

Article

Theoretical Models and QSRR in Retention Modeling of Eight Aminopyridines

Anja Tumpa¹, Marko Kalinić², Predrag Jovanović³, Slavica Erić², Tijana Rakić¹, Biljana Jančić-Stojanović^{1,*}, and Mirjana Medenica⁴

¹Department of Drug Analysis, Faculty of Pharmacy, University of Belgrade, Vojvode Stepe 450, Belgrade 11000, Serbia, ²Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Belgrade, Vojvode Stepe 450, Belgrade 11000, Serbia, ³Department of Organic Chemistry, Faculty of Pharmacy, University of Belgrade, Vojvode Stepe 450, Belgrade 11000, Serbia, and ⁴Department of Physical Chemistry and Instrumental Methods, Faculty of Pharmacy, University of Belgrade, Vojvode Stepe 450, Belgrade 11000, Serbia

*Author to whom correspondence should be addressed. Email: jancic.stojanovic@pharmacy.bg.ac.rs

Received 22 July 2015; Revised 14 September 2015

Abstract

In this article, retention modeling of eight aminopyridines (synthesized and characterized at the Faculty of Pharmacy) in reversed-phase high performance liquid chromatography (RP-HPLC) was performed. No data related to their retention in the RP-HPLC system were found. Knowing that, it was recognized as very important to describe their retention behavior. The influences of pH of the mobile phase and the organic modifier content on the retention factors were investigated. Two theoretical models for the dependence of retention factor of organic modifier content were tested. Then, the most reliable and accurate prediction of $\log k$ was created, testing multiple linear regression model—quantitative structure–retention relationships (MLR-QSRR) and support vector regression machine—quantitative structure–retention relationships (SVM-QSRR). Initially, 400 descriptors were calculated, but four of them (P_{OM} , $\log D$, $M-S_{ZX}/R_{ZX}$ and $m-RPCG$) were included in the models. SVM-QSRR performed significantly better than the MLR model. Apart from aminopyridines, four structurally similar substances (indapamide, gliclazide, sulfamethoxazole and furosemide) were followed in the same chromatographic system. They were used as external validation set for the QSRR model (it performed well within its applicability domain, which was defined using a bounding box approach). After having described retention of eight aminopyridines with both theoretical and QSRR models, further investigations in this field can be conducted.

Introduction

At an early stage of drug discovery, a vast number of new chemical substances pass through high-throughput screening to reveal their properties. In pharmaceutical industry, getting as many information as possible from relatively small set of substances is of great importance. As for retention behavior investigation, reversed-phase high performance liquid chromatography (RP-HPLC) is the most widely used technique. Even though it is well described, retention mechanism is not yet fully understood. Therefore, there is always a need for investigation of newly synthesized entities in order to describe their retention behavior. In this article, retention behavior of eight aminopyridines is

investigated for the first time. All aminopyridines have been synthesized in the laboratory at the Faculty of Pharmacy. Two of the substances (PJ47 and PJ51) represent completely new entities (never been synthesized before), while six others are already known (PJ11, PJ13, PJ15, PJ44, PJ45 and PJ46). However, retention behavior of all eight of them is for the first time investigated in this article. Among studied aminopyridine structures, there were benzene sulfonamide, benzene carboxamide and *N*-phenylurea derivatives, and their chemical structures are presented in Figure 1.

Aminopyridines as pharmacophores are present in many biologically active compounds or aminopyridines themselves show a various

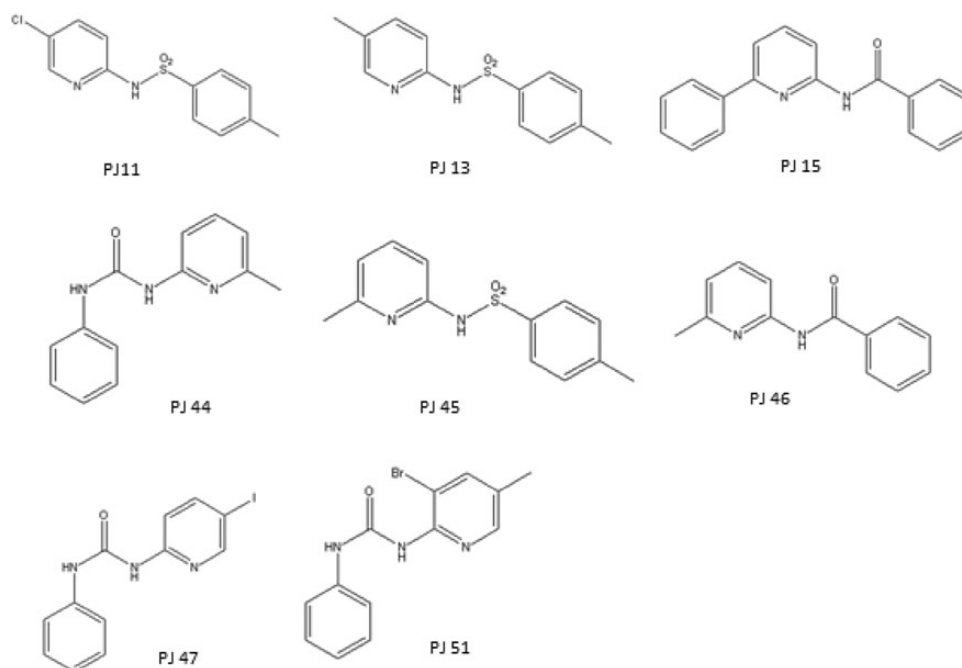


Figure 1. Structures of the newly synthesized substances.

physiological properties. They are capable of modulating the activity of various kinases most likely by mimicking their natural ligand, adenosine triphosphate (ATP). As planar compounds with potential to develop hydrogen bonding interactions which resemble ATP, they represent a good starting point to design novel ATP competitive inhibitors.

Taking into account ionizable characteristics of the substances, and the fact that they have never been investigated in RP-HPLC system before, it was important to establish relationships between retention and mobile phase properties. In the literature, there are several articles dealing with retention behavior of ionizable substances in RP-HPLC systems (1–4). They point out the importance of describing the type of dependence between retention factor and pH of the mobile phase, as well as the concentration of the organic modifier. In the first part of this report, several theoretical models that explain these dependences were tested, while in the second part, a quantitative structure–retention relationships (QSRR) model was created. It provided insights into the relationship between retention behavior of substances and their molecular properties.

