Understanding the Chromatographic Properties and Cytotoxicity of Hidrazinoselenazole Compounds by Computational Study

ALINA C. COZMA[1], ADRIANA GROZAV[2], IOAN C. TARCA[3], VALENTIN ZAHARIA[2], SORANA D. BOLBOACA[4]*, LORENTZ JANTSCHI[5]

[1] University of Oradea, Department of Chemistry, 1 Universitatii Str., 410087, Oradea, Romania

[2] Iuliu Hatieganu University of Medicine and Pharmacy, Department of Organic Chemistry, 6 Louis Pastuer, 400269, Cluj Napoca, Romania

[3] University of Oradea, Department of Mechanical Engineering and Automotives, 1 Universitatii Str., 410087, Oradea, Romania

[4]Iuliu Hatieganu University of Medicine and Pharmacy, Department of Medical Informatics and Biostatistics, 6 Louis Pastuer, 400269, Cluj Napoca, Romania

[5]Technical University of Cluj-Napoca, Department of Physics and Chemistry, 103-105 Muncii Blvd., 400641 Cluj Napoca, Romania & Babes-Bolyai University, Institute for Doctoral Studies, 1st Kogalniceanu Str., 400084, Cluj Napoca, Romania

Selenium compounds have been proven to possess anti-inflammatory, anti-cancer, anti-bacterial and anti-viral activities. A series of fifteen synthesized hidrazinoselenazole was investigated to assess the chromatographic properties as function of structural features and the cytotoxicity as function of chromatographic properties and/or structural descriptors. The investigated chromatographic properties were retention factor, specific surface area of the solvent and chromatographic hydrophobicity index. The 3D model of the compounds was optimized using Hartree-Fock DFT/B3LYP method, 6-31+G basis set both in vacuum and water with Spartan software (v.8). Furthermore, several quantitative structure-activity relationship (QSAR) descriptors were calculated with Spartan and Dragon (v. 5.5) software. Full search approach was used to construct simple and multiple linear regression models. No reliable model was identified for specific surface area of the solvent. The models with higher performances in estimation and prediction for retention factor and chromatographic hydrophobicity index proved the ones with Dragon descriptors and molecules optimized in water (retention factor: r_{loo}^2 (loo = leave-one-out analysis) = 0.9244; r_{tr}^2 (tr = training set) = 0.9652; r_{ts}^2 (ts = test set) = 0.9606; chromatographic hydrophobicity index: r_{loo}^2 = 0.9489; r_{tr}^2 = 0.9592; r_{ts}^2 = 0.9669). The cytotoxicity proved related neither to chromatographic properties nor with compounds' structural characteristics.*

Keywords: quantitative structure-property relationship (QSPR); quantitative property-cytotoxicity relationship (QPCR); selenazole; computational modeling

Quantitative structure-activity/property relationship (QSAR/QSPR) models are regression or classification models used to link the structure of chemical compounds with associated pharmacological activity/properties in a quantitative manner [1]. The approaches are based on the assumption that the structure of chemical compound (such as geometric, topologic, steric, electronic properties, etc.) contains features responsible for its physical, chemical and biological properties [2,3]. Corwin Herman Hansch developed one of the earliest models that linked the property of compounds with their structures [4]. Quantitative structure-property/activity relationships (QSPR/QSPR) is a term used when a chemical property/ activity is modeled as the response variable [5]. The mathematical expression, if carefully validated, could be further used to predict the modeled response of other similar chemical structures [6-8]. According with Tropsha et al. [6], any QSPR/QSAR model is considered reliable and predictive if: (1) is statistically significant and robust; (2) is validated by making accurate predictions for external data sets that were not used in the model development; and (3) have defined application boundaries.

Pharmaceutical experiments were done to investigate the link between biological activity and/or properties of the heterocyclic selenium compounds and their structures [9]. Selenium is a controversial chemical element because is considered an essential nutrient by its presence in some selenoproteins [10] but its accumulation in cultivated plants had been proven to causes poisoning [11]. Furthermore, selenium is involved in detoxification processes [12] and proved to decrease the inflammation [13]. The heterocyclic selenium compounds also possess anti-cancer [14,15], anti-bacterial and anti-viral activities [16].

A group of chemists from the Department of Organic Chemistry, Iuliu Haþieganu University of Medicine and Pharmacy Cluj-Napoca synthesized a series of 15 hidrazinoselenazole compounds [15,17]. The structures of the synthesized compounds were confirmed using the IR, ^{1}H-RMN, ^{13}C-RMN, COSY, HMQC, HMBC, ^{77}Se-RMN and SM spectra [17]. The compounds were obtained by a Hantzsch condensation reaction using a series of ariliden-selenosemicarbazide and aroyl-selenosemicarbazide as selenoamidic component [17]. This study was conducted on this sample of hidrazinoselenazole compounds and aimed to investigate (i) the chromatographic properties as function of structural features and (ii) their citotoxicity as function of chromatographic properties and/or structural descriptors involving computational modeling.

Experimental part

Materials and methods

Dataset of hidrazinoselenazole

The structures of fifteen selenazole compounds investigated in this study are presented in the supplementary file. This set of compounds was previously analyzed using RP-HPLC (Reverse Phase - High-Performance Liquid Chromatography) and the chromatographic properties represented by retention factor, chromatographic hydrophobicity index and specific surface area of the solvent are presented in table 1. The experimental values of cytotoxicity investigated on liver hepatocelular carcinoma (HepG2) cell line expressed in µg/mL are also given in Table 1 [18].

The logarithm of retention factor R_f ($\log(R_f)$) has been computed using the following formula:

$$\log(R_f) = \log(t_R - t_0)/t_0 \quad (1)$$

where R_f = retention factor, t_R = retention time of the solute, and t_0 = dead time.

