

QSPR Models for Chromatographic Retention of Some Azoles with Physicochemical Properties

Yulia Polyakova, Long Mei Jin, and Kyung Ho Row*

Center for Advanced Bioseparation Technology and Dept. of Chem. Eng., Inha University, Incheon 402-751, Korea

*E-mail: rowkho@inha.ac.kr

Received August 3, 2005

This work deals with 24 substances composed of nitrogen-containing heterocycles. The relationships between the chromatographic retention factor (k) and those physicochemical properties which are relevant in quantitative structure-properties relationship (QSPR) studies, such as the polarizability (α), molar refractivity (MR), lipophilicity ($\log P$), dipole moment (μ), total energy (E_{tot}), heat of formation (ΔH_f), molecular surface area (S_M), and binding energy (E_b), were investigated. The accuracy of the simple linear regressions between the chromatographic retention and the descriptors for all of the compounds was satisfactory (correlation coefficient, $0.8 \leq r \leq 1.0$). The QSPR models of these nitrogen-containing heterocyclic compounds could be predicted with a multiple linear regression equation having the statistical index, $r = 1.000$. This work demonstrated the successful application of the multiple linear approaches through the development of accurate predictive equations for retention factors in liquid chromatography.

Key Words : Nitrogen-containing heterocycles, Retention factor, QSPR, Correlation

Introduction

The investigation of the quantitative structure-properties relationship (QSPR) of substances is an important issue in chromatographic science and medicinal chemistry, as well as in drug discovery. One of the most central areas of chemistry, QSPR provides information that is useful for organic chemistry, physical chemistry, molecular design and medicinal chemistry. This information is composed of mathematical equations relating the chemical structure of compounds to a wide variety of their physical, chemical, biological and technological properties. The derived relationships between molecular descriptors and activity are used to estimate the properties of other molecules and/or to find the parameters affecting their biological activity.¹

Over the past few years, different methods have been used to correlate chromatographic retention behavior with various molecular structure and physicochemical parameters, such as the Van der Waals and molecular volume,² the molecular connectivity index,^{3,4} the number of carbon atoms,⁵ the n -octanol/water partition coefficient ($\log P$),^{6,7} the hydrophobic substituent constant ($\log S$)^{8,9} and the solubility parameters.^{10,11} Correlations between chromatographic retention and molecular parameters provide significant information on the effect of the molecular structure on the retention time and on the possible mechanism of absorption and elution. These correlations can be investigated using QSPR, which provides statistical equations that relate the molecular structure to the retention phenomena.

The incorporation of the heterocyclic nuclei, imidazole and triazole, into various compounds therapeutic agents is the main synthetic strategy in drugs discovery.¹² The high therapeutic properties of the resulting drugs have encouraged medicinal chemists to synthesize large numbers of novel

chemotherapeutic agents. Imidazole and triazole drugs have a broad scope in remedying various disorders in clinical medicine.¹³ Pharmaceutical properties of the imidazole and benzimidazole derivatives include a wide spectrum of biological activities such as antifungal, antimycotic, antineoplastic, antiulcer, antihistaminic, antiallergic, anesthetic, hypnotic, antihypertensive, anthelmintic, neuroleptic, antipsychotic, and thromboxane synthetase inhibition. At present, there is a real perceived need for the discovery of new compounds endowed with biological activity.

In the present paper, a quantitative structure-properties relationship analysis was conducted to investigate the quantitative effect of the physicochemical properties of various nitrogen-containing molecules on their chromatographic retention. The retention factor of investigated substances was assumed to be governed by their descriptors such as polarizability (α), molar refractivity (MR), lipophilicity ($\log P$), dipole moment (μ), total energy (E_{tot}), heat of formation (ΔH_f), molecular surface area (S_M), and binding energy (E_b). The investigation of the linear and multiple linear regression equations between descriptors and retention factors was the purpose of this work.

Experimental Section

Reagents. Twenty-four 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-Aldrich Corp. (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. 5 mg of each of the twenty-four substances was dissolved in 1 mL of water or the 1 mL of

mixture (water/DMSO). A constant injection volume (5 μL) was used for the mixture solution throughout.

Apparatus and Methods. The assay 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 equipment was used with UV-Vis diode array detection (DAD), and evaporative light scattering detection (ELSD) and mass spectrometric (MS) detection. The liquid chromatographic separations were performed using a Betasil 10 μm octadecyl silica (C_{18}) column (4.6 \times 150 mm) (Phenomenex, USA). The column temperature was kept constant at 293 K. The mobile phase 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. An Agilent 1100 LC-MS system operated by Windows NT based HP ChemStation software (Rev 8.04) was used.