A number of models have been proposed for predicting the retention behavior of ionizable solutes at varying mobile phase compositions (5). One of the earliest approaches of this kind suggested introducing additional terms into the linear solvation energy relationship (LSER), which would account for either the apparent solute pK_a under the conditions studied (6), or the degree of ionization combined with an additional retention-derived factor, which is specific for a given chromatographic system (7). As an alternative to models requiring several empirical parameters, QSRR models developed using computational molecular descriptors present a convenient way of estimating retention times and can also provide mechanistic insights into the retention mechanism in the chromatographic system under study (8). In a conventional QSRR approach, prior to descriptor calculation, the major microspecies at pH of interest are identified and optimized. However, if the pH is <2 units apart from the pK_a value, the distribution of microspecies may be such that the solute is represented in several forms in approximately equal amounts. The choice of a single microspecies to use in subsequent QSRR analysis may therefore be

highly arbitrary, and given that values of most descriptors deviate considerably depending on the molecule's formal charge, the quality of the derived models could be significantly affected. To address this issue, in this study we attempted to take into account the structural properties of both the major microspecies at a given pH and the next most dominant microspecies. By using this approach, we were able to derive a simple, interpretable QSRR model which is applicable in a range of pH values and mobile phase compositions, to solutes containing both basic and acidic ionizable groups. After the model was established and its applicability domain defined, external validation was carried out using retention data for four additional compounds: furosemide, indapamide (benzene sulfonamide, benzene carboxamide), gliclazide (benzene sulfonamide, urea) and sulfamethoxazole (benzene sulfonamide).

Because there are no data in the literature dealing with retention behavior of aminopyridines and they have vast biological potential, it was recognized as very important to describe retention behavior of these potentially active pharmaceutical compounds and provide deep insight into their behavior in reverse phase system. That means that for the first time, some theoretical models were investigated, and relationships between retention and mobile phase properties were established. Furthermore, by creating a simple QSRR model with good predictive capabilities, further investigations in this field are much facilitated, because for any novel, chemically similar substance (that fits into applicability domain) this QSRR model can be used in order to predict retention behavior and eventually reduce time needed for method development.

Methodology

Synthesis of eight aminopyridines

Synthesis of benzamide derivatives (PJ15 and PJ46)

Compound PJ15 (9) was synthesized by the Buchwald–Hartwig amidation of 2-chloro-6-phenylpyridine with benzamide in dioxane using

$\text{Pd}(\text{OAc})_2$, Xantphos as a ligand and K_2CO_3 as a base (Scheme 1). The product was isolated in good yield (71%).

Related amide PJ46 (10) was obtained via a standard procedure employing 6-methylpyridin-2-amine and benzoyl chloride in the presence of pyridine as a base (Scheme 2). The amide was isolated after column chromatography in 66% yield.

Synthesis of benzenesulfonamide derivatives (PJ11, PJ13 and PJ45)

Compounds PJ11 (11), PJ13 (12) and PJ45 (13) were synthesized by the condensation reaction of *p*-toluene sulfonyl chloride with different substituted amino pyridines in the presence of pyridine in dioxane as solvent (Scheme 3). All products were isolated in moderate yield.

Synthesis of urea derivatives (PJ44, PJ47 and PJ51)

Urea derivatives PJ44 (14), PJ47 and PJ51 were prepared by a reaction of amino pyridines and phenyl carbamate in tetrahydrofuran (THF) in the presence of 4-dimethylaminopyridine (DMAP) (Scheme 4). The products were isolated in 52–77% yields.

Influence of pH and solvent composition on retention factors of analyzed substances

Influence of pH

Horvath *et al.* (15) first described sigmoidal dependence between retention factor $\log k$ and pH in a specific RP-HPLC column. That function was later verified experimentally in several articles (2, 16–18). The following equation shows the relationship between retention factor of a monoprotic solute HA and the mobile phase pH:

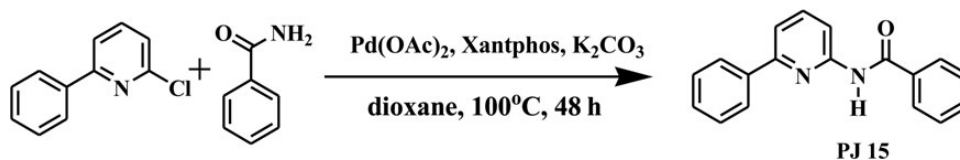
$k = (k_{\text{HA}} + k_{\text{A}} \times 10^{(\text{pH}-\text{p}k_{\text{A}})}) / (1 + 10^{(\text{pH}-\text{p}k_{\text{A}})})$, where k_{HA} and k_{A} stand for retention factor of protonated and deprotonated substance, respectively.

Influence of solvent composition

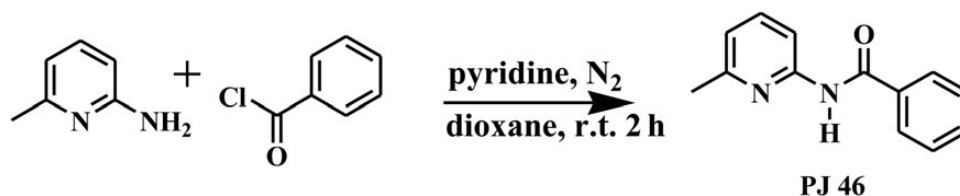
There are several models in the literature describing influence of organic solvent concentration on retention behavior of a substance (19). Usually, it is suspected that the dependence between retention factor and organic modifier concentration is linear: $\log k = \log k_w - S\varphi$, where $\log k_w$ is the intercept (hypothetical retention factor in the mobile phase consisting solely from water), S is the slope of the equation and φ is concentration of the organic modifier (20).

But it was proved several times that this relationship is linear only in very narrow range of organic modifiers. The more complex nonlinear model proposed in the literature is called “polarity parameter model” (21–24), which creates relationship between $\log k$ and the mobile phase parameter (P_m^N), a solute parameter (p) and two chromatographic system parameters [P_s^N and $(\log k)_0$]. This model is supposed to show better correlation than the previous one (25), because the relationship between organic modifier and polarity parameter is not linear, which is why it is included in our article.

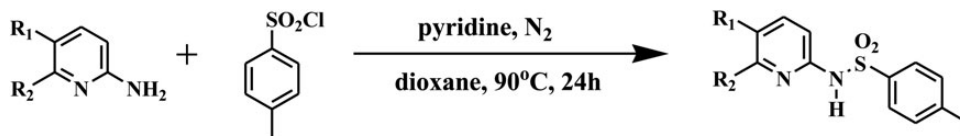
This model is based on the equation: $\log k = (\log k)_0 + p(P_m^N - P_s^N)$, where $(\log k)_0$ is the retention factor of a solute eluting from a hypothetical mobile phase which has the same polarity as the stationary phase ($P_m^N = P_s^N$). P_m^N and P_s^N are normalized polarity parameters for the mobile phase and the stationary phase, respectively. A variation of this model was proposed (25). In it, the solute is characterized by two parameters (p and q), according to the following



Scheme 1. Coupling of 2-chloro-6-phenylpyridine with benzamide catalyzed by $\text{Pd}(\text{OAc})_2$.