* email: sbolboaca@umfcluj.ro; Phone: +40750774506

Retention factor is defined as the ratio of the distance traveled by the centre of the spot to the distance simultaneously traveled by the mobile phase [19].φ_0) was calculated using the formula given in eq (2):

$$\varphi_0 = - \log(R_f)/S \quad (2)$$

where R_f is the value extrapolated for a concentration of organic solvent equal to 0% of the retention factor; and S is the specific surface area of the solvent.

Dataset of hidrazinoselenazole

The computational study has conducted to identify the link between structure and properties, respectively properties and cytotoxicity of the investigated selenazole, and the following steps were applied:

- Draw the 2D (two-dimensional) structure of compounds with ChemDraw software (v. 6.0).
- Build and optimize the 3D model of the compounds with Spartan software (v. 8) using equilibrium geometry at ground state with Hartree-Fock DFT/B3LYP method and 6-31+G* basis set, both in vacuum and water.
- Extract structural information from optimized compounds by computing the theoretical structural descriptors with:

-Spartan software (v.8): area (surface area, Δ2), volume (Δ3), PSA (polar surface area, Δ2), E-Homo (Energy-Highest occupied molecular orbital, eV; 1 eV = 1.602 × 10^{-19}J), E-Lumo (Energy-Lowest unoccupied molecular orbital, eV), dipole-moment (Debye), ovality (adimensional), AccArea (Δ), MinEIPot (kJ/mol), MinLocIonPot (kJ/mol), P-Area (Δ2), ΔccP-Area (Δ2), MaxEIPot (kJ/mol), logP, polarizability, ZPE (zero-point energy, kJ/mol), H (enthalpy, au; atomic units; 1 au = 2625 kJ/mol), CV (heat capacity at constant volume, J/mol), S (entropy, J/mol), and G (free enthalpy, au).

-Dragon software (v. 5.5): compute several classes of 2D descriptors such as Gateway, WHIM, Morse, information, topological, geometrical, and constitutional descriptors [20].

Select non-redundant descriptors: degenerated descriptors defined as those with identical values for more than three compounds were excluded from the pool of predictors in searching the models.

Identification of linear models

A full search approach was applied to identify simple and multiple linear models (LMs). Identification of the multiple linear models was done using the following criterion regarding the maximum number of descriptors to be included in the model: $k = n/5$, where k = maximum number of descriptors in the model, n = number of compounds in the sample. The steps applied in identification of performing LMs were:

-Step 1: Test the normal distribution of chromatographic data. Data were considered normal distributed if an agreement of Anderson-Darling, Kolmogorov-Smirnov, and Chi-Square test [21] was obtained according to Fisher's combined probability test [22].

Table 1

COMPOUNDS AND THEIR CHROMATOGRAPHIC PROPERTIES AND MEASURED CYTOTOXICITY

Comp.	log(R_f)	S	φ_0	IC_{50} (μg/mL)
S01	5.21	-0.06	94	>25
S02	4.31	-0.05	83	>25
S03	3.99	-0.05	85	20.18
S04	4.76	-0.06	85	17.98
S05	4.72	-0.06	76	>25
S06	4.87	-0.06	89	>25
S07	3.54	-0.05	75	24.49
S08	4.22	-0.05	90	19.56
S09	2.82	- 0.04	64	24.49
S10	3.81	-0.05	75	23.69
S11	2.31	-0.04	53	8.84
S12	2.2	-0.04	61	**3.38**
S13	1.68	-0.03	53	2.9
S14	3.08	-0.05	67	**2.84**
S15	4.15	-0.06	75	8.17

log(k_{ow}) = logarithm of retention factor k; S = specific surface area of the solvent;φ_0 = chromatographic hydrophobicity index; IC_{50} = half maximal inhibitory concentration

-Step 2: Search for and evaluate the best performing linear regression models using performance criteria in estimation and prediction. The applied performances in estimation were determination coefficient (r^2) and its adjusted form (r^2_{adj}), rate of the variance explained by the model and its significance (F statistic and its associated p-value), significance of model coefficients (t-statistics and its significance), measures of residuals errors (e.g. MAE = Mean Absolute Error, MAPE = Mean Absolute Percentage Error, SEP = Standard Error of Prediction; REP(%) = Relative Error of Prediction, TES = Total Square Error, CCC = Concordance Correlation Coefficient (CCC). The applied performances in prediction: prediction power (PP, Fisher approach [23]), leave-one-out analysis (loo), and training vs. test analysis (n_{TR} = 2×15/3 = 10, where n_{TR} = number of compounds in training set).

-Step 3: Compare the models able to link the property/ cytotoxicity data with theoretical descriptors/ chromatographic properties in regards of goodness-of-fit: Steiger's test [24].

- Step 4: Identify and assess the models able to estimate/ predict the cytotoxicity of investigated compounds using as input data the values of chromatographic measurements and the values of theoretical descriptors.

