ADMET

Original scientific paper

Modelling of chromatographic and electrophoretic behaviour of imidazoline and alpha adrenergic receptors ligands under different acid-base conditions

Slavica Oljacic[1],*, Mitja Križman[2], Marija Popovic-Nikolic[1], Irena Vovk[2], Katarina Nikolic[1] and Danica Agbaba[1]

[1]*Department of Pharmaceutical Chemistry, Faculty of Pharmacy, University of Belgrade, Vojvode Stepe 450, 11000 Belgrade, Serbia*

[2]*Laboratory of Food Chemistry, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia*

Corresponding Author: E-mail: slavica.oljacic@pharmacy.bg.ac.rs; Tel.: +381 11 3951 335; Fax: +381 11 3974 349

Abstract

Background and Purpose: The ligands of the imidazoline and α-adrenergic receptors are mainly imidazoline and guanidine derivatives, known as centrally-acting antihypertensives and compounds with potential use in various neurological disorders. The extent of their ionisation has a major influence on their behaviour in the different analytical systems. The main objective of this work was to compare the mechanism of chromatographic retention and electrophoretic mobility under acidic, neutral and basic conditions. **Experimental Approach**: Multiple Linear Regression and Partial Least Squares Regression were applied for the QSRR (quantitative structure-retention relationship) and QSMR (quantitative structure-mobility relationship) modelling and to select the most important molecular parameters describing the chromatographic and electrophoretic behaviour of the investigated compounds. **Key Results:** The most important molecular descriptors, such as the chemical composition of the compounds, lipophilicity, polarizability and molecular branching, in the selected QSRR models showed that an important insight into the retention behaviour can be derived from the 0D-, 1D- and 2D-descriptors. The electrophoretic mobility could be explained by 2D- and 3D-descriptors, which provide information on the molecular mass, size and complexity, as well as on the influence of charge transfer and electronic properties on the migration behaviour. **Conclusion:** All created QSRR/QSMR models met the stringent validation criteria and showed high potential in describing the chromatographic and electrophoretic behaviour of investigated compounds.

Keywords

Quantitative structure-retention relationship; quantitative structure-mobility relationship; liquid chromatography; capillary electrophoresis

Introduction

Imidazoline and α-adrenergic receptor ligands are mainly imidazoline and guanidine derivatives with different pharmacological effects and target tissues [1]. The pharmacological effects of these ligands are generally the results of interaction with three types of imidazoline receptors (I_1-IR, I_2-IR, and I_3-IR) and α_2-adrenoreceptors (α_2-AR) [2,3]. Clonidine is a well-known first-generation centrally acting antihypertensive drug that lowers blood pressure by stimulating α_2-AR and I_1-IR. On the other hand, moxonidine and rilmenidine, which belong to the second generation of antihypertensives, act mainly at the I_1-IR and produce

fewer α_2-adrenoceptor-mediated side effects [3-5]. Idazoxan and its analogs have a high affinity for the I_2-IR whose involvement in different psychiatric disorders, opiate withdrawal, pain, and Parkinson's and Alzheimer's diseases has been investigated [6-8]. Efaroxan is able to induce insulin secretion from pancreatic β-cells by activating I_3-imidazoline receptors (I_3-IR) [9].

Considering the chemical structures of the above-mentioned ligands, the design of new compounds as drug candidates for the treatment of various cardiovascular or neurological diseases will be directed to compounds that contain one or more basic centers in their chemical scaffold and have the ability to ionize. For ionizable compounds, the extent of ionization together with lipophilicity is of great importance for their pharmacokinetic properties, such as absorption, distribution, metabolism and elimination (ADME), as well as for the drug-receptor interaction [10].

In addition, the extent of ionisation of compounds at different pH values has a strong effect on their behaviour in different separation systems, such as chromatographic and electrophoretic systems. As the degree of ionisation of basic compounds increases, their mobility in capillary electrophoresis increases, while the retention of unionised form of the analyte can be 10–20 times greater than that of the corresponding ionised form for a given mobile phase composition. Currently, high-performance liquid chromatography (HPLC) and capillary electrophoresis (CE) are very useful techniques for the rapid and efficient separation as well as examination of the individual imidazoline and α-adrenergic receptor ligands [11-14]. Therefore, accurate quantitative relationships between chromatographic retention or electrophoretic mobility and pH, which can be established through quantitative structure-retention relationships (QSRR) and quantitative structure-mobility relationships (QSMR) modelling can be very useful for predicting the retention behaviour and electrophoretic mobility of the new compounds, understanding the separation mechanism in a given analytical system and identifying important molecular descriptors [15]. In addition, the processes of drug action are believed to share many similarities with the processes underlying chromatographic separations, as the similar elementary intermolecular interactions are important for the behaviour of chemical compounds in both biological and chromatographic environments [16]. Therefore, the chromatographic retention constant log K_w has been successfully used to evaluate the lipophilicity of compounds [17] as well as in correlation studies with various physicochemical and biological properties to establish appropriate QSPR (quantitative structure-property relationship) and QSAR (quantitative structure-activity relationship) models [18,19]. Chromatographic and electrophoretic data have also been successfully used for the pharmacological classification of compounds and prediction of their ADME properties [20]. HPLC was used to determine the effective permeability of 40 compounds (18 IRs/a-ARs ligands and 22 CNS drugs) using the parallel artificial membrane permeability assay (PAMPA) to predict brain penetration [21]. The electrophoretic behaviour of a compound is indicative of its acid-base properties. *i.e.,* pK_a values and molecular mass, since the charge-to-mass ratio of compounds largely determines their electrophoretic mobility. Thus, all these models, QSPR, QSRR, QSAR and QSMR, can provide us with valuable information about the activity, physicochemical and ADME properties of drug candidates and could, therefore, significantly reduce the number of drug candidates synthesised and tested in the early phase of drug discovery.

The main objective of this work was to investigate and compare the mechanism of chromatographic retention and electrophoretic mobility of 29 compounds with main or side effects to alpha adrenergic/imidazoline receptors. The study was performed at different ionization levels, including physiological pH, and applied to QSRR and QSMR models using multiple linear regression (MLR) and partial least squares regression (PLS). Rigorous validation criteria as proposed by Golbraikh and Tropsha [22,23] and Roy [24], were also applied as an essential part of QSRR/QSMR model development. The developed QSRR/QSMR models can be used for the fast and reliable prediction of the chromatographic and electrophoretic behavior of structurally related

imidazoline and alpha adrenergic receptor ligands. In addition, the usefulness of chromatographic and electrophoretic data for predicting permeation across the blood-brain barrier (BBB) was also examined.

Experimental

Reagents and chemicals

Methanol (HPLC grade, 99.8 %) and glacial acetic acid (≥99.7 %) were purchased from J.T. Baker (Deventer, Netherlands). Ammonium acetate, sodium hydroxide, sodium dihydrogen phosphate, and disodium hydrogen phosphate of ACS reagent grade were purchased from Merck (Darmstadt, Germany). Boric acid (ACS reagent, 99.5 to 100.5 %) was purchased from Sigma-Aldrich (St. Louis, MO, USA). Ammonium hydroxide solution 30 % (analytical grade) was purchased from Carlo Erba (Milan, Italy). Aqueous solutions of background electrolytes and buffers for the preparation of mobile phases were prepared with water (HPLC grade).

Clonidine hydrochloride (≥98 %), moxonidine hydrochloride (≥98 %), guanfacine hydrochloride (≥98 %), brimonidine hydrochloride (≥98 %), efaroxan hydrochloride (≥98 %), idazoxan hydrochloride (≥98 %), rilmeni-dine hemifumarate (≥98 %), harmane (98 %), harmine hydrochloride (≥98 %), tizanidine hydrochloride (≥98 %), triamterene (≥99 %), clopamide (≥98 %), indapamide (≥99 %), naphazoline hydrochloride (≥98 %), xylome-tazoline hydrochloride (≥99 %), tetrahydrozoline hydrochloride (≥98 %), oxymetazoline hydrochloride (≥99 %), ephedrine hydrochloride (99 %), pseudoephedrine hydrochloride (99 %), maprotiline hydrochloride (>99 %), tamsulosin hydrochloride (≥98 %), mianserin hydrochloride (≥98 %), carvedilol (≥98 %), clozapine (≥98 %) and olanzapine (≥98 %) were purchased from Sigma–Aldrich (St. Louis, MO, USA). Tramazoline hydrochloride (≥98 %) and doxazosin mesilat (≥98 %) were obtained from Zdravlje-Actavis (Leskovac, Serbia). Amiloride hydro-chloride (≥98 %) was kindly donated by Galenika (Belgrade, Serbia). Phenilephrine hydrochloride (98 %) was obtained from Ivančić i sinovi (Belgrade, Serbia).

HPLC conditions

Solutions for HPLC analysis

The working solutions were prepared by dissolving the substances in methanol to obtain a concentration of 0.7 mg/mL for rilmenidine and 0.1 mg/mL for the other compounds.

HPLC equipment and working conditions

The HPLC working conditions are described in our previously published work [20]. HPLC analysis was performed using an Agilent Technologies 1200 (Wilmington, DE, USA) liquid chromatography system equipped with an online degasser, a binary pump, a column oven and a photodiode array detector. The data was recorded and analyzed using Agilent's ChemStation software (Chemstation for LC 3D system (Rev. B. 02. 01-SR2 [260]).

Separation was performed on the XTerra® RP18 column, 4.6×100 mm, particle size 3.5 µm (Waters Corporation, Milford, MA, USA) using three different constant ionic strength buffer solutions (I = 25 mmol/L) prepared at pH values 4.4, 7.4 and 9.1, as the aqueous component in the mobile phase while methanol added as an organic modifier. All measurements were performed at 25°C with a constant flow rate of 0.8 mL min^{-1} and UV detection in the range of 200-280 nm. The injection volume was 20 µL. Retention times were determined in the isocratic elution mode using at least six mobile phases of methanol/buffer (pH 4.4, 7.4 or 9.1) with a methanol concentration of 75-2 % depending on the retention properties of compounds. The retention factor k was calculated for each mobile phase composition according to Equation (1):

$$k = (t_r - t_0)/t_0 \tag{1}$$

where t_r is the retention time of the tested compound and t_0 is the dead volume measured with KNO_3 as a non-retentive marker. The values corresponding to 100 % of the buffered eluent (log k_w) were determined by extrapolation according to Equation (2):

$$\log k = \log k_w + S\varphi \tag{2}$$

where log k_w is the intercept, S is the slope of the regression line, and φ is the volume fraction of the organic modifier in the mobile phase. Log k_w of the analyte at three different experimental conditions, methanol/buffer pH 4.4, methanol/buffer pH 7.4 and methanol/buffer pH 9.1 (log $k_{w4.4}$, log $k_{w7.4}$ and log $k_{w9.1}$, respectively) were calculated for all investigated compounds and used for the QSRR analysis.