Scheme 2. Synthesis of benzamide PJ46 from 2-aminopyridine.

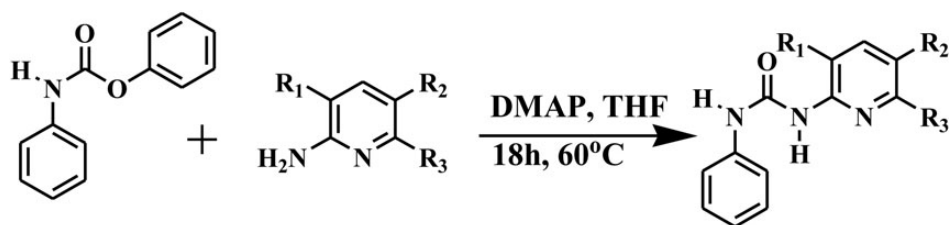


PJ11 $R_1\text{-Cl}$, $R_2\text{-H}$, 45%

PJ13 $R_1\text{-Me}$, $R_2\text{-H}$, 42%

PJ45 $R_1\text{-H}$, $R_2\text{-Me}$, 38%

Scheme 3. Reactions of substituted amino pyridines with *p*-toluene sulfonyl chloride.



PJ44 R₁-H, R₂-H, R₃-Me, 69%

PJ47 R₁-Me, R₂-I, R₃-H, 52 %

PJ51 R₁-Br, R₂-Me, R₃-H, 77%

Scheme 4. Synthesis of urea derivatives.

equation: $\log k = q + pP_m^N$. For the acetonitrile–water mobile phases, the polarity parameter of the mobile phase can be calculated from the following equation (25, 26): $P_m^N = 1.00 - 2.13\phi/(1 + 1.42\phi)$.

Computational methods

Structure optimization and descriptor calculation

Protonation states of each analyte were analyzed using the Calculator Plugin of MarvinSketch 6.0.3 (27), which provides predictions of microspecies distribution in a range of pH values. If the percentage of the major microspecies was <90% at any of the considered pH values, structures of both the major microspecies and the next most prevalent form of the analyte were considered in further calculations. After protonation state adjustment, a conformational search was conducted by using distance geometry constraints to generate 300 random conformers of each structure, as implemented in RDKit (28). Each conformer was optimized to energy convergence using the MMFF94s forcefield. The lowest energy conformer was then selected and further optimized in MOPAC 7 (29), using the AM1 Hamiltonian.

Optimized structures were used as input for calculating over 400 molecular descriptors in CODESSA (30). Values of all descriptors were exported from CODESSA, apparent distribution coefficient ($\log D$) predictions from MarvinSketch were added and further QSRR analysis and modeling were then carried out using NumPy (31), SciPy (32) and Scikit-learn (33) Python libraries. For each observation (i.e., $\log k$ measurement), descriptors of both relevant microspecies were treated as independent variables. All descriptor values were first range scaled to an interval between 0 and 1. The initial descriptor pool was reduced by: (i) removing descriptors with missing or all-zero values; (ii) removing descriptors whose coefficient of variation did not exceed 0.05; (iii) removing descriptors which had a coefficient of correlation to the dependent variable smaller than 0.1 and (iv) by removing highly correlated descriptors ($R > 0.8$); if several descriptors were highly inter-correlated, only the one with the best correlation to the dependent variable was kept. After this preprocessing step, 50 descriptors remained.

QSRR modeling

QSRR model development was based on 168 $\log k$ measurements for the eight aminopyridines. This set was divided into training ($n = 125$, $\approx 75\%$) and test sets ($n = 43$, $\approx 25\%$) using the k -means clustering algorithm, as implemented in SciPy. The number of clusters was set to 42, equaling the number of test set points. After clustering, the point

nearest the centroid of each cluster was placed into the test set, and all the remaining points were assigned to the training set.

To select the descriptors that would be used in the final model, an iterative multiple linear regression approach was used. Two independent variables were included in all descriptor combinations that were tested: percentage of cosolvent in the mobile phase and $\log D$ at pH corresponding to the mobile phase pH at which the measurement was made. First, all possible 3-parameter MLR models were established using training set data points. Ten top-ranked descriptor combinations, as evaluated by coefficient of determination (R^2) values, were retained and then combined with an additional descriptor to give all possible 4-parameter MLR models. This cycle was repeated up to five descriptors in a model. The final choice of descriptors was made by comparison of adjusted R^2 values from best 3-, 4- and 5-parameter models.

The final model was established using Scikit-learn's implementation of epsilon support vector regression with a radial basis function (RBF) kernel. Optimal choice of SVM hyperparameters was made by comparing 5-fold cross-validation root mean squared errors for models obtained by exhaustively varying the soft-margin constant (C) and the RBF gamma (34). C was varied on a logarithmic scale in the range between 10^{-2} and 10^5 , and gamma on a logarithmic scale from 10^{-5} to 10. Performance of the final model was validated using an external test set of substances.

Applicability domain definition

The applicability domain of the final model was defined using a bounding box approach (35). Ranges of descriptor values used in training the final model were used as reference. If any of the relevant descriptor values for analytes from the external test set deviated by more than 15% from the reference range, it was considered that the prediction is outside the model's applicability domain. Having said that, it is evident that the QSRR model created in this article could be used in further chromatographic investigations of aminopyridines or chemically similar substances that are within applicability domain, with the great certainty of good prediction of retention behavior.

Materials and methods

Physico-chemical characterization of synthesized compounds

The NMR spectra were recorded on a Varian Gemini 2000 (200 MHz) spectrometer. The chemical shifts are given in parts per

million (δ) downfield from tetramethylsilane as the internal standard, deuteriochloroform were used as solvents. The mass spectral data were recorded using an Agilent MSD TOF spectrometer coupled with Agilent 1200 HPLC or an Agilent Technologies 5975C MS coupled with Agilent Technologies 6890N GC. The IR spectra were recorded on an IR Thermo Scientific NICOLET iS10 (4950) spectrometer. Silica gel 60 (230–400) mesh was employed for the flash chromatography while thin layer chromatography was realized using alumina plates with 0.25 mm silica layer (Kieselgel 60 F254, Merck).

