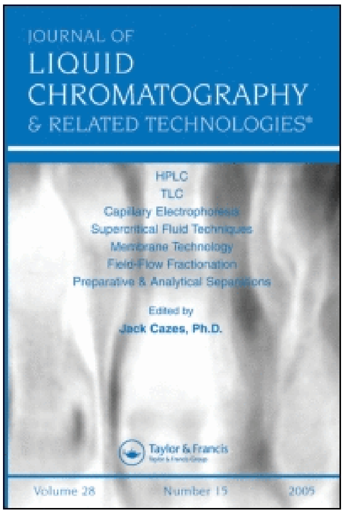
This article was downloaded by:
On: *23 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Liquid Chromatography & Related Technologies

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597273

Linear Regression Based QSPR Models for the Prediction of the Retention Mechanism of Some Nitrogen Containing Heterocycles

Yulia Polyakova[a]; Long Mei Jin[a]; Kyung Ho Row[a]
[a] Center for Advanced Bioseparation Technology and Department of Chemical Engineering, Inha University, Incheon, Korea

To cite this Article Polyakova, Yulia , Jin, Long Mei and Row, Kyung Ho(2006) 'Linear Regression Based QSPR Models for the Prediction of the Retention Mechanism of Some Nitrogen Containing Heterocycles', Journal of Liquid Chromatography & Related Technologies, 29: 4, 533 — 552

To link to this Article: DOI: 10.1080/10826070500479062

URL: http://dx.doi.org/10.1080/10826070500479062

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Journal of Liquid Chromatography & Related Technologies®, 29: 533–552, 2006

ISSN 1082-6076 print/1520-572X online
DOI: 10.1080/10826070500479062

Linear Regression Based QSPR Models for the Prediction of the Retention Mechanism of Some Nitrogen Containing Heterocycles

Yulia Polyakova, Long Mei Jin, and Kyung Ho Row
Center for Advanced Bioseparation Technology and Department
of Chemical Engineering, Inha University, Incheon, Korea

Abstract: This study evaluates retention factors of 29 nitrogen containing heterocycles using QSPR models in liquid chromatography. Some structure properties, such as the molecular connectivity indices ($^0\chi \sim {}^5\chi$), Wiener index (W), Kier flexibility index (ϕ), Harary index (H), Balaban indices ($JX \sim JY$), and Zagreb indices ($M_1 \sim M_2$) were obtained by theoretical molecular descriptors derived from information of chemical structures of substances. The relationship between the chromatographic retention factors and the structure descriptors were predicted using a mathematical method, regression analysis. All substances were divided into five groups based on similar structures with functional groups. For each group, the simple linear regression between each structural index and the retention factor showed good regression coefficients. The linear regression between retention factor and various orders of connectivity indices showed good regression coefficient ($r^2 = 0.8 \sim 1$) and, with multiple structural properties, showed $r^2 = 1$. This study has demonstrated the successful linear regression approaches to prediction of the retention factors and some molecular descriptors of the substances.

Keywords: Linear regression, QSPR, Nitrogen containing heterocycles, Retention factor

INTRODUCTION

During recent years, the nitrogen containing heterocyclic compounds have been developed as medical and biological agents to combat various

---

Address correspondence to Prof. Kyung Ho Row, Center for Advanced Bioseparation Technology and Department of Chemical Engineering, Inha University, Incheon 402-751, Korea. E-mail: rowkho@inha.ac.kr

diseases. During the 20th century, they were introduced as active compounds in medicine and the number of such heterocyclic compounds used as drugs is expected to expand in the near term outlook.[1] They represent a successful approach for chemotherapeutic development but, as the occurrence of the different infections has increased, coupled to prolonged use of the drugs, the almost inevitable emergence of resistance has occurred. This group of drugs, with activity against most bacteria and filamentous microorganisms, consist of nitrogen containing heterocyclic compounds and have been commercially available since the middle part of the last century. Currently, these form the greatest part of the world market for healing agents. For these reasons, constant attempts to synthesize and design the heterocyclic drugs has been made in recent years.[2–6]

In particular, nitrogen containing heterocyclic compounds (imidazole and triazole derivatives) has led to many helpful drugs presently in clinical utilization. Newer "azoles" with expanded spectra of biological activities are, at the present moment, in continuous development.

An interdisciplinary approach to their chemistry is currently moving towards medicinal chemistry, supramolecular chemistry, and advanced organic materials. In most cases, these tasks can be achieved by realization of the technique of forecasting of properties of compounds.

A QSPR approach (quantitative structure–property relationship), as one of the most essential areas in modern chemistry, gives information that is practical and necessary for drug design and for combinatorial and medicinal chemistry. In addition, the quantitative structure-property relationship is an alternative approach for estimating their physicochemical properties.

The basis of the QSPR is that physicochemical properties of organic compounds are governed by their molecular structural characteristics (geometric and electronic), expressed in terms of the appropriate molecular descriptors.[7,8] This technique requires only the knowledge and description of the chemical structures. When the significant QSPR models are obtained, they may also provide explicitness and comprehension into which aspect of the molecular structural influences on the properties.

In this study, the QSPR approach was used to investigate the structural descriptors of some nitrogen containing heterocycles and their chromatographic retention.

THEORETICAL BACKGROUND

Retention Factor

The retention factor (k) can be calculated according to Eq. (1).

$$k = (t_R - t_M)/t_M \tag{1}$$

where t_M is the hold-up time and t_R is the retention time.