Results and discussions

Vacuum and water optimization

The method used in optimization of 3D structure of compounds influence the value of Spartan's descriptors,

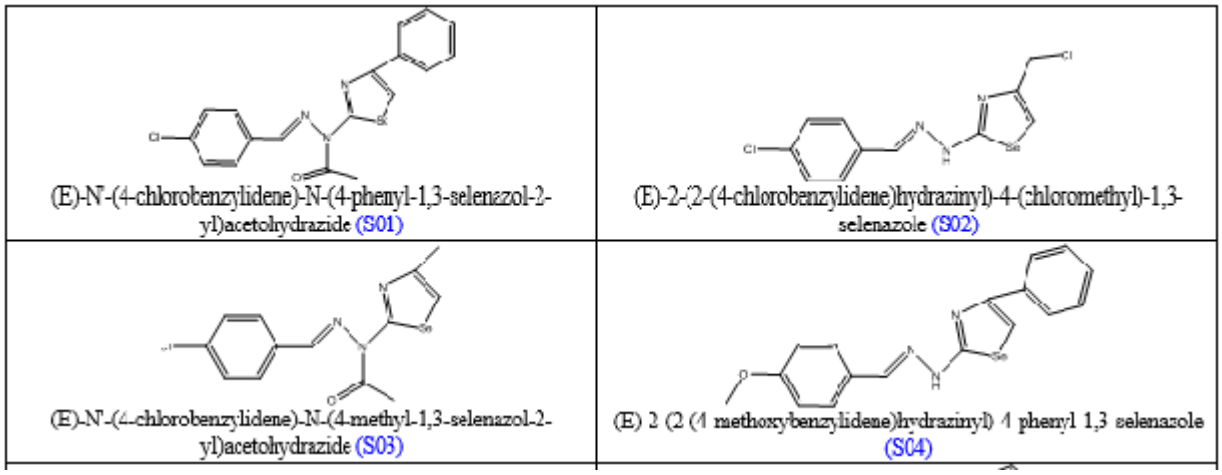


Table 2

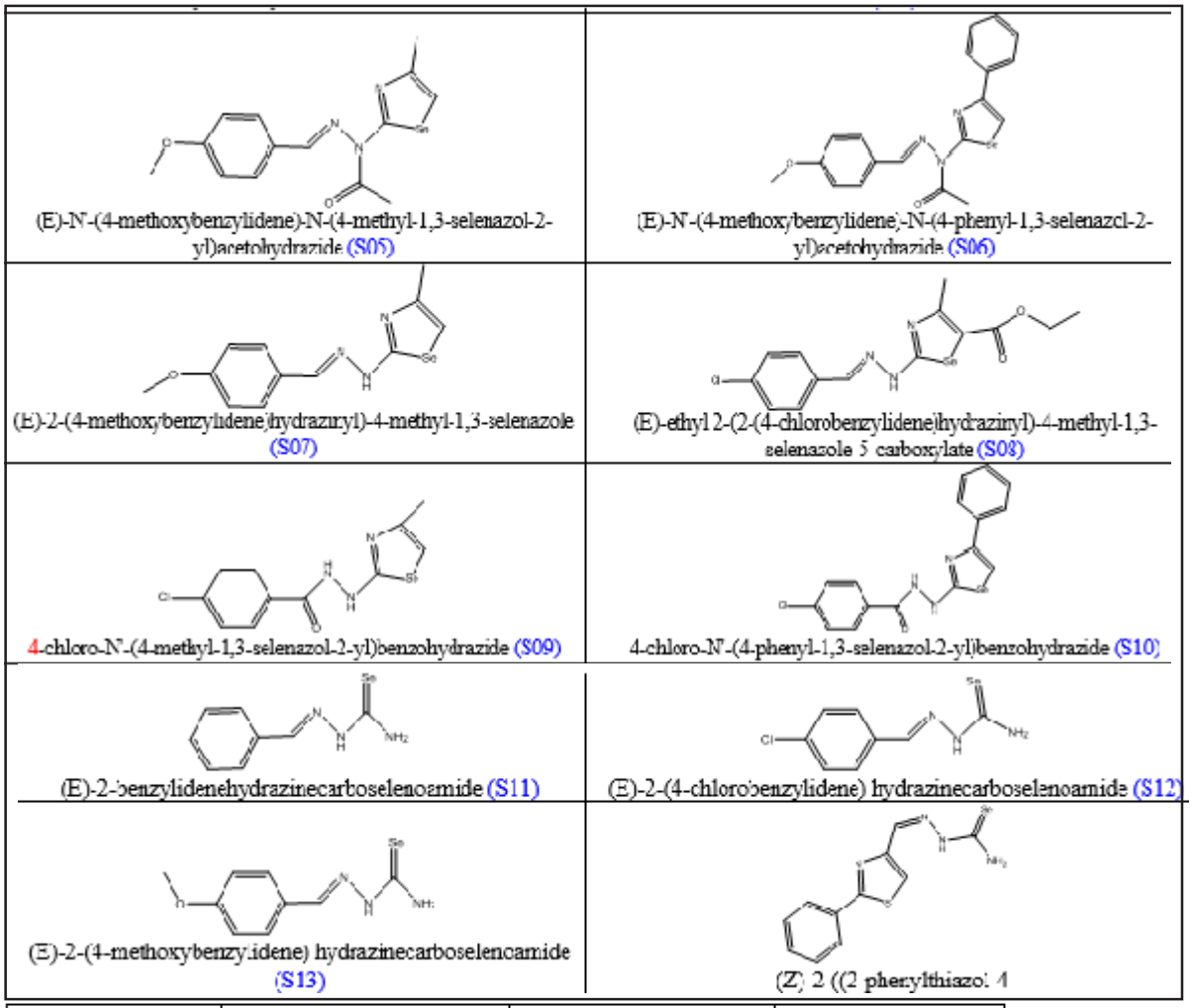
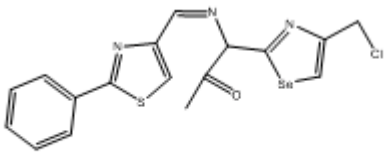


Table 2
(continued)

