Quantitative Structure–Capillary Column Gas Chromatographic Retention Time Relationships for Natural Sterols (trimethylsilyl ethers) from Olive Oil

Rocío Acuña-Cueva[a], Francisco Hueso-Ureña[a], Nuria A. Illán Cabeza[a], Sonia B. Jiménez-Pulido[a], Miguel N. Moreno-Carretero[a,*], and José M. Martínez Martos[b]

[a]Departamento de Química Inorgánica y Orgánica, and [b]Area de Fisiología, Campus Universitario "Las Lagunillas," Universidad de Jaén, 23071 Jaén, Spain

ABSTRACT: A study of the relationship between the capillary column gas chromatographic retention times on two different columns (SE-54 and SE-52) for a group of trimethylsilyl ethers derived from 16 natural sterols found in olive oil and an initial set of 60 molecular descriptors was made. Three kinds of molecular descriptors were used: conventional, topological, and quantum-chemical parameters. By using multivariable regression, two empirical functions for each column were obtained, which were selected on the basis of their respective statistical parameters. The first model relates the retention index with the quantum-chemical descriptors and the second with both topological and conventional descriptors. In all cases, the correlation coefficients of the empirical functions were higher than 0.9880, and the mean relative errors range between 2.88 and 3.24%. In any case, both models could be used, although the second model could be more useful and preferable because it has higher R^2 values and a lower standard error percentage, giving slightly more exact results.

Paper no. J9323 in *JAOCS 77*, 627–630 (June 2000).

KEY WORDS: Olive oil, QSAR, QSRR, sterols.

Sterols are important nonglyceridic constituents that appear in many plants and animal species. It is widely believed that virgin olive oils present a higher nutritional and biological value compared to other vegetable oils. As they can be consumed without any previous refinement, it is important to know their purity, the sterolic fraction being characteristic for each oil. For this reason, the determination of sterols has been used to check the authenticity and purity of virgin olive oil (1).

Four types of sterols may be found in olive oil: 4α-desmethylsterols, 4α-methylsterols, 4,4-dimethylsterols, and triterpene dialcohols (2). The main sterols in olive oil are sitosterol, Δ^5-avenasterol, and campesterol. Also, stigmasterol, cholesterol, 24-methylene-cholesterol, Δ^7-campesterol, $\Delta^{5,23}$-stigmastadienol, chlerosterol, sitostanol, $\Delta^{5,24}$-stigmastadienol, Δ^7-stigmasterol, and Δ^7-avenasterol may be found in small quantities. In olive oils, these compounds can be present in either the free form or esterified with fatty acids. The standard method proposed by the European Union legislation (3) for the determination of sterols, which also is the most frequently used, is based on the isolation of the unsaponifiable fraction from the oil and later separation of the sterol fraction by basic silica gel plate chromatography. The sterols recovered from silica gel are transformed into trimethylsilyl ethers and are analyzed by capillary column gas chromatography (GC).

The prediction of the retention times (hereafter denoted as RT) using quantitative structure–retention relationships (QSRR) (4) is an important and easy method to corroborate the structural identification of single components in complex mixtures. In a previous paper, we successfully used this method to calculate the RT of natural phenols in olive oil (5). The present work was focused to make an analogous study with the natural sterols found in olive oil as target, which may help to elucidate the structure of unknown sterolic fractions.

EXPERIMENTAL PROCEDURES

Experimental capillary column GC RT, and the corresponding experimental details, for the trimethylsilyl ethers derived from 16 natural sterols used in this work were taken from the literature (3). In Table 1, the studied sterols together with the experimental and calculated RT values for their trimethylsilyl ethers in two different chromatographic columns (SE 54 and SE 52) are given.

The procedure used in the present study comprised two fundamental stages: (i) molecular descriptors generation and (ii) statistical analysis.

Molecular descriptors generation. In the present work, three types of descriptors[1] were used: conventional, topological, and quantum-chemical descriptors. Both conventional and topological descriptors were calculated by means of the DESCRIPTOR program developed by us for a PC system using GW-BASIC language. Conventional descriptors are basically related to the number and types of atoms and bonds in each molecule. Topological descriptors include valence and nonvalence molecular connectivity indices calculated from the formula of suppressed hydrogens of the molecule, accord-

*To whom correspondence should be addressed.
E-mail: mmoreno@ujaen.es.

[1]Tables with the values of the used descriptors are available from the correspondence author.