Theoretical Background

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

$$k = (t_R - t_M)/t_M \quad (1)$$

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

Descriptors: The descriptors of the molecules were drawn using HyperChem software (Hypercube Inc. FL, USA, version 7) on a personal computer, using semi empirical AMI calculations for the determination of the molecular dynamics (293 K) of the molecules. Extending the empirical approach can produce methods with good accuracy and excellent computational performance. In this extended method, the calculation is generally sufficiently reliable and the results in good agreement with those obtained experimentally.

To obtain the quantitative effects of the structural parameters of the azole derivatives on their chromatographic retention, a QSPR analysis with the molecular descriptors, including their electronic and physicochemical adjectives, was operated.¹⁴ The polarizability (α), molar refractivity (MR), lipophilicity ($\log P$), dipole moment (μ), total energy (E_{tot}), heat of formation (ΔH_f), molecular surface area (S_M) and binding energy (E_b) are the properties that can be calculated for each molecule. Therefore, these data were used to determine the QSPR models. The values of the calculated descriptors and retention factors of the substances are summarized in Table 2.

Polarizability: The polarizability (α) of a molecule, which is an important physical property, is currently the focus of our research in the area of QSPR for chemical interactions. The electrons and nuclei of a molecule are mobile and free to move to a limited degree. Thus, small

charge displacements can take place in polar or non-polar (apolar) molecules in an electric field, and which results in a dipole being introduced into the molecule, in addition to the permanent one that may already exist. Thus, the polarizability (α) of a molecule is a measure of its overall electronic charge distribution that can be distorted by an external electric field. The polarizability is defined as:

$$\mu' = \mu + \alpha E + 1/2bE^2 + \dots \quad (2)$$

where μ is the permanent dipole moment of the molecule, μ' is the induced dipole moment of the molecule, and E is the external electric field.¹⁵

Molar Refractivity: The molar refractivity (MR) is one of the oldest and most successful additive-constitutive physico-chemical properties of a compound, which has a strong correlation with the molecular polarizability. Thus, the MR is known to be a measure of the polarizability and is calculated by the well-documented Lorentz-Lorenz equation:

$$MR = (n^2 - 1)/(n^2 + 2) MW/\rho - 4\pi N\alpha/3 \quad (3)$$

In this equation, n is the refractive index, MR is the molar refractivity, MW is the molecular weight, and ρ is the density of the compound. N is Avogadro's number, α is the polarizability and $\pi = 3.14$. The relationship between the molar refractivity and the number of electrons is well established and is represented by the following equation:

$$MR = (4Na/3)(sv_e^2/v_s^2 - v_i^2) \quad (4)$$

In this equation, v_e , v_s , and v_i are the frequencies of vibration of the electron, the electric oscillator, and light, respectively, and s is the number of dispersion electrons per molecule.

Lipophilicity: Lipophilicity was the first physicochemical property to be defined, and remains the only one for which methods of prediction has been developed and widely accepted in the field of pharmaceutical research. The n -octanol-water partition coefficient, or its common logarithm ($\log P$), has become the standard scale for lipophilicity, largely as the result of the work of Hansch and Leo.^{16,17} The interactions between water and a drug have been extensively studied and there are several ways to estimate this part of the problem. The simplest method is to use $\log P$, since Yalkowsky¹⁸ showed that $\log P$ can reasonably stand in for the activity coefficient in the overall solubility equation and provide an estimate of the strength of the interaction of the compound with water.

Dipole Moment: The mean square dipole moment (μ) is one of the fundamental conformation-dependent physical properties of a polar molecule.¹⁹ Even though the total charge on a molecule is zero, the nature of its chemical bonds is such that the positive and negative charges do not completely overlap in most molecules. Such molecules are said to be polar, because they possess a permanent dipole moment, which can be written as:

$$\mu = - \sum_{i=1}^{ooc} \int \phi_i^* r \phi_i dv + \sum_{a=1}^M Z_a R_a \quad (5)$$

where ϕ_i is the molecular orbital, \hat{r} is the electron position operator, Z_a is the a -th atomic nuclear charge, and R_a is the position vector of the a -th atomic nucleus.¹⁵

