STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, LIV, 3, 2009

STUDY OF THE CHROMATOGRAPHIC RETENTION OF SOME NEW ORGANOSELENIUM AND ORGANOTELLURIUM COMPOUNDS CONTAINING INTRAMOLECULAR INTERACTIONS BY HPTLC

SILVIA LENUȚA DUNCA[a], MONICA KULCSAR[a], ANCA SILVESTRU[a], CRISTIAN SILVESTRU[a], COSTEL SÂRBU[a]

ABSTRACT. The chromatographic behaviour of some new metal complexes of organoselenium and organotellurium compounds containing intramolecular interactions, were investigated by means of different HPTLC systems with polyamide, cellulose, normal and modified silica gel thin-layers and various organic solvents of relatively high polarity. Recommendable phase system for the separation of metal-complexes is a combination of fluorescent polyamide thin-layer with a methanol-water (8:2 v/v) mixture when well-defined compact spots come out and migrate. Dark zones appeared on fluorescent green background under UV lamp (λ = 254 nm). R_F values were determined by using the one-dimensional ascending technique and modelling by using different molecular descriptors calculated by using efficient software. It has been concluded that a successful analysis will be executable for the compounds studied with the possibility of modelling the chromatographic retention.

Keywords*: organoselenium and organotellurium compounds, lipophilicity, TLC, QSRR, MLR, PCA, PCR*

INTRODUCTION

Quantitative structure-activity relationships (QSAR) describe how the molecular structure, in terms of descriptors – lipophilic, electronic and steric – affects the biological activity of a compound [1-4]. Similarly, quantitative structure-retention relationships (QSRR) relate these descriptors to chromatographic retention. Finally, the quantitative retention-activity relationships (QRAR) imply that conclusions concerning biological activity can be based on chromatographic experiments [5-11]. In this regard of QRAR it is considering that the same basic intermolecular interactions determine the behaviour of chemical compounds in both biological and chromatographic environments. As a consequence, the

[a] *Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, Arany Janos 11, 400028 Cluj-Napoca, Romania; E-mail: costelsrb@yahoo.co.uk*

chromatographic approach has been quite successful, for example, in duplicating Log P data derived by traditional “shake-flask” technique or other procedures. The relationships themselves are usually based on correlation analysis.

Another form of computational analysis used for the correlation of chemical or biological activity and chromatographic retention with different molecular descriptors are Multiple Linear Regression (MLR) [11-13], Principal Component Analysis (PCA) [14-16], Partial Least Squares (PLS) [17-19], or Artificial Neural Networks (ANN) [20-22]. In the case of PCA and PLS, for example, starting from a multidimensional space described by different variables, a quantitative model is derived that transforms the axes of the hypersystem. The first principal component (PC1) defines as much of the variation in the data as possible. The second principal component (PC2) describes the maximum amount of residual variation after the first PC has been taken into consideration, and so on. By using only a limited number of PCs, the dimensionality of the data space is reduced, thereby simplifying further analysis.

In this paper we discuss and apply three multivariate regression methods to develop comparative studies and to provide a QSAR-QSRR model for the characterization and classification of some new organo-selenium and organotellurium compounds with potential applications for asymmetric synthesis in organic and organometallic chemistry, catalytic antioxidant activity, enzyme mimics and chemotherapeutic agents.

PRINCIPAL COMPONENT ANALYSIS

Principal components analysis (PCA) is also known as *eigenvector* analysis, *eigenvector* decomposition or Karhunen-Loéve expansion. Many problems from chemistry and other scientific fields are strongly related to PCA. The main purpose of PCA is to represent in an economic way the location of the samples in a reduced coordinate system where instead of *m*-axes (corresponding to *m* characteristics) only *p* ($p < m$) can usually be used to describe the data set with maximum possible information.

Principal component analysis practically transforms the original data matrix ($\mathbf{X}_{nxm}$) into a product of two matrices, one of which contains the information about the objects ($\mathbf{S}_{nxm}$) and the other about the variables ($\mathbf{V}_{mxm}$). The **S** matrix contains the scores of the *n* objects on *m* principal components (the scores are the projection of the objects on principal components). The **V** matrix is a square matrix and contains the loadings of the original variables on the principal components (the loadings are the weights of the original variables in each principal component).