Molecular Descriptors

The molecular descriptors of the nitrogen containing heterocycles were drawn by the preADME program software. The preADME consists of two main parts: a program called TOPOMOL, for the calculation of important descriptors in characterizing ADME properties, and a neural network program for the construction of the drug absorption prediction system. The ADME/Tox properties are closely related to physicochemical descriptors, such as the lipophilicity (logP), molecular weight, polar surface area, and water solubility. The TOPOMOL program calculates more than 125 molecular descriptors, including the constitutional, topological, physicochemical, and geometrical descriptors for ADME prediction, from 2D chemical structures. The TOPOMOL reads MDL.* mol or.* sd types of files and provides a rapid means to calculate 125 descriptors of 1,000,000 compounds in less than 1 hour using a 1.6 GHz PC.[9]

To investigate the relationship between structural descriptors of the azoles and their retention characteristics, the QSPR approach was used to obtain topological descriptors. The connectivity index ($^0\chi \sim {}^5\chi$), Winner index (W), Kier flexibility index (ϕ), Harary index (H), Balaban index ($JX \sim JY$), and Zagreb index ($M_1 \sim M_2$) topological descriptors were obtained by the preADME, and are listed in Table 2.

Molecular Connectivity Index (χ)

The molecular connectivity index was originally proposed as a branching index by Randic (in 1975) and was later extensively developed and formalized by Kier and Hall (in 1976). The molecular connectivity index is the topological descriptor of the molecular structure based on a count of the skeletal atom groupings.[10] It reflects the atoms, combination environment, and the number of combined to hydrogen. Hall and Kier showed the connectivity index based on the Subgraph Count Indices; also, this includes information about the circumferential environment. The value of the connectivity index is able to be expressed, firstly, by $\delta_i = \sigma_I - h_i$. δ shows the combined number of σ combination, which accepts hydrogen in the *ith* atom. It regards the structure of the chemical compound simply, as a line by exception hydrogen, ignoring the combination number. The case of order 0, 1, 2 only has the path form, but from order 3 to higher orders, it also has the cluster form. In this study, we just calculate the path form for each order form. Against each order, when it appears, Eq. (2) $\sim$ Eq. (4) is expressed.

Order $= 0$

$$^0\chi_P = \sum_{j=1}^{A} (\delta_j)^{-1/2} \tag{2}$$

where A is the number of skeletal atoms in the molecule; δ_i is the δ value of each skeletal atom. The nonvalence first-order term ${}^1\chi$ is calculated after assigning a δ value to the adjacent atoms i and j.[11]

Order = 1

$$ {}^1\chi_P = \sum_{j=1}^{\tau}\sum_{i=1}^{\tau}(\delta_j\delta_i)^{-1/2} \tag{3}$$

where τ is the number of bonds in the molecule, δ_i, δ_j are the values of two adjacent atoms. Calculations of the third-, fourth-, fifth-, and sixth-order path terms are performed in a similar way.

Higher Order

$$ {}^m\chi_t = \sum_{i=1}^{\tau}\prod_{j=1}^{N}[(\delta_j)_i]^{-1/2} \tag{4}$$

where m indicates the order, t indicates the fragment type, e.g., paths of one or more bonds, clusters (a tripod-like structure) where three atoms are adjacent to a single atom, and path clusters that are similar to a cluster with one or more atoms connected to any of the three atoms that form the cluster.[12]

Wiener Index (*W*)

The Wiener index was introduced by Wiener in 1947; it is defined as in Eq. (5). The equation, which reflects the Wiener index, is half the sum of the total element of the distance matrix.

$$ W = \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N}D_{ij} \tag{5}$$

where N is the number of the atom which accepts the hydrogen inside one element; distance D_{ij} is the distance matrix which is expressed by the combination number in the shortest course which atom connects i and j.[11]

Flexibility Index (ϕ)

The flexibility index is one of the structural characteristics; it affects the chemical and biological phenomena. Kier regards a completely flexible molecule to an endless carbon sp^3 atom chain; by this assumption, he defined several flexibility indices using several factors. The factor which is considered is listed below.

1. The number of atom
2. The existence of cycle

3. The existence of branching
4. The existence of the atom which has the covalent radii smaller than sp^3.

Flexibility considered with above factors is defined by the kappa alpha index 1 which shows information about the relative cyclicity and the branching of the molecule, the relative spatial of the element density and kappa alpha index 2 as shown in Eq. (6).

$$\Phi = {}^1k_a{}^2k_a/A \tag{6}$$

Harary Index (*H*)

The Harary index, $H = H(G)$ of a molecular graph G with N vertices is based on the concept of reciprocal distance and is defined, in parallel to the Wiener index, as the half-sum of the off-diagonal elements of the reciprocal molecular distance matrix $D^r = D^r(G)$:

$$H = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}[D^r]_{ij} \tag{7}$$

The reciprocal distance matrix D^r can be simply obtained by replacing all off-diagonal elements of the distance matrix $[D]_{ij}$ by their reciprocals.[13]

$$[D^r]_{ij} = \frac{1}{[D]_{ij}} \tag{8}$$

Balaban Index (*J*)

It was introduced by Balaban (in 1982) and it is the better identified descriptor than the Wiener index. The two most important distance matrix based topological indices are the Wiener index and the Balaban index. The asymptotic behavior of the Balaban index for linear polymers (i.e., finite and repetitive) branching patterns along the main chain has been previously analyzed. Unlike the Wiener index, which diverges, the behavior of the Balaban index mimics the behavior of the melting temperatures and glass transition temperatures for linear macromolecules, which possess asymptotic limits for these physical properties. Unfortunately, for highly branched molecular graphs (i.e., for structures in which the branching pattern involves the formation of branches or other branches) the Balaban index diverges too.