Total Energy: The total energy (E_{tot}) of a molecule is the sum of its different energy contributions. Thus, the total energy can be described as follows:

$$E_{tot} = E_{el} + \sum_{A \neq B} Z_A Z_B / R_{AB} \quad (6)$$

where E_{el} is the total electronic energy of the molecule, Z_A and Z_B are the nuclear charges of atoms A and B , respectively, and R_{AB} the distance between nuclei A and B .^{20,21}

Standard Heat of Formation: The fuel heat of formation ΔH_f is the amount of energy per kilomole of substance required to form the molecule in the state it is while in the fuel tank from its elemental molecules under standard conditions. A positive value of the fuel heat of formation means that energy is required to form the molecule and that the molecule so formed will release energy when reverting to standard state of its elemental components. A negative value means that energy is released when forming the molecule and that the molecule so formed will absorb energy when reverting to standard state of its elemental components. Generally, the higher the value, the more reactive or volatile the substance will be. From the point of view of theory, the heat of formation can be calculated as:

$$\Delta H_f = H_f - \sum H_f^A \quad (7)$$

where H_f is the calculated total energy of the molecule, and H_f^A is the calculated energies of isolated atoms, A .¹⁵

Surface Area: This descriptor (S_M) allows the molecular surface of a molecule to be calculated. The molecular surface resembles the Van der Waals surface of a molecule, except that crevices between atoms are smoothed over and interstices too small to accommodate the probe are eliminated. The S_M includes cavities in the interior of the molecule, even if they are not accessible to a solvent molecule coming from the outside. The surface area of a molecule is defined as:

$$S_M = \sum S_{VW}^{(i)} - S_{ov} \quad (8)$$

where $S_{VW}^{(i)}$ is the Van der Waals area of the i -th constituent atom of the molecule, and S_{ov} is the Van der Waals of any atoms inside overlapping atomic envelopes.^{22,23}

Binding Energy: The binding energy (E_b) of a molecule is the minimum energy required to dissociate it into its component atoms. To break the molecule apart, energy has to be supplied, usually in the form of heat. This energy represents the binding energy of the molecule.²⁴

B. Multiple Linear Regression. The mathematical foundation of the quantitative structure-property relationship is based on the principle of polylinearity.²⁵ Multiple linear regression is a common method used in QSPR studies. The QSRR equations were obtained by forward stepwise multiple regression techniques using the multilinear forms:

$$P = f(D) = a_0 + a_1 D_1 + a_2 D_2 + \dots + a_n D_n \quad (9)$$

where P is the property (in our case, P is the retention factor k), D_1 , D_2 and D_n are the descriptors and n is the number of descriptors. The intercept (a_0) and regression coefficients of the descriptors ($a_1, a_2 \dots a_n$) were determined using the least squares method. The statistical evaluation of the data was performed using the Origin program package.²⁶ To test the quality of the regression equation, the correlation coefficient (r) was utilized as the statistical parameter.

Results and Discussion

In this paper, the retention factors of the azoles listed in Table 1 were estimated using the mathematical relationships expressed as a linear relationship in terms of their physicochemical descriptors. We collect the retention times obtained from ELSD, DAD, and MS detectors for each substance. The display one of them (DAD) was taken for calculation of retention factors, because instrumentation indications from ELSD and MS were very close with DAD's data.

According to previous experience in QSPR studies,^{1,16,22} not all of the theoretical molecular structural descriptors are necessary to model the physicochemical properties (including chromatographic retention) being studied. To investigate the relationship between the chromatographic retention factor and the molecular structural descriptors, several sets of experiments were performed in gradient mode. Typical chromatograms of one of the azoles are shown in Figure 1.

The absence of strict homology in a sequence (set) of investigated substances poses a serious problem in the creation of a "structure-properties" model. In such cases, the linear correlation equation is not always applicable to the description of the chromatographic retention within the limit of a full nonhomologous set. From this point of view, the best results can be obtained by the formal division of all of the investigated substances into groups of substances having some common features.

In the present work, the principle of structural analogy was used for the division of the substances into groups. Molecules having similar structural fragments were classified into the same group. In this way, all of the investigated molecules were classified into 5 groups depending upon their chemical structures. The data in Table 2 served as the basis for statistical analysis.

Using QSPR methodology, the authors^{1,27-29} revealed relationships between the experimental parameter (k) and theoretical descriptors. QSPR studies have meant that, once a correlation between experimental and calculated property is found, it is possible to screen any number of compounds, including those not yet synthesized, on the computer for the selection of those structures with the desired properties. Usually, QSPR studies involve three components; these are a relevant description of molecular properties, informative data on properties, and correct and meaningful correlations.