TABLE 1
Values of Experimental (RT1 and RT2, min) and Calculated Retention Times (RT) from the Two Proposed Models

	RT1 (SE 54 column)					RT (SE 52 column)				
Compound (trimethylsilyl ether derivative)	Experim. value	Calcd. model 1	Res.[a] (%)	Calcd. model 2	Res. (%)	Experim. value	Calcd. model 1	Res. (%)	Calcd. model 2	Res. (%)
Cholesterol, Δ^5-cholesten-3β-ol	0.67	0.67	0.00	0.65	3.07	0.63	0.64	–1.56	0.61	3.27
Cholestanol, 5α-cholestan-3β-ol	0.68	0.67	1.49	0.70	–2.85	0.64	0.63	1.58	0.65	–1.53
Brassicasterol, (24*S*)-24-methyl-$\Delta^{5,22}$-cholesten-3β-ol	0.73	0.77	–5.19	0.74	–1.35	0.71	0.74	–4.05	0.72	–1.38
24-Methylene-cholesterol, 24-methylene $\Delta^{5,24}$-cholesten-3β-ol	0.82	0.78	5.12	0.78	5.12	0.80	0.76	5.26	0.76	5.26
Campesterol, (24*R*)-24-methyl-Δ^5-cholesten-3β-ol	0.83	0.82	1.21	0.83	0.00	0.81	0.80	1.25	0.81	0.00
Campestanol, (24*R*)-24-methyl-cholestan-3β-ol	0.85	0.83	2.40	0.86	–1.16	0.82	0.80	2.50	0.83	–1.20
Stigmasterol, (24*R*)-24-methyl-$\Delta^{5,22}$-cholestadien-3β-ol	0.88	0.90	–2.22	0.90	–2.22	0.87	0.90	–3.33	0.90	–3.33
Δ^7-Campesterol, (24*R*)-24-methyl-Δ^7-cholesten-3β-ol	0.93	0.92	1.08	0.92	1.08	0.92	0.90	2.22	0.90	2.22
$\Delta^{5,23}$-Stigmastadienol, (24*R,S*)-24-ethyl-$\Delta^{5,23}$-cholestadien-3β-ol	0.95	0.95	0.00	0.93	2.15	0.95	0.94	1.06	0.93	2.15
Chlerosterol, (24*S*)-24-ethyl-$\Delta^{5,25}$-cholestadien-3β-ol	0.96	0.93	3.22	0.97	–1.03	0.96	0.92	4.34	0.97	–1.03
Sitosterol, (24*R*)-24-ethyl-Δ^5-cholesten-3β-ol	1.00	1.02	–1.96	0.98	2.04	1.00	1.02	–1.96	0.97	3.09
Sitostanol, (24*R*)-24-ethyl-cholestan-3β-ol	1.02	1.02	0.00	1.00	2.00	1.02	1.02	0.00	1.00	2.00
Δ^5-Avenasterol, (24*Z*)-24-ethylidene-5-cholesten-3β-ol	1.03	1.04	–0.96	1.06	–2.83	1.03	1.04	–0.96	1.06	–2.83
$\Delta^{5,24}$-Stigmastadienol, (24*R,S*)-24-ethyl-$\Delta^{5,24}$-cholestadien-3β-ol	1.08	1.03	4.85	1.05	2.85	1.08	1.03	4.85	1.04	3.84
Δ^7-Stigmastenol, (24*R,S*)-24-ethyl-Δ^7-cholesten-3β-ol	1.12	1.13	–0.88	1.14	–1.75	1.12	1.13	–0.88	1.14	–1.75
Δ^7-Avenasterol, (24*Z*)-24-ethylidene-Δ^7-cholesten-3β-ol	1.16	1.16	0.00	1.14	1.75	1.16	1.16	0.00	1.14	1.75

[a]Residual (%) = $\{[(RT_{exp} - RT_{calc}) \times 100]/RT_{calc}\}$.

ing to the method proposed by Randic (6) and Kier and Hall (7), and encode information about the size and the degree of branching of a molecule. Quantum-chemical descriptors include information about binding and formation energies, energy levels in the molecule, vibrational energies, and inertia moments. To obtain quantum-chemical descriptors, the semiempirical method AM1 as implemented in HYPERCHEM 4.0 software package was used. All calculations were carried out at restricted Hartree Fock level for the singlet lowest energy state with no configuration interaction. The molecular structures were generated with the molecular builder inside HYPERCHEM and optimized following the Polak-Ribiere algorithm until the root mean square gradient was 0.1 $Kcal \cdot Å^{-1} \cdot mol^{-1}$. In order to check the goodness of the resulting structures, the infrared spectrum was also calculated, the nonappearance of negative frequencies being assumed as unequivocal evidence that the generated structure represents a potential energy minimum and not a saddle point.