Electrophoretic conditions

Solutions for CE analysis

Stock solutions of harman, harmine, clopamide, indapamide, rilmenidine, mianserin, doxazosin, carvedilol, clozapine, olanzapine and triamterene were prepared in methanol (c = 1 mg/mL) and diluted with water to a final concentration of 5.8 µg/mL for triamterene, 60 µg/mL for rilmenidine and 30 µg/mL for the other substances. Brimonidine was dissolved in 0.1 % formic acid, while the remaining substances were dissolved in water. Acetone 2 vol.% was used as electroosmotic flow (EOF) marker. Depending on UV responsiveness and solubility, samples were prepared at different concentrations, from 5.8 to 60 µg/mL.

CE equipment and working conditions

The CE working conditions are described in our previously published paper [20]. All experiments were performed using the SpectraPhoresis 500-capillary electrophoresis system (Spectra Physics Analytical, USA) with UV detector. Data were recorded and analyzed using ChromQuest software version 4.0 (Thermo Finnigan, USA).

The separations were performed with an uncoated fused capillary (total length 31.5 cm, 50 µm id, effective length 24 cm, Polymicro Technologies, USA) applying a voltage of 11 kV and at 25 °C. The samples were injected hydrodynamically and detection was performed in the 200-280 nm range. The new capillary was rinsed stepwise with 0.1 M NaOH (15 min), water (10 min) and a running buffer (10 min). Finally, a 10 min high voltage was applied through the background electrolyte-filled capillary to equalize the inner surface, stabilize the electroosmotic flow and ensure good reproducibility of injections from run to run. Background solutions with constant ionic strength (I = 25 mmol/L) were prepared at three different pH values by mixing the corresponding amounts of $CH_3COOH/NaOH$, $NaH_2PO_4/NaOH$ and $H_3BO_3/NaOH$ for pH 4.4, 7.4 and 9.1, respectively. All compounds were injected in three replicates at each pH (4.4; 7.4 and 9.1). Between runs, the capillary was rinsed with the background electrolyte for 1 minute. The effective electrophoretic mobility, μ_{eff}, of the analyte at three different pH values 4.4; 7.4 and 9.1 ($\mu_{eff4.4}$, $\mu_{eff7.4}$ and $\mu_{eff9.1}$, respectively) is expressed in cm^2 / V s and was calculated using the Equation (3):

$$\mu_{eff} = \frac{L_t\, L_{eff}}{U}\left(\frac{1}{t_m} - \frac{1}{t_{eof}}\right) \tag{3}$$

where L_t is the total length of the capillary (cm), L_{eff} is the effective length of the capillary (cm), U is the applied voltage (V), t_m is the migration time of the analyte (s) and t_{eof} is the migration time of the electroneutral EOF marker acetone (s). The calculated μ_{eff} values ($\mu_{eff4.4}$, $\mu_{eff7.4}$ and $\mu_{eff9.1}$) were used as dependent variables in QSMR analysis.

Theory/calculation

Computational methods

    Clonidine, moxonidine, tizanidine, brimonidine, rilmenidine, and tramazoline are cyclic guanidines or amidine structures that can exist in two main tautomeric forms (amino and imino). The stability of their amino and imino tautomers was investigated at the B3LYP/6–31G(d, p) level of density functional theory-(DFT) [25,26] using the Gaussian 98 program [27]. Based on the Self Consistent Field Energy calculated for neutral amino and imino tautomers of the compounds, it was shown that clonidine, moxonidine, tizanidine, brimonidine and tramazoline exist as more stable imino tautomers, while the predominant tautomeric form of rilmenidine is the amino form. The calculation of pK_a and the selection of the dominant molecule/cation species under experimental conditions (pH 4.4, 7.4 and 9.1) were performed for 29 α-adrenergic and imidazoline receptor ligands using the Marvin 5.5.1.0 ChemAxon program [28].

    Structural optimization (selected tautomers and molecules/cations species at three different pH values) was performed using the B3LYP/3–21(d,p) levels of the DFT in the Gaussian 98 program [27]. Molecular descriptors were calculated for all investigated compounds in the programs Dragon [29], Gaussian 98 (B3LYP/3-21G(d,p) basis set) [27], Marvin 5.5.1.0 ChemAxon [28], and Chem3D Ultra 7.0.0 [30]. In addition, descriptors such as hardness (η), global softness, chemical potential (μ), electronegativity (χ) and electrophilicity index (ω), which have been extensively used to interpret various aspects of chemical bonding and reaction mechanism, were included in the set of calculated properties [31]. The partial least squares regression analysis started with 3445 computed descriptors. The descriptors with constant values were excluded from the modelling, as were descriptors with an intercorrelation greater than 0.996. For a pair of correlated descriptors, the one with the better correlation to Y was retained, while the other was excluded from the analysis. In the case of the MLR analysis, the cut-off value for the intercorrelation was set at 0.9 [32].

Quantitative structure-retention relationship and quantitative structure-mobility relationship modelling

    QSRR and QSMR modelling studies were performed for 29 imidazoline and α-adrenergic receptors ligands using PLS and stepwise MLR statistical methods at pH values of 4.4, 7.4 and 9.1 and on both HPLC and CE systems. An overview of the dataset was performed for each pH value and each chromatographic or electrophoretic system was examined in order to identify and exclude possible outliers before modelling. For this purpose, the scatter plots in the Simca P 12+ program were used [33], in which the pool of calculated molecular descriptors and the response variables are organised as an *X*-matrix. The scatter plot t_1 *vs.* t_2 is a window in *X*-space, showing how the *X* observations relate to each other, where t_1 and t_2 are scores (one vector for each model dimension) representing new variables computed as linear combinations of *X*. The score t_1 (first component) explains the largest variation of the *X* space, followed by t_2 and so on. For a two-dimensional score plot, the tolerance ellipse is presented based on Hotelling's T2. Observations that lie outside the ellipse are outliers. The *t-u* plot (t_1 *vs.* u_1) shows the relationship between *X* and *Y*. In addition, the degree of fit (a good fit corresponds to a small scatter around the straight line), indications of curvature and outliers can also be seen. The *u* score (one vector for each model dimension) is new variable summarizing *Y*, so as to maximize the correlation with the scores t. Based on the plots (t_1 *vs.* t_2 and t_1 *vs.* u_1), the remaining compounds were distributed among the training and test sets such that each chemical group (*i.e.*, guanidine, 2-aminoimidazoline, 2-arylmethylimidazoline, phenylethylamine etc.) has a representative in the test set and the log $k_{w4.4}$, log $k_{w7.4}$, log $k_{w9.1}$, $\mu_{eff4.4}$, $\mu_{eff7.4}$ and $\mu_{eff9.1}$ of the compounds in the test set are homogeneously distributed over the entire range of values of log $k_{w4.4}$, log $k_{w\,7.4}$, log $k_{w9.1}$, $\mu_{eff4.4}$, $\mu_{eff7.4}$ and $\mu_{eff9.1}$, respectively. The selected training and test sets contained at least 20 compounds used for model building and at least 5 compounds used for external validation. In order to perform a meaningful comparison between the different

methods (MLR and PLS) used to develop the QSRR models, the same training and test set was used for each pH value. Therefore, six independent training and test sets were created, one for each dependent variable (log $k_{w4.4}$, log $k_{w7.4}$, log $k_{w9.1}$, $\mu_{eff4.4}$, $\mu_{eff7.4}$ and $\mu_{eff9.}$1). The predictive power and robustness of the models were estimated using the cross-validated parameter R^2 (Q^2), which was calculated from Equation (4):

$$Q^2 = 1 - \frac{\text{PRESS}}{\sum \left(Y_{obs(training)} - \overline{Y}_{training} \right)^2} \tag{4}$$

In Equation (4), PRESS is a parameter calculated using the leave-one-out (LOO) approach where each compound from the training set was deleted once and a new model was created with the remaining compounds and used to predict the Y-value of the deleted compound. The procedure was repeated until all compounds had been deleted once. For all created models, the squared sum of the differences between observed and LOO-predicted values ($e(i)$) (PRESS) was calculated according to Equation (5):

$$\text{PRESS} = \sum_{i=1}^{n} e_i^2 \tag{5}$$

The models with $Q^2 \geq 0.5$ can be considered to have good predictive capability [34]. The root mean squared error of estimation (RMSEE) was calculated to characterize the predictive ability of the models for the compounds in the training set, Equations (6):

$$\text{RMSEE} = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{obs(training)} - Y_{pred(training)} \right)^2}{n}} \tag{6}$$

where n is the number of compounds in the training set, while $Y_{obs(training)}$ and $Y_{pred(training)}$ denote the experimental and predicted values for the compounds in the training set.

In addition to internal validation, the predictive power of all created models was estimated by external validation using the following parameters in Equations (7) and (8):

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} \left(Y_{obs(test)} - Y_{pred(test)} \right)^2}{n}} \tag{7}$$

where RMSEP is the root mean square error of prediction, n is the number of compounds in the test set, while $Y_{obs(test)}$ and $Y_{pred(test)}$ are the experimental and predicted values for the compounds in the test set.

$$R^2_{pred} = 1 - \frac{\sum \left(Y_{obs(test)} - Y_{pred(test)} \right)^2}{\sum \left(Y_{obs(test)} - \overline{Y}_{training} \right)^2} \tag{8}$$

where $Y_{pred(test)}$ and $Y_{obs(test)}$ are the predicted and observed values of the dependent variables of the compounds in the test set, respectively, and $\overline{Y}_{training}$ indicates the mean values of the dependent variables of the compounds in the training set. For a predictive QSRR/QSMR model, the R^2_{pred} value should be higher than 0.5 [35].

The rigorous external validation parameters proposed by Roy ($r_m^2, r_m'^2, \overline{r}_m^2 \Delta r_m^2$) [24] and Golbraikh and Tropsha [22,23] were also applied to all created models and were used to select the most reliable models for the prediction of chromatographic and electrophoretic behaviour at each pH value.

According to Ojha and Roy [35], r_m^2 is validation metric calculated according to Equation (9) using observed (*y*-axis) and predicted (*x*-axis) values:

$$r_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2} \right) \tag{9}$$

where r^2 is the determination coefficient for the least squares regression line (with intercept) correlating observed (*y*-axis) and predicted (*x*-axis) values and r_0^2 is the determination coefficient for the least squares regression line (without intercept) correlating observed (*y*-axis) and predicted (*x*-axis) values, $r_m'^2$ is validation metric calculated according to Equation (10) using observed (*x*-axis) and predicted (*y*-axis) values

$$r_m'^2 = r^2 \left(1 - \sqrt{r^2 - r_0'^2}\right) \tag{10}$$

where $r_0'^2$ is the determination coefficient for the least squares regression line (without intercept) correlating observed (*x*-axis) and predicted (*y*-axis) values. $\overline{r}_m^2$ is average of r_m^2 and $r_m'^2$, while Δr_m^2 is the absolute difference between r_m^2 and $r_m'^2$.