To obtain the optimal QSPR model using the subset of the molecular structural descriptors the next followed procedures were realized in this work. Firstly, a linear mathematical analysis with all eight theoretical physicochemical

Table 1. Chemical structures of azoles

N	Structure	Name	N	Structure	Name
1		methyl 4-((1H-imidazol-4-yl)methyl)benzoate	13		2-((1H-imidazol-4-yl)methyl)-N'-(2-methoxybenzylidene)-benzohydrazide
2		methyl 4-((1H-1,2,4-triazol-1-yl)methyl)benzoate	14		2-((1H-imidazol-4-yl)methyl)-N'-(3-hydroxy-4-methoxybenzylidene)-benzohydrazide
3		methyl 4-((3H-benzimidazol-5-yl)methyl)benzoate	15		2-((1H-imidazol-4-yl)methyl)-N'-(4-hydroxy-3-methoxybenzylidene)-benzohydrazide
4		methyl 4-((3H-benzotriazol-5-yl)methyl)benzoate	16		2-((1H-imidazol-4-yl)methyl)-N'-(4-methoxybenzylidene)-benzohydrazide
5		4-((1H-imidazol-4-yl)methyl)benzohydrazide	17		4-((1H-imidazol-4-yl)methyl)-N'-(4-hydroxy-3-methoxybenzylidene)-benzohydrazide
6		2-((1H-imidazol-4-yl)methyl)benzohydrazide	18		4-((1H-imidazol-4-yl)methyl)-N'-(3-hydroxy-4-methoxybenzylidene)-benzohydrazide
7		4-((3H-benzimidazol-5-yl)methyl)benzohydrazide	19		4-((1H-imidazol-4-yl)methyl)phenol
8		4-((3H-benzotriazol-5-yl)methyl)benzohydrazide	20		2-((4-phenyl-1H-imidazol-1-yl)methyl)phenol

Table 1. Continued

N	Structure	Name	N	Structure	Name
9		4-((1H-imidazol-4-yl)methyl)-N'-benzylidenebenzohydrazide	21		2-((3H-benzimidazol-5-yl)methyl)phenol
10		2-((1H-imidazol-4-yl)methyl)-N'-benzylidenebenzohydrazide	22		2-((1H-imidazol-4-yl)methyl)phenol
11		4-((1H-imidazol-4-yl)methyl)-N'-(4-hydroxybenzylidene)benzohydrazide	23		4-((4-phenyl-1H-imidazol-1-yl)methyl)phenol
12		2-((1H-imidazol-4-yl)methyl)-N'-(4-hydroxybenzylidene)benzohydrazide	24		2-((2-methyl-1H-imidazol-1-yl)methyl)phenol