TABLE 2
Definition of the Total Set of Descriptors Used

Descriptor code	Definition
Quantum-chemical descriptors:	
HOMO	HOMO energy (eV)
ENLACE	Binding energy ($kcal \cdot mol^{-1}$)
MAXPOB	Highest atomic orbital electron population
IX	Inertia moment in the *x* axis ($g \cdot cm^2 \cdot 10^{-40}$)
IY	Inertia moment in the *y* axis ($g \cdot cm^2 \cdot 10^{-40}$)
IZ	Inertia moment in the *z* axis ($g \cdot cm^2 \cdot 10^{-40}$)
ACM	Rotational constant in the *x* axis (cm^{-1})
NEL	Number of electrons
ET	Total energy ($kcal \cdot mol^{-1}$)
POINTCHG	Total dipole moment module (Debyes)
CORE	Core-core interaction energy ($kcal \cdot mol^{-1}$)
ELECTRON	Electronic energy ($kcal \cdot mol^{-1}$)
Conventional descriptors:	
NH	Number of hydrogen atoms
C1_PESO	Relative weight of C atoms
CH_ENL	Ratio between number of C-H bonds and total number of bonds
CC_ENL	Ratio between number of C-C bonds and total numbere of bonds
C1	Relative number of carbon atoms
CEF1_PES	Relative weight of effective carbon atoms
POLARIZA	Polarizability ($Å^3$)
Topological descriptors:	
$^0\chi^v$	Valence-corrected zero-order molecular connectivity
$^1\chi^v$	Valence-corrected first-order molecular connectivity
$^2\chi^v$	Valence-corrected second-order molecular connectivity
$^3\chi_p{}^v$	Valence-corrected third-order molecular connectivity (A-B-C)
$^4\chi_p{}^v$	Valence-corrected fourth-order molecular connectivity (A-B-C-D)

TABLE 3
Regression Model 1 for RT1[a] and RT2[b]

	RT1			RT2		
Variable[c]	Regression coefficient	Standard error	Contribution to R-SQ	Regression coefficient	Standard error	Contribution to R-SQ
POINTCHG	–0.24083	0.09135	0.02059	–0.24003	0.09396	0.01725
MAXPOB	–2.75992	0.43330	0.12020	–2.82586	0.44570	0.10628
IZ	0.00019	0.00008	0.01700	0.00020	0.00008	0.01571
IY	–0.00014	0.00005	0.01793	–0.00015	0.00006	0.01656
ET	–0.00110	0.00021	0.07909	–0.00116	0.00022	0.07345
ENLACE	–0.01642	0.00330	0.07300	–0.01712	0.00340	0.06690
NEL	–1.40882	0.27924	0.07541	–1.47141	0.28723	0.06938
Intercept	0.09962	1.25659		–0.42574	1.29255	

[a]n = 16; Mallows' C_P = 8.00; mean relative error = 3.15%; R = 0.9880; R^2 = 0.9763; $F(7,8)$ = 47.07.
[b]n = 16; Mallows' C_P = 8.00; mean relative error = 3.24%; R = 0.9893; R^2 = 0.9788; $F(7,8)$ = 52.88.
[c]For explanation of variables, see Table 2. See Table 1 for abbreviation.

Statistical analysis. A total of 60 molecular descriptors was used to explain the behavior of the dependent variable liquid chromatographic RT. Two sets of descriptors were considered. The first set included the quantum-chemical descriptors, and the second one was constituted by the conventional and topological descriptors. In both sets, a stepwise regression was carried out, in order to select the best independent variables subset, following as criterion a minimum value for the Mallows' C_P. For this purpose, the multicolinearity effect inside each set was eliminated. From this, we considered the following independent variables, whose definitions are given in Table 2: Set 1: HOMO, ENLACE, MAXPOB, IX, IY, IZ, ACM, NEL, ET, POINTCHG, CORE, ELECTRON; Set 2: NH, C1_PESO, CH_ENL, CC_ENL, C1, CEF1_PES, POLARIZA, ${}^0\chi^v$, ${}^1\chi^v$, ${}^2\chi^v$, ${}^3\chi_p{}^v$, ${}^4\chi_p{}^v$.

Then, using the 9R program in the BMDP statistical package, the best subset of regression may be selected. The data and statistics obtained for each set and column are given in Tables 3 and 4. When both sets are considered simultaneously, good correlation was not found.