Descriptor	Optimization: vaccum	Optimization: water	Stat (p-value)
Area[a]	298.37±54.94	298.56±54.92	-2.16 (0.0482)
Volume[a]	273.11±54.94	273.26±54.92	-2.71 (0.0171)
PSA[a]	36.81±8.83	37.19±8.47	-2.72 (0.0165)
Ovality[a]	1.47±0.06	1.46±0.06	1.00 (0.3343)
AccArea[a]	220.82±31.95	221.20±31.36	-0.34 (0.7396)
MinEIPot[a]	-219.12±35.56	-262.41±42.29	15.28 ($4.00 \cdot 10^{-10}$)
MinLocIonPot[b]	44.09 (39.11−44.44)	45.05 (42.99−45.44)	3.24 (0.0012)
P-Area[a]	101.25±20.12	123.80±20.86	-19.79 ($1.24 \cdot 10^{-11}$)
AccP-Area[a]	73.40±18.28	89.61±18.31	-15.27 ($4.01 \cdot 10^{-10}$)
MxEIPot[a]	210.64±65.00	259.38±83.78	-8.21 ($1.01 \cdot 10^{-6}$)
Polarizability[a]	61.12±4.46	61.03±4.52	3.71 (0.0023)
HBDCount[b]	0.00 (0.00−1.50)	0.00 (0.00−1.00)	0.00 (0.9999)
HBACount[b]	3.00 (3.00−4.00)	3.00 (3.00−3.50)	0.00 (0.9999)
E-HOMO[a]	-7.96±0.36	-8.37±0.30	3.83 (0.0018)
E-LUMO[a]	2.29±0.37	2.34±0.27	-0.90 (0.3821)
Dipole moment[a]	5.17±2.55	7.12±3.31	-8.84 ($4.19 \cdot 10^{-7}$)
ZPE[a]	662.95±143.32	664.28±143.63	-3.93 (0.0015)
H[0a]	-3471.12±350.45	-3471.94±349.74	1.34 (0.2019)
Cv[a]	224.23±49.18	256.34±52.05	-3.24 (0.0059)
S0[a]	491.67±44.75	490.89±45.96	1.33 (0.2041)
G0[a]	-3471.78±349.76	-3471.80±349.76	1.94 (0.0724)

a: mean±standard deviation & paired t-test; b: median (Q1−Q3) & Wilcoxon Matched Pairs Test

Table 3



(Z)-1-(4-(chloromethyl)-1,3-selenazol-2-yl)-1-(((2-phenylthiazol-4-yl)methylene)amino)propan-2-one (S15)

Fig. 1. Compounds, associated name and abbreviation

frequently obtaining higher values when the compounds were optimized in water (table 2,3 and fig. 1).

QSPR analysis: chromatography parameters

The computational study was carried out for compounds optimized in vacuum and respectively for compounds optimized in water. The descriptor(s) selection process gave different significant descriptors ($p<0.05$). The best performing models identified by the full search were as follows:

$$\hat{Y}_{\log(Rf)} = -124.57 - 0.26 \cdot Volume - 0.95 \cdot HBDCount + 3.27 \cdot Polarizability \quad (3)$$

$$\hat{Y}_S = -0.02 - 0.00011 \cdot Volume \quad (4)$$

$$\hat{Y}_{\varphi 0} = 83.41 - 12.62 \cdot HBDCount \quad (5)$$

-Water & Spartan descriptors (eq.(6)-(8)):

$$\hat{Y}_{log(Rf)} = -4.54 - 0.05 \cdot PSA + 0.17 \cdot Polarizability \quad (6)$$

$$\hat{Y}_S = 0.0843 - 0.0030 \cdot MinLocIonPot \quad (7)$$

$$\hat{Y}_{\varphi 0} = 69.82 + 0.23 \cdot P\text{-}Area - 2.43 \cdot DipoleMoment - 10.13 \cdot HBDCount \quad (8)$$

where $\log(R_f)$ = logarithm of retention factor; S = specific surface area; φ_0 = chromatographic hydrophobicity index; HBDCount = no. of hydrogen-bond donor; Polarizability = alpha polarizability parameter; P-Area = polar areas, defined as the area for which the absolute value of the electrostatic potential is >100 kJ/mol; DipoleMoment = dipole moment (debyes); large value indicates large separation of charge.

- Vacuum & Dragon descriptors (eq(9)-(11)):

$$\hat{Y}_{log(Rf)} = 12.92 + 4.13 \cdot H8u - 13.42 \cdot REIG \quad (9)$$

$$\hat{Y}_S = -0.21 + 0.77 \cdot G2v + 0.25 \cdot R5m + ?????????? \quad (10)$$

$$\hat{Y}_{\varphi 0} = 98.15 - 172.11 \cdot HATS1u + 159.80 \cdot H7e \quad (11)$$

where $\log(R_f)$ = logarithm of retention factor; S = specific surface area; φ_0 = chromatographic hydrophobicity index; H8u = H autocorrelation of lag 8 / unweighted, REIG = first eigenvalue of the R matrix, G2v = 2nd component symmetry directional WHIM index / weighted by van der Waals volume, R5m = R autocorrelation of lag 5 / weighted by mas, HATS1u = leverage-weighted autocorrelation of lag 1 / unweighted, H7e = H autocorrelation of lag 7 / weighted by Sanderson electronegativity.

- Water & Dragon descriptors (eq(.12)-(14)):

$$\hat{Y}_{log(Rf)} = 12.74 + 3.91 \cdot H8u - 13.25 \cdot REIG \quad (12)$$

$$\hat{Y}_S = -0.05 - 0.19 \cdot PW4 + 0.02 \cdot Mor30e + 0.22 \cdot Gm \quad (13)$$

$$\hat{Y}_{\varphi 0} = 205.72 \cdot H7e + 31.22 \cdot Mor12p + 229.19 \cdot G3m \quad (14)$$

where H8u = H autocorrelation of lag 8 / unweighted, REIG = first eigenvalue of the R matrix, PW4 = path/walk 4 - Randic shape index, Mor30e = signal 30 / weighted by Sanderson electronegativity, Gm = total symmetry index / weighted by mas, H7e = H autocorrelation of lag 7 / weighted by Sanderson electronegativity, *Mor12p* = signal 12 / weighted by polarizability, *G3m* = 3rd component symmetry directional WHIM index / weighted by mass.

