Bull. Environ. Contam. Toxicol. (2006) 77:793–798
© 2006 Springer Science+Business Media, Inc.
DOI: 10.1007/s00128-006-1213-6

Bulletin of Environmental Contamination and Toxicology

Prediction of Gas Chromatographic Retention Indices of Benzene Dicarboxylic Diesters Using Novel Topological Indices

C. H. Lu,[1] A. F. Jalbout,[2,3] L. Adamowicz,[2] Y. Wang,[1] C. S. Yin[1]

[1] School of Environmental Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China
[2] NASA Astrobiology Institute and the Department of Chemistry, The University of Arizona, Tucson, AZ 85721, USA
[3] Chemical Institute, National Autonomous University of Mexico, Mexico City, Mexico

Received: 11 May 2005/Accepted: 16 November 2006

Benzene dicarboxylic diesters including phthalates, isophthalates, and terephthalates are abundant chemicals commonly used in industry. Excessive use of benzene dicarboxylic diesters in industrial applications, mainly as plasticizers, has given rise to their persistent presence in consumer goods, and has raised numerous questions about their health effects or damage to the environment (Adams et al. 1995; Staples et al. 1997; Gomez-Hens and Aguilar-Caballos 2003; Balafas et al. 1999). Their widespread applications mean that they have become ubiquitous environmental contaminants. Benzene dicarboxylic diesters are often found in water, soil, air, food products and the human body, and this has led to extensive testing of benzene dicarboxylic diesters because of their possible health effects on humans or damage to the environment. More recent studies in several laboratories, have demonstrated that several benzoic esters have surprising 'anti-androgenic' activity (Li et al. 2003; Mylchreest et al. 1998; Sharpe 2001). Therefore the properties of the benzene dicarboxylic diesters call for special attention.

A high-performance gas chromatography (GC) has played an important role as an analytical technique in separation and analysis of organic compounds. Therefore the high-performance GC retention index (I_R) is a very useful property for the compounds under study. However, there has been a general lack of chromatographic data on benzene dicarboxylic diesters due to the availability of reference compounds, therefore methods that can be employed to predict chromatographic retention indices of benzene dicarboxylic diesters from their structure are important. A quantitative structure-property relationship (QSPR) is a model that relates chemical structure to molecular property such as the chromatographic retention index. A close relationship often exists between the molecular structures of organic compounds and many of their properties. Many of these relationships have been investigated using the topological descriptors of molecular structures (Ren 2003; Krawczuk et al. 2003; Amboni et al. 2002).

Correspondence to: C. S. Yin

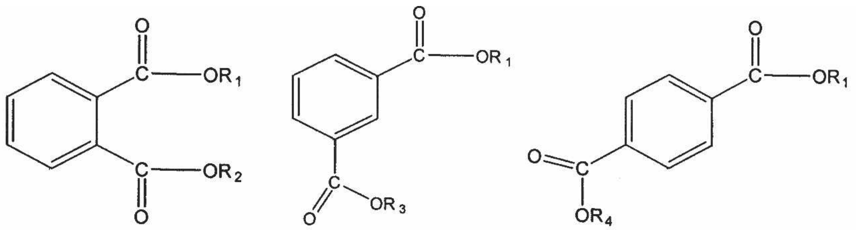


Figure. 1 General sketch of benzene dicarboxylic diester, R_1, R_2, R_3 and R_4 represent carbon chains.

In this study, gas chromatographic retention indices of 32 benzene dicarboxylic diesters were taken into consideration. The quantitative relationships are set up between our recently proposed topological indices and the I_R of 32 benzene dicarboxylic diesters using a multiple linear stepwise regression procedure.

MATERIALS AND METHODS

The general structure of a benzene dicarboxylic diester is shown in Fig. 1. A total of 32 benzene dicarboxylic diester compounds were used as the investigated objects for the QSPR studies in this work. The high-performance GC retention indices for benzene dicarboxylic diester compounds were taken from Friocourt's work (1979). The data set used in this study, consisting of 32 benzene dicarboxylic diesters, is summarized in the Table1 and Table 2.

The *Lu* index is defined as follows:

$$Lu = n^{1/2} \log(1/2\sum_{i}^{n}\sum_{j}^{n} D_{ij}) \qquad 1$$

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 between two adjacent vertices in the shortest path.

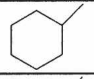
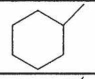
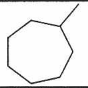
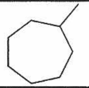
For any atom i that belongs to the jth atom-type in a graph, the novel distance-based atom-type topological index $DAI_i(j)$ is expressed as follows:

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

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

where the parameter Φ is considered as a perturbing term of the ith atom reflecting the effects of its structural environment.

Table 1. Structural characteristics of 32 benzene dicarboxylic diesters.