RESULTS AND DISCUSSION

In model 1, whose statistics are given in Table 3, MAXPOB, ET, ENLACE, and NEL indexes are significant above the 99% level, whereas POINTCHG, IZ, and IY indexes are significant above 95% for both RT1 and RT2. In model 2 (Table 4), all coefficients are significant above the 99.9% level, except the ${}^0\chi^v$ index, which is significant above the 99.5% level, also for both RT1 and RT2.

Plots of experimental vs. calculated values of 16 cases, for each regression equation, are depicted in Figures 1 and 2. Residuals vs. experimental RT1 and RT2 values, following the two models, have been plotted in Figures 3 and 4.

The residuals are normally distributed and independent, and there is no autocorrelation between them. In the same way, the Mahalanobis distance shows that extremely high values do not exist, at a confidence level of 95%. If we consider leverage values about the influence of a sample value, there are no sample values with leverage values greater than three times that of an average data point for RT1 and RT2 in the two studied models.

The largest studentized residuals in absolute value among cases are 2.60 for RT1 and 2.77 for RT2 for the first model regression equation. In the second model regression equation, those values are 2.21 and 2.19. In conclusion, the two models clearly make good predictions. Furthermore, both models have the same number of descriptors; therefore, it is not necessary to find a compromise between the exactitude of the prediction and the total number of descriptors used.

TABLE 4
Regression Model 2 for RT1[a] and RT2[b]

	RT1			RT2		
Variable[c]	Regression coefficient	Standard error	Contribution to R-SQ	Regression coefficient	Standard error	Contribution to R-SQ
${}^0\chi^v$	–1.81136	0.36249	0.06197	–1.89413	0.39096	0.05715
${}^4\chi_p{}^v$	–1.03704	0.17641	0.08577	–1.09121	0.19026	0.08047
${}^3\chi_p{}^v$	–2.11210	0.31303	0.11299	–2.16115	0.33761	0.09978
${}^2\chi^v$	–3.34951	0.48616	0.11782	–3.49335	0.52433	0.10809
${}^2\chi^v$	–13.8331	1.75051	0.15500	–14.2056	1.88796	0.13787
POLARIZA	1.73925	0.24673	0.12334	1.80156	0.26610	0.11161
NH	3.72927	0.47578	0.15249	3.84696	0.51313	0.13686
Intercept	19.4024	3.41748		19.7536	3.68621	

[a]n = 16; Mallows' C_P = 8.00; mean relative error = 2.88%; R = 0.9900; R^2 = 0.9801; $F(7,8)$ = 56.41.
[b]n = 16; Mallows' C_P = 8.00; mean relative error = 3.11%; R = 0.9902; R^2 = 0.9805; $F(7,8)$ = 57.52.
[c]For explanation of variables, see Table 2. See Table 1 for abbreviation.

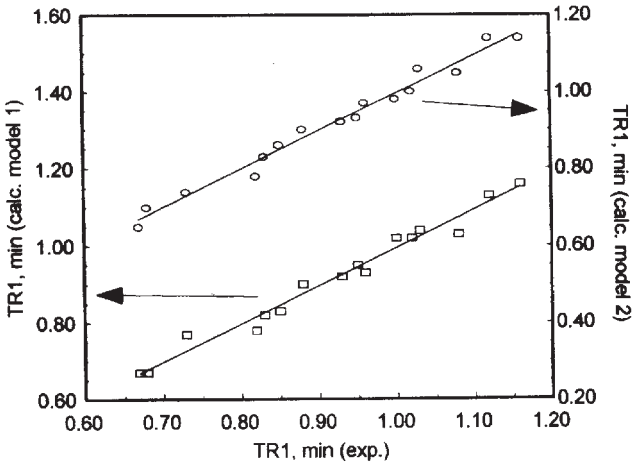

FIG. 1. Plot of experimental vs. expected values of RT1.

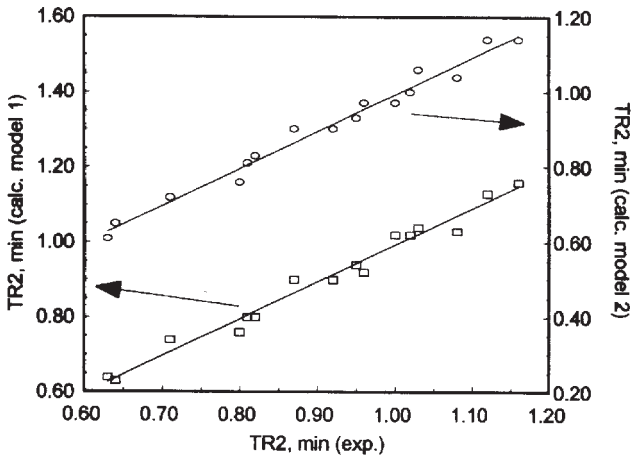

FIG. 2. Plot of experimental vs. expected values of RT2.