PJ15, *N*-(6-phenyl-2-pyridinyl)benzamide (9)

A mixture of 2-chloro-6-phenylpyridine (40 mg, 0.2 mmol), benzamide (31 mg, 0.25 mmol), Pd(OAc)₂ (5 mg, 0.02 mmol), Xantphos (22 mg, 0.04 mmol) and K₂CO₃ (41 mg, 0.3 mmol) in dioxane (10 mL) was refluxed under a nitrogen atmosphere for 48 h. After solvent evaporation, the residue was purified by flash column chromatography (SiO₂).

Flash chromatography (SiO₂, 8 : 2 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 75°C) in 71% yield. IR: 2159, 2030, 1977, 1678, 1567, 1441 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 8.73 (bs, 1H, NH), 8.36 (d, *J* = 7.8 Hz, 1 H), 7.98–7.91 (m, 4 H), 7.86–7.78 (t, *J* = 7.8 Hz, 1H), 7.61–7.37 (m, 7H). ¹³C NMR (CDCl₃, 50 MHz) δ 165.8, 156.0, 151.3, 139.2, 138.7, 134.3, 132.2, 129.1, 128.8, 128.7, 127.2, 126.8, 116.7, 112.4.

PJ46, *N*-(6-methyl-2-pyridinyl)benzamide (10)

To a mixture of 6-methylpyridin-2-amine (200 mg, 1.85 mmol) and pyridine (0.18 mL, 2.2 mmol) in 5 mL of dioxane was added benzoyl chloride (0.24 mL, 2.0 mmol) and the mixture was stirred under a nitrogen atmosphere for 2 h at room temperature. Dichloromethane (10 mL) was then added, and the resulting mixture was extracted with water (3 × 10 mL). The organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated *in vacuo*, the residue was purified by flash column chromatography (SiO₂).

Flash chromatography (SiO₂, 7 : 3 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 79–80°C) in 66% yield. IR: 2526, 2159, 1670, 1576, 1452, 1302 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 8.88 (bs, 1H, NH), 8.23 (d, *J* = 8.0 Hz, 1H), 7.96–7.90 (m, 2H), 7.67–7.59 (t, *J* = 7.8 Hz, 1H), 7.54–7.41 (m, 3H), 6.92 (d, *J* = 7.8 Hz, 1 H), 2.40 (s, 3 H). ¹³C NMR (CDCl₃, 50 MHz) δ 165.7, 156.8, 150.9, 138.7, 134.4, 132.0, 128.6, 127.2, 119.4, 111.0, 23.7.

General procedure for the synthesis of benzenesulfonamide derivatives

To a mixture of pyridin-2-amine (1 mmol) and pyridine (1.2 mmol) in 5 mL of dioxane was added tosyl chloride (1.2 mmol) and the mixture was stirred under a nitrogen atmosphere for 24 h at 90°C. Water (10 mL) was then added, and the resulting mixture was extracted with dichloromethane (3 × 10 mL). The organic layers were dried over anhydrous Na₂SO₄, filtered and concentrated *in vacuo*, and the residue was purified by flash column chromatography (SiO₂).

PJ11, 4-Methyl-*N*-(5-chloro-2-pyridinyl)-benzenesulfonamide (11)

Flash chromatography (SiO₂, 4 : 6 v/v, petroleum ether–ether) afforded the product as a white solid (m.p. 173°C) in 45% yield. IR: 2159, 2028, 1976, 1592, 1494, 1377 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 11.12 (bs, 1H, NH), 8.49 (d, *J* = 2.2 Hz, 1H), 7.72–7.65 (m, 2H), 7.61 (m, 1H), 7.46 (d, *J* = 9.0 Hz, 1H), 7.25 (d, *J* = 8.4 Hz, 2H), 2.39 (s, 3H). ¹³C NMR (CDCl₃, 50 MHz) δ 149.7, 147.0, 144.3, 139.1, 136.3, 129.9, 127.1, 126.9, 112.9, 21.5.

PJ13, 4-Methyl-*N*-(5-methyl-2-pyridinyl)-benzenesulfonamide (12)

Flash chromatography (SiO₂, 2 : 8 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 190–191°C) in 42% yield. IR: 2159, 2027, 1976, 1605, 1362, 1274 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 14.00 (bs, 1H, NH), 8.18 (s, 1H), 7.81 (d, *J* = 8.4 Hz, 2H), 7.52–7.47 (m, 1H), 7.36 (d, *J* = 9 Hz, 1H), 7.25 (d, *J* = 7.8 Hz, 2H), 2.37 (s, 3H), 2.19 (s, H). ¹³C NMR (CDCl₃, 50 MHz) δ 153.5, 143.8, 142.5, 139.1, 138.9, 129.4, 126.7, 123.9, 114.6, 21.3, 17.1.

PJ45, 4-Methyl-*N*-(6-methyl-2-pyridinyl)-benzenesulfonamide (13)

Flash chromatography (SiO₂, 2 : 8 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 107°C) in 38% yield. IR: 2159, 2027, 1977, 1607, 1360, 1260 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 11.00 (bs, 1H, NH), 7.84 (d, *J* = 8.0 Hz, 2H), 7.50 (m, 1 H), 7.26 (d, *J* = 8.4 Hz, 2 H), 7.17 (d, *J* = 8.4 Hz, 1 H), 6.63 (d, *J* = 7.4 Hz, 1 H), 2.47 (s, 3H), 2.38 (s, 3H). ¹³C NMR (CDCl₃, 50 MHz) δ 153.3, 152.2, 143.0, 140.8, 138.6, 129.4, 126.8, 115.2, 112.3, 21.7, 21.4.

General procedure for the synthesis of urea derivatives

A mixture of phenyl-*N*-phenyl carbamate (0.5 mmol), pyridin-2-amine (0.55 mmol) and DMAP (0.1 mmol) in THF (5 mL) was refluxed for 18 h. After solvent evaporation, the residue was purified by flash column chromatography (SiO₂).

PJ44, *N*-(6-methyl-2-pyridinyl)-*N'*-phenyl-urea (14)

Flash chromatography (SiO₂, 4 : 6 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 182–183°C) in 69% yield. IR: 2994, 2159, 2028, 1652, 1578, 1410 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 12.09 (bs, 1H, NH), 7.58–7.54 (m, 2 H), 7.43–7.18 (m, 3H), 7.12–6.98 (m, 1H), 6.81 (d, *J* = 8 Hz, 2H), 2.53 (s, 3H). ¹³C NMR (CDCl₃, 50 MHz) δ 154.9, 152.1, 138.8, 128.7, 123.2, 122.6, 119.8, 119.7, 119.1, 116.4, 108.8, 23.6.