Models were considered acceptable if they met all of the following conditions:

    a. the values of r_m^2 and $r_m'^2$ should be close to each other and $\overline{r}_m^2$ should be greater than 0.5;

    b. Δr_m^2 should be less than 0.2 [24].

According to criteria proposed by Golbraikh and Tropsha, following conditions should be achieved: (i) $Q^2 > 0.5$; (ii) $R^2 > 0.6$, (iii) $[(R^2 - R_0^2) / R^2] < 0.1$ or $[(R^2 - R'_0^2) / R^2] < 0.1$; (iv) $0.85 \le k \le 1.15$ or $0.85 \le k' \le 1.15$ where Q^2 is the cross-validated correlation coefficient calculated for the training set, but all other criteria are calculated for the test set [22,23]. R_0^2 is the coefficient of determination (predicted versus observed values) for regressions through the origin. R'_0^2 is the coefficient of determination (observed versus predicted values) for regressions through the origin. *K* and *K′* are slopes of regression lines through the origin.

Multiple linear regression

Stepwise multiple linear regression was performed with the program STATISTICA [36]. A stepwise multiple linear regression procedure was used to select the molecular descriptors with the greatest influence on the dependent variables. The models were created stepwise according to the following procedure: identification of an initial model, then iterative „stepping" in which descriptors were added or removed from the model based on the previously established stepping criteria (*F* to enter and *F* to remove) according to their statistical significance (*F*-test), and finally termination of the search when no more stepping was possible or when a specified maximum number of steps was reached. Stepping criteria were set to a narrow range, and the predictive abilities of the models were tested by leave-one-out cross-validation and external test set prediction methods to avoid obtaining an overfitting model.

Partial least squares regression

The PLS analysis was carried out using the SIMCA P+ 12.0 (soft independent modelling of class analogy) program [33]. The PLS approach is able to analyse data with noisy, collinear and even incomplete independent variables. In PLS modelling, the variable importance in the projection (VIP) parameter was used to assess the significance of the independent variables in the course of PLS model building. The VIP parameter represents a summary of the importance of each variable (*X*) for both the *Y* and *X* matrices. Molecular descriptors with VIP values greater than 1 are most important in explaining the regression model, those with $1.0 > VIP > 0.5$ have a moderate influence, while the independent variables with VIP values less than 0.5 are not relevant for the model [33,37]. Therefore, during model generation, the descriptors with the lowest VIP values were successively removed from the model, while statistical parameters such as the squares of the multiple correlation coefficients R^2, Q^2 (a cross-validated version of R^2), the root mean square error of estimation for the training set, and the root mean square error of prediction for the test set were calculated for each new PLS model and compared with the previous one. The procedure was repeated until the best models for the investigated experimental conditions (pH 4.4, 7.4 and 9.1) and analytical systems (HPLC and CE) were formed.

The response permutation test (*Y* scrambling), as a measure of model overfitting, was used to examine the statistical significance of the R^2 and Q^2 values and overfitting due to the chance correlation [37]. In this test, the *Y*-variables were randomly re-ordered 100 times while the *X*-matrix was remained unchanged. The PLS model was fitted to the permuted data and the new parameters R^2, Q^2 and VIP were calculated. All model selection steps were repeated with the scrambled *Y* response data. Regression lines were fitted through the R^2 and Q^2 values to obtain two separate intercepts. The values of the obtained intercepts for valid QSRR and QSMR models should be below 0.05 for the Q^2 intercept and not above 0.4 for the R^2 intercept [37].

Results and discussion

The influence of the degree of ionization (under acidic, neutral and basic conditions) of 29 imidazoline and/or α-adrenergic receptor ligands (Figure S1, Supplementary material) on their retention in reversed-phase liquid chromatography and electrophoretic mobility in capillary electrophoresis was investigated. A pH value of 4.4 was chosen for the work in an acidic medium at which the majority of the tested compounds reached the highest degree of ionization, pH = pK_a - 2 (based on calculated pK_a values in [28] and experimentally determined pK_a values found in the literature, Table S1, Supplementary material). For operation in a neutral medium a pH value of 7.4 was used, which simulates physiological conditions, while a pH of 9.1 was chosen as a compromise where 9 compounds move together with the electroosmotic flow marker and the effective mobility is zero such as *e.g.* brimonidine (experimental pK_a determined by potentiometry 7.50), clozapine (experimental pK_a determined by HPLC 7.719 ± 0.015), doxazosine (experimental pK_a determined by voltammetry 6.89 ± 0.57), guanfacine (calculated basic pK_a 8.65), harman (experimental pK_a determined by HPLC 7.21), mianserin (experimental pK_a determined by potentiometry 7.40), moxonidine (experimental pK_a determined by HPLC 7.84), tizanidine (calculated pK_a 7.49) and triamterene (experimental pK_a determined by pH spectrophotometric method 7.16) (Tables S1 and S4, Supplementary material); about 1/3 of the compounds reach their mean mobility (pH close to the pK_a value, amiloride (experimental pK_a determined by potentiometry 8.72 ± 0.05), idazoxan (experimental pK_a determined by HPLC 9.04), phenylephrine (experimental pK_a determined by spectrophotometry 9.17), tamsulosin (calculated pK_a 9.28), clopamide (calculated acid pK_a 8.85), indapamide (experimental acid pK_a determined by potentiometry 8.8 ± 0.2) (Tables S1 and S4, Supplementary material) and the remaining compounds with pK_a values around and above 10 are still ionized to a high percentage, such as *e.g.* efaroxan (experimental pK_a determined by HPLC 10.02), ephedrine (experimental pK_a determined by spectrophotometry 9.65), maprotiline(experimental pK_a determined by potentiometry 10.45 ± 0.02), naphazoline (experimental pK_a determined by pH spectrophotometric method 10.81), oxymetazoline (experimental pK_a determined by pH spectrophotometric method 10.62), tetrahyrozoline (experimental pK_a determined by pH titration 10.51 ± 0.05), tramazoline (experimental pK_a determined by pH titration 10.66 ± 0.05)(Table S1 and S4, Supplementary material).

Two different linear statistical methods (MLR and PLS) were applied and compared for each experimental condition and examined systems in order to identify the most reliable models enabling accurate prediction of dependent variables.

Quantitative structure-retention relationship modelling

Determination of log k_w values

The retention behaviour in HPLC system was monitored by calculating extrapolated retention factors log k_w corresponding to pure buffer as mobile phase. The log k_w values are considered to be better lipophilicity indices compared to the isocratic log *k*, as their values are of the same order of magnitude as the log *P* / log *D* octanol-water [38]. The values of log k_w (intercepts) and slopes (*S*) at all investigated pH values: 4.4, 7.4 and 9.1 are

shown in Table S2 (Supplementary material). In an acidic environment (pH 4.4), the obtained log k_w values are in the range from 0.09 (phenilephrine) to 3.75 (carvedilol). Under the given conditions, only two compounds (clopamide and indapamide) are present in neutral (molecular) form, while the majority of the other compounds are completely ionized (over 99 %) [28]. This is also the reason for the lower log k_w values of the investigated compounds at pH 4.4 compared to pH 7.4 and 9.1, where the compounds are present in ionized form to a lesser extent or are predominantly non-ionized (pH 9.1). Therefore, as the pH increases, the lipophilicity of the compounds and, consequently, their log k_w values increase (Table S2, Supplementary material).

Quantitative structure-retention relationship models

As a result of the QSRR studies applied to the log k_w data obtained at three pH values (log $k_{w4.4}$, log $k_{w7.4}$ and log $k_{w9.1}$), six models were created: MLR-QSRR (log $k_{w4.4}$), PLS-QSRR (log $k_{w4.4}$), MLR-QSRR (log $k_{w7.4}$), PLS-QSRR (log $k_{w7.4}$), MLR-QSRR (log $k_{w9.1}$), and PLS-QSRR (log $k_{w9.1}$). The score plot of the first two components (t_1 *vs.* t_2) at each pH shows a uniform data distribution in all four quartiles. Tamsulosin is in group with oxymetazolin, clopamide, indapamide, carvedilol, doxazosin and oxymetazoline. These compounds are characterized by the highest descriptor values with tamsulosin lying outside or on the Hoteling T2 ellipse. Thus, tamsulosin was identified as an outlier in all chromatographic models.

For the generation of QSRR models at pH 4.4, 21 compounds were used in the training set, while the remaining 7 compounds (clopamide, clozapine, ephedrine, harman, mianserin, naphazoline, tizanidine) were used in the test set.

Using the stepwise MLR method, 4 descriptors with the greatest influence on the dependent variable log $k_{w4.4}$ were selected, Equation (11):

MLR-QSRR (log $k_{w4.4}$) = -5.24598 + 0.64963×SM06_EA − 9.89747×VE2_B(s) −
- 0.12794×SpMax8_Bh(v) + 1.51417×SpMax5_Bh(p)　　　　　　　　　　　　　　(11)

while statistically, the most significant PLS-QSRR model was formed with 6 descriptors, Equation (12):

PLS-QSRR (log $k_{w4.4}$) = f(H_Dz(p);VE2_Dz(p);SpMax5_Bh(e);SpMax5_Bh(i);SM02_EA(ri);SM04_EA(ri))　(12)

Better regression and validation parameters (Tables 1 and 2) were obtained for the PLS-QSRR(log $k_{w4.4}$) model (Q^2 = 0.865, RMSEP = 0.325, F= 57.524, p = 1.52E-08) compared to the MLR-QSRR(log $k_{w4.4}$) model (Q^2 = 0.764, RMSEP = 0.496, F = 23.502, p = 1.57E-06), so that the PLS-QSRR (log $k_{w4.4}$) model was selected as optimal for the prediction of retention behaviour at pH 4.4.