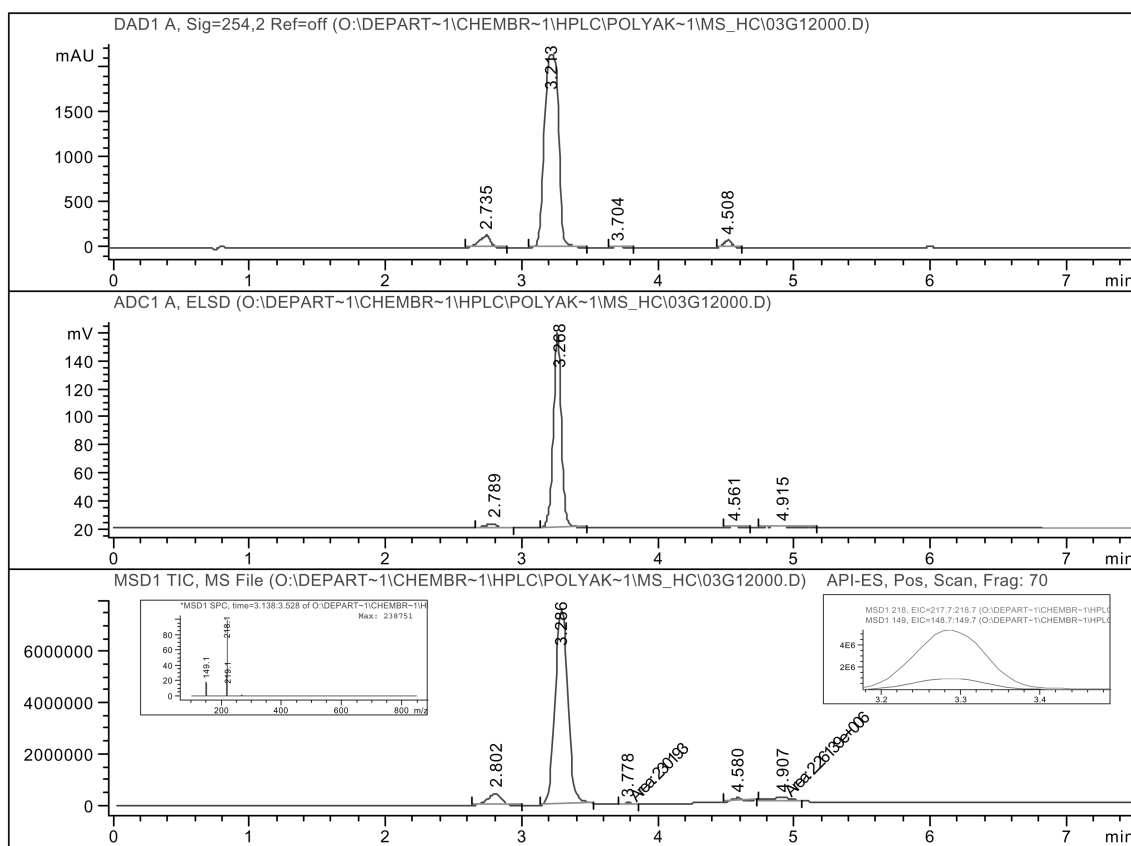


Figure 1. Chromatograms of azole N2 in Table 1 in gradient mode with ELSD, DAD, and MS detectors.

Table 2. The values of the retention factors (k) and descriptors of the azoles

No	k	MW	$\log P$	MR, A^3	α, A^3	S_M, A^3	$E_t, \text{kcal mol}^{-1}$	μ, D	$E_b, \text{kcal mol}^{-1}$	$\Delta H_f, \text{kcal mol}^{-1}$
Group I										
1	2.663	216	-0.60	65.32	23.49	429	-2991	4.61	-3016	5.99
2	3.526	217	0.23	62.82	22.78	426	-2839	2.76	-2862	48.61
3	3.810	266	-0.29	84.64	29.67	495	-3748	4.19	-3772	28.90
4	5.686	267	0.75	86.22	28.96	485	-3602	1.79	-3629	69.15
Group II										
5	0.818	216	-1.51	66.86	23.72	432	-2909	4.47	-2935	81.54
6	1.386	216	-1.51	66.86	23.72	411	-2907	5.18	-2932	84.11
7	1.724	266	-1.20	86.18	29.90	495	-3671	3.24	-3699	104.40
8	2.979	267	-0.16	87.76	29.19	482	-3526	2.89	-3553	140.37
Group III										
9	3.386	304	-0.16	94.65	34.79	583	-4246	3.68	-4284	137.09
10	3.752	304	-0.16	94.65	34.79	550	-4245	5.54	-4285	135.85
11	2.737	320	-1.19	96.25	35.43	609	-4348	3.84	-4386	93.99
12	2.797	320	-1.13	100.47	35.43	581	-4347	6.27	-4385	95.63
Group IV										
13	3.682	334	-1.10	105.24	37.27	613	-4512	4.57	-4544	211.82
14	3.150	350	-2.13	106.84	37.90	556	-4711	4.42	-4748	67.06
15	3.128	350	-2.13	106.84	37.90	622	-4719	2.60	-4760	55.05
16	3.860	334	-1.10	105.24	37.27	614	-4611	3.84	-4649	106.79
17	2.900	350	-2.13	106.84	37.90	621	-4714	1.85	-4753	62.13
18	3.014	350	-2.13	106.84	37.90	634	-4711	2.87	-4755	60.03
Group V										
19	1.520	174	-1.05	56.16	19.73	368	-2249	3.51	-2471	43.97
20	3.953	250	-0.04	85.06	29.39	475	-3634	4.48	-3665	84.14
21	3.476	224	-0.74	75.48	25.91	428	-3211	1.98	-3236	66.75
22	2.265	174	-1.05	56.16	19.37	362	-2449	3.88	-2468	46.85
23	3.628	250	-0.04	85.06	29.39	489	-3630	3.35	-3658	90.53
24	2.623	188	-0.53	59.34	21.56	388	-2717	4.44	-2743	47.27

molecular descriptors as predictor variables were performed (Table 2). These data served as a basis for further statistical analysis.