Moreover, it may well turn out that usually two or three principal components provide a good summary of all the original variables. Loading and respectively score plots are very useful as a display tool for examining the relationships between characteristics and between compounds, looking for trends, grouping or outliers.

MULTIPLE LINEAR REGRESSION

Multiple linear regression (MLR) is an extension of simple linear regression consisting of two or more independent variables (*e.g.* chemical descriptors or properties) and a numeric dependent variable (*e.g.* chromatographic retention index). MLR attempts to model the relationship between the independent variables and a response variable (R) by fitting a linear equation to observed data in the following equation:

$$R = a_o + \sum_{i=1}^{k} a_i x_i \ , \qquad (1)$$

where a_o, a_i are the estimated regression parameters.

PRINCIPAL COMPONENT REGRESSION

Principal component regression (PCR) is a two-step multivariate calibration method: in the first step, a principal component analysis of the data matrix **X** is performed. The measured or calculated variables (*e.g.* descriptors) are converted into new ones (scores on latent variables). This is followed by a multiple linear regression step, MLR, between the scores obtained in the PCA step and the characteristic R to be modelled.

RESULTS AND DISCUSSION

By reducing the number of features from 9 original descriptors (including retention indices R_f and R_m) to three principal components (latent variables), the information preserved is enough to permit a primary examination of the similarities and differences between descriptors and organoselenium and organotellurium compounds. The contribution of the first component represents 42.81% of the total variance and a two components model accounts for 68.52% of the total variance. The first three components reproduce approximately 81% of the total variance and the first six even 99.31%, and the *eigenvalues* become negligible after the seventh component.

All the statements above are well supported by the 2D- and 3D-representations of the loadings (Figures 1 and 2). The projection of the 3D-representation gives a more complete pattern: it is clear, for example, that the majority of the descriptors considered in this study form two close clusters: the first one includes *M*, *MW*, *V* and *MR*, the second one encompasses *CLogP*, *H* and *GS*; R_f/(R_m) and *GS* appear more or less as outliers.

The scatter plot of scores onto the plane defined by PC1 and PC2 (Figure 3) and in the space described by PC1, PC2 and PC3 (Figure 4) shows interesting results. Two clusters appear to be well defined and in a good agreement to the structure of compounds: one of them corresponds to the

compounds **3**, **4**, **5**, **6**, **8**, **12**, **13**, **14**, **17**, **18** and **19** (the largest molecules in the series) in the above right part of the graph, the second include the group of fluorine derivatives (**9**, **10**, **15**, **16**), with the exception of compound **1**, **2**, **7** and **11**, is located in the middle-bottom of the graph.