Table 1. The most important statistical parameters obtained for MLR-QSRR and PLS-QSRR models

Model	Q^2	R^2Y	F	p	RMSEE	RMSEP	$R^2_{intercept}$	$Q^2_{intercept}$
MLR-QSRR (log $k_{w4.4}$)	0.764	0.855	23.502	$1.57×10^{-6}$	0.369	0.496		
PLS-QSRR (log $k_{w4.4}$)	0.865	0.876	57.524	$1.52×10^{-8}$	0.341	0.325	-0.0969	-0.210
MLR-QSRR (log $k_{w7.4}$)	0.846	0.907	54.995	$5.82×10^{-9}$	0.365	0.446		
PLS-QSRR (log $k_{w7.4}$)	0.883	0.903	28.978	$3.76×10^{-7}$	0.371	0.378	0.0091	-0.279
MLR-QSRR (log $k_{w9.1}$)	0.931	0.953	113.538	$1.91×10^{-11}$	0.226	0.353		
PLS-QSRR (log $k_{w9.1}$)	0.915	0.924	96.874	$2.32×10^{-10}$	0.286	0.359	-0.0781	-0.224

Table 2. External validation parameters for QSRR models, based on Golbraikh, Tropsha, and Roy [22-24]

Model	R^2	R_0^2	$(R^2-R_0^2)/R^2$	k	R'_0^2	$(R^2-R'_0^2)/R^2$	k'	R^2_{pred}	r_m^2	$r_m'^2$	$\overline{r_m^2}$	Δr_m^2
MLR-QSRR (log$k_{w4.4}$)	0.635	0.618	0.028	0.981	0.578	0.090	0.950	0.617	0.551	0.484	0.517	0.067
PLS-QSRR (log$k_{w4.4}$)	0.835	0.835	0.000	1.003	0.806	0.035	0.968	0.835	0.830	0.693	0.762	0.138
MLR-QSRR (log$k_{w7.4}$)	0.829	0.782	0.056	0.943	0.827	0.002	1.015	0.766	0.650	0.793	0.722	0.144
PLS-QSRR (log$k_{w7.4}$)	0.854	0.839	0.018	0.962	0.854	0.000	1.006	0.832	0.748	0.843	0.796	0.095
MLR-QSRR (log$k_{w9.1}$)	0.910	0.877	0.036	0.960	0.802	0.118	1.022	0.872	0.745	0.611	0.678	0.134
PLS-QSRR (log$k_{w9.1}$)	0.874	0.872	0.002	0.960	0.868	0.007	1.022	0.868	0.838	0.805	0.822	0.033

The QSRR models at pH 7.4 were created with a training set of 21 compounds, while the remaining 7 compounds (clonidine, ephedrine, maprotiline, olanzapine, tizanidine, triamterene, xylometazoline) were used in the test set. The best MLR-QSRR model was obtained with 3 descriptors, Equation (13):

MLR-QSRR(log $k_{w7.4}$) = 5.91899 - 22.05690×VE2_Dt + 0.00274×QXXp + 0.02905×F02[C-C]   (13)

The PLS analysis resulted in the model with five important descriptors, Equation (14):

PLS-QSRR (log $k_{w7.4}$) = f(Ho_Dt; SM3_Dt; VE2_Dt; H_D/Dt; log$D$ 7.4)   (14)

Although the MLR-QSRR (log $k_{w7.4}$) model has statistically more significant values for *F*, *p* and R^2 than the PLS-QSRR (log $k_{w7.4}$) model, better validation parameters (Q^2 = 0.883; R^2_{test}=0.854; RMSEP = 0.378) for the PLS model compared to the MLR model ((Q^2 = 0.846; R^2_{test} = 0.829; RMSEP = 0.446) (Tables 1 and 2) were allocateed this regression model as more reliable for predicting the retention behaviour of related compounds at pH 7.4.

At pH 9.1, the test set consisted of 7 compounds (clozapine, efaroxan, ephedrine, maprotiline, tizanidine, triamterene, xylometazoline) and a training set of 21 compounds.

The most significant MLR model describing the retention behaviour at pH 9.1 consists of 3 molecular descriptors, Equation (15):

MLR-QSRR (log $k_{w9.1}$)= -0.53081+0.09255×nBM + 0.70852×SpMin6_Bh(p) + 0.58125×log *D* 9.1   (15)

The statistically best PLS model was formed with 5 descriptors, Equation (16):

PLS-QSRR (log $k_{w9.1}$) = f(VE2_Dt; H_D/Dt ; F04[C-C]; ALOGP; log *D* 9.1)   (16)

The analysis of the calculated statistical parameters (Tables 1 and 2) obtained for the created models at pH 9.1, suggests that the MLR-QSRR (log $k_{w9.1}$) model can be selected as an optimal and reliable model for predicting the retention behaviour of α-adrenergic and imidazoline receptor ligands in an alkaline environment.

In addition, it should be emphasised that all created MLR-QSRR and PLS-QSRR models meet the external validation criteria proposed by Golbraikh, Tropsha and Roy (Table 2) [22-24].

Plots of observed *vs*. predicted values for compounds in the training and test sets in selected QSRR models are depicted in Figure 1. The values of molecular descriptors in selected QSRR models are listed in Table S3 .
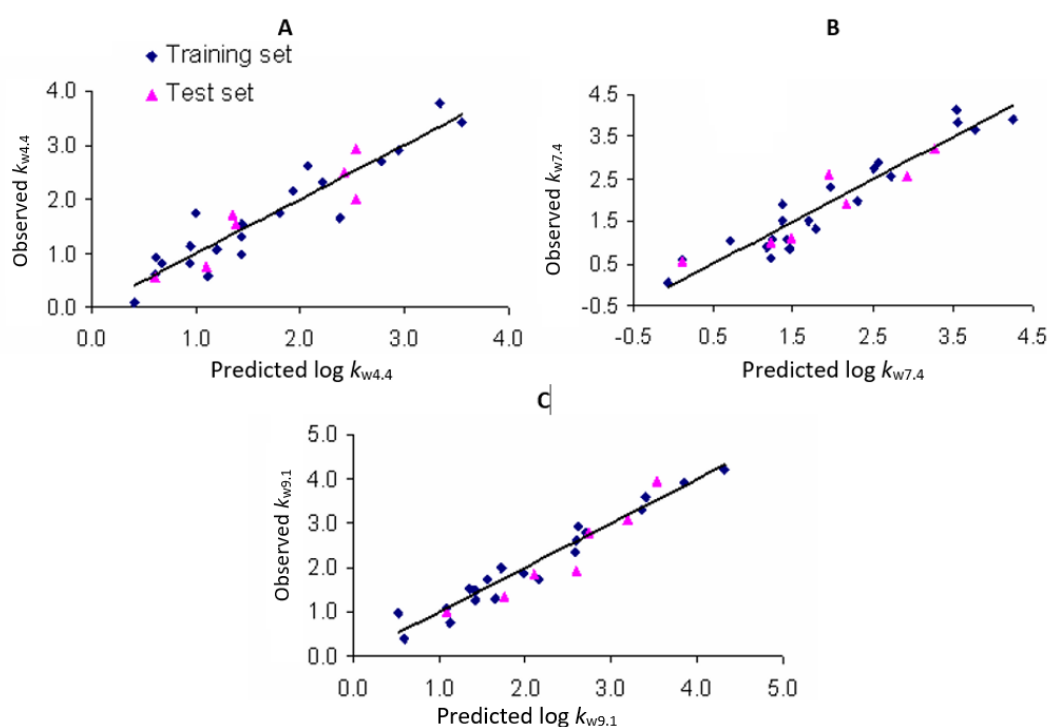

Figure 1. Observed *vs*. predicted values in selected QSRR models: A - PLS-QSRR (log $k_{w4.4}$);
B - PLS-QSRR (log $k_{w7.4}$); C - MLR-QSRR (log $k_{w9.1}$)

Quantitative structure-mobility relationship modelling

Determination of effective electrophoretic mobility

The effective electrophoretic mobility of the investigated compounds was determined in the presence of acetone as a neutral marker, and the results obtained under the examined conditions (pH 4.4, 7.4 and 9.1) are presented in Table S4 (Supplementary Material). At pH 4.4, 27 compounds that move faster than the neutral marker are in the form of cations and reach their maximum effective mobility, while the effective mobility of clopamide and indapamide, which are in neutral form under the studied conditions, is zero. The range of the obtained $\mu_{eff4.4}$ values is between 0 (indapamide, clopamide) and 28.19×10^{-5} cm^2 / V s (harmane). At a physio-logical pH of 7.4, compounds with pK_a values around 7.4 have reached their intermediate mobility, indapamide and clopamide co-migrate with the EOF marker (μ_{eff} = 0), and compounds with a pK_a value above 10 (Table S1, Supplementary material), a high percentage of which are in the form of cations under the given conditions, have maintained an effective mobility similar to that in an acidic environment. The resulting range of effective mobility is between 0 and 24.18×10^{-5} cm^2 / V s (naphazoline). In an alkaline environment at pH 9.1, negatively charged ions of clopamide and indapamide migrate to the cathode more slowly than EOF markers and exhibit negative values of effective mobility. Uncharged compounds migrate together with the EOF marker and their mobility is zero, while the derivatives of 2-methylimidazoline with pK_a values above 10 maintain a high degree of mobility (μ_{eff} = 25.05×10^{-5} cm^2 / V s, tetrahydrozoline) (Table S1 and Table S4).

Quantitative structure-mobility relationship models

As a result of the QSMR studies applied to the μ_{eff} data obtained at three pH values ($\mu_{eff4.4}$, $\mu_{eff7.4}$ and $\mu_{eff9.1}$), six models were created: MLR-QSMR ($\mu_{eff4.4}$), PLS-QSMR ($\mu_{eff4.4}$), MLR-QSMR ($\mu_{eff7.4}$), PLS-QSMR ($\mu_{eff7.4}$), MLR-QSMR ($\mu_{eff9.1}$), and PLS-QSMR ($\mu_{eff9.1}$).

The QSMR models, which describe the mobility of the investigated compounds at a pH of 4.4, were created with 20 compounds in the training set. Three compounds, one with the highest electrophoretic mobility (olanzapine $\mu_{eff4.4}$ = 33.632 ± 0.036 10^{-5} cm^2 / V s) and two with the lowest electrophoretic mobility (clopamide $\mu_{eff4.4}$ = 0.000 ± 0.000 and indapamide $\mu_{eff4.4}$ = 0.000 ± 0.000) clearly showed a significant deviation from the regression line on the t_1 *vs.* u_1 scater plot and were therefore identified as outliers. The remaining 6 compounds (amiloride, carvedilol, ephedrine, mianserin, tizanidine and xylometazoline) were used for external validation.

The most significant MLR-QSMR model at pH 4.4 was formed with 4 descriptors, Equation (17):

MLR-QSMR ($\mu_{eff\ 4.4}$) = 45.51010 - 3.93697×ATS6s - 0.12728×ATSC6m +
1.09327×Mor30i + 0.03226×ATSC8m   (17)

The PLS analysis also selected the model with the four descriptors that showed the greatest influence on the dependent variable $\mu_{eff\ 4.4}$, Equation (18)

PLS-QSMR ($\mu_{eff\ 4.4}$) = f(Ho_Dz(e); ATS6s; ATS7s; ATSC6m)   (18)

According to the better statistical parameters of the PLS-QSMR model (Q^2 = 0.931, F = 115.508, p = 1.28×10^{-10}, R^2_{test} = 0.897 and RMSEP = 0.984) compared to the MLR-QSMR model (Q^2 = 0.880, F = 64.338, p = 2.92×10^{-9}, R^2_{test} = 0.876 and RMSEP = 1.063) (Tables 3 and 4), the PLS-QSMR model can be selected as the optimal and reliable model for predicting the electrophoretic mobility of related compounds at pH 4.4.

QSMR models at pH 7.4 were formed with 20 compounds in the training set, while 5 compounds (ephedrine, olanzapine, oxymetazoline, tamsulosin and tizanidine) were retained in the test set. Considering the fact that the investigated data sets are very different in their acid-base properties, the resulting differences in effective electrophoretic mobility are significant (Table S1, Supplementary material).