The linear regression equations were obtained for each of the groups and a summary of these models is given in Table 3. Some of the results had a high value of r , which indicate that these equations represent a satisfactory model for some of the calculated descriptors. It could be seen that $\log P$, MR and α were well correlated with the retention factor. As can be observed in this experiment, the linear models were adequate ($0.8 \leq r \leq 1.0$). Most of the regression equations showed that the retention was primarily influenced by the size, steric factors, and polar effects. A comparison between $\log P$ and the retention factors of the azoles using linear equations is shown in Figure 2. In all of the groups, the retention factor increased with increasing $\log P$. Thus, it is possible to assert that the investigated azoles are eluted in increasing order of their hydrophobicity.

The theoretical descriptors calculated in this study include a wide spectrum of molecular properties, which all have high values, because each of them influences the chromatographic retention to some extent.

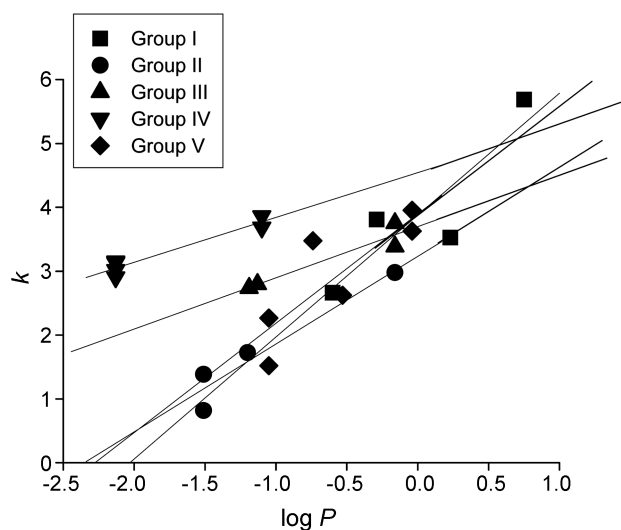
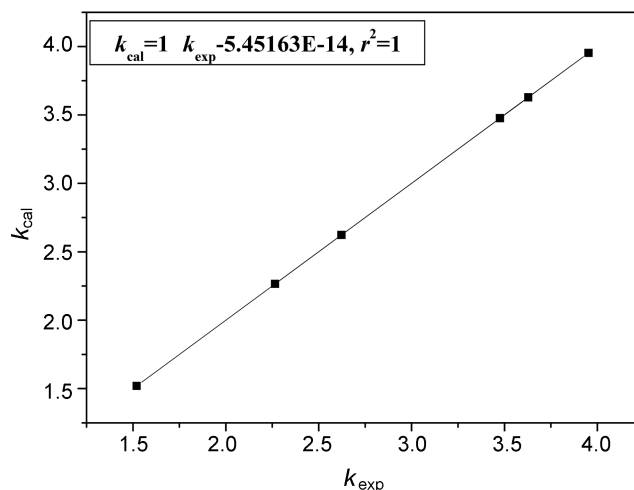
In this paper for QSPR studies involving descriptors, the physicochemical parameters are of utmost importance, as basic-model.²² Electronic effects, in contrast to global molecular properties such as lipophilicity ($\log P$), normally refer to a certain atom or group. Steric parameters describe and parameterize the molecules in terms of, for example, molecular surface area (S_M). They significantly influence physicochemical interactions in chromatography that involve weak bonds, in particular, electrostatic and dispersion interactions. These descriptors provide information about the fundamental properties of molecules that is not available by other means. Undoubtedly, these effects have been of interest to chromatographer and other chemists.

Chromatographic retention results from the solvation and partition of individual compounds in a stationary phase. The retention is affected by various intermolecular forces, which include hydrogen bonding, ion-dipole, dipole-dipole, specific and dispersion (nonspecific) interactions, etc.

A common practice employed in the development of QSPR methodology has been to identify relationships employing self-chosen single or more physicochemical descriptors. However, regardless of which type of variable is chosen,

Table 3. Linear mathematical analysis ($k = a + bD$) of the descriptors (D) for the five groups