Compound	$-R_1$	$-R_2$	$-R_3$	$-R_4$
1	$-CH_2CH_3$	$-CH_2CH_3$	——	——
2	$-CH_2CH_2CH_3$	$-CH_2CH_2CH_3$	——	——
3	$-CH_2(CH_2)_2CH_3$	$-CH_2(CH_2)_2CH_3$	——	——
4	$-CH_2(CH_2)_3CH_3$	$-CH_2(CH_2)_3CH_3$	——	——
5	$-CH_2(CH_2)_4CH_3$	$-CH_2(CH_2)_4CH_3$	——	——
6	$-CH_2(CH_2)_5CH_3$	$-CH_2(CH_2)_5CH_3$	——	——
7	$-CH_2(CH_2)_6CH_3$	$-CH_2(CH_2)_6CH_3$	——	——
8	$-CH_2(CH_2)_7CH_3$	$-CH_2(CH_2)_7CH_3$	——	——
9	$-CH_2(CH_2)_8CH_3$	$-CH_2(CH_2)_8CH_3$	——	——
10	$-CH_2C_2H_5CH(CH_2)_3CH_3$	$-CH_2C_2H_5CH(CH_2)_3CH_3$	——	——
11			——	——
12			——	——
13	$-CH_3$	$-CH_2(CH_2)_9CH_3$	——	——
14	$-CH_2CH_3$	$-CH_2(CH_2)_8CH_3$	——	——
15	$-CH_2CH_2CH_3$	$-CH_2(CH_2)_7CH_3$	——	——
16	$-CH_2(CH_2)_2CH_3$	$-CH_2(CH_2)_6CH_3$	——	——
17	$-CH_2(CH_2)_3CH_3$	$-CH_2(CH_2)_5CH_3$	——	——
18	$-CH_2(CH_2)_4CH_3$	$-CH_2(CH_2)_5CH_3$	——	——
19	$-CH_2(CH_2)_2CH_3$	$-CH_2(CH_2)_7CH_3$	——	——
20	$-CH_2(CH_2)_3CH_3$	$-CH_2(CH_2)_8CH_3$	——	——
21	$-CH_3$	——	$-CH_2(CH_2)_9CH_3$	——
22	$-CH_2CH_3$	——	$-CH_2(CH_2)_8CH_3$	——
23	$-CH_2CH_2CH_3$	——	$-CH_2(CH_2)_7CH_3$	——
24	$-CH_2(CH_2)_2CH_3$	——	$-CH_2(CH_2)_6CH_3$	——
25	$-CH_2(CH_2)_3CH_3$	——	$-CH_2(CH_2)_5CH_3$	——
26	$-CH_2(CH_2)_4CH_3$	——	$-CH_2(CH_2)_4CH_3$	——
27	$-CH_3$	——	——	$-CH_2(CH_2)_9CH_3$
28	$-CH_2CH_3$	——	——	$-CH_2(CH_2)_8CH_3$
29	$-CH_2CH_2CH_3$	——	——	$-CH_2(CH_2)_7CH_3$
30	$-CH_2(CH_2)_2CH_3$	——	——	$-CH_2(CH_2)_6CH_3$
31	$-CH_2(CH_2)_3CH_3$	——	——	$-CH_2(CH_2)_5CH_3$
32	$-CH_2(CH_2)_4CH_3$	——	——	$-CH_2(CH_2)_4CH_3$

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.:

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

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.

For each property, multiple linear regression using the Lu index and several DAI indices is used to develop the final models correlating the three physical properties of aldehydes and ketones. The final model is obtained in the form of Eq. 5.

$$\text{Property} = a_0 + a_1 Lu + \sum b_j DAI(j) \qquad 5$$

where a_0 is a constant, a_1 is the contribution coefficient of the Lu index, and b_j is the contribution coefficient of jth group (atom type).

RESULTS AND DISCUSSION

The Lu index, along with DAI indices, was used to select a set of useful descriptive variables and correlate the experimental chromatographic retention indices of these compounds using a multiple linear stepwise regression procedure. As the multiple linear stepwise regression procedure was manipulated, a variety of correlation models were obtained. A linear model was developed by using the above variable selection method for the calculation of the chromatographic retention indices of 32 benzene dicarboxylic diesters as follows:

$$I_R = -13515.33 + 273.11 Lu - 136.37 DAI(CH_3{-})$$

$$-60.41 DAI({-}CH_2{-}) + 6066.96 DAI({-}CH{\approx}) \qquad 6$$

$$N=32,\ R^2=0.987,\ R^2_{cv}=0.979,\ s=32,\ F=501,\ p<0.0001$$

The t-values are -4.24, 8.49, -8.68, -2.97, and 3.97, respectively. All indices in the model are statistically significant according to the t-values at the level of $p<0.0001$. This model produces a standard error of 32 and explains more than 98.7% (R^2) of the variances in the experimental I_R values for these compounds.

On the other hand, the model is further validated using the leave-one-out cross-validation procedure. The correlation coefficient of the cross-validation demonstrates the model to be statistically significant. The I_R estimated values for the above model are listed in Table 2.

Furthermore, in order to demonstrate the good predictive power of the constructed model, the model (Eq. 6) is used to predict the randomly selected I_R values of 10

Table 2. The experimental and calculated I_R for 32 compounds.