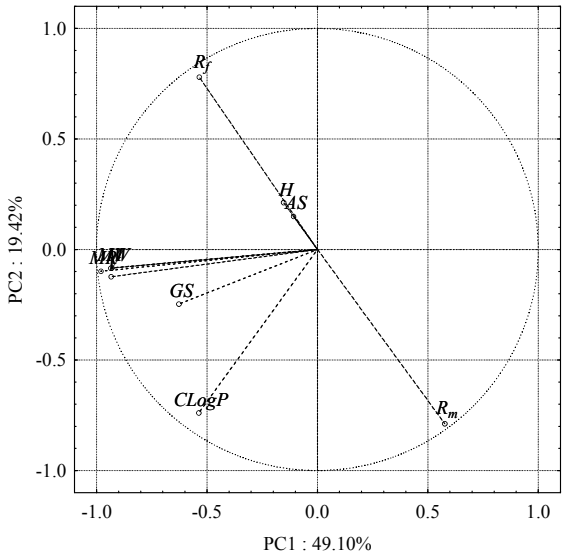


Figure 1. PC1 and PC2 loading plot of the autoscaled data in Table 2

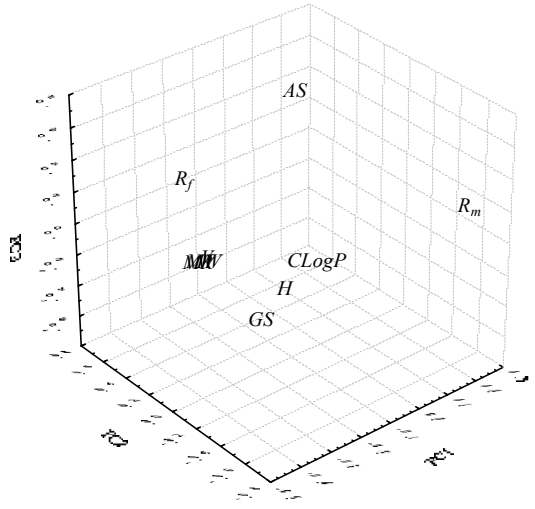


Figure 2. PC1, PC2 and PC3 loading plot of the autoscaled data in Table 2

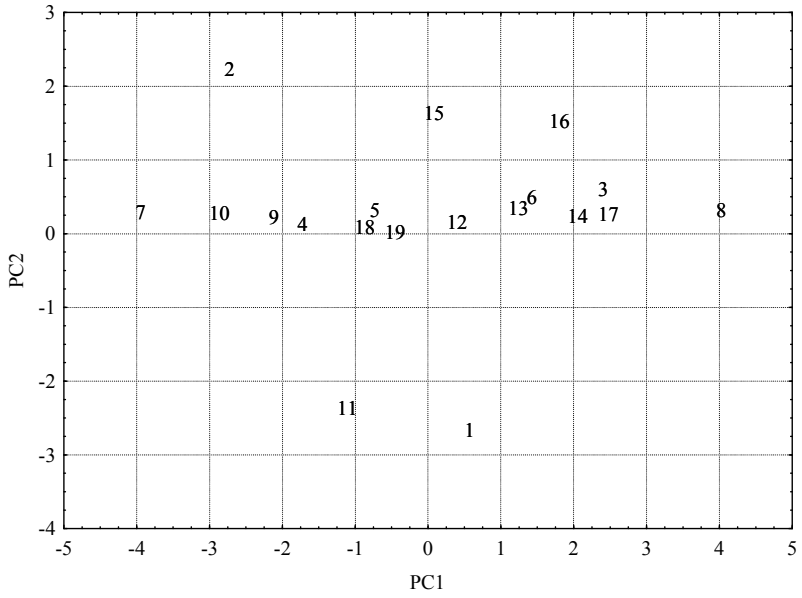


Figure 3. PC1and PC2 score plot of the autoscaled data in Table 2

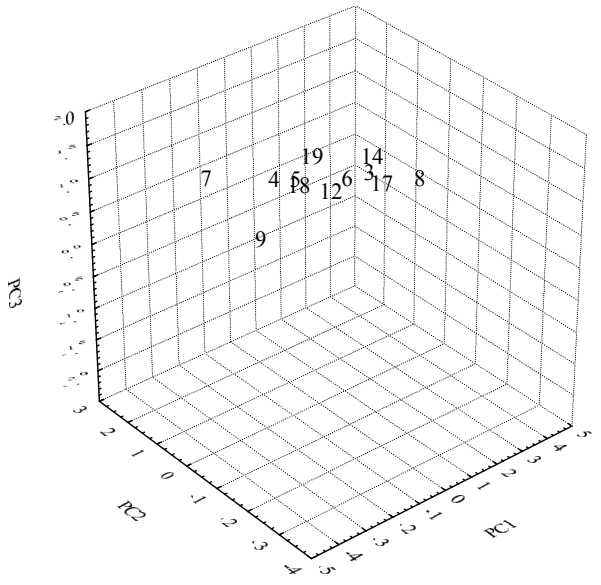


Figure 4. PC1, PC2 and PC3 score plot of the autoscaled data in Table 2

In order to describe the relationship between the chromatographic retention indices of the tested compounds (R_{Mo} and b, intercept and slope values, in equation 7) and the calculated structural descriptors, a multivariate regression analysis was performed. By forward stepwise multiple regression analysis, the following high-quality regression equations were obtained:

$$R_{Mo} = 0.119 - 0.404MR + 0.264CLogP + 0.006M \quad (2)$$
$$(r = 0.9153,\ n = 19,\ F = 26,\ p < 0.0000,\ s = 0.225)$$

$$R_f = 0.500 + 0.173RM - 0.105CLogP - 0.003M \quad (3)$$
$$(r = 0.8894,\ n = 19,\ F = 19,\ p < 0.0000,\ s = 0.105)$$

where n is the number of compounds, r the correlation coefficient, F the F-test value, p is the significance level of the all equation and s is standard error of estimates. The F and p values of equation (2) and (3) show that the multiple regression equations are very significant having high correlation coefficient and relatively small s values.