The metrics for characterization of the power of estimation are presented in table 4, and the parameters related with residual errors for each identified model (eq.(3)-(14)) in table 5. The analysis of the estimation parameters of the models presented in eq.(3)-eq.(14) revealed that without any exception, the model with the best goodness-of-fit is a model obtained on molecules optimized in water for all investigated properties (retention factor Eq(12) & specific surface area of the solvent eq.(13) & chromatographic hydrophobicity index eq(14), table 4). The models relating the specific surface area of the solvent with structure of compounds proved smallest goodness-of-fit, with the highest determination coefficient equal with 0.8868 (table 4).

Internal validity of models was tested in leave-one-out

Table 4
QSPR MODELS: CHARACTERISTICS OF POWER OF ESTIMATION

Eq.	r^2	r^2_{adj}	RMSE	F (p)	$t_{coefficients}$ (p)
3	0.9264	0.9064	0.3292	46 ($1.60 \cdot 10^{-6}$)	Int: -2.82 (0.0167); X_{Volume}: -5.01 (0.0004); $X_{HBDCount}$: 2.86 (0.0156); $X_{Polarizability}$: -2.76 (0.0187)
4	0.5890	0.5574	0.0052	19 (0.0008)	Int: -2.63 (0.0209); X_{Volume}: -4.32 (0.0008)
5	0.7531	0.7341	6.7464	40 ($2.76 \cdot 10^{-5}$)	Int: 38.00 ($1.04 \cdot 10^{-14}$); $X_{HBDCount}$: -6.30 ($2.76 \cdot 10^{-5}$)
6	0.8829	0.8634	0.3976	45 ($2.58 \cdot 10^{-6}$)	Int: -2.38 (0.0345); X_{PSA}: -3.52 (0.0042); $X_{Polarizability}$: 6.28 ($4.05 \cdot 10^{-5}$)
7	0.5927	0.5613	0.0052	19 (0.0008)	Int: 2.76 (0.0164); $X_{MinLocIonPot}$: -4.35 (0.0008)
8	0.8642	0.8271	5.4396	23 ($4.53 \cdot 10^{-5}$)	Int: 7.29 ($1.57 \cdot 10^{-5}$); $X_{P\text{-}Area}$: 2.34 (0.0394); $X_{DipoleMoment}$: -3.73 (0.0033); $X_{HBDCount}$: -5.28 (0.0003)
9	0.9457	0.9366	0.2708	104 ($2.57 \cdot 10^{-8}$)	Int: 11.95 ($5.07 \cdot 10^{-8}$); X_{H8u}: 5.08 (0.0003); X_{REIG}: -9.22 ($8.52 \cdot 10^{-7}$)
10	0.6070	0.5415	0.0171	9 (0.0037)	Int: -4.03 (0.0017); X_{G2v}: 2.77 (0.0170); X_{R5m+}: 3.23 (0.0072)
11	0.9070	0.8915	4.3102	58 ($6.49 \cdot 10^{-7}$)	Int: 5.79 ($8.63 \cdot 10^{-5}$); X_{HATS1u}: -2.75 (0.0175); X_{H7e}: 5.92 ($6.99 \cdot 10^{-5}$)
12	0.9518	0.9437	0.2552	118 ($1.26 \cdot 10^{-8}$)	Int: 12.58 ($2.85 \cdot 10^{-8}$); X_{H8u}: 5.04 (0.0003); X_{REIG}: -9.70 ($4.95 \cdot 10^{-7}$)
13	0.8868	0.8560	0.0030	29 ($1.67 \cdot 10^{-5}$)	Int: -2.44 (0.0331); X_{PW4}: -3.14 (0.0094); X_{Mor30e}: 2.77 (0.0182); X_{Gm}: 2.49 (0.0301);
14	0.9646	0.8754	2.6570	109 ($1.82 \cdot 10^{-8}$)	X_{H7e}: 13.76 ($1.03 \cdot 10^{-8}$); X_{Mor12p}: 6.79 ($1.94 \cdot 10^{-5}$); X_{G3m}: 29.37 ($1.51 \cdot 10^{-12}$)

Int = the intercept of the models; X = theoretical descriptors (eq.(3)-(14))

Table 5
MEASURES OF RESIDUAL ERRORS ON MODELS FOR CHROMATOGRAPHIC CHARACTERISTICS

Eq	MAE	MAPE	SEP	REP(%)	CCC [95%CI]	TSE
3	0.2484	0.0766	0.2918	7.86	0.9618 (0.8919−0.9868)	4
4	0.0037	5.3035	0.0050	10.37	0.7413 (0.4234−0.8966)	2
5	5.5725	0.0035	6.5010	8.67	0.8591 (0.6454−0.9480)	2
6	0.2742	0.0738	0.3681	9.92	0.9378 (0.8289−0.9782)	3
7	0.0041	5.2312	0.0050	10.33	0.7442 (0.4282−0.8980)	2
8	3.4955	0.0033	4.8217	6.43	0.9271 (0.8020−0.9743)	4
9	0.2043	0.0726	0.2507	6.75	0.9720 (0.9200−0.9904)	3
10	0.0145	5.1300	0.0198	40.81	0.2340 (-0.1640−0.5668)	3
11	3.2615	0.0034	3.9904	5.32	0.9512 (0.8636−0.9830)	3
12	0.1868	0.0735	0.2363	6.37	0.9752 (0.9288−0.9915)	3
13	0.0018	6.7375	0.0026	5.44	0.9400 (0.8346−0.9790)	4
14	1.8326	0.0034	2.4599	3.28	0.9823 (0.9485−0.9940)	5