Clopamide and indapamide, which carry a weakly acidic sulfonamide group, are still predominantly unioniz-ed at pH 7.4 and move together with the electroosmotic flow, resulting in an electrophoretic mobility of zero.

Table 3. The most important statistical parameters obtained for MLR-QSMR and PLS-QSMR models

Model	Q^2	R^2	F	p	RMSEE	RMSEP	$R^2_{intercept}$	$Q^2_{intercept}$
MLR-QSMR ($\mu_{eff\,4.4}$)	0.880	0.945	64.338	2.92E-09	0.739	1.063		
PLS-QSMR ($\mu_{eff\,4.4}$)	0.931	0.939	115.508	1.28E-10	0.776	0.984	-0.111	-0.228
MLR-QSMR ($\mu_{eff\,7.4}$)	0.858	0.932	51.412	1.4E-08	1.569	2.216		
PLS-QSMR ($\mu_{eff\,7.4}$)	0.855	0.870	50.098	7.47E-08	2.170	3.244	-0.0245	-0.225
MLR-QSMR ($\mu_{eff\,9.1}$)	0.891	0.936	58.736	2.32E-09	2.640	3.794		
PLS-QSMR ($\mu_{eff\,9.1}$)	0.882	0.916	29.464	3.35E-07	3.033	4.290	0.046	-0.329

Table 4. External validation parameters for QSMR models, based on Golbraikh, Tropsha, and Roy [22-24]

Model	R^2	R_0^2	$(R^2-R_0^2)/R^2$	k	R'^2_0	$(R^2-R'^2_0)/R^2$	k'	R^2_{pred}	r_m^2	$r_m'^2$	$\overline{r_m^2}$	Δr_m^2
MLR-QSMR ($\mu_{eff\,4.4}$)	0.876	0.862	0.016	0.999	0.876	0.000	0.999	0.869	0.772	0.865	0.819	0.094
PLS QSMR ($\mu_{eff\,4.4}$)	0.893	0.882	0.013	1.002	0.893	0.000	0.996	0.888	0.798	0.888	0.843	0.089
MLR-QSMR ($\mu_{eff\,7.4}$)	0.788	0.771	0.022	1.003	0.781	0.009	0.980	0.799	0.685	0.722	0.703	0.037
PLS-QSMR ($\mu_{eff\,7.4}$)	0.686	0.563	0.178	0.938	-0.818	2.192	1.030	0.570	0.446	-0.155	0.145	0.601
MLR-QSMR ($\mu_{eff\,9.1}$)	0.879	0.860	0.022	0.843	0.814	0.075	1.096	0.802	0.758	0.654	0.706	0.104
PLS-QSMR ($\mu_{eff\,9.1}$)	0.873	0.834	0.045	0.813	0.750	0.141	1.118	0.747	0.701	0.567	0.634	0.134

A very low effective electrophoretic mobility is also obtained for triamterene ($\mu_{eff7.4}$ = 1.140±0.053× ×10^{-5} cm^2 / V s) as a weak base and guanfacine ($\mu_{eff7.4}$ = 5.083 ± 0.025×10^{-5} cm^2 / V s). The Y-values of these compounds in relation to the calculated X-values showed a significant deviation in the scatter plot t_1 *vs.* u_1 and were identified as outliers. The most significant MLR-QSMR model is composed of 4 descriptors, Equation (19):

MLR-QSMR ($\mu_{eff7.4}$) = 34.68883 - 31.30360×GGI6 -158.70600×R7u+ +

1.21557×Mor28s - 10.57450×HATS2i   (19)

Using PLS analysis, a model was constructed with 5 molecular descriptors that significantly affect the electrophoretic mobility of the studied compounds at pH 7.4, Equation (19):

PLS-QSMR ($\mu_{eff\,7.4}$) = f(GGI6; SpMaxA_EA(dm); G1; RDF045e; Mor32m)   (20)

For the MLR-QSMR model, lower values for RMSEE and RMSEP, slightly better parameters Q^2 and F value as well as significantly higher values for $R^2_{trening}$ and R^2_{test} were obtained compared to the PLS-QSMR model (Tables 3 and 4). In addition, the PLS-QSMR model did not meet external validation criteria established by Golbraikh and Tropsha [22] [$(R^2 - R_0^2) / R^2$ = 0.178 > 0.1; R'^2_0= -0.818; $(R^2- R'^2_0) / R^2$ = 2.192 > 0.1] (Table 4); and Roy [23,24] [r_m^2 =0.446 and $r_m'^2$ = -0.155 are not closed; $\overline{r_m^2}$ = 0.145 < 0.5; Δr_m^2 = 0.601 > 0.2] (Table 4) indicating that the MLR-QSMR regression equation can be reliably applied to predict the electrophoretic mobility of related compounds at pH 7.4.

At a pH of 9.1, nine compounds migrate together with EOF and have an effective mobility of 0. Five compounds with an effective mobility of zero remain in the training set, while four compounds move to the test set taking into account that all chemical clusters have a representative in the test set. Since all created models were unable to predict the mobility of harman and mianserin (μ_{eff} = 0) and taking into account that these two compounds had the greatest deviation on the t_1 *vs.* u_1 plot, among others from the training set with zero mobility, these two compounds were excluded from further analysis.

QSMR models at pH 9.1 were created with 21 compounds in the training set, while 6 compounds (doxazosin, ephedrine, idazoxan, olanzapine, tizanidine, xylometazoline) were used for external validation.

By MLR analysis, four significant molecular descriptors were selected, Equation (21):

MLR-QSMR ($\mu_{eff\,9.1}$) = 60.98426 - 15.11770×SM1_Dz(Z) + 1.00148×SM15 EA(dm) -

- 29.9053×HOMO[B3LYP/3-21G(d,p)] – 60.3708×ISH   (21)

While the statistically best PLS model was formed with 5 important molecular parameters, Equation (22):

PLS-QSMR ($\mu_{eff9.1}$) = f(ZM1MulPer; AAC; SM1_Dz(Z); SM11_EA(dm); ISH)   (22)

Appropriate validation criteria were met for both MLR-QSMR/PLS-QSMR models created (Tables 3 and 4), with the exception of the value $(R^2 - R'^2_0)/R^2$ 0.141 > 0.1 of the PLS-QSMR ($\mu_{eff9.1}$) model (Table 4). Due to the higher values of Q^2, F, $R^2_{trening}$, R^2_{test} (Q^2 = 0.891, F = 58.736, $R^2_{traning}$ = 0.936 and R^2_{test} =0.879) for the MLR model than for the PLS model (Q^2 = 0.882, F = 29.464, $R^2_{trening}$ = 0.916 and R^2_{test} = 0.873) and the lower RMSEE and RMSEP (RMSEE = 2.640 and RMSEP = 3.794) for the MLR model than for the PLS model (RMSEE=3.033 and RMSEP = 4.290), the MLR-QSMR model: MLR-QSMR ($\mu_{eff9.1}$) = f(SM1_Dz (Z); SM15_EA (dm); HOMO [B3LYP/3--21G (d, p)]; ISH) was selected as the optimal model for predicting the effective mobility of related compounds at a given pH of 9.1.

Plots of observed *vs.* predicted values for compounds in the training and test sets of the selected QSMR models are depicted in Figure 2.

It is interesting to note that in Figure 2B, which shows the correlation between the calculated and predicted values of the QSMR model at pH 7.4, two clusters of data can be observed. The upper cluster includes compounds whose experimentally determined effective mobility is greater than 19×10^{-5} cm^2 / V s and experimental pK_a value is greater than 8, and the lower cluster includes compounds with experimental effective mobility of less than 15×10^{-5} cm^2 / V s and pK_a value less than 8 (Table S1 and S4, Supplementary material). The selected MLR-QSMR model could be used for a rough prediction of the pK_a values of new imidazoline analogs.
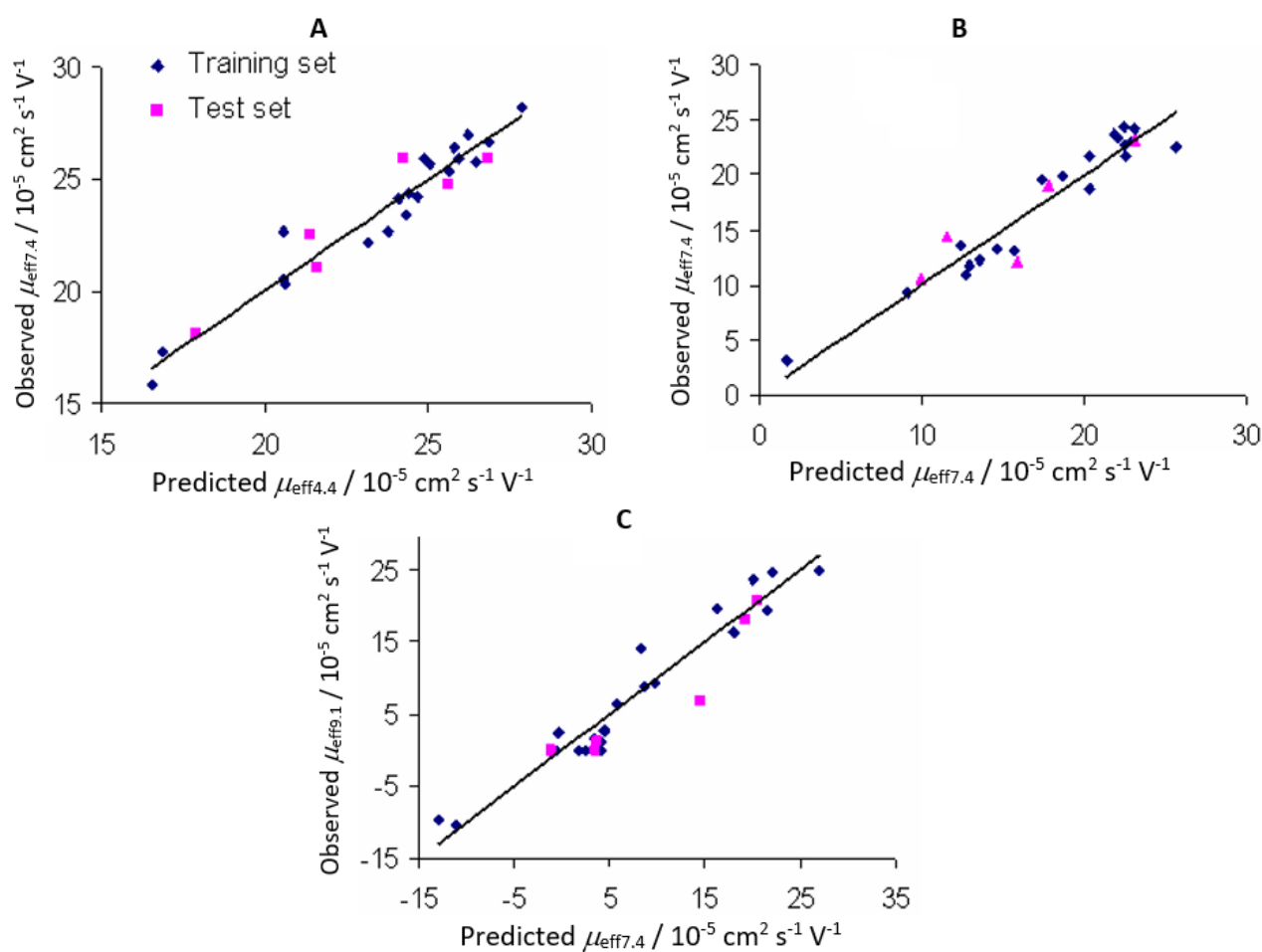

Figure 2. Observed *vs.* predicted values in selected QSMR models: A - PLS-QSMR ($\mu_{eff4.4}$);
B - MLR-QSMR ($\mu_{eff7.4}$); C - MLR-QSMR ($\mu_{eff9.1}$)

The values of the molecular descriptors in the selected QSMR models are listed in Table S5 (Supplementary Material).