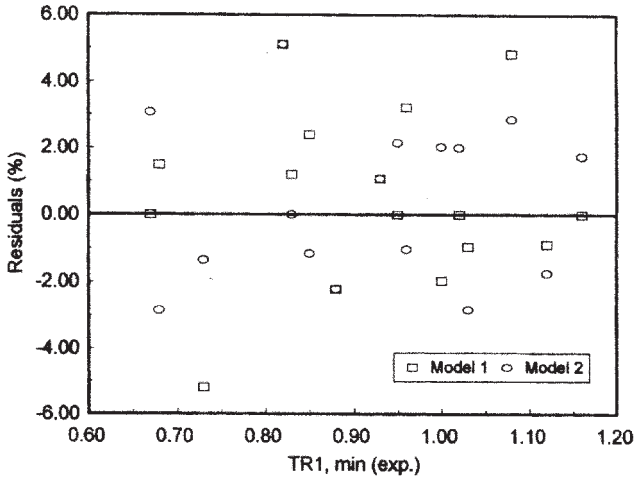

FIG. 3. Plot of residuals vs. experimental values of RT1.

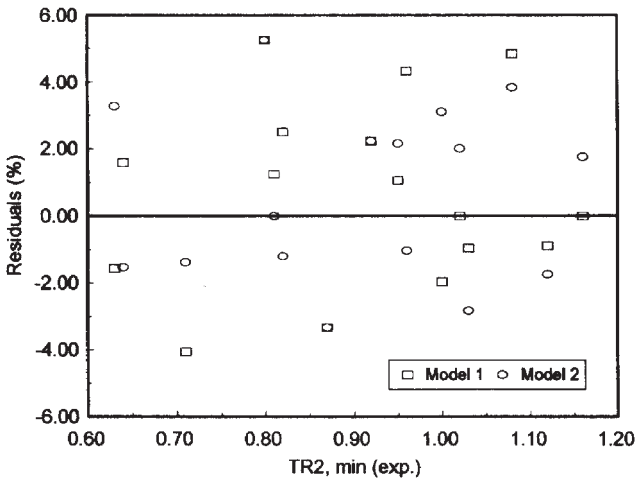

FIG. 4. Plot of residuals vs. experimental values of RT2.

In any case, both models could be used, although the second model could be more useful and preferable because it has higher R^2 values and a lower standard error percentage, giving a little more exact results.

However, the lack of reproducibility of the chromatographic columns may be a major problem in applying the results reported here; obviously, these models are valid only with the same experimental conditions in which the RT values (from which the statistical models have been calculated) have been measured. Finally, the prediction of RT for new related compounds also will depend on the degree of similarity between the query molecules and those in the data set.

ACKNOWLEDGMENT

Thanks are due to Professor Joaquin Altarejos for his kind assistance.

REFERENCES

1. de Blas, O. Jiménez, and A. del Valle González, Determination of Sterols by Capillary Column Gas Chromatography. Differentiation Among Different Types of Olive Oil: Virgin, Refined, and Solvent-Extracted, *J. Am. Oil Chem. Soc. 73*:1685–1689 (1996).
2. Boskou, D., Olive Oil Composition, in *Olive Oil: Chemistry and Tecnology*, edited by D. Boskou, AOCS Press, Champaign, 1996, pp. 52–83.
3. Reglamento CEE 2568/91 de la Comisión, Relativo a las características de los aceites de oliva y de los aceites de orujo de oliva y sobre sus métodos de análisis, *Diario Oficial L248*: pp. 1–48, (1991).
4. Kaliszan, R., *Quantitative Structure–Chromatographic Retention Relationship*, John Wiley & Sons, New York, 1987.
5. Hueso-Ureña, F., S.B. Jiménez-Pulido, M.N. Moreno-Carretero, and J. Rodríguez-Avi, Quantitative Structure–Liquid Chromatographic Retention Time Relationships on Natural Phenols Found in Olive Oil, *J. Am. Oil Chem. Soc. 75*:793–799 (1998).
6. Randic, M., On Characterization of Molecular Branching, *J. Am. Chem. Soc. 97*:6609–6615 (1975).
7. Kier, L.B., and L.H. Hall, *Molecular Connectivity Analysis*, Wiley, New York, 1986.

[Received July 19, 1999; accepted March 16, 2000]