MAE= Mean Absolute Error; MAPE = Mean Absolute Percentage Error;
SEP = Standard Error of Prediction; REP = Relative Error of Prediction;
CCC = Concordance Correlation Coefficient; 95% CI = 95% confidence interval;
TSE = Total Square Error

Eq.	Leave-one-out			Training vs. test (n_{TR} = 10, n_{TS} = 5)			
	r^2_{loo}	s_{loo}	F_{loo} (p)	r^2_{TR}	F_{TR} (p)	$t_{coefficients}$ (p)	r^2_{TS}
3	0.8698	0.44	24 ($3.9 \cdot 10^{-5}$)	0.9415	32 (<0.001)	Int: -94.97 (0.297); X_{Volume}: -0.20 (0.310); $X_{HBDCount}$: -0.82 (0.019); $X_{Polarizability}$: 2.50 (0.289)	0.9274
4	0.4640	0.01	11 (0.006)	0.5242	9 (0.018)	Int: -0.03 (0.020); X_{Volume}: $-9.09 \cdot 10^{-5}$ (0.018)	0.8335
5	0.6718	8.17	12 (0.001)	0.7694	27 (0.001)	Int: 84.83 ($5.7 \cdot 10^{-10}$); $X_{HBDCount}$: -13.5 (0.001)	0.6935
6	0.8239	0.49	28 ($3.1 \cdot 10^{-5}$)	0.9340	50 ($7.4 \cdot 10^{-5}$)	Int: -4.86 (0.018); X_{PSA}: -0.05 (0.010); $X_{Polarizability}$: 0.17 (<0.001)	0.8619
7	0.4446	0.01	10 (0.008)	0.6529	15 (0.005)	Int: 0.098 (0.034); $X_{MinLocIonPot}$: -0.003 (0.005)	0.5177
8	0.7726	7.07	12 (0.001)	0.8344	10 (0.009)	Int: 70.77 (0.011); $X_{P\text{-}Area}$: 0.19 (0.261); $X_{DipoleMoment}$: -1.81 (0.2989); $X_{HBDCount}$: -10.29 (0.0365)	0.9390
9	0.9160	0.34	65 ($3.6 \cdot 10^{-7}$)	0.9587	81 ($1.4 \cdot 10^{-5}$)	Int: 11.83 ($2.5 \cdot 10^{-5}$); X_{H8u}: 4.12 (0.004); X_{REIG}: -11.77 (<0.001)	0.9542
10	0.0040	0.03	-2.32 (0.140)	0.0609	0.23 (0.803)	Int: -0.08 (0.080); X_{G2v}: 0.12 (0.524); X_{R5m+}: 0.02 (0.7247)	0.9481
11	0.8549	5.41	35 ($9.9 \cdot 10^{-6}$)	0.8843	27 (0.001)	Int: 99.24 (0.004); X_{HATS1u}: -167.43 (0.102); X_{HTe}: 147.55 (0.007)	0.9635
12	0.9244	0.32	73 ($1.9 \cdot 10^{-7}$)	0.9652	97 ($7.9 \cdot 10^{-6}$)	Int: 11.56 ($1.4 \cdot 10^{-5}$); X_{H8u}: 4.02 (0.003); X_{REIG}: -11.7 ($9.6 \cdot 10^{-5}$)	0.9606
13	0.8086	0.004	15 (0.0003)	0.9252	25 (0.001)	Int: -0.11 (0.010); $X_{P(V)4}$: -0.10 (0.154); X_{Mor30e}: 0.02 (0.068); X_{Gm}: 0.45 (0.012);	0.6913
14	0.9489	3.40	65 ($2.7 \cdot 10^{-7}$)	0.9592	54 ($9.9 \cdot 10^{-5}$)	X_{HTe}: 188.19 ($8.0 \cdot 10^{-5}$); X_{Mor12p}: 35.33 (0.001); X_{G3m}: 233.62 ($3.0 \cdot 10^{-07}$)	0.9669

Int = intercept; X = molecular descriptors as presented in eq.(1)-(12)

Table 6
MODELS PREDICTION PERFORMANCES IN LEAVE-ONE-OUT (LOO) AND TRAINING *vs.* TEST ANALYSES

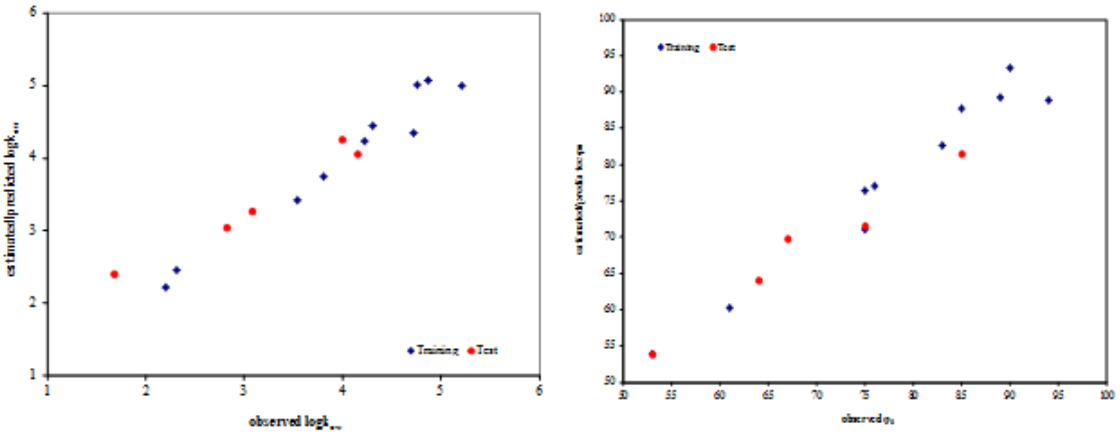


Fig. 2. Best performing models in estimation and prediction of chromatographic parameters

analysis, and with one exception represented by eq.(10), all models proved internally valid. The analysis of the determination coefficients in leave-one-out analysis proved that the best performing models are obtained on molecules optimized in water using Dragon descriptors (eq.(12)-(14), table 6). The model presented in eq.(10) proved not significant in training test and the residuals of 3 models (eq.(3), eq.(4) and v(8)) proved significantly different by expected value (zero).