After making the reciprocal distance matrix, the sum of element value S_i of each row or column can be calculated as in Eq. (9). Relative electronegativity X and relative covalent radii Y are defined by Eqs. (10) and (11).

$$S_i = \sum_{i}^{N} a_{ij}^{D} \tag{9}$$

$$X = 0.4196 - 0.0078Z_i + 0.1567G_i \tag{10}$$

$$Y = 1.1191 + 0.0160Z_i - 0.0537G_i \tag{11}$$

where Z_i is the number of the atom; G_i is the group number of Mendeleev's short periodic table. The Balaban index JX and/or JY is given by Eq. (9).

$$J = \frac{M}{M - N + 2} \sum \frac{1}{\sqrt{S_i^a S_j^a}} \tag{12}$$

where M is the number of the bond; N is the number of the atom. In the Balaban index JX, $S_i^a = XS_i$; In the Balaban index JY, $S_i^a = YS_i$.[14]

Zagreb Indices M_1 and M_2

This pair of indices was introduced in 1972 in Zagreb, Croatia. It has been given different names in the literature, such as the Zagreb group indices, the Zagreb group parameters and, most often, the Zagreb indices. Zagreb index is the sum of square values of all atom degrees and it is related to molecular branching. These indices are denoted by M_1 and M_2 and are defined as per the following equations:

$$M_1 = \sum_{vertices} d(i)d(j) \tag{13}$$

$$M_2 = \sum_{edges} d(i)d(j) \tag{14}$$

where *d(i)* is the degree of vertex *i* and *d(j)* is the weight of edge *i–j*.[15]

EXPERIMENTAL

Reagents

Twenty-nine substances were kindly provided by Professor A. Bulanova (Russia). HPLC grade acetonitrile was obtained from Panreac Corp. (Spain). Trifluoroacetic acid (TFA) and dimethylsulfoxide (DMSO) were purchased from Sigma (St. Louis, MO, USA). All water used in the experiments was filtered by a Millipore ultra pure water system (Millipore, Belford, MA, USA).

Sample Preparation

Five mg of each of the twenty-four substances was dissolved in 1 mL of water or 1 mL of mixed solvents (water/DMSO). A constant injection volume (5 μL) was used for the mixture solutions throughout this work.

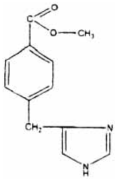
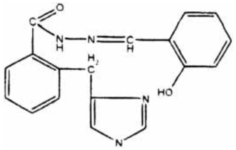
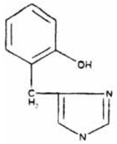
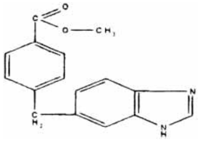
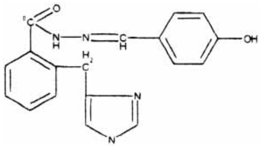
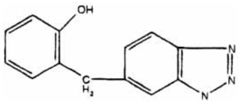
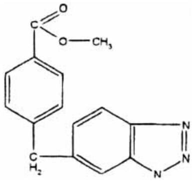
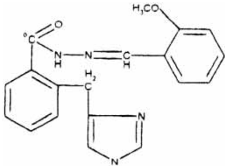
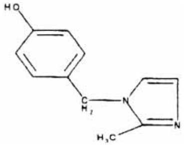
Apparatus and Methods

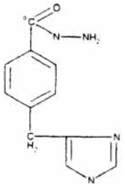
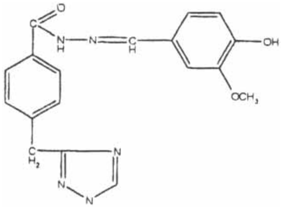
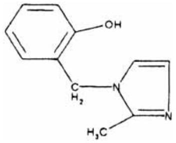
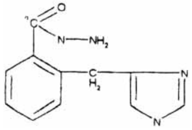
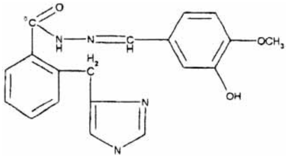
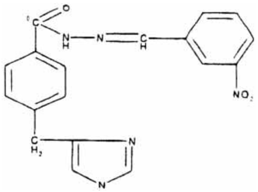
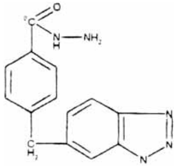
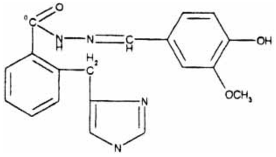
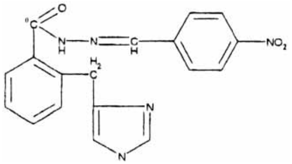
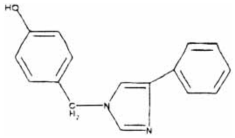
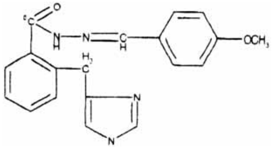
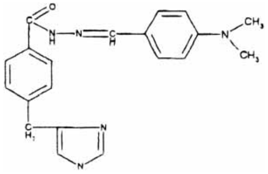
The research was performed using an Agilent 1100 series HPLC–MS system (Agilent technologies, USA). The HPLC system included an Agilent (Wilmington, Delaware, USA) 1100 binary pump with an online degasser, and an autosampler. The HPLC system was used with UV-Vision diode array detection (DAD), evaporative light scattering detection (ELSD), and mass spectrometric (MS) detection. The chromatographic separations were performed using a Betasil 10 μm octadecyl silica (C_{18}) column (4.6 × 150 *mm*) (Phenomenex, USA). The column temperature was 293 K. The eluent consisted of a mixture of 0.1% TFA in water (A) with 0.1% TFA in acetonitrile (B) and was delivered at a flow-rate of 2.5 mL/min. The gradient program consisted of increasing the concentration of B from 5% to 90% after 0.5 min, over a period of 5.5 min, and then holding it at this concentration for another 1.5 min. The run time, from injection to injection, was 9.5 min. The single quadrupole MS was equipped with an electrospray source and controllable in the selected ion-monitoring (SIM) mode. The electrospray was operated in positive ion mode to produce, mainly, protonated molecular mass ions $[M - H]^{+1}$. The scanning mass to charge values ranged from 90 to 900 m/z with a scanning speed of 13000 m/z per second. The MS parameters were set to facilitate the ionization process and achieve the optimal sensitivity. These conditions allowed for minimal fragmentation of the target compound, optimum yield of its quasi-molecular ion, and minimum interference from the matrix. An Agilent 1100 LC–MS system, managed by Windows NT based HP ChemStation software (Rev 8.04), was used. The chromatograms were integrated using the HP ChemStation software package.