PJ47, *N*-(5-iodo-3-methyl-2-pyridinyl)-*N'*-phenyl-urea

Flash chromatography (SiO₂, 1 : 1 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 207°C) in 52% yield. IR: 2432, 2159, 2028, 1668, 1500, 1411 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 11.79 (bs, 1H, NH), 8.33 (d, *J* = 1.6 Hz, 1H), 7.76 (m, 1H), 7.56–7.49 (m, 2H), 7.40–7.28 (m, 2H), 7.14–7.06 (m, 1H), 2.26 (s, 3H). ¹³C NMR (CDCl₃, 50 MHz) δ 196.0, 149.1, 146.6, 137.7, 128.8, 123.7, 122.0, 120.3, 120.2, 82.4, 16.4. HRMS (ESI): calculated for C₁₃H₁₂IN₃O (M + H)⁺ 354.00978, found 354.01007.

PJ51, *N*-(3-bromo-5-methyl-2-pyridinyl)-*N'*-phenyl-urea

Flash chromatography (SiO₂, 1 : 1 v/v petroleum ether–ether) afforded the product as a white solid (m.p. 137°C) in 77% yield. IR: 3291, 2972, 2159, 2027, 1677, 1595, 1557, 1479 cm⁻¹. ¹H NMR (CDCl₃, 200 MHz) δ 11.58 (bs, 1H, NH), 7.95 (s, 1H), 7.63 (s, 1H), 7.47–7.43 (m, 2H), 7.34–7.17 (m, 2H), 7.04–6.97 (m, 1H), 2.19 (s, 3H). ¹³C NMR (CDCl₃, 50 MHz) δ 224.5, 144.4, 142.3, 137.5, 128.7, 128.2, 123.6, 120.2, 120.1, 118.9, 106.2, 16.8. HRMS (ESI): calculated for C₁₃H₁₂BrN₃O (M + H)⁺ 306.02365, found 306.02353.

Chemicals

Working standards of furoseamide, indapamide, sulfamethoxazole and gliclazide were used. All reagents used were of an analytical grade. Acetonitrile–HPLC gradient grade (J. T. Baker, The Netherlands)

Table I. Comparison of Two Theoretical Models for the Influence of Organic Modifier on the Retention Behavior of the Substances

Substance	$\log k = \log k_w - S\varphi$			$\log k = q + pP_m^N$		
	$\log k_w$	S	R	Q	p	R
PJ11	2.39	-0.04	0.9832	49.96	103.27	0.9971
PJ13	1.33	-0.03	0.9824	35.70	74.62	0.9968
PJ15	3.37	-0.05	0.9768	62.01	127.29	0.9971
PJ44	1.73	-0.03	0.9879	38.66	80.19	0.9977
PJ45	1.04	-0.02	0.9850	32.11	67.48	0.9960
PJ46	1.15	-0.02	0.9793	28.10	58.50	0.9967
PJ47	3.15	-0.04	0.9797	56.38	115.54	0.9970
PJ51	2.87	-0.04	0.9809	52.04	103.75	0.9971
Sulfamethoxazole	1.03	-0.03	0.9913	35.87	75.68	0.9918
Furosemide	1.69	-0.04	0.9860	47.31	99.06	0.9967
Gliclazide	2.37	-0.04	0.9797	49.15	101.61	0.9862
Indapamide	1.81	-0.03	0.9848	46.12	96.22	0.9968

and water–HPLC grade were used to prepare a mobile phase. *Ortho*-phosphoric acid (J. T. Baker, The Netherlands) and 0.1 M sodium-hydroxide (Zorka Pharma, Sabac, Serbia) were used to adjust pH of the mobile phase. The prepared mobile phases were filtered through a Nylon membrane filter (0.45 μm Whatman, England).

Solutions

Stock solutions were prepared by dissolving each of the aminopyridines, as well as furosemide, indapamide, sulfamethoxazole and gliclazide in acetonitrile, to the concentration of 1 mg mL⁻¹. Furthermore, the stock solutions were diluted up to the concentration of 100 $\mu\text{g mL}^{-1}$ in the adequate mobile phase to obtain solutions that underwent the analysis.

Chromatographic conditions

The chromatographic system Waters Breeze consisted of a Waters 1525 Binary HPLC Pump, a Waters 2487 UV/VIS dual absorbance detector and Breeze Software Windows XP for data collection. Separations were performed under RP-HPLC mode on a ZORBAX Extend-C18 column (150 mm \times 4.6 mm, 5 μm particle size), Agilent, USA. The samples were introduced through a Rheodyne injector valve with a 20 μL sample loop. The flow rate was 1 mL min⁻¹ and the column temperature was 30°C. UV detection was performed at 254 nm.

The mobile phase consisted of acetonitrile and water, pH adjusted with *ortho*-phosphoric acid or 0.1 M NaOH. The content of acetonitrile in the mobile phase varied from 40 to 70%, with the increment of 5%. Another factor varied was the pH of the mobile phase, which was 3.0, 7.0 and 10.0. In total, 21 mobile phases were prepared for the analysis, according to the plan of experiments given in Supplementary Table SI.

Results

After running all the experiments, the retention factors k and $\log k$ values were calculated. The plan of experiments and obtained results for k and $\log k$ values are given in Supplementary Table SI. For the polarity parameter model, the P_m^N values also had to be calculated. Table I shows obtained coefficients as well as correlation coefficients for both models (linear solvation and polarity parameter) describing relationship between retention factor and acetonitrile content. After investigating theoretical models, QSRR study was performed. For that purpose, first the descriptors that would be used in final models had

Table II. Summary of the Final SVR Model Performance, with MLR Model Performance for Comparison

Dataset	SVR model RBF kernel, $C = 1,000$, $\gamma = 0.1$	MLR model
Training set ($n = 125$)		
R^2	0.97	0.92
RMSE	0.1054	0.1591
Test set ($n = 43$)		
R^2	0.94	0.88
RMSE	0.1382	0.1841
External test set ($n = 28$)		
R^2	0.93	0.77
RMSE	0.2032	0.2950

Table III. Four Descriptors Included in the Final QSRR Model, with the Corresponding MLR Coefficients

Descriptor symbol	Description	MLR coefficients
P_{OM}	Percentage of organic modifier in the mobile phase	-0.9724
$\log D$	Apparent octanol–water distribution coefficient	1.1362
$M-S_{ZX}/R_{ZX}$	ZX shadow/ZX rectangle (dominant microspecies)	0.5469
$m-RPCG$	Relative positive charge (second microspecies)	0.3345

to be chosen, out of many that were calculated at the beginning. The chosen descriptors and the corresponding MLR coefficients are given in Table III. At the end, several models were created and compared to find the one that has best performances. Table II shows comparison of SVR and the corresponding MLR model.