Top 3 models according with predictive power defined as high value of determination coefficient in test sets are as follows: eq.(12) for *logarithm of retention factor* ($\log(R_f)$), eq.(13) for *specific surface area of the solvent* (S), and eq.(14) for *chromatographic hydrophobicity index* (φ_0). The lowest predictive performance is obtained for the specific surface area of the solvent even so two models (eq.(10) and eq.(4)) were with good predictive performances in terms of determination coefficients but with coefficients not significantly different by zero.

The analysis of the models for each investigated chromatographic property identified as best performing models those who comprise descriptors used by eq.(12) for logarithm of retention factor and respectively eq.(14) for chromatographic hydrophobicity index (fig. 2).

Two models with prediction performances were obtained for the logarithm of retention factor and chromatographic hydrophobicity index on the investigated set of hidrazinoselenazole. However, identification of performing models will not assure a high accuracy when new compounds are analyzed. Testing the model on external molecules is desired but was not performed in this analysis due to the limited number of hidrazino-selenazole compounds. The small number of investigated compounds due to their availability gives another limitation that is link with the previous one. The true accuracy of the models for retention factor and chromatographic hydrophobicity index is investigated when new compounds from the same class will be identifies and will be characterized.

QPCR analysis: cytotoxicity

The analysis of the relation between cytotoxicity as outcome variable (eight compounds out of fifteen) and chromatographic measurements as predictors identified just one significant linear model (eq. (15)):

$$\hat{Y}_{(1/\log(IC50))} = 2.52 - 0.41 \cdot log(Rf) \qquad (15)$$

Characteristics of the model presented in eq. (15) are as follows:

-Estimation characteristics: r^2 = 0.4312; r^2_{adj} = 0.3680; RMSE = 0.4807; F(p) = 6.82 (0.0282); $t_{coefficients}$ (p): Int. = 4.71 (0.0011), $X_{log(Rf)}$ = -2.61 (0.0282); n = 11

-Measures of residual errors: MAE = 0.3596, MAPE = 0.3359, SEP = 0.4807, REP(%) = 244.28

-Prediction performance in leave-one-out analysis: r^2_{loo} = 0.2387, s_{loo} = 0.5664, F_{loo} (p) = 0.1560

Theoretical calculated on the investigated sample of the compounds proved not to be able to explain the cytotoxicity of investigated compounds (correlation coefficients < 0.25 and/or p-values associated to F-statistic > 0.05). The QPCR model presented in eq. (15) was low estimation abilities being able to explain just 43.12% from the variability in cytotoxicity as linearity with $\log(R_f)$.

The absence of a significant linear model able to explain cytotoxicity of the investigated hidrazinoselenazole compounds could have at least two explanations. First, the method used to calculate theoretical descriptors based on the structure of the compound were not able to extract those structural characteristic that could explain the activity of interest. In this case, the use of other methods or approaches (such as Molecular Descriptors Family [26,27], correlation weights [28], etc.) could lead to significant models and this is under analysis in our laboratory. Second, cytotoxicity could be an activity that is not linearly related neither to chromatographic properties nor with the structure of investigated compounds. Solution to the absence of linearity is given by investigation of cytotoxicity using non-linear methods such as exponential, power, polinomial function, or other approaches such as GLM (general linear model).

The assessment of the QPCR model presented by eq. (15) could also be done by investigation of the power of the model to identify those compounds with desired toxicity. Different values of IC_{50} are reported as desired when HepG2 cell line is investigated [29,30]. The desire cytotoxicity is observed on compounds with IC_{50} values smaller than 4 μg/mL [31], respectively around 10μM [32]. Two compounds presented in table 1 accomplish simultane-ously this criterion and could be considered with desired cytotoxicity, S12 and S14 (for details related with IC_{50} expressed in μM see [15]). The model in eq. (15) proved ability in proper identification of compounds without desired cytotoxicity, but was not able to classify correctly those compounds with desired cytotoxicity.

In another paper newheterocyclic compounds with potential cytotoxic activity were studied [32].

Conclusions

Two out of three chromatographic properties proved linearly related with structural characteristics of investigated hidrazinoselenazole compounds. No valid model was obtained for the specific surface area of the solvent while models with good estimation and prediction performances were obtained for the logarithm of retention factor and chromatographic hydrophobicity index, both of them using Dragon descriptors and hidrazinoselenazole compounds optimized in water.

Our results showed that the retention factor can be expressed based on topologic measurements and suggests that inertia is a determinant factor of dynamics of separation. The presence of electronegativity (partial charge, first ionization energy) in the expression of chromatographic hydrophobicity index indicates that the charge of ions dissolved in the mobile phase and molecular polarizability are determinants of the dynamics of separation.

The investigation of cytotoxicity as function of chromatographic parameters and/or theoretical descriptors proved no significant linearity, suggesting that cytotoxicity is not an activity linearly neither related with chromatographic parameters nor with investigated structural descriptors.