RESULTS AND DISCUSSION

In this paper, the retention factors of the nitrogen containing heterocyclic compounds from Table 1 were estimated by the mathematical relationships expressed as a linear relationship in terms of some topological descriptors. To investigate the relationship of the chromatographic retention factor in terms of descriptors, several experimental sets were performed in gradient mode. A typical chromatogram of one of the substances is shown in Figure 1.

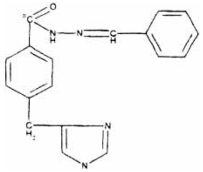
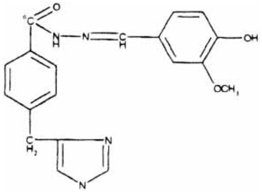
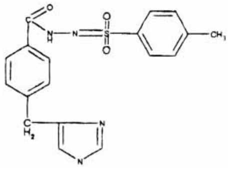
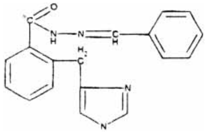
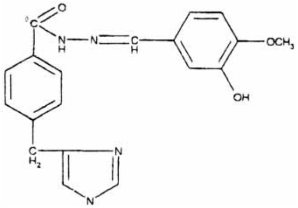
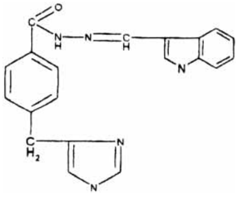
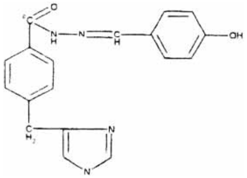
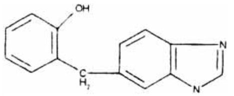
Table 1. Chemical structures of nitrogen-containing heterocycles compounds

No	Structure	*No*	Structure	*No*	Structure
1		11		21	
2		12		22	
3		13		23	

4		14		24	
5		15		25	
6		16		26	
7		17		27	

(*continued*)

Table 1. Continued

No	Structure	*No*	Structure	*No*	Structure
8		18		28	
9		19		29	
10		20			

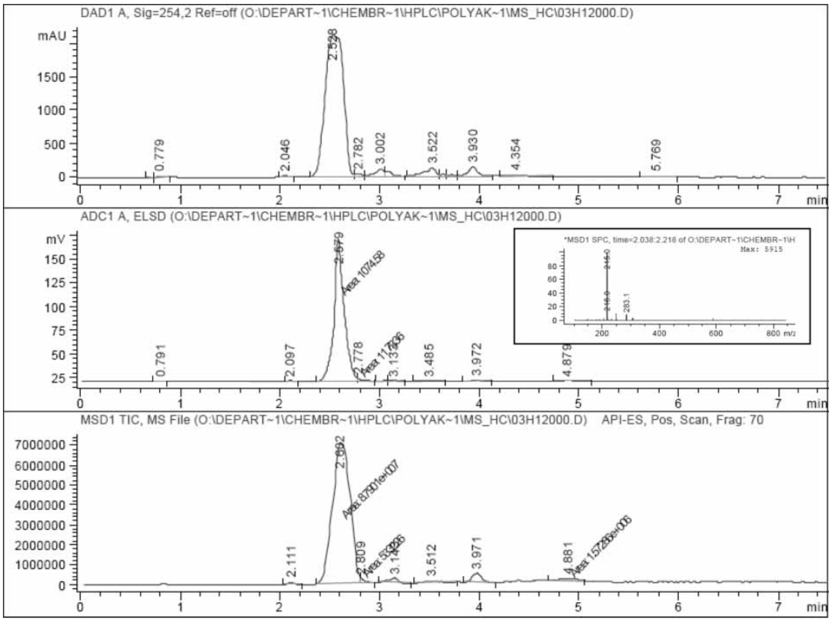


Figure 1. Chromatogram of substance No. 1 in Table 1 in gradient mode.

The deficiency of homology in chemical structures of the nitrogen containing derivatives made it complex to investigate the relationships between topological models. In this case, linear regression analysis is not always valid for the description of the chromatographic retention within the limit of a full nonhomologous set of the substances. From this point of view, the best results can be obtained by the formal division of all of the investigated substances into batches (groups) of substances having some common features.