Group	D	a	b	r
I	$\log P$	3.879	1.911	0.891
	MR	-1.876	0.078	0.755
	α	-2.346	0.239	0.674
	S_M	-6.813	0.024	0.668
	E_t	-1.870	-0.002	-0.618
	μ	6.694	-0.830	-0.848
	E_b	-1.917	-0.002	-0.619
	ΔH_f	2.317	0.04204	0.892
II	$\log P$	3.232	1.375	0.962
	MR	-3.226	0.065	0.819
	α	-3.597	0.200	0.738
	S_M	-4.879	0.015	0.635
	E_t	-3.422	-0.002	-0.698
	μ	4.280	-0.649	-0.772
	E_b	-3.45	-0.002	-0.697
	ΔH_f	-1.642	0.033	0.976
III	$\log P$	3.698	0.803	0.952
	MR	15.297	-0.126	-0.710
	α	47.166	-1.254	-0.951
	S_M	13.044	-0.017	-0.843
	E_t	37.035	0.008	0.953
	μ	3.110	0.012	0.032
	E_b	37.523	0.008	0.949
	ΔH_f	0.957	0.020	0.945
IV	$\log P$	4.544	0.702	0.963
	MR	51.327	-0.452	-0.963
	α	46.543	-1.148	-0.963
	S_M	3.586	-0.001	-0.035
	E_t	21.208	0.004	0.839
	μ	2.466	0.244	0.691
	E_b	20.534	0.004	0.840
	ΔH_f	2.854	0.005	0.730
V	$\log P$	3.896	1.712	0.842
	MR	-1.412	0.063	0.932
	α	-1.658	0.189	0.936
	S_M	-3.545	0.016	0.904
	E_t	-1.612	-0.002	-0.973
	μ	3.320	-0.113	-0.112
	E_b	-1.898	-0.002	-0.944
	ΔH_f	0.328	0.041	0.895

**Figure 2.** Simple linear regressions between experimental k and $\log P$ for the five groups (k was calculated from DAD data).**Figure 3.** Comparison between the calculated retention factors from DAD data and the corresponding experimental values in group V.

it is usually difficult to predict in advance the descriptor variables of value in describing the properties variation. Understandably, it is very difficult, or even impossible, to handle a large number of descriptors while aiming to produce an unambiguous and understandable selection. The latter fact is one of the disadvantages of QSPR. As a result, in this study, we attempted to consider the entire spectrum of

Table 4. The multiple linear models $k = f(D) = a_0 + a_1 \log P + a_2 MR + a_3 \alpha + a_4 S_M + a_5 E_t + a_6 \mu + a_7 E_b + a_8 \Delta H_f$ containing eight descriptors

Group	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	r
I	8.525	-6.006	0.084	-0.292	0.003	0.003	-3.461	-0.005	-0.005	1.000
II	-2.654	3.293	-0.043	0.130	-0.001	-0.002	0.924	-0.001	-0.037	1.000
III	-1.837	7.338	-0.099	0.862	0.003	-0.001	0.143	0.001	-0.147	1.000
IV	-287.47	3.041	-3.768	0.997	0.022	-0.151	0.560	0.016	0.122	1.000
V	-6.607	-1.500	-0.066	-0.375	-0.004	-0.003	0.147	-0.006	-0.009	1.000

the calculated descriptors. To overcome the difficulty, however, software program like HyperChem have been used that can calculate the standard descriptor values. However, some of the simple linear correlations shown in Table 3 gave a low (inadequate) regression coefficient, r , of less than 0.80.

Therefore, the multiple linear statistical analysis technique described by Eq. (9) was used, in order to obtain a reliable expression, which considers all of the descriptors calculated in this work. Eq. (9) was examined and the numeric coefficients were determined by linear regression. The multiple linear regressions of the retention factors with the physico-chemical properties of the azoles and their regression coefficients are listed in Table 4. The result of chromatographic retention was estimated with calculated values of regression coefficients, and the result showed clearly that each obtained descriptor with different molecular properties has different effect. It is not difficult to see that coefficients a_4 , a_5 , and a_7 are small and close to zero. This testifies, that the physico-chemical parameters such as surface area, total energy, and binding energy weakly influence the retention of the investigated sorbates. However, even such low-informative descriptors should be considered also, since their use unambiguously increases the predictive ability of the equations. The obtained multiple linear equations showed that they possess the best predictive ability, in comparison with the linear equations.

The calculated retention times are compared with the corresponding experimental values in Figure 3. This plot clearly shows the relationships between the observed and calculated retention times. The multi-linear correlations obtained for the observed and calculated retention factors are of excellent quality.