Interpretation and comparison of selected quantitative structure-retention relationship and quantitative structure-mobility relationship models

Based on the presented statistical results (Tables 1-4) obtained after internal and rigorous external validation, optimal QSRR and QSMR models were selected for each investigated pH value. The classes of descriptors in the selected models and their descriptions are presented in Tables 5 and 6.

The descriptors selected for the optimal PLS-QSRR (log $k_{w4.4}$) model belong to three different descriptor classes (Table 5), such as 2D matrix-based descriptors (H_Dz(p), VE2_Dz(p)), Burden eigenvalues (SpMax5_Bh(e), SpMax5_Bh(i)) and Edge adjacency indices (SM02_EA(ri), SM04_EA(ri)).

Table 5. The most important descriptors in selected QSRR models

Selected descriptors	Description	Class of descriptors
H_Dz(p)	Harary-like index from Barysz matrix weighted by polarizability	2D matrix-based descriptors
VE2_Dz(p)	Average coefficient of the last eigenvector from Barysz matrix weighted by polarizability	2D matrix-based descriptors
SpMax5_Bh(e)	Largest eigenvalue n. 5 of Burden matrix weighted by Sanderson electronegativity	Burden eigenvalues
SpMax5_Bh(i)	Largest eigenvalue n. 5 of Burden matrix weighted by ionization potential	Burden eigenvalues
SM02_EA(ri)	Spectral moments of order 2 from edge adjacency mat. weighted by resonance integral	Edge adjacency indices
SM04_EA(ri)	Spectral moments of order 4 from edge adjacency mat. weighted by resonance integral	Edge adjacency indices
Ho_Dt	Hosoya-like index (log function) from detour matrix	2D matrix-based descriptors
SM3_Dt	Spectral moment of order 3 from detour matrix	2D matrix-based descriptors
VE2_Dt	Average coefficient of the last eigenvector from detour matrix	2D matrix-based descriptors
H_D/Dt	Harary-like index from distance/detour matrix	2D matrix-based descriptors
log $D_{7.4}$	Logarithm of distribution coefficient at pH 7.4	Molecular properties
nBM	Number of multiple bonds	Constitutional indices
SpMin6_Bh(p)	Smallest eigenvalue n. 6 of Burden matrix weighted by polarizability	Burden eigenvalues
log $D_{9.1}$	Logarithm of distribution coefficient at pH 9.1	Molecular properties

The coefficient plot (Figure 3) shows that all descriptors have a positive influence, with the exception of VE2_Dz(p), which has a negative influence on the log $k_{w4.4}$ value.
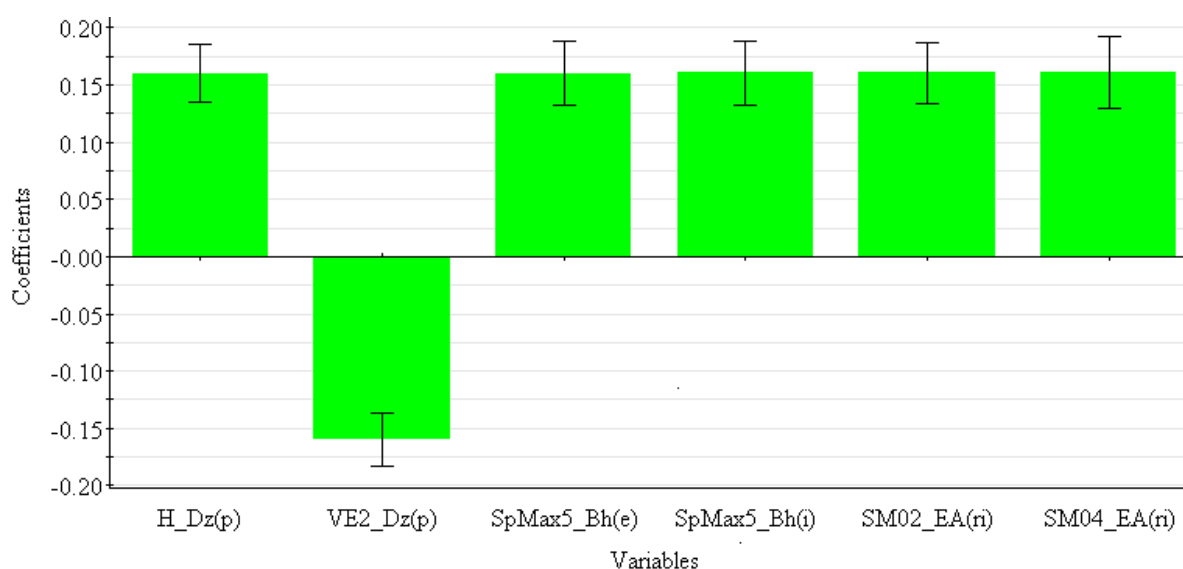

Figure 3. Coefficient plot of descriptors in PLS-QSRR (log $k_{w4.4}$) model.

2D matrix-based descriptors are topological indices computed by applying a set of basic algebraic operators to different graph-theoretical matrices representing an H-depleted molecular graph [29]. Depending on which matrix they are derived from (detour matrix Dt, Barysz matrices (Dz(w)), *etc.*), they can be divided into different sub-blocks. H_Dz(p) (Harary-like index from Barysz matrix weighted by polarizability) and VE2_Dz(p) (average coefficient of the last eigenvector from the Barysz matrix weighted by polarizability) are derived from the Barysz matrices. Barysz matrices (Dz(w)) are weighted distance matrices that take into account the presence of heteroatoms and multiple bonds in the molecule. Polarizability is an important atomic property included in both the VE2_Dz(p) and H_Dz(p) descriptors, which differ in their influence on log $k_{w4.4}$. Higher values of VE2_Dz(p) are associated with less lipophilic compounds in the studied group that have lower log $k_{w4.4}$ values (ephedrine, rilmenidine), while lower values are seen for the structures of carvedilol, clozapine and maprotiline with the highest log $k_{w4.4}$ values. The opposite effect can be observed for H_Dz(p). SpMax5_Bh(e) and SpMax5_Bh(i) are the largest eigenvalue n. 5 of the Burden matrix weighted by Sanderson electronegativity and ionization potential, respectively. Burden eigenvalues descriptors were originally proposed for searching for chemical similarity/diversity in large databases. They are calculated from the Burden matrix as an ordered sequence of the largest positive and smallest negative eigenvalues that reflect relevant aspects of molecular structure and can be useful for similarity searches [29,39].

SM02_EA(ri) and SM04_EA(ri) (spectral moments of order 2 and 4 from edge adjacency mat. weighted by resonance integral) belong to the Edge adjacency indices descriptors. These indices are based on the edge adjacency matrix of a graph. They encode information about the connectivity between graph edges, taking into account bonds between non-hydrogen atom pairs. Dragon calculates the Edge adjacency matrices using the following bond properties: edge degree (*ed*), dipole moment (*dm*), conventional bond order (*bo*) and parameters related to the resonance integral (*ri*). The spectral moment represents the linear combination of different structural fragments in the graph [29,40,41].

In contrast to chromatographic behaviour, electrophoretic mobility in an acidic medium is primarily determined by 2D autocorrelation descriptors (ATS6s, ATS7s, ATSC6m) (Table 6), which describe how the observed property (in this case, the intrinsic state in ATS6s and ATS7s and the mass in ATSC6m) is distributed along the topological structure.

Table 6. The most important descriptors in selected QSMR models

Selected descriptors	Description	Class of descriptors
Ho_Dz(e)	Hosoya-like index (log function) from Barysz matrix weighted by Sanderson electronegativity	2D matrix-based descriptors
ATS6s	Broto-Moreau autocorrelation of lag 6 (log function) weighted by I-state	2D autocorrelations
ATS7s	Broto-Moreau autocorrelation of lag 7 (log function) weighted by I-state	2D autocorrelations
ATSC6m	Centred Broto-Moreau autocorrelation of lag 6 weighted by mass	2D autocorrelations
GGI6	Topological charge index of order 6	2D autocorrelations
R7u+	R maximal autocorrelation of lag 7 / unweighted	GETAWAY descriptors
Mor28s	Signal 28 / weighted by I-state	3D-MoRSE descriptors
HATS2i	leverage-weighted autocorrelation of lag 2 / weighted by ionization potential	GETAWAY descriptors
HOMO [B3LYP/3-21G(d,p)]	Highest occupied molecular orbital	Quantum chemical
SM1_Dz(Z)	Spectral moment of order 1 from Barysz matrix weighted by atomic number	2D matrix-based descriptors
ISH	Standardized information content on the leverage equality	GETAWAY descriptors
SM15_EA(*dm*)	Spectral moment of order 15 from edge adjacency mat. weighted by dipole moment	Edge adjacency indices

This class of descriptors represents the homogeneity of the molecular structure and can be weighted by the intrinsic state (*s*), the Sanderson electronegativity (*e*), the atomic mass (*m*) and the van der Waals volume (*v*) [42]. H0_Dz(e) (Hosoya-like index (log function) from the Barysz matrix, weighted by the Sanderson electronegativity) is 2D-matrix-based descriptor and represents a measure of molecular size and branching [43]. From the coefficient plot (Figure 4), It can be concluded that all selected descriptors have a negative influence on the $\mu_{eff4.4}$ values.