The results suggest also that the molecular refractivity and partition coefficient seem to be dominant in the retention mechanism and, as a consequence, control the lipophilicity of the investigated compounds.

For the PCR method, the original 7 descriptors were used for the selection of the optimum number of factors (principal components) by using also the statistics discussed above. The obtained multiple regression equations are also highly significant:

$$R_m = -0.247 - 0.774PC6 + 0.091PC1 + 0.184PC3 + 0.226PC4 + 0.857PC7 + 40.179PC8 - 0.111PC5 \quad (4)$$
$$(r = 0.9360,\ n = 19,\ F = 11,\ p < 0.0003,\ s = 0.229)$$

$$R_f = 0.599 + 0.353PC6 - 0.103PC4 - 0.033PC1 - 0.337PC7 - 0.054PC3 - 23.392PC8 + 0.038PC5 \quad (5)$$
$$(r = 0.9557,\ n = 19,\ F = 17,\ p < 0.0000,\ s = 0.079)$$

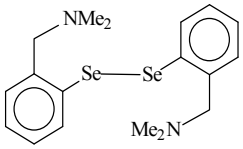
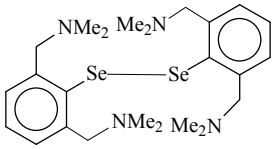
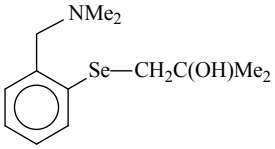
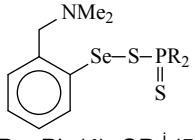
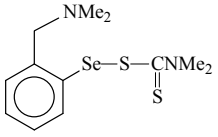
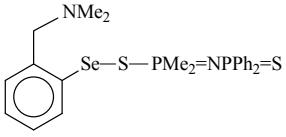
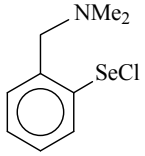
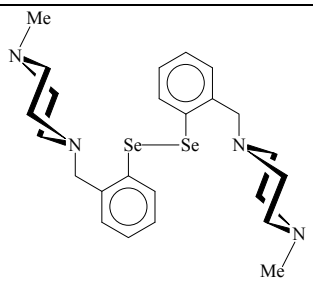
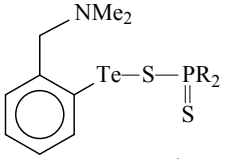
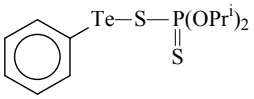
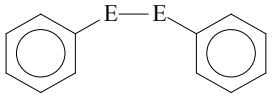
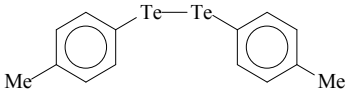
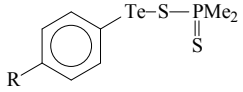
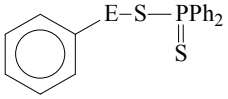
CONCLUSIONS

Correlation obtained between chromatographic retention indices and structure descriptors for organoselenium and organotellurium compounds are high significant and might be used to predict the retention behaviour and, as a consequence, the lipophilicity of other members of the series. By comparing the multivariate regression methods used in this study, the forward stepwise MLR appeared to be the most effective in predicting retention indices for the investigated compounds. The molecular refractivity and the partition coefficient seem to be dominant in the retention mechanism and hence these descriptors control the lipophilicity.