In our study, the idea of the structural analogy was used for division of the substances into groups. Molecules having similar structural fragments were classified into the same group. Thus, all investigated substances were placed into 5 groups, depending upon their structural analogies. The data from Table 2 served as the basis for statistical analysis.

The chemical structures of the nitrogen containing derivatives are listed in Table 1 and the calculated retention factors and topological descriptors are listed given in Table 2. To investigate the relationships between them, simple and multiple linear regression analysis was used. Simple linear regressions ($k = a + b\ P_i$, P_i indicates topological index) of each group are listed in Table 3. The regression coefficient (r^2) of the group I ($0.788 \sim 0.803$), group III ($0.531 \sim 0.963$, except 0.289), and group IV ($0.796 \sim 0.872$, except 0.347) are high, but group II and group V showed low regression coefficients. This indicates that these linear equations

Table 2. The value of retention factors (k) and structural descriptors of nitrogen-containing heterocycles compounds

No	k	$^0\chi$	$^1\chi$	$^2\chi$	$^3\chi$	$^4\chi$	$^5\chi$	W	ϕ	H	JX	JY	M_1	M_2
Group I														
1	2.663	11.380	7.792	6.553	5.482	4.044	2.677	491	3.238	45.542	2.041	2.124	78	88
2	3.810	13.949	9.758	8.532	7.405	5.812	4.385	909	3.412	64.491	1.769	1.823	104	121
3	5.686	13.949	9.758	8.532	7.405	5.812	4.385	909	3.367	64.491	1.761	1.826	104	121
4	0.818	11.380	7.792	6.553	5.482	4.044	2.677	491	3.208	45.542	2.046	2.122	78	88
5	1.386	11.380	7.809	6.480	5.454	4.255	2.885	443	3.208	46.743	2.230	2.316	78	89
6	2.979	13.949	9.758	8.532	7.405	5.812	4.385	909	3.341	64.491	1.764	1.824	104	121
7	3.628	13.079	9.310	8.199	6.700	5.398	4.296	787	3.228	59.613	1.760	1.814	98	112
Group II														
8	3.386	15.908	11.310	9.579	7.834	6.150	4.352	1487	4.831	75.223	1.539	1.593	114	128
9	3.752	15.908	11.327	9.507	7.806	6.371	4.523	1355	4.831	77.184	1.662	1.722	114	129
10	2.737	16.778	11.704	10.201	8.245	6.292	4.636	1689	5.024	79.868	1.527	1.586	120	135
11	3.842	16.778	11.737	10.025	8.333	6.690	4.719	1511	5.024	82.468	1.671	1.738	120	137
12	2.797	16.778	11.720	10.129	8.217	6.513	4.807	1545	5.024	81.893	1.644	1.709	120	136
Group III														
13	3.682	17.485	12.275	10.216	8.635	7.070	4.988	1691	5.582	87.369	1.666	1.739	124	142
14	3.168	18.355	12.652	10.900	9.105	7.018	5.024	2090	5.718	89.922	1.530	1.608	130	148

15	3.150	18.355	12.699	10.827	9.061	7.334	5.028	1939	5.773	91.901	1.635	1.711	130	149
16	3.128	18.355	12.669	10.827	9.077	7.239	5.196	1922	5.773	92.087	1.645	1.722	130	149
17	3.860	17.485	12.258	10.298	8.625	6.823	4.905	1759	5.582	86.367	1.619	1.688	124	141
18	2.900	18.355	12.652	10.900	9.105	7.018	5.024	2090	5.773	89.222	1.535	1.607	130	148
19	3.014	18.355	12.652	10.900	9.089	7.114	4.857	2107	5.773	89.746	1.527	1.597	130	148
Group IV														
20	3.476	11.665	8.327	7.388	6.323	5.185	3.979	533	2.475	52.187	1.932	1.978	90	105
21	2.265	9.096	6.360	5.408	4.400	3.417	2.268	253	2.206	34.236	2.180	2.252	64	72
22	5.084	11.665	8.327	7.388	6.323	5.185	3.979	533	2.435	52.187	1.921	1.982	90	105
23	1.710	9.966	6.754	6.030	4.811	3.565	2.626	320	2.433	38.073	2.128	2.209	70	79
24	2.623	9.966	6.771	5.926	4.934	3.704	2.552	306	2.433	38.529	2.207	2.296	70	80
Group V														
25	3.831	17.769	12.776	11.205	8.933	7.169	5.110	2091	5.635	90.289	1.290	1.361	136	157
26	4.027	17.769	12.793	11.121	8.973	7.315	5.214	1974	5.635	91.841	1.354	1.428	136	158
27	2.518	18.355	12.614	11.100	8.944	6.818	4.821	2143	5.804	89.205	1.508	1.567	130	147
28	3.439	18.571	12.437	11.628	9.203	6.791	5.046	1983	5.287	92.009	1.537	1.624	134	153
29	3.489	17.769	12.793	11.121	9.554	7.864	5.651	2049	4.614	91.689	1.359	1.408	136	158

Table 3. Linear equations ($k = a + b\ P_i$) of the structural descriptors P_i with groups