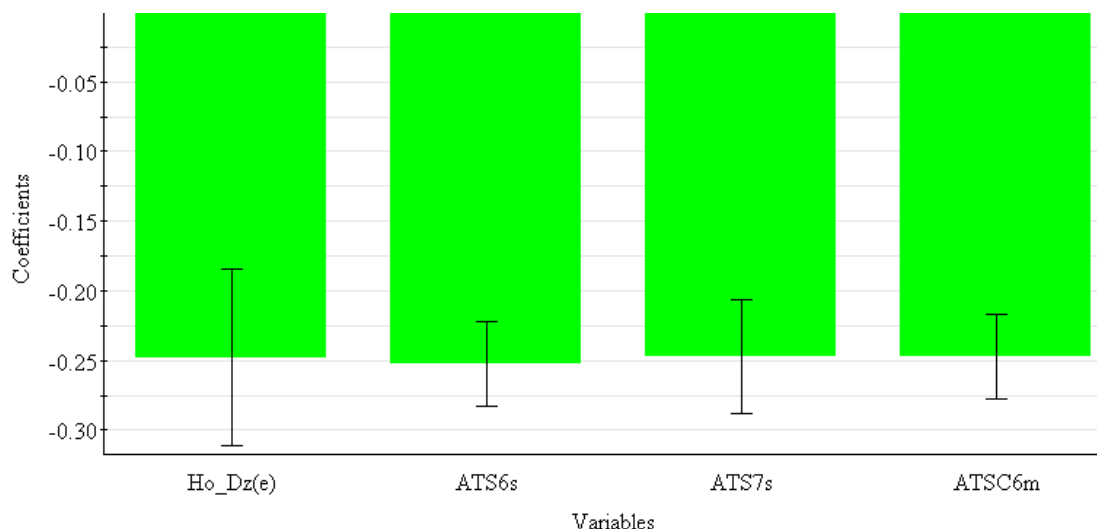**Figure 4**. Coefficient plot of descriptors in PLS-QSMR ($\mu_{eff4.4}$) model

At a physiological pH of 7.4, retention behaviour is a function of 2D matrix-based descriptors and molecular properties (Table 5).

Selected 2D matrix-based descriptors are derived from the detour matrix (Ho_Dt, SM3_Dt, VE2_Dt) and the distance/detour matrix (H_D/Dt), in contrast to the 2D matrix-based descriptors at pH 4.4. Ho_Dt is a Hosoya-like index (log function) from the detour matrix and represents a molecular descriptor that depends on the size of the molecule as well as its voluminosity [43-45]. Molecules with a simple structure, such as ephedrine and phenilephrine have the lowest value of Ho_Dt, while complex structures, such as those found in doxazosin and tetracyclic (mianserin) and tricyclic compounds (carvedilol, clozapine) have the highest value of this descriptor. H_D/Dt (Harary-like index from distance/detour matrix) is also a parameter sensitive to steric effects and whose value increases with increasing molecule size and branching [46]. Both descriptors are positively correlated with the log $k_{w7.4}$ data (Figure 5.), which means that more lipophilic compounds are characterised by a higher value of this descriptor.
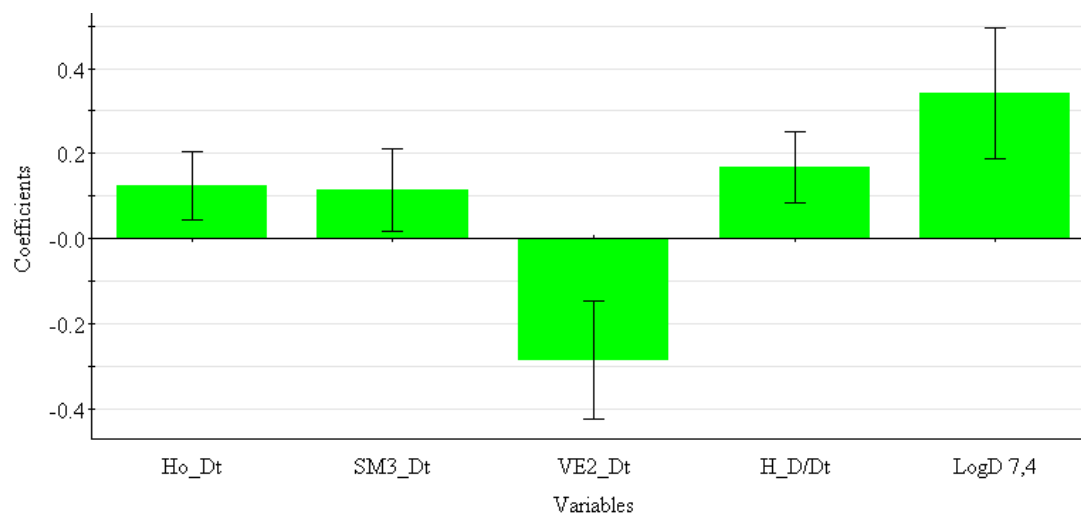

Figure 5. Coefficient plot of descriptors in PLS-QSRR (log $k_{w7.4}$) model

The logarithm of distribution coefficient at pH 7.4, log $D_{7.4}$ [28], is a measure of the lipophilicity of compounds at pH 7.4 and indicates the importance of considering the presence of ionic species when evaluating the lipophilicity of compounds at pH 7.4. The selected model shows that calculations of all proposed descriptors allow a more reliable prediction of compound lipophilicity and that the design of compounds with the desired properties is influenced by the presence of groups responsible for steric effects and the size of the molecule.

In capillary electrophoresis, the electrophoretic mobility of investigated compounds at pH 7.4 can be discussed on the basis of selected 2D autocorrelations (GGI6), GETAWEY and 3D-MoRSE descriptors (Table 6). In contrast to the PLS-QSMR ($\mu_{eff4.4}$) model, in which the class of 2D autocorrelation descriptors dominates, the 3D descriptors appear to be more important in the selected MLR-QSMR ($\mu_{eff7.4}$) model.

GETAWAY descriptors combine the 3D-molecular geometry provided by the molecular influence matrix and atomic relatedness through molecular topology with chemical information by using different atomic weights (atomic mass, polarizability, van der Waals volume and electronegativity) together with unit weights [47]. GETAWAYs (HATS2i, R7u+) are geometric descriptors that are sensitive to molecular branching and cyclicity, encode information about the effective position of substituents and fragments in molecular space, and may be suitable to describe differences in congeneric molecular series [29]. Equation 19 shows that both descriptors (HATS2i, R7u+) have a negative influence on the $\mu_{eff7.4}$ values.

3D-MoRSE descriptors (Mor28s) represent the 3D structure of molecules based on electron diffraction [48]. Mor28s is a signal 28 / weighted by I-state and, in contrast to the GETAWAY descriptors, has a positive influence on the $\mu_{eff7.4}$ parameter (Equation 19).

All molecular descriptors selected with the MLR-QSRR (log $k_{w9.1}$) model (Equation (15)) have a positive influence on log $k_{w9.1}$ and belong to constitutional indices, Burden eigenvalues and molecular properties classes of descriptors (Table 5).

 Constitutional descriptors such as the selected nBM (number of multiple bonds), which counts a molecule's double, triple and aromatic bonds, are the simplest and most commonly used descriptors. They represent the chemical composition of a compound without information about the molecular geometry or the connectivity of the atoms [29]. SpMin6_Bh(p) is the smallest eigenvalue n.6 of the Burden matrix weighted by polarizability and belongs to the class of Burden eigenvalue descriptors. The lipophilicity of compounds at pH 9.1 (log $D_{9.1}$) [28] appears again as an important molecular property considering all forms of the molecule (ionised and unionised), and its high positive correlation with log $k_{w9.1}$ suggests that more lipophilic compounds have a higher affinity for the apolar stationary phase and greater retention in HPLC.

In capillary electrophoresis at pH 9.1, the descriptors selected using the MLR method (Equation (21)) belong to the 2D matrix-based descriptors, GETAWAY, Edge adjacency indices and quantum chemical descriptors (Table 6).

SM15_EA(dm) is the spectral moment of order 15 from the edge adjacency matrix, weighted by the dipole moment, and is the only descriptor in a model that shows a positive correlation with $\mu_{eff9.1}$. In contrast, toSM15_EA(dm), SM1_Dz(Z) (spectral moment of order 1 from the Barysz matrix weighted by atomic number) belongs to the 2D matrix-based descriptors. ISH (standardised information content on leverage equality) is a class of GATAWEY descriptors and mainly encodes information about molecular symmetry. If all atoms have different leverage values, the molecule has no symmetry and ISH=1, otherwise a theoretically perfectly symmetric molecule has ISH=0. In the case of the tested compounds, all ISH values are above 0.82 and show a negative correlation with electrophoretic mobility, suggesting that compounds with a higher degree of symmetry have lower mobility under the given experimental conditions of background electrolyte pH 9.1. This

descriptor also provides information about molecular entropy, so it can be useful in modelling physicochemical properties related to entropy and symmetry [47]. The highest occupied molecular orbital (HOMO) energy calculated using the B3LYP/3-21G(d,p) method [27] is negatively correlated with $\mu_{eff9.1}$ (Equation (21)), implying that compounds with a lower HOMO energy have greater mobility because they interact less with positively charged buffer cations against the wall of the capillary.

Conclusions

Evaluation of the chromatographic and electrophoretic behaviour of a series of 29 imidazoline and alpha adrenergic receptors ligands at pH 4.4, 7.4, and 9.1 using QSRR and QSMR models allows several conclusions to be drawn. When comparing the statistical PLS and MLR methods used for model building, it is not possible to give preference to only one method, as the optimal models were formed by both. All selected QSRR models: PLS-QSRR (log $k_{w4.4}$), PLS-QSRR (log $k_{w7.4}$), MLR-QSRR (log $k_{w9.1}$) and QSMR models: PLS-QSMR ($\mu_{eff4.4}$), MLR-QSMR ($\mu_{eff7.4}$), MLR-QSMR ($\mu_{eff9.1}$) displayed high accuracy in retention/migration prediction for internal and external test set compounds. Based on the established QSRR and QSMR models, it can be seen that different classes of descriptors appear in the models created at different pH values and for different analytical systems, such as HPLC and CE. Chemical composition of compounds, lipophilicity, and voluminosity have the strongest influence on the retention mechanism, while molecular size, mass, complexity, charge and electronic properties are valuable structural determinants for the electrophoretic mobility of the examined ligands. The created QSRR and QSMR models are a very useful predictive tools for retention and electro-phoretic behaviour of new imidazoline/alpha adrenergic receptors ligands at three pH values, which could be used not only during the RP-HPLC/CE method development but also in the initial phase of design of novel imidazoline and alpha adrenergic receptors ligands with optimized physicochemical properties, such as lipophilicity (log k_w values) and charge-to-mass characteristics (μ_{eff} values).