References

1.HAMMETT, L.P., J. Am. Chem. Soc., 59, no. 1, 1937, p. 96.
2.CRUM-BROWN, A., FRASER, T.R. Phil. Trans. R. Soc. London, 25, 1868, p. 151.
3.RICHARDSON, B.J., Med. Times Gaz., 2, 1868, p. 703.
4.HANSCH, C., Acc. Chem. Res., 2, 1969, p. 232.
5.NANTASENAMAT, C., ISARANKURA-NA-AYUDHYA, C., NAENNA, T., PRACHAYASITTIKUL, V., Excli. J., 8, 2009, p. 74.
6.TROPSHA, A., GRAMATICA, P., GOMBAR, V.K., QSAR & Comb. Sci., 22, 2003, p. 69.
7.gRAMATICA, P. QSAR & Comb. Sci., 26, 2007, p. 694.
8.CHIRICO, N., GRAMATICA, P., J. Chem. Inf. Model, 52, no. 8, 2012, p. 2044.
9.PAPP, L.V., LU, J., HOLMGREN, A., KHANNA, K.K. Antioxid. Redox. Signal., 9, no. 7, 2007, p. 775.
10.RAYMAN, M.P., Lancet, 356, no. 9225, 2000, p. 233.
11.HASANUZZAMAN, M., HOSSAIN, M.A., FUJITA, M., J. Plant. Sci., 5, no. 4, 2010, p. 354.
12.YAMASHITA, M., YAMASHITA, Y., SUZUKI, T., KANI, Y., MIZUSAWA, N., IMAMURA, S., TAKEMOTO, K., HARA, T., HOSSAIN, T.A., YABU, T., TOUHATA, K., Mar. Biotechnol., 15, no. 5, 2013, p. 559.
13.DUNTAS, L.H., Horm. Metab. Res., 41, no. 6, 2009, p. 443.
14.SRIVASTAVA, P.C., ROBINS, R.K. J. Med. Chem., 26, no. 3, 1983, p. 445.
15.ZAHARIA, V., IGNAT, A., NGAMENI, B., KUETE, V., MOUNGANG, M.L., FOKUNANG, C.N., VASILESCU, M., PALIBRODA, N., CRISTEA, C., SILAGHI-DUMITRESCU, L., NGADJUI, B.T., Med. Chem. Res., 22, no. 12, 2013, p. 5670.
16.DEIDDA, D., LAMPIS, G., MAULLU, C., POMPEI, R., ISAIA, F., LIPPOLIS, V., VERANI, G., Pharmacol. Res., 36, no. 3, 1997, p. 193.
17.GROZAV, A., GAINA, L., KUETE, V., SILAGHI-DUMITRESCU, L., EFFERTH, T., ZAHARIA, V., Molecules, 18, no. 4, 2013, p. 4679.
18.COZMA, A., VLASE, L., IGNAT, A., ZAHARIA, V., GOCAN, S., MARUTOIU, C., FODOR, A., Rev. Chim.(Bucharest), **63**, no. 7, 2012, p. 651.
19.***IUPAC. Compendium of Chemical Terminology, 2nd ed. (the Gold Book). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by Nic M, Jirat J, Kosata B; updates compiled by Jenkins A. ISBN 0-9678550-9-8. doi:10.1351/goldbook. (Accessed June 2014)
20.TETKO, I.V., GASTEIGER, J., TODESCHINI, R., MAURI, A., LIVINGSTONE, D., ERTL, P., PALYULIN, V.A., RADCHENKO, E.V., ZEFIROV, N.S., MAKARENKO, A.S., TANCHUK, V.Y., PROKOPENKO, V.V., J. Comput. Aid. Mol. Des., 19, 2005, p. 453.
21.BOLBOACA, S.D., JANTSCHI, L., Bulletin UASVM Horticulture, 66, no. 2, 2009, p. 698.
22.FISHER, R.A., Am. Statist., 2, no. 5, 1948, p. 30.
23.FISHER, R.A., J. R. Stat. Soc., 85, no. 4, 1922, p. 597.
24.STEIGER, J.H. Tests for comparing elements of a correlation matrix. Psychol. Bull., 1980, 87, 245-251. doi:10.1037/0033-2909.87.2.245
25.BOLBOACA, S.D., ROSCA, D.D., JANTSCHI, L., MATCH-Commun. Math. Co., 71, no. 1, 2014, p. 149.
26.BOLBOACÃ, S.D., JANTSCHI, L., DIUDEA, M.V., Curr. Comput.-Aid. Drug, 9, no. 2, 2013, p. 195.
27.HARSA, A.M., HARSA, T.E., BOLBOACA, S.D., DIUDEA, M.V., Curr. Comput.-Aid. Drug, 10, 2014, p. 115.
28.MACHANA, S., WEERAPREEYAKUL, N., BARUSRUX, S., NONPUNYA, A., SRIPANIDKULCHAI, B., THITIMETHAROCH, T., Chin. Med., 6, no. 1, 2011, 39.
29.ELDEHNA, W.M., ALTOUKHY, A., MAHROUS, H., ABDEL-AZIZ, H.A., Eur. J. Med. Chem., 90, 2015, p. 684.
30.BOIK, J. Natural compounds in cancer therapy. Oregon MedicalPress, Princeton, 2001, pp. 24-25.
31.BRAHEMI, G., KONA, F.R., FIASELLA, A., BUAC, D., SOUKUPOV, J., BRANCALE, A., BURGER, A.M., WESTWELL, A.D., J. Med. Chem., 53, 2010, p. 2757.
32. SOCEA, I., BARBUCEANU, S.F., SOCEA, B., DRAGHICI, C. APOSTOL, T.V., PAHONTU, E.M., OLARU, O.T., Rev. Chim. (Bucharest), **68**, no. 11, 2017, p. 2503

Manuscript received: 15.11.2017