Group	P_i	a	b	r^2	P_i	a	b	r^2
I	$^0\chi$	−9.712	0.999	0.798	W	−1.169	0.006	0.804
	$^1\chi$	−8.485	1.297	0.801	ϕ	−42.108	13.726	0.720
	$^2\chi$	−6.599	1.258	0.803	H	−4.650	0.137	0.797
	$^3\chi$	−5.612	1.329	0.797	JX	15.473	−6.532	0.779
	$^4\chi$	−4.443	1.480	0.793	JY	15.009	−6.072	0.771
	$^5\chi$	−2.408	1.472	0.788	M_1	−6.007	0.098	0.801
					M_2	−5.261	0.078	0.799
II	$^0\chi$	11.681	−0.510	0.469	W	8.312	−0.003	0.764
	$^1\chi$	14.742	−0.990	0.421	ϕ	14.674	−2.299	0.469
	$^2\chi$	13.346	−1.016	0.631	H	5.826	−0.032	0.190
	$^3\chi$	8.899	−0.692	0.331	JX	−3.305	4.108	0.553
	$^4\chi$	−1.411	0.736	0.294	JY	−2.923	3.729	0.531
	$^5\chi$	7.567	−0.926	0.316	M_1	11.999	−0.074	0.469
					M_2	8.442	−0.039	0.312
III	$^0\chi$	17.819	−0.803	0.955	W	6.977	−0.002	0.893
	$^1\chi$	25.534	−1.774	0.952	ϕ	24.965	−3.799	0.958

	$^2\chi$	15.283	−1.123	0.950	H	15.311	−0.134	0.800
	$^3\chi$	17.040	−1.537	0.963	JX	−2.322	3.510	0.597
	$^4\chi$	11.395	−1.146	0.531	JY	−2.277	3.328	0.571
	$^5\chi$	8.063	−0.958	0.289	M_1	18.217	−0.117	0.956
					M_2	17.649	−0.098	0.937
IV	$^0\chi$	−6.756	0.935	0.815	W	−0.167	0.008	0.837
	$^1\chi$	−5.610	1.183	0.850	ϕ	−7.088	4.223	0.347
	$^2\chi$	−4.598	1.187	0.819	H	−2.610	0.131	0.849
	$^3\chi$	−3.610	1.240	0.851	JX	19.092	−7.745	0.809
	$^4\chi$	−2.362	1.281	0.872	JY	17.747	−6.866	0.796
	$^5\chi$	−1.078	1.334	0.843	M_1	−3.894	0.090	0.844
					M_2	−3.265	0.071	0.850
V	$^0\chi$	20.567	−0.948	0.633	W	14.818	−0.006	0.685
	$^1\chi$	−18.541	1.735	0.468	ϕ	4.538	−0.200	0.164
	$^2\chi$	1.218	0.200	0.077	H	−25.753	0.321	0.673
	$^3\chi$	2.690	0.085	0.039	JX	8.769	−3.766	0.695
	$^4\chi$	−0.673	0.575	0.434	JY	8.179	−3.193	0.617
	$^5\chi$	−1.348	0.930	0.490	M_1	−24.706	0.210	0.941
					M_2	−14.061	0.113	0.922

represent satisfactory models for some group compounds, but not for all of the groups. It could be seen that some order of the connectivity indices ($^0\chi$, $^1\chi$), Wiener index (W), Harary index (H), Balaban index (J), and Zagreb index (M) had good correlation with the retention factors. For example, Figure 2 shows the retention factor and the zero order connectivity index of each group. The retention factor of group I and group IV increased with a increases in connectivity index ($^0\chi$), but the groups II, III, and V showed the opposite tendency. Figure 3 showed another example of the interrelationships of the retention factor and the Wiener index of each group. In this case, this plot also has the same tendency as Figure 2.

However, some simple linear correlations in Table 3 gave a very low regression coefficients. The descriptors calculated in the paper express the wide spectrum of the values of the molecular properties. All these properties have high values because each of them influences, to larger or smaller degree, the chromatographic retention. The chromatographic retention results from the solvation and partition of the individual compounds in a stationary phase. The retention is affected by the various intermolecular forces, which include hydrogen bonding, ion–dipole, dipole–dipole, specific, and dispersion (non-specific) interactions, etc.

The linear regression of the retention factors and different orders of the connectivity index with groups is shown in Table 4. The regression coefficients of

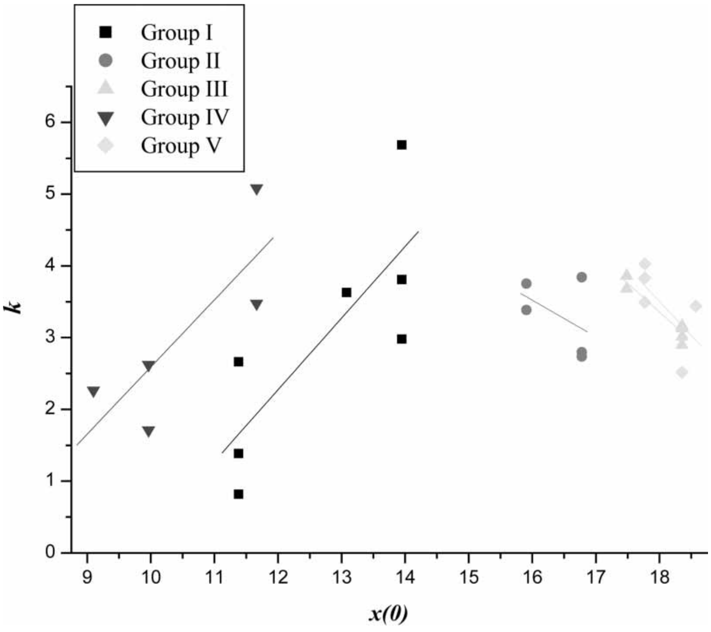


Figure 2. Simple linear regressions between experimental k and 0 order connectivity ($^0\chi$) for the five groups.