Conflict of interest: *The authors declare no conflict of interest.*

References

[1] C. Dardonville, I. Rozas. Imidazoline binding sites and their ligands: an overview of the different chemical structures. *Medicinal Research Reviews* **24** (2004) 639-661. https://doi.org/10.1002/med.20007

[2] G.A. Head, D.N. Mayorov. Imidazoline receptors, novel agents and therapeutic potential. *Hematological Agents in Medicinal Chemistry* **4** (2006) 17-32. https://doi.org/10.2174/187152506775268758

[3] P. Bousquet, A. Hudson, J.A. García-Sevilla, J.X. Li. Imidazoline receptor system: the past, the present, and the future. *Pharmacological Reviews* **72** (2020) 50-79. https://doi.org/10.1124/pr.118.016311

[4] K. Nikolic, D. Agbaba. Imidazoline antihypertensive drugs: Selective I1-Imidazoline receptors activation. *Cardiovascular Therapeutics* **30** (2012) 209-216. https://doi.org/10.1111/j.1755-5922.2011.00269.x

[5] G.A. Head, S.L. Burke. I1 imidazoline receptors in cardiovascular regulation: the place of rilmenidine. *American Journal of Hypertension* **13** (2000) 89S-98S. https://doi.org/10.1016/S0895-7061(00)00224-7

[6] J.X. Li, Y. Zhang. Imidazoline I2 receptors: target for new analgesics? *European Journal of Pharmacology* **658** (2011) 49-56. https://doi.org/10.1016/j.ejphar.2011.02.038

[7]   J.H. Li. Imidazoline I2 receptors: An update. *Pharmacology & Therapeutics* **178** (2017) 48-56.
        http://dx.doi.org/10.1016/j.pharmthera.2017.03.009

[8]   A. Bagán, S. Rodriguez-Arévalo, T. Taboada-Jara, C. Griñán-Ferré, M. Pallàs, I. Brocos-Mosquera, L. F.
        Callado, J.A. Morales-García, B. Pérez, C. Diaz, R. Fernández-Godino, O. Genilloud, M. Beljkas, S. Oljacic,
        K. Nikolic, C. Escolano. Preclinical Evaluation of an Imidazole-Linked Heterocycle for Alzheimer's
        Disease*. Pharmaceutics* **215** (2023) 2381. https://doi.org/10.3390/pharmaceutics15102381

[9]   N.G. Morgan, S.L. Chan. Imidazoline binding sites in the endocrine pancreas: can they fulfill their
        potential as targets for the development of new insulin secretagogues? *Current Pharmaceutical
        Design* **7** (2001) 1413-1431. https://doi.org/10.2174/1381612013397366

[10]   H. van de Waterbeemd, E. Gifford. ADMET in silico modelling: towards prediction paradise? *Nature
        Reviews Drug Discovery* **2** (2003) 192-204. https://doir.org/10.1038/nrd1032

[11]   K. Vucićević, G. Popović, K. Nikolić, I. Vovk, D. Agbaba. An experimental design approach to selecting
        the optimum HPLC conditions for the determination of 2-arylimidazoline derivatives. *Liquid
        Chromatography & Related Technologies* **32** (2009) 656-667.
        https://doi.org/10.1080/10826070802711113

[12]   S. Filipić, D. Ružić, J. Vucićević, K. Nikolić, D. Agbaba. Quantitative structure-retention relationship of
        selected imidazoline derivatives on α1-acid glycoprotein column. *Journal of Pharmaceutical and
        Biomedical Analysis* **12** (2016) 101-111. http://dx.doi.org/10.1016/j.jpba.2016.02.053

[13]   D. Obradović, S. Oljačić, K. Nikolić, D. Agbaba. Investigation and prediction of retention
        characteristics of imidazoline and serotonin receptor ligands and their related compounds on mixed-
        mode stationary phase. *Journal of Chromatography A* **1585** (2019) 92-104.
        https://doi.org/10.1016/j.chroma.2018.11.051

[14]   S. Filipic, K. Nikolic, I. Vovk, M. Krizman, D. Agbaba. Quantitative structure-mobility relationship
        analysis of imidazoline receptor ligands in CDs-mediated CE. *Electrophoresis* **34** (2013) 471-482.
        https://doi.org/10.1002/elps.201200171

[15]   R. Kaliszan. QSRR: Quantitative Structure-(Chromatographic) Retention Relationships. *Chemical
        Reviews* **107** (2007) 3212–3246. https://doi.org/10.1021/cr068412z

[16]   R. Kaliszan. Chromatography and capillary electrophoresis in modelling the basic processes of drug
        action. *TrAC Trends in Analytical Chemistry* **18** (1999) 400-410. https://doi.org/10.1016/S0165-
        9936(99)00120-X

[17]   J.X. Soares, Á. Santos, C. Fernandes, M.M.M. Pinto. Liquid chromatography on the different methods
        for the determination of lipophilicity: an essential analytical tool in medicinal chemistry.
        Chemosensors **10** (2022) 10, 340. https://doi.org/10.3390/chemosensors10080340

[18]   D. Obradović, Ł. Komsta, V.M. Petrović, I. Stojković, S. Lazović. An alternative biomimetic tool – Dual
        hydrophilic/reversed-phase interaction mode. *Microchemical Journal* **193** (2023) 108967
        https://doi.org/10.1016/j.microc.2023.108967

[19]   M. Koba, T. Bączek, M.P. Marszałł. Importance of retention data from affinity and reverse-phase
        high-performance liquid chromatography on antitumor activity prediction of imidazoacridinones
        using QSAR strategy. *Journal of Pharmaceutical and Biomedical Analysis* **64-65** (2012) 87-93.
        https://doi.org/10.1016/j.jpba.2012.02.010

[20]   K. Nikolic, S. Filipic, A. Smoliński, R. Kaliszan, D. Agbaba. Partial least square and hierarchical
        clustering in ADMET modelling: prediction of blood – brain barrier permeation of α-adrenergic and
        imidazoline receptor ligands. *The Journal of Pharmacy & Pharmaceutical Sciences* **16** (2013) 622–647.
        https://doi.org/10.18433/j3jk5p

[21]   J. Vucićević, K. Nikolić, V. Dobričić, D. Agbaba. Prediction of blood–brain barrier permeation of a-
        adrenergic and imidazoline receptor ligands using PAMPA technique and quantitative-structure
        permeability relationship analysis. *European Journal of Pharmaceutical Sciences* **68** (2015) 94-105.
        http://dx.doi.org/10.1016/j.ejps.2014.12.014

[22]   A. Golbraikh, A. Tropsha. Beware of q2! *Journal of Molecular Graphics and Modelling* **20** (2002) 269-
        276. https://doi.org/10.1016/S1093-3263(01)00123-1

[23] A. Tropsha, Best practices for QSAR model development, validation and exploitation, Mol. Inform. **29** (2010) 476–488. https://doi.org/10.1002/minf.201000061

[24] P.K. Ojha, I. Mitra, R.N. Das, K. Roy. Further exploring r_m^2 metrics for validation of QSPR models data set. *Chemometrics and Intelligent Laboratory Systems* **107** (2011) 194–205. https://doi.org/10.1016/j.chemolab.2011.03.011

[25] A.D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **98** (1993) 5648-5652. https://doi.org/10.1063/1.464913

[26] C. Lee, W. Yang, R.G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B* **37** (1988) 785-789. https://doi.org/10.1103/physrevb.37.785

[27] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A.Jr. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, , C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B.G. Johnson, W. Chen, M.W. Wong, J.L. Andres, M. Head-Gordon, E.S. Replogle, J.A. Pople. *Gaussian 98* (revision A.7), Gaussian, Inc., Pittsburgh, PA 1998. https://gaussian.com/

[28] ChemAxon Marvin 5.5.1.0 program, Budapest, Hungary, 2011. www.chemaxon.com/products.html

[29] Dragon 6, TALETE srl, Via V. Pisani, 13 - 20124 Milano – Italy. http://www.talete.mi.it

[30] CS Chem3D Ultra 7.0, Cambridge Soft Corporation, (Property Picker ActiveX Control), 100 Cambridge Park Dr. Cambridge, MA 02140-2317 U.S.A., 2001. http://www.cambridgesoft.com/

[31] R.G. Parr, R.G. Pearson. Absolute hardness: companion parameter to absolute electronegativity. *Journal of the American Chemical Society* **105** (1983) 7512-7516. https://doi.org/10.1021/ja00364a005

[32] A. Rácz, D. Bajusz, K. Héberger. Intercorrelation Limits in Molecular Descriptor Preselection for QSAR/QSPR. *Molecular Informatics* **38** (2019) 1800154. https://doi.org/10.1002/minf.201800154

[33] Umetrics AB, SIMCA P+ program, Version 12.0.0.0, Umeå, May 20, 2008, www.umetrics.com.

[34] S. Wold, E. Johansson, M. Cocchi, in: H. Kubinyi, (Ed.), 3D QSAR in Drug Design, Theory, Methods, and Applications, ESCOM Science Publishers, Leiden 1993, pp. 523-550.

[35] P.K. Ojha, K. Roy. Comparative QSARs for antimalarial endochins:importance of descriptor thinning and noise reduction prior to feature selection. *Chemometrics and Intelligent Laboratory Systems* **109** (2011) 146-161. https://doi.org/10.1016/j.chemolab.2011.08.007

[36] STATISTICA Neural Networks 4.0, StatSoft, Inc., Tulsa, OK, 1998.

[37] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikstrom, S. Wol (eds.), *Multi-and Mega-variate Data Analysis. Basic Principles and Applications* I, 2nd edn. Umetrics Academy, Umeå, 2001.

[38] C. Giaginis, A. Tsantili-Kakoulidou. Current state of the art in HPLC methodology for lipophilicity assessment of basic drugs. A Review. *Journal of Liquid Chromatography & Related Technologies* **31** (2008) 79-96. https://doi.org/10.1080/1082607070166562679

[39] F. Burden. Molecular Identification Number for Substructure Searches. *The Journal for Chemical Information and Computer scientists* **29** (1989) 225-227. https://doi.org/10.1021/ci00063a011

[40] E. Estrada. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. *The Journal for Chemical Information and Computer scientists* **36** (1996) 844-849. https://doi.org/10.1021/ci950187r

[41] S. Marković, I. Gutman. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. Benzenoid Hydrocarbons. *The Journal for Chemical Information and Computer scientists* **39** (1999) 289-293. https://doi.org/10.1021/ci980032u

[42] K. Bodzioch, A. Durand, R. Kaliszan, T. Baczek, Y. Vander Heyden. Advanced QSRR modelling of peptides behaviour in RPLC. *Talanta* **81** (2010) 1711–1718. https://doi.org/10.1016/j.talanta.2010.03.028

[43]   M. Randić. Wiener-Hosoya Index A Novel Graph Theoretical Molecular Descriptor. *The Journal for Chemical Information and Computer scientists* **44** (2004) 373-377. https://doi.org/10.1021/ci030425f

[44]   M. Randić. On Characterization of Cyclic Structures. *The Journal for Chemical Information and Computer scientists* **37** (1997) 1063-1071. https://doi.org/10.1021/ci9702407

[45]   R. Todeschini, V. Consonni. Handbook of Molecular Descriptors, Wiley-VCH, Weinheim, 2000.

[46]   B. Lučić, A. Miličević, S. Nikolić, N. Trinajstić. Harary Index – Twelve Years Later. *Croatica Chemica Acta* **75** (2002) 847-868. https://hrcak.srce.hr/131738

[47]   V. Consonni, T. Todeschini, M. Pavan. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *The Journal for Chemical Information and Computer scientists* **42** (2002) 682-692. https://doi.org/10.1021/ci015504a