EXPERIMENTAL SECTION

The chromatographic behaviour of the compounds were investigated by means of different HPTLC systems with polyamide, cellulose, normal and modified silica gel thin-layers and various organic solvents of relatively high polarity. Recommendable phase system for the separation of metal complexes is a combination of fluorescent polyamide thin-layer with a methanol-water (8:2 v/v) mixture when well-defined compact spots come out and migrate. Dark zones appeared on fluorescent green background under UV lamp (λ = 254 nm). Glass HPTLC plates (20 x 20 cm) were obtained from Macherey-Nagel (Düren, Germany) and methanol for chromatography was supplied from Reactivul (Bucharest, Romania). Solutions in chloroform of each compound (Table 1)

Table 1. Chemical structure of the investigated organometallic compounds

 **1**	 **2**	 **3**
 R = Ph (**4**), OPr^i (**5**)	 **6**	 **7**
 **8**	 **9**	 R = Ph (**10**), OPr^i (**11**)
 **12**	 E = Te (**13**), Se (**14**)	 **15**
 R = Me (**16**), H (**17**)	 E = Te (**18**), Se (**19**)	

were prepared at a concentration of approximative 1 mg mL^{-1}. Chromatograms were developed by ascending technique at room temperature (~20 ^{0}C); the developing distance being 10 cm. After being developed, the dried plates were examined under UV lamp (λ = 254 nm). The R_M values of each compound were obtained by using the following well-known equation

$$R_M = \log (1/R_f - 1) \quad (6)$$

The R_M values are measured at several compositions of binary mobile phase systems and linearly extrapolated (interpolated) on the basis of the relationship between the R_M and the mobile phase organic modifier as was described by the TLC adapted Soczewiński-Wachtmeister equation:

$$R_M = R_{M0} + bC \quad (7)$$

where R_{M0} indicates the extrapolated value to the pure water as mobile phase and it is the HPTLC descriptor most frequently used into QSAR analysis. *b* is frequently associated to the specific surface area of the stationary phase, while *C* represents the volume fraction of the organic modifier in the mobile phase. The specific surface area is considered an alternative descriptor of lipophilicity.

Table 2. The descriptors and retention indices computed for the organo-selenium and organotellurium compounds investigated in this paper

Nr.	*M*	*H*	*CLogP*	*MR*	*AS*	*GS*	*V*	R_f	R_m
1	428.03	0.97	4.71	10.93	69046	160.38	584.13	0.726	-0.423
2	542.14	1.44	4.38	14.45	3680	377.84	650.25	0.973	-1.562
3	287.08	0.97	2.36	7.563	155	324.16	486.52	0.873	-0.839
4	463.01	0.97	8	12.95	21	353.86	603.18	0.455	0.079
5	427.03	0.97	6.38	11.02	3281	362.82	573.15	0.647	-0.264
6	334.01	0.97	3.51	9.013	187	330.52	506.56	0.652	-0.273
7	538.03	0.97	9.51	15.19	186	415.57	712.31	0.729	-0.430
8	248.98	0.97	1.76	6.046	164	261.12	381.15	0.737	-0.447
9	538.11	0.97	5.54	14.1	129	352.33	588.69	0.799	-0.598
10	638.09	0.97	3.68	14.84	685	353.81	591.81	0.925	-1.089
11	477.02	0.97	5.45	11.39	115180	353.89	556.98	0.650	-0.269
12	419.96	0.97	5.61	9.628	2918	320.02	487.19	0.554	-0.094
13	413.89	0.97	3.19	8.154	123	307.05	471.81	0.390	0.194
14	313.91	0.97	5.04	7.412	125	299.23	462.66	0.436	0.112
15	441.92	1.28	4.18	9.081	164	335.87	526.59	0.316	0.335
16	345.93	1.28	3.94	7.93	224	298.46	447.89	0.334	0.300
17	331.91	0.97	3.44	7.466	119	284.98	414.53	0.403	0.171
18	455.94	0.97	7.24	11.56	4	331.76	544.93	0.392	0.191
19	405.95	0.97	8.16	11.19	1	326.12	541.16	0.385	0.203

DESCRIPTION OF ORGANOSELENIUM AND ORGANOTELLURIUM COMPOUNDS

The investigated organoselenium and organotellurium compounds were synthesized by procedures described earlier [24-26]. The molecular structure of the organoselenium and organotellurium compounds studied in this paper is depicted in Table 1.

In order to define the character of the compound structure, the following descriptors available in the ChemDraw Pro program were taken into consideration and used as independent variables: exact mass (*M*), partition coefficient (*CLogP*), molar refractivity (*MR*), Henry's law constant (*H*), surface area (*AS*), surface area (Grind)(*GS*), volume (*V*), molecular polarizability (*MP*). The obtained values are presented in Table 2. We have been computing only so few descriptors because the conventional software does not recognize Se and Te. Also, in the case of polarizability, the software does not have the ability to calculate the dithiophosphinates ligands.