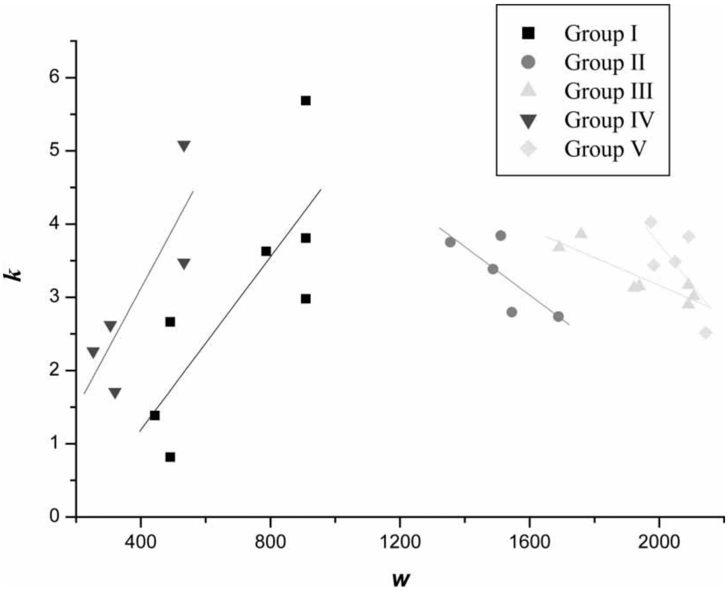


Figure 3. Simple linear regressions between experimental k and Wiener index (W) for the five groups.

the connectivity indices with retention factors were very high ($0.805 \sim 1$). For example, for group I, the regression equation is as shown below.

$$\begin{aligned} k = {} & -11.8 + 0.871\,{}^{0}\chi + 0.575\,{}^{1}\chi + 0.949\,{}^{2}\chi \\ & - 0.160\,{}^{3}\chi - 1.730\,{}^{4}\chi + 0.315\,{}^{5}\chi \end{aligned}$$

Table 4. Linear equations of the connectivity index (${}^{0}\chi \sim {}^{5}\chi$) with groups

	Equation: $k = a_0 + \Sigma a_i(P_i)$ Group				
	I	II	III	IV	V
a_0	−11.8	−765.850	2057.159	−1.110	−83.511
a_1	0.871	56.690	447.987	2.141	2.311
a_2	0.575	306.000	−550.625	0.860	0.525
a_3	0.949	−468.023	−223.962	−6.642	4.866
a_4	−0.16	239.360	−65.561	1.245	−6.341
a_5	−1.73	−333.567	−30.735	1.372	6.561
a_6	0.315	244.952	−11.267	1.842	−1.043
r^2	0.805	1	0.976	0.902	1

In our paper, we have attempted to consider entire spectrum of the calculated descriptors. Therefore, multiple linear statistical analysis was used in order to find another reliable expression which considers all descriptors calculated in the work. The equation, $k = a_0 + \Sigma a_i(P_i)$, was examined and the numeric coefficients were determined by linear regression. Multiple linear regressions of the retention factors with topological properties of nitrogen containing heterocyclic compounds and regression coefficients are listed in Table 5.

Thus, the calculated retention times are compared with the corresponding experimental values in Figure 4 and, as one of the examples, based on linear regression presented in Table 4. This plot clearly shows the relationships between observed retention times and calculated retention times. The obtained multi-linear correlations for observed and calculated retention factors are of good quality. The small errors of prediction make it possible for these models to be applied for the peak identification on a chromatogram, particularly for various retention times. It is probably possible to further improve our ability to identify the unknown azole compounds. Unfortunately, in these cases, the small data sets limit the ability to predict retention data, due to inevitable statistical irregularities. The retention time and structural correlations used in tandem will offer further significant progress in this regard.

Table 5. Multiple linear equations of the various structural descriptors with groups

	Equation: $k = a_0 + \Sigma a_i(P_i)$ Group				
	I	II	III	IV	V
a_0	46408	−43.627	−4.712	−28.636	−4.985
a_1	2008	−0.235	−5.019	−4.416	0.103
a_2	−1806	7.003	5.356	3.072	0.580
a_3	−8305	6.946	−13.186	12.976	−0.203
a_4	−1723	−6.338	5.272	12.407	−1.601
a_5	6938	1.126	−5.870	−6.017	−0.276
a_6	5877	−18.453	−4.795	−0.999	0.674
a_7	54.3	0.012	0.040	−0.034	−0.001
a_8	7.429	5.634	−7.108	−34.703	−0.563
a_9	−554	0.662	0.860	−0.718	−0.045
a_{10}	445.9	−8.233	31.001	−13.438	−10.843
a_{11}	1926	40.137	32.061	24.320	12.017
a_{12}	23.8	−0.664	0.693	−1.027	0.033
a_{13}	−494	−0.337	−0.785	0.247	0.103
r^2	1	1	1	1	1

