi.org/10.1016/j.cherd.2017.07.029)
29. Y. Yoon, G. Swales, T. M. Margavio, *J. Oper. Res. Soc.* **44** (1993) 51 (https://doi.org/10.1057/jors.1993.6)
30. N. Radulović, B. Zlatković, R. Palić, G. Stojanović, *Nat. Prod. Comm.* **2** (2007) 453 (https://journals.sagepub.com/doi/pdf/10.1177/1934578X0700200417)
31. M. Mohammadhosseini, *Anal. Chem. Lett.* **3** (2013) 226 (https://doi.org/10.1080/22297928.2013.861164)
32. M. Nekoei, M. Mohammadhosseini, E. Pourbasheer, *Med. Chem. Res.* **24** (2015) 3037 (https://doi.org/10.1007/s00044-015-1354-4)
33. R. Todeschini, V. Consonni, *Molecular descriptors for chemoinformatics*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2009, p. 27 (ISBN: 978-3-527-31852-0)
34. L. B. Kier, L. H. Hall, *Molecular connectivity in chemistry and drug research*, Academic Press, New York, 1976, p. 1214 (https://doi.org/10.1002/jps.2600660852)
35. K. Roy, G. Ghosh, *J. Chem. Inf. Comput. Sci.* **44** (2004) 559-567. (https://doi.org/10.1021/ci0342066)
36. A. Roon, J.R. Parsons, H. A. J. Govers, *J. Chromatogr., A* **955** (2002) 105 (https://doi.org/10.1016/S0021-9673(02)00200-5)
37. M. T. Scotti, M. B. Fernandes, M. J. P. Ferreira, V. P. Emerenciano, *Bioorgan. Med. Chem*. **15** (2007) 2927 (https://doi.org/10.1016/j.bmc.2007.02.005)
38. B. Hollas, *MATCH* **45** (2002) 27 (http://match.pmf.kg.ac.rs/electronic_versions/Match45/match45_27-33.pdf)
39. R. García-Domenech, J. GálvezJesus, J.V. de Julián-Ortiz, L. Pogliani, *Chem. Rev*. **108** (2008) 1127 (https://doi.org/10.1021/cr0780006).