Discussion

Influence of pH and solvent composition on retention factors of analyzed substances

For the description of chromatographic behavior, wide ranges of investigated factors were taken into consideration. As for the pH of the mobile phase, the whole range that was acceptable considering

the column in use (ZORBAX Extend-C18) was investigated. For the acetonitrile content, it was important that the substances are retainable, but do not exhibit excessive retention times. After preliminary investigations, acetonitrile content from 40 to 70% was chosen.

Knowing the theoretically described dependence of retention factors of ionizable substances of pH of the mobile phase, it was the first factor to be investigated. Even though the investigated substances were all ionizable, most of them (PJ13, PJ15, PJ45, PJ47 and PJ51) had the same retention behavior under these chromatographic conditions (Figure 2A, the example of PJ51). On the other hand, PJ11, PJ44 and PJ46 did show different behavior under different pH values, which can be seen in Figure 2B (the example of PJ44). At pH 3.0, $\log k$ values of PJ44 and PJ46 are significantly different than those at pH 7.0 and 10.0. As for PJ11, their $\log k$ values differ at pH 10.0. This is in compliance with the characteristics of the analytes, namely their pK_a values. It can be explained with the fact that PJ11 exists mainly as non-ionized molecule at pH 3.0 and pH 7.0, but as anion at pH 10.0, which causes different retention. Unlike PJ11, PJ44 and PJ46 exist as neutral molecules at pH 7.0 and 10.0, but as cations at pH 3.0. This finding can be very helpful during the method development process.

For describing the relationship between $\log k$ and organic solvent concentration, two models were used (described in detail in 'Methodology'). The linearity of the dependence was proved, with the multiple R values above 0.97 for all 12 substances and both models (Table I). As for the comparison between these two models, higher linearity was discovered in a relationship between polarity parameter and retention factors, then between acetonitrile concentration and retention factor (for example on the substance PJ11, R for acetonitrile concentration and polarity parameter were 0.9832 and 0.9971, respectively). By

obtaining better correlation coefficients for the polarity parameter model, the non-linear relationship was once again proved in this article. Figure 2 shows dependence of $\log k$ of all 12 substances, of acetonitrile content, φ (C) and polarity parameter, P_m^N (D).

After these relations have been discussed, it is possible to make some conclusions which should help the future RP-HPLC analysis of these newly synthesized aminopyridines or some novel entities with similar structural characteristics. For example, after having calculated p and q values, the polarity parameter model can be used for predictions of retention behavior of structurally similar substances on the same chromatographic column.

In the next part of the study, the QSRR model was created. The model was created not only to describe retention behavior of aminopyridines as a function of their molecular characteristics, but also as a tool for prediction of retention behavior of similar structures as investigated substances.

QSRR modeling

The iterative MLR procedure suggested several descriptor combinations which have good predictive capabilities. Considering all the top-performing MLR models, 16 of 30 contained descriptors calculated for the second most dominant microspecies, suggesting that accounting for properties of all microspecies at a given pH has an important effect on the predictive capabilities of the resulting models. By comparing adjusted R^2 values and the frequency of occurrence of individual descriptors in the best performing MLR models, a 4-parameter model was chosen for further study. Although several 5-parameter models had slightly higher adjusted R^2 values (0.02–0.03 better), this study was confined to a simpler model, given that the number of structures

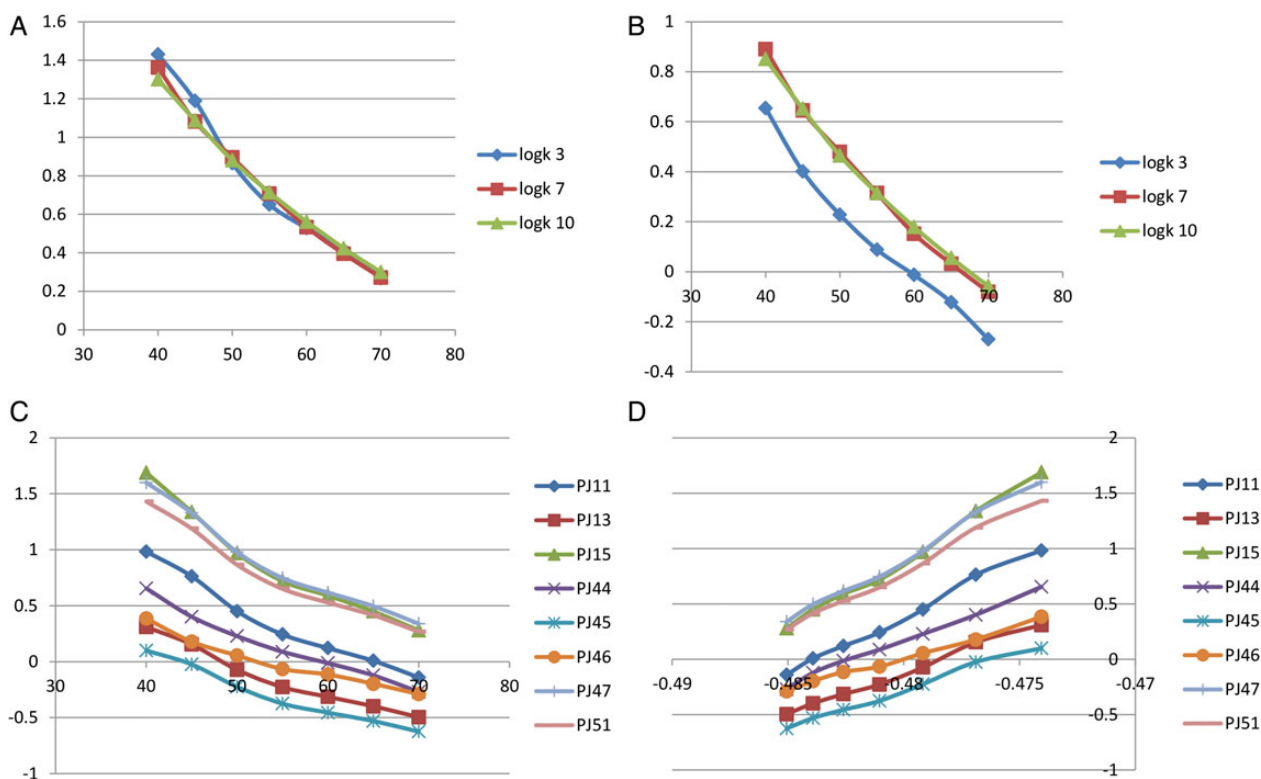


Figure 2. Plots of logarithm of retention factors of (A) PJ51 and (B) PJ44 as a function of φ at different pH values; logarithm k values for all eight PJ substances as a function of (C) acetonitrile content, φ , and (D) polarity parameter, P_m^N . This figure is available in black and white in print and in color at JCS online.