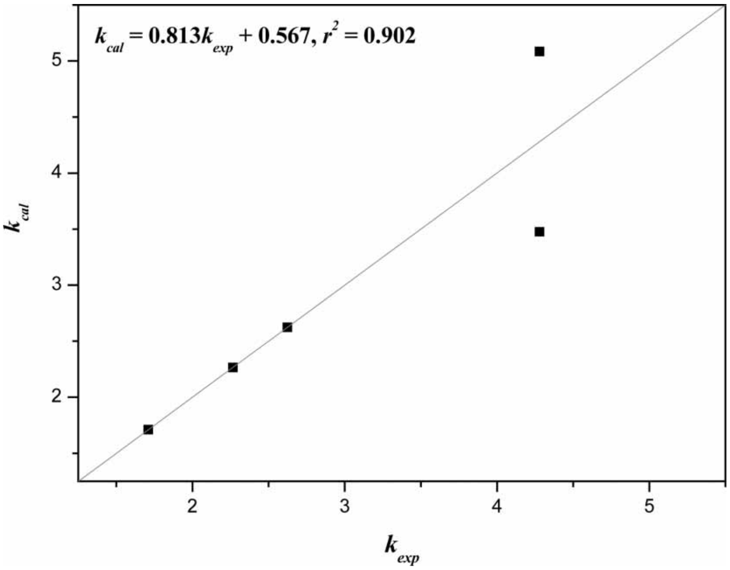


Figure 4. Comparison between the calculated retention times and the corresponding experimental values in group IV.

CONCLUSIONS

The retention factors of 29 nitrogen containing heterocyclic compounds and their topological properties were compared using multi-linear regression equations based on QSPR models. The obtained molecular topological models showed good correlations and were successfully utilized for prediction of the retention characteristics of the investigated nitrogen containing heterocyclic compounds. The resulting fourteen-parameter multiple linear regressions, including all of the descriptors for retention factors of nitrogen containing heterocyclic compounds, can be used to predict the chromatographic retentions of unknown substances and structurally similar compounds with a considerable degree of confidence. Also, this study clearly demonstrates that the QSPR models can be used to predict the retention times and indices without the need for chemical standards.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of the Center for Advanced Bioseparation Technology, Inha University.

REFERENCES

1. Kleeman, A.; Engel, J.; Kutscher, B.; Reichert, D. *Pharmaceutical Substances*, 3rd Ed.; Stuttgart: New York, 1999.
2. Nakano, H.; Inoue, T.; Kawasaki, N.; Miyataka, H.; Matsumoto, H.; Taguchi, T.; Inagaki, N.; Nagai, H.; Satoh, T. Synthesis and biological activities of novel antiallergic agents with 5-lipoxygenase inhibiting action. Bioorg. Med. Chem. **2000**, *8*, 373.
3. Delmas, F.; Timon-David, P. Functional derivatives of 5-benzo[1,3]dioxol-5-yl-1-methyl-1H-imidazole-2-carbaldehyde and evaluation of leishmanicidal activity. Eur. J. Med. Chem. **2000**, *35*, 157.
4. Kendi, E.; Stables, J.P. Synthesis of some 1-(2-naphthyl)-2-(imidazole-1-yl)ethanone oxime and oxime ether derivatives and their anticonvulsant and antimicrobial activities. Eur. J. Med. Chem. **2001**, *36*, 421.
5. Khabnadideh, S.; Rezaei, Z.; Khalafi-Nezhad, A.; Bahrinajafi, R.; Mohamadi, R.; Farrokhroz, A.A. Synthesis of N-alkylated derivatives of imidazole as antibacterial agents. Bioorg. Med. Chem. Lett. **2003**, *13*, 2863.
6. Antolini, M.; Boaaoli, A.; Ghiron, C.; Kennedy, G.; Rossi, T.; Ursini, A. Analogues of 4,5-bis(3,5-dichlorophenyl)-2-trifluoromethyl-1H-imidazole as potential antibacterial agents. Bioorg. Med. Chem. Lett. **1999**, *9*, 1023.
7. Feng, Y.Q.; Liu, M.C.; Hu, Z.C. Chin. J. Chromatogr. **1986**, *4*, 259.
8. Roca, R.M.S.; Albarez, F.J.G. Molecular topology and chromatographic retention parameters for benzodiazepines. J. Chromatogr. A. **1992**, *607*, 91.
9. Lee, S.K.; Lee, I.H.; Kim, H.J.; Chang, G.S.; Chung, J.E.; No, K.T. The 14th European Symposium on Quantitative Structure-Activity Relationships, 2002, Bournemouth, U.K.
10. Zhao, H.; Zhang, Q.; Chen, J.; Xue, X.; Liang, X. Prediction of octanol-air partition coefficients of semivolatile organic compounds based on molecular connectivity index. Chemosphere **2005**, *59*, 1421.
11. Bajaj, S.; Sambi, S.S.; Madan, A.K. Topological models for prediction of anti-HIV activity of acylthiocarbamates. Bioorg. Med. Chem. **2005**, *13*, 3263.
12. Lee, S.K.; Polyakova, Y.; Row, K.H. Interrelation of retention factor of amino-acids by QSPR and linear regression. Bull. Korean Chem. Soc. **2003**, *24* (12), 1757.
13. Lučić, B.; Miličević, A.; Nikolić, S.; Trinajstić, N. Harary index-twelve years later. C.C.A.C.A.A. **2002**, *75* (4), 847.
14. Thakur, A.; Thakur, M.; Khadikar, P.V.; Supuran, C.T.; Sudele, P. QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological approach using Balaban index. Bioorg. Med. Chem. **2004**, *12*, 789.
15. Bajaj, S.; Sambi, S.S.; Madan, A.K. Topological models for prediction of anti-inflammatory activity of N-arylanthranilic acids. Bioorg. Med. Chem. **2004**, *12*, 3695.

Received August 12, 2005
Accepted September 25, 2005
Manuscript 6704