The advantages of approach, used in this paper, are that they are not restricted to closely related compounds, they can be easily obtained, and describe clearly defined molecular properties. Because of their advantages, they are widely used in the QSAR studies. The small errors of prediction make it possible for these models to be applied to the problem of peak identification in a chromatogram, particularly in the case where multiple retention time are observed. It is undoubtedly possible to further improve the ability of this technique to identify unknown azole's compounds. Unfortunately, in the present study, the small data sets limited our ability to predict the retention data, due to the inevitable statistical irregularities. Another approach to identifying unknown azoles peaks in a chromatogram is to use a similar QSPR approach to correlate the structural properties of various azoles compounds. The use of the retention times and structural correlations should offer the possibility of making further significant progress in this regard.

Conclusion

The retention times of 24 azoles were modeled using multi-linear models based on the calculated physico-

chemical descriptors. The obtained models showed perfect correlation and predictive ability. The resulting eight-parameter multiple linear regressions for the retention times of the azoles can be used to predict the retention times and indices of unknown but structurally similar compounds with a considerable degree of confidence. Also, this study clearly demonstrates that 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 Center for Advanced Bioseparation Technology, Inha University. Also professor Angela Bulanova of Samara State University, Russia gave a valuable technical advice for this article.

References

1. Kaliszan, R. *Quantitative Structure-Chromatographic Retention Relationships*; Wiley: New York, 1987.
2. Feng, Y.; Zhu, P.; Hu, Z. *J. Chromatogr.* **1988**, 25, 382.
3. Feng, Y. Q.; Liu, M. C.; Hu, Z. *Chin. J. Chromatogr.* **1986**, 4, 259.
4. Roca, R. M. S.; Alvarez, F. J. G. *J. Chromatogr.* **1992**, 607, 91.
5. Hinze, W. L.; Weber, S. G. *Anal. Chem.* **1991**, 63, 1808.
6. Yamagami, C.; Oguda, T.; Takao, N. *J. Chromatogr.* **1990**, 514, 123.
7. Kaibara, A.; Hohda, C.; Hirata, N.; Hirose, M.; Nakagawa, T. *Chromatographia* **1990**, 29, 275.
8. Smith, R. M. *J. Chromatogr. A* **1993**, 656, 381.
9. Law, B.; Weir, S. *J. Chromatogr. A* **1993**, 657, 17.
10. Smith, P. L.; Cooper, W. T. *Chromatographia* **1988**, 25, 55.
11. Billiet, H. A. H.; Schoenmakers, P. J.; Galan, L. D. *J. Chromatogr.* **1981**, 218, 443.
12. Townsend, L. B. *Chem. Rev.* **1976**, 67, 533.
13. Kleeman, A.; Engel, J.; Kutscher, B.; Reichert, D. *Pharmaceutical Substances*, 3rd ed.; Stuttgart: New York, 1999.
14. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim (Germany), 2000.
15. Atkins, P. W. *Quanta*; Oxford University Press: Oxford, 1991.
16. Hansch, C.; Leo, A. J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley: New York, 1979.
17. Leo, A. J. *Chem. Rev.* **1993**, 93, 1281.
18. Sanghvi, T. *QSAR & Combinatorial Science* **2003**, 22, 258.
19. Diaz-Calleja, R.; Riande, E., In *Dielectric Spectroscopy of Polymeric Materials*; Runt, J. P.; Fitzgerald, J. J., Eds.; American Chemical Society: Washington, DC, 1997; pp 139-173.
20. Csizmadia, I. G. *Theory and Practice of MO Calculations on Organic Molecules*; Elsevier: Amsterdam, 1976.
21. Bodor, M.; Gabanyi, Z.; Wong, C.-K. *J. Am. Chem. Soc.* **1989**, 111, 3783.
22. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; J. Wiley & Sons: New York, 2000.
23. Wang, R.; Gao, Y.; Lai, L. *Perspect. Drug Discov. Des.* **2000**, 19, 47.
24. Bruket, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monograph, American Chemical Society: Washington, 1982.
25. Palm, V. A. *Fundamentals in Quantitative Theory of Organic Chemistry*; Khimiya: Russia (in Russian), 1967.
26. Kubinyi, H.; In *QSAR: Hansch Analysis and Related Approaches*; Mannhold, R.; Krogs Gaard-Larsen, P.; Timmerman, H., Eds.; VCH: Weinheim, 1993.
27. Gami-Yilinkou, G.; Kaliszan, R. *J. Chromatogr.* **1991**, 550, 573.
28. Dai, J.; Jin, L.; Yao, S.; Wang, L. *Chemosphere* **2001**, 42(8), 899.
29. Breneman, C. M.; Rhem, M. *J. Comput. Chem.* **1997**, 18, 182.