ELSEVIER

Quantitative structure-retention relationship study of the gas chromatographic retention indices of saturated esters on different stationary phases using novel topological indices

Chunhui Lu, Weimin Guo, Chunsheng Yin *

School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China

Abstract

The gas chromatographic retention data of 90 saturated esters obtained isothermally at 150 °C on seven stationary phases (SE-30, OV-7, DC-710, OV-25, XE-60, OV-225 and Silar-5CP) are well-correlated with recently introduced *Lu* index and novel distance-based atom-type *DAI* topological indices using multiple linear regression (MLR) method. For the retention indices (RI) based on each stationary phase, the correlation coefficient *r* of the final models is larger than 0.994, and particularly, the decrease in the standard error is within the range of 53.0–64.8% as compared with the simple linear models with *Lu* index alone. The results indicate the efficiency of these indices in the structure–retention index correlations of complex compounds. The role of each of the molecular size and individual groups in the molecules is illustrated by analyzing the relative or fraction contributions of individual indices. The results also indicate that the molecular size makes a dominant contribution to retention behaviors, while some atomic type or groups have smaller influence dependent on the polarity of the column. The leave-one-out cross-validation demonstrates the final models to be statistically significant and reliable.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Gas chromatographic retention indices; *Lu* index; *DAI* indices; Esters; QSRR

1. Introduction

The study of the quantitative structure–retention relationship (QSRR) of solutes is an important topic in chromatographic thermodynamics. As it is well-known, the chromatographic retention is based on the interaction between the solute and the stationary phase, which included directional force, induction force, dispersion force, hydrogen bond, and so on. The difference of the retention is the macrocosmic reflection of the molecular structure of the solute and the properties of the stationary phase. Correlations between retention indices (RI) and molecular topological indices provide a feasible and useful methodology to study possible mechanism of solute and stationary phase. In the past few years, numerous inves-

tigators have reported very good correlation between experimental chromatographic RI and various topological indices [1–5].

Most recently, two kinds of novel topological indices were introduced by our laboratory and both are meaningful [6,7]. One topological index is called *Lu*, which codes the information of molecular size; the other is distance-based atom-type topological index *DAI*, which characterizes the influence of molecular groups, and both can be easily calculated. The main advantage of the use of these two kinds of topological indices is that the above topological indices are so meaningful that can be used to effectively interpret the possible mechanism of solute and stationary phase retention.

The aim of our study was to construct highly correlative models between gas chromatographic RI of 90 saturated esters on different stationary phases (SE-30, OV-7, DC-710, OV-25, XE-60, OV-225 and Silar-5CP) and our recently presented topological indices (*Lu* and *DAI*) [6,7]. Furthermore, we also wish to demonstrate the structural features or groups that are likely

* Corresponding author. Tel.: +86 21 54748994; fax: +86 21 54740825 608.
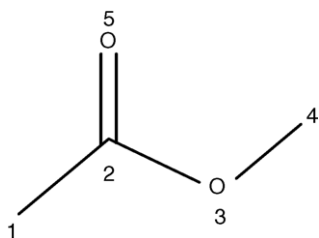*E-mail address:* csyin@sjtu.edu.cn (C. Yin).

Fig. 1. The labeled molecular graph of methyl acetate.

to be important to these RI on different stationary phases with increasing polarity.

2. Methods

2.1. The definition of the Lu index

The *Lu* index is defined as follows:

$$Lu = n^{1/2} \log \left[\frac{1}{2} \left(\sum_{i}^{n} \sum_{j}^{n} D_{ij} + \sum_{i}^{n} \sum_{j}^{n} D_{ij}^2 \right) \right] \quad (1)$$

where n is the number of vertices in a molecular topological graph. D_{ij} is the shortest distance between vertices i and j and is calculated by summing the relative bond length [8] (take C—C bond length 0.154 nm as 1) between two adjacent vertices in the shortest path.

As an illustration, Fig. 1 depicts the labeled molecular graph of methyl acetate. The shortest distance matrix D is expressed as follows:

$$D = \begin{bmatrix} 0 & 1 & 1.928 & 2.928 & 1.792 \\ 1 & 0 & 0.928 & 1.856 & 0.792 \\ 1.928 & 0.928 & 0 & 0.928 & 1.720 \\ 2.928 & 1.856 & 0.928 & 0 & 2.648 \\ 1.792 & 0.792 & 1.720 & 2.648 & 0 \end{bmatrix}$$

According to the definition above, the *Lu* index is calculated as

$$Lu = 5^{1/2} \log \left[\frac{1}{2} \left(\sum_{i=1}^{5} \sum_{j=1}^{5} D_{ij} + \sum_{i=1}^{5} \sum_{j=1}^{5} D_{ij}^2 \right) \right]$$

$$= 5^{1/2} \log \left[\frac{1}{2} (33.0412 + 64.5321) \right]$$

$$= 5^{1/2} \log(48.7866) = 3.775$$

2.2. The definition of the DAI index

For any atom i that belongs to the jth atom-type in a graph, the novel distance-based atom-type topological index $DAI_i(j)$ is expressed as follows:

$$DAI_i(j) = 1 + \Phi_i(j) \quad (2)$$

$$\Phi_i(j) = n \frac{\sum_{j}^{n} D_{ij}}{\sum_{i}^{n} \sum_{j}^{n} D_{ij}} \quad (3)$$

where the parameter Φ is considered as a perturbing term of the ith atom reflecting the effects of its structural environment; n the number of total vertices in molecular topological graph; and D_{ij} the shortest distance between vertices i and j and is calculated by summing the relative bond length between two adjacent vertices in the shortest path.

According to this definition, for jth atom-type in a molecular graph, the corresponding distance-based atom-type topological index, $DAI(j)$, is the sum of all $DAI_i(j)$ values of the same atom-type in a molecular graph.:

$$DAI(j) = \sum_{i=1}^{m} DAI_i(j) = m + \sum_{i=1}^{m} \Phi_i(j) \quad (4)$$

where m is the count of atoms of the same type. Therefore, the value of $DAI(j)$ is equal to the number of jth atom-type plus total perturbation terms and is closely related to its structural environment.

Similarly, we choose the compound of methyl acetate to calculate its *DAI* indices, which are listed as follows:

$$DAI(CH_3-) = DAI(1) + DAI(4) = \left(1 + 5 \times \frac{7.6480}{33.040}\right)$$

$$+ \left(1 + 5 \times \frac{8.3610}{33.040}\right) = 4.411$$

$$DAI(\rangle C=) = DAI(2) = 1 + 5 \times \frac{4.5760}{33.040} = 1.692$$

$$DAI(-O-) = DAI(3) = 1 + 5 \times \frac{5.5040}{33.040} = 1.833$$

$$DAI(=O) = DAI(5) = 1 + 5 \times \frac{6.9520}{33.040} = 2.052$$

2.3. Multiple linear regression

For each stationary phase, multiple linear regression (MLR) using the *Lu* index and several *DAI* indices is used to develop the final models correlating the RI of esters. The final model is obtained in the form of Eq. (5).

$$RI = a_0 + a_1 Lu + \sum b_j DAI(j) \quad (5)$$

where a_0 is a constant, a_1 the contribution coefficient of the *Lu* index, and b_j the contribution coefficient of jth group (atom-type). Each coefficient describes the sensitivity of a stationary phase to each of the individual indices, so the coefficient of these parameters would reflect the relative importance of each index. Once the set of descriptors containing *Lu* and *DAI* has been calculated, the main problem is the selection of that restricted set which contains the most informative descriptors for modeling the retention index under different stationary phases. As topological descriptors are added and removed, changes in the statistics