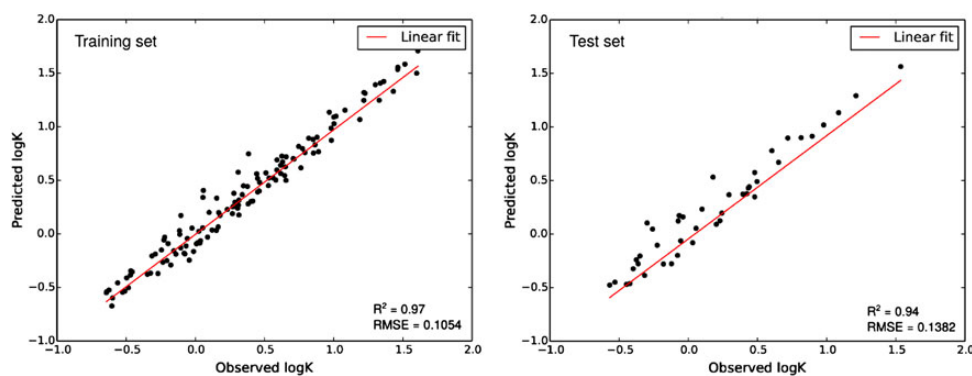


Figure 3. The predictive capabilities of the final SVR model. This figure is available in black and white in print and in color at *JCS* online.

used in model training was relatively small. The descriptors involved in the final models are summarized in Table III, along with the coefficients from the MLR model. These four descriptors account for both the properties of the mobile phase and those of the solutes under study (eight aminopyridines synthesized at the Faculty of Pharmacy). The significance of the percentage of organic modifier in the mobile phase and of the apparent distribution coefficient is readily appreciable in the context of retention mechanisms in RP-HPLC. Descriptor $M-S_{ZX}/R_{ZX}$ is a geometrical descriptor calculated as the surface ratio of a molecule's projection on the ZX plane and the smallest rectangle surrounding the projection. This descriptor accounts for the shape of the solutes, with its values increasing with increasing sphericity of a molecule. Consistent with the sign of the MLR coefficient for $M-S_{ZX}/R_{ZX}$, this suggests that irregularly shaped, asymmetrical molecules are less efficiently retained in the stationary phase, whereas molecules with a more spherical surface projection tend to be retained longer (e.g., PJ11 anion, $S_{ZX}/R_{ZX} = 0.6426$ vs. PJ47 neutral form, $S_{ZX}/R_{ZX} = 0.8146$). Finally, $m-RPCG$ descriptor is computed as the ratio of the maximal partially positive charge in the molecule and the sum of all partially positive atomic charges. Structures with a well-distributed positive charge would have smaller values of this descriptor and consequently, larger partially positive surface areas (e.g., PJ45 cation, $RPCG = 0.4919$ vs. PJ47 cation, $RPCG = 0.1285$). It is, therefore, likely that this descriptor accounts for the lower retention of solutes which have a resonantly spread charge which explains the positive sign of the corresponding MLR coefficient. The predictive capabilities of the final SVR model are summarized in Table II and are illustrated in Figure 3.

As can be seen, the model performs significantly better than the corresponding MLR model and has sustained predictive capabilities in the test set. Its performance was also validated against four structurally related drugs, specifically sulfamethoxazole, furosemide, gliclazide and indapamide. Of the 42 measurements that were available for these compounds, 28 were within the defined applicability domain of the 4-parameter SVR model. Within its applicability domain, the SVR model performs well. A larger and more diverse dataset would allow for further exploration of the possibility to establish a globally applicable QSRR model for retention of ionizable solutes with varying pH and mobile phase content.

Conclusion

The substances investigated in this article were eight aminopyridines, namely PJ11, PJ13, PJ15, PJ44, PJ45, PJ46, PJ47 and PJ51. They are potentially active pharmaceutical compounds, and their retention

behavior in the most commonly used RP-HPLC system was thoroughly investigated and described. The content of the organic modifier and pH of the mobile phase were the investigated factors. Being all ionizable substances, the pH of the mobile phase was expected to have big influence on the retention, but the obtained results showed relatively uniform behavior under these chromatographic conditions, except for three substances which was explained using their pK_a values. Dependence of the retention factor of polarity parameter was proven to be linear. The polarity parameter model better fitted the behavior according to obtained correlation coefficient (multiple R) values. After general conclusions about retention behavior of the substances were established, a simple SVR-QSRR model was created, using computational molecular descriptors. Its predictive abilities were tested using four external substances which were all within the applicability domain, and the model performed well. Therefore, this article not only addressed theoretical groundings of retention behavior in RP-HPLC, but also created the QSRR model which can be used for accurate prediction of retention behavior of the substances that are within the applicability domain.

Supplementary material

Supplementary materials are available at *Journal of Chromatographic Science* (<http://chromsci.oxfordjournals.org>).

Acknowledgments

The authors thank the Ministry of Education, Science and Technological Development of the Republic of Serbia for supporting these investigations in Projects 172052 and 172009.

References

- Canals, J., Portal, J.A., Bosch, E., Rosés, M.; Retention of ionizable compounds on HPLC. 4. Mobile-phase pH measurement in methanol/water; *Analytical Chemistry*, (2000); 72: 1802–1809.
- Espinosa, S., Bosch, E., Rosés, M.; Retention of ionizable compounds on HPLC. 5. pH scales and the retention of acids and bases with acetonitrile–water mobile phases; *Analytical Chemistry*, (2000); 72: 5193–5200.
- Rosés, M., Bosch, E.; Influence of mobile phase acid–base equilibria on the chromatographic behaviour of protolytic compounds; *Journal of Chromatography A*, (2002); 982: 1–30.
- Rosés, M.; Determination of the pH of binary mobile phases for reversed-phase liquid chromatography; *Journal of Chromatography A*, (2004); 1037: 283–298.