ACKNOWLEDGEMENTS

This work was supported by the Romanian Ministery of Education and Research (PNCDI-CERES Program, grant: 4-62/2004).

REFERENCES

1. A. Leo, C. Hansch, D. Elkins, *Chem. Rev*., **1971**, *71*, 525.
2. R. F. Rekker, R. Mannhold, *Calculation of Drug Lipophilicity*, VCH, Weinheim,**1992**.
3. M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley & Sons, New York, **2000**.
4. A. Chiriac, D. Ciubotariu, Z. Simon, *QSAR-Quantitative Relationships*, Mirton Publishing House, Timişoara, **1995**.
5. R. Kaliszan, *Anal. Chem*., **1992**, *64*, 619A.
6. A. Pyka, M. Miszczyk, *Chromatographia*, **2005**, *61*, 37.
7. R. Kaliszan, *Trends Anal. Chem*., **1999**, *18*, 401.
8. Q. S. Wang, L. Zhang, *J. Liq. Chrom. & Rel. Technol*., **1999**, *22*, 1.
9. M. C. Garcia Alvarez-Coque, J. R. Torres Lapasio, *Trends Anal. Chem*., **1999**, *18*, 533.
10. T. L. J. Djaković-Sekulić, C. Sârbu, N. U. Perišić-Janjić, *J. Planar Chromatogr*., **2005**, *18*, 432.
11. J. Dai, L. Jin, S. Yao, L. Wang, *Chemosphere*, **2001**, *42*, 899.

12. A. Detroyer, Y. V. Heyden, S. Carda-Broch, M. C. Garcia-Alvarez-Coque, D. L. Massart, *J. Chromatogr. A*, **2001**, *912*, 211.
13. C. Sârbu, D. Casoni, M. Dărăbanțu, C. Maiereanu, J*. Pharm. Biomed. Anal*., **2004**, 35, 213.
14. C. Sârbu, S. Todor, *J. Chromatogr. A*, **1998**, 822, 263.
15. C. Sârbu, K. Kuhajda, S. Kevresan, *J. Chromatogr. A*, **2001**, 917, 361.
16. A. Detroyer, V. F. Schoonjans, F. Questier, Y. V. Heyden, A. P. Borosy, Q. Guo, D. L. Massart, *J. Chromatogr. A*, **2000**, 897, 23.
17. H. Martens, T. Naes, *Multivariate Calibration*, Wiley, Chichester, **1991**.
18. Y. Zhou, L. Xu, Y. Wu, B. Liu, *Chemometr. Intell. Lab. Syst*., **1999**, 45, 95.
19. T. Li, H. Mei, P. Cong, *Chemometr. Intell. Lab. Syst*., **1999**, 45, 177.
20. R. Zhang, A. Yan, M. Liu, H. Liu, Z. Hu, *Chemometr. Intell. Lab. Syst*., **1999**, 45, 113.
21. R. H. Zhao, B. F. Yue, J. Y. Ni, H. F. Zhou, Y. K. Zhang, *Chemometr. Intell. Lab. Syst*., **1999**, 45, 163.
22. Y. Chen, D. Chen, C. He, S. Hu, *Chemometr. Intell. Lab. Syst*., **1999**, 45, 267.
23. C. Sârbu, H. F. Pop, *Fuzzy Soft-Computing Methods and Their Applications in Chemistry*. In *Reviews in Computational Chemistry*. K. B. Lipkowitz, D. B. Boyd, T. R. Cundari, Eds., Wiley-VCH, **2004**, 249-332.
24. M. Kulcsar, A. Silvestru, C. Silvestru, J. E. Drake, C. L. B. Macdonald, M. B. Hursthouse, M. E. Light, *J. Organomet. Chem*., **2005,** 690, 3217.
25. J. E. Drake, M. B. Hursthouse, M. Kulcsar, M. E. Light, A. Silvestru, *J. Organomet. Chem*., **2001**, 623, 153.
26. C. Deleanu, J. E. Drake, M. B. Hursthouse, M. Kulcsar, M. E. Light, A. Silvestru, *Appl. Organometal. Chem*., **2002**, 16, 727.