Compound	Expt	Cald	No.	Expt	Cald
1	1583	1574	17	2310	2327
2	1758	1755	18	2404	2417
3	1940	1943	19	2415	2443
4	2122	2136	20	2602	2623
5	2308	2322	21	2459	2471
6	2497	2512	22	2423	2434
7	2685	2693	23	2419	2405
8	2876	2876	24	2419	2382
9	3067	3052	25	2416	2369
10	2509	2514	26	2417	2364
11	2475	2525	27	2483	2503
12	2735	2689	28	2455	2469
13	2389	2437	29	2459	2440
14	2343	2398	30	2465	2420
15	2325	2360	31	2466	2408
16	2317	2342	32	2469	2404

Table 3. Structural characteristics, the predicted and experimental I_R for the test set of 10 benzene dicarboxylic diesters.

Compound	$-R_1$	$-R_2$	$-R_3$	Expt	Cald
1	$-CH_2CH_3$	$-CH_2CH_3$	——	1639	1639
2	$-CH_2CH_2CH_3$	$-CH_2CH_2CH_3$	——	1829	1812
3	$-CH_2(CH_2)_2CH_3$	$-CH_2(CH_2)_2CH_3$	——	2030	1995
4	$-CH_2(CH_2)_3CH_3$	$-CH_2(CH_2)_3CH_3$	——	2222	2181
5	$-CH_2(CH_2)_4CH_3$	$-CH_2(CH_2)_4CH_3$	——	2417	2367
6	$-CH_2CH_3$	——	$-CH_2CH_3$	1650	1696
7	$-CH_2CH_2CH_3$	——	$-CH_2CH_2CH_3$	1851	1863
8	$-CH_2(CH_2)_2CH_3$	——	$-CH_2(CH_2)_2CH_3$	2060	2041
9	$-CH_2(CH_2)_3CH_3$	——	$-CH_2(CH_2)_3CH_3$	2261	2223
10	$-CH_2(CH_2)_5CH_3$	——	$-CH_2(CH_2)_5CH_3$	2665	2588

compounds not involved in regression analysis. The structural characteristics and experimental I_R of the 10 compounds are listed in Table 3. The predictive standard error s_{pred} is 16 for the test set, showing a good predictive power of the model.

The quantitative structure-property relationship model was successfully developed in the estimation and prediction of GC retention indices by using novel descriptors including *Lu* and *DAI* indices having high predictive ability for benzene dicarboxylic diesters. It obviously seems that QSPR is a feasible approach to estimate the GC index of benzene dicarboxylic diesters. The results will certainly be valuable in estimating the I_R of suchlike chemicals.

Acknowledgments. The project was supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, and the National Science Foundation of China (NNSFC) (No.30470918).

REFERENCES

Amboni RDC, Junkes BDS, Heinzen VEF, Yunes RA, (2002) Semi-empirical topological method for prediction of the chromatographic retention of esters. J Mol Struct 579: 53–62

Adams WJ, Biddinger GR, Robillard KA, Gorsuch JW (1995) A summary of the acute toxicity of 14 phthalate esters to representative aquatic organisms. Environ Toxicol Chem 14: 1569–1574

Balafas D, Shaw KJ, Whitfield FB (1999) Phthalate and adipate esters in Australian packaging materials. Food Chem 65: 279–287

Friocourt MP, Berthou F, Picart D, Dreano Y, Floch, HH (1979) Glass capillary column gas chromatography of phthalate esters. J Chromatog 172: 261–271

Gomez-Hens A, Aguilar-Caballos MP, (2003) Social and economic interest in the control of phthalic acid esters. TRAC-Trend Anal Chem 22: 847–857

Heinzen VEF, Yunes RA (1996) Using topological indices in the prediction of gas chromatographic retention indices of linear alkylbenzene isomers. J Chromatog 719: 462–467

Krawczuk A, Voelkel A, Lulek J, Urbaniak R, Szyrwinska K (2003) Use of topological indices of polychlorinated biphenyls in structure–retention relationships. J Chromatog 1018: 63–71

Li LH, Jester WF, Laslett Jr, Laslett AL, Orth JM (2000) A single dose of Di-(2-ethylhexyl) phthalate in neonatal rats alters gonocytes, reduces sertoli cell proliferation, and decreases cyclin D2 expression. Toxicol Appl Pharmacol 166: 222–229

Mylchreest E, Cattley RC, Foster PMD (1998) Male Reproductive Tract Malformations in Rats Following Gestational and Lactational Exposure to Di(*n*-butyl) Phthalate: An Antiandrogenic Mechanism? Toxicol Sci 43: 47–60

Ren BY (2003) Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones. Chemometr Intel Lab 66: 29–39

Staples CA, Peterson DR, Parkerton TF, Adams WJ (1997) The environmental fate of phthalate esters: a literature review. Chemosphere 35: 667–749

Sharpe RM (2001) Hormones and testis development and the possible adverse effects of environmental chemicals. Toxicol Lett 120: 221–232