- Rosés, M., Subirats, X., Bosch, E.; Retention models for ionizable compounds in reversed-phase liquid chromatography: effect of variation of mobile phase composition and temperature; *Journal of Chromatography A*, (2009); 1216: 1756–1775.
- Bolliet, D., Poole, C.F., Rosés, M.; Conjoint prediction of the retention of neutral and ionic compounds (phenols) in reversed-phase liquid chromatography using the solvation parameter model; *Analytica Chimica Acta*, (1998); 368: 129–140.
- Rosés, M., Bolliet, D., Poole, C.F.; Comparison of solute descriptors for predicting retention of ionic compounds (phenols) in reversed-phase liquid chromatography using the solvation parameter model; *Journal of Chromatography A*, (1998); 829: 29–40.
- Héberger, K.; Quantitative structure–(chromatographic) retention relationships; *Journal of Chromatography A*, (2007); 1158: 273–305.
- Heitman, L.H., van Veldhoven, J.P.D., Zweemer, A.M., Ye, K., Brussee, J., Ijzerman, A.P.; False positives in a reporter gene assay: identification and synthesis of substituted *N*-pyridin-2-ylbenzamides as competitive inhibitors of firefly luciferase; *Journal of Medicinal Chemistry*, (2008); 51: 4724–4729.
- Sun, W.H., Zhang, W., Gao, T., Tang, X., Chen, L., Li, Y., et al.; Synthesis and characterization of *N*-(2-pyridyl)benzamide-based nickel complexes and their activity for ethylene oligomerization; *Journal of Organic Chemistry*, (2004); 69: 917–929.
- Hamdouchi, C., Sanchez, C., Ezquerro, Z.; Chemoselective arylsulfenylation of 2-aminoimidazo[1,2-*a*]pyridines by phenyliodine(III) bis(trifluoroacetate) (PIFA); *Journal of Synthesis*, (1998); 6: 867–872.
- Klagkou, K., Pullen, F., Harrison, M., Organ, A., Firth, A., Langley, G.J.; Fragmentation pathways of sulphonamides under electrospray tandem mass spectrometric conditions; *Rapid Communications in Mass Spectrometry*, (2003); 17: 2373–2379.
- Beloso, I., Borrás, J., Castro, J., Garcia-Vazques, J.A., Perez-Louride, P., Romero, J., et al.; Flexidentate behaviour of 2-pyridylsulfonamides—direct electrochemical synthesis and spectroscopic and X-ray characterisation of neutral copper(II) complexes of *N*-(2-pyridyl)sulfonamides; *European Journal of Inorganic Chemistry*, (2004); 2004: 635–645.
- Sudha, L.V., Sathyanarayana, D.N.; Molecular conformation of 1,3-pyridylphenylureas by ¹H and ¹³C NMR study; *Journal of Molecular Structure*, (1985); 131: 141–146.
- Horvath, C., Melander, W., Molnar, I.; Liquid chromatography of ionogenic substances with nonpolar stationary phases; *Analytical Chemistry*, (1977); 49: 142–154.
- Schoenmakers, P.J., van Molle, S., Hayes, C.M.G., Uunk, L.G.M.; Effects of pH in reversed-phase liquid chromatography; *Analytica Chimica Acta*, (1991); 250: 1–19.
- Sykora, D., Tesarova, E., Popl, M.; Interactions of basic compounds in reversed-phase high-performance liquid chromatography influence of sorbent character, mobile phase composition, and pH on retention of basic compounds; *Journal of Chromatography A*, (1997); 758: 37–51.
- Berges, R., Sanz-Nebo, V., Barbosa, J.; Modelling retention in liquid chromatography as a function of solvent composition and pH of the mobile phase; *Journal of Chromatography A*, (2000); 869: 27–39.
- Nikitas, P., Pappa-Louisi, A.; Retention models for isocratic and gradient elution in reversed-phase liquid chromatography; *Journal of Chromatography A*, (2009); 1216: 1737–1755.
- Snyder, L.R., Kirkland, J.J., Glajch, J.L.; *Practical HPLC Method Development*. John Wiley & Sons, New York, NY, (1997).
- Bosch, E., Bou, P., Rosés, M.; Linear description of solute retention in reversed-phase liquid chromatography by a new mobile phase polarity parameter; *Analytica Chimica Acta*, (1994); 299: 219–229.
- Torres-Lapasio, J.R., Garcia-Alvarez-Coque, M.C., Rosés, M., Bosch, E.; Prediction of the retention in reversed-phase liquid chromatography using solute–mobile phase–stationary phase polarity parameters; *Journal of Chromatography A*, (2002); 955: 19–34.
- Bosque, R., Sales, J., Bosch, E., Rosés, M., Garcia-Alvarez-Coque, M.C., Torres-Lapasio, J.R.; A QSPR study of the *p* solute polarity parameter to estimate retention in HPLC; *Journal of Chemical Information and Computer Sciences*, (2003); 43: 1240–1247.
- Izquierdo, P., Rosés, M., Bosch, E.; Polarity parameters of the Symmetry C18 and Chromolith Performance RP-18 monolithic chromatographic columns; *Journal of Chromatography A*, (2006); 1107: 96–103.
- Agrafiotou, P., Rafols, C., Castells, C., Bosch, E., Rosés, M.; Simultaneous effect of pH, temperature and mobile phase composition in the chromatographic retention of ionizable compounds; *Journal of Chromatography A*, (2011); 1218: 4995–5009.
- Lazaro, E., Izquierdo, P., Rafols, C., Rosés, M., Bosch, E.; Prediction of retention in reversed-phase liquid chromatography by means of the polarity parameter model; *Journal of Chromatography A*, (2009); 1216: 5214–5227.
- Marvin 6.0.3 [Internet]. ChemAxon; 2013. <http://www.chemaxon.com> (last accessed 26 June 2015).
- RDKit: Open-source cheminformatics [Internet]. <http://www.rdkit.org> (last accessed 26 June 2015).
- Stewart, J.; MOPAC 7 [Internet]. 1996. <http://openmopac.net> (last accessed 26 June 2015).
- Katritzky, A.R., Karelson, M., Lobanov, V.S.; CODESSA, version 2.20. Gainesville, FL, University of Florida, (1996).
- van der Walt, S., Colbert, S.C., Varoquaux, G.; The NumPy array: a structure for efficient numerical computation; *Computational Science and Engineering*, (2011); 13: 22–30.
- Jones, E., Oliphant, T., Peterson, P.; SciPy: Open Source Scientific Tools for Python [Internet]. (2001). <http://www.scipy.org/>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al.; Scikit-learn: machine learning in Python; *Journal of Machine Learning Research*, (2011); 12: 2825–2830.
- Ben-Hur, A., Weston, J.; A user's guide to support vector machines. In Carugo, O., Eisenhaber, F. (eds). *Data mining techniques for the life sciences*. Humana Press, New York, NY, (2009), pp. 223–239. http://link.springer.com/protocol/10.1007/978-1-60327-241-4_13 (accessed 22 January 2015).
- Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., Todeschini, R.; Comparison of different approaches to define the applicability domain of QSAR models; *Molecules*, (2012); 25: 4791–4810.