Talanta 100 (2012) 329–337

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Quantitative structure–retention relationships of azole antifungal agents in reversed-phase high performance liquid chromatography

Jelena Golubović [a], Ana Protić [a], Mira Zečević [a], Biljana Otašević [a,*], Marija Mikić [b], Ljiljana Živanović [a]

[a] University of Belgrade—Faculty of Pharmacy, Department of Drug Analysis, Vojvode Stepe 450, 11 152 Belgrade, Serbia
[b] University of Belgrade—Faculty of Mathematics, Department of Differential Equations, Studentski Trg 16, 11 000 Belgrade, Serbia

ARTICLE INFO

Article history:
Received 23 March 2012
Received in revised form
24 July 2012
Accepted 27 July 2012
Available online 11 August 2012

Keywords:
QSRR
Artificial neural networks
Antifungal agents
Azoles
HPLC

ABSTRACT

Artificial neural network (ANN) is a learning system based on a computational technique which can simulate the neurological processing ability of the human brain. It was employed for building of the quantitative structure–retention relationships (QSRRs) model of antifungal agents—imidazoles or triazoles by structure. Computed molecular descriptors together with the percentage of acetonitrile in mobile phase (v/v) and buffer pH, being the most influential HPLC factors, were used as network inputs, giving the retention factor as model output. The multilayer perceptron network with a 9-5-1 topology was trained by using the back propagation algorithm. Good correlation between experimentally obtained data and ones predicted by using QSRR-ANN on previously unseen data sets indicates good predictive ability of the model.

1. Introduction

The azole antifungal agents in clinical use contain either two or three nitrogens in the azole ring and are thereby classified as imidazoles or triazoles. They inhibit ergosterol synthesis, the main sterol constituent of fungal membranes, by blocking the cytochrome P450-dependent enzyme lanosterol 14-α-demethylase. Lack of ergosterol and accumulation of 14-α-methylated precursors result in dysfunction of membrane fluidity and the activities of several enzymes located in the membrane (e.g. chitin synthase) [1,2].

Azoles can be analyzed by many different techniques, including spectrophotometric [3], electrochemical [4], and gas chromatographical techniques [5]. Reference pharmacopoeial elaborations [6–8] require for azoles, in most cases, titrimetric method in waterless environments—acidimetric determination using perchloric acid as a titrant.

It is important to note that titrimetric and spectrophotometric methods are suitable for analyzing active pharmaceutical ingredients (API) per se, but are not proper for their determination in complex matrices such as drug products (DP). In such cases, HPLC methods are recommended. There are many published RP-HPLC methods for determination of azoles [9–12] which served as the starting point for this study in terms of experimental conditions.

Quantitative structure–retention relationships (QSRRs) represent a powerful technique for relating the chromatographic retention parameters of groups of analytes and their descriptors, which are quantities encoding the structural characteristics [13–15]. The QSRR approach as a potential tool for optimizing separation of complex mixtures has been largely proved by accurate prediction of solute retention for many compound classes. Further, QSRR studies can significantly contribute to clarify molecular mechanisms of chromatographic retention [9,16,17].

A molecular descriptor is the final result of a logical and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment. The choice of descriptors was based on the widely accepted assumption that retention is governed by intermolecular interactions, as suggested by the fundamental theory of liquid chromatography. In this view, quantum chemical calculations were used to derive electronic and geometrical properties able to describe dispersive, polar and hydrogen bonding interactions widely recognized, together with cavity effect, as driving forces for solute partition between the chromatographic phases. In addition to the effect of the solute molecular structure on the retention, that of the mobile phase composition is investigated by including the organic modifier concentration as an independent variable of the QSRR model [15,16].

The QSRR model has been built employing an artificial neural network (ANN). The ANNs are difficult to describe with a simple definition. Maybe the closest description would be a comparison

* Corresponding author. Tel.: +381 11 39 51384; fax: +381 11 39 72840.
E-mail address: bjocic@pharmacy.bg.ac.rs (B. Otašević).

0039-9140/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.talanta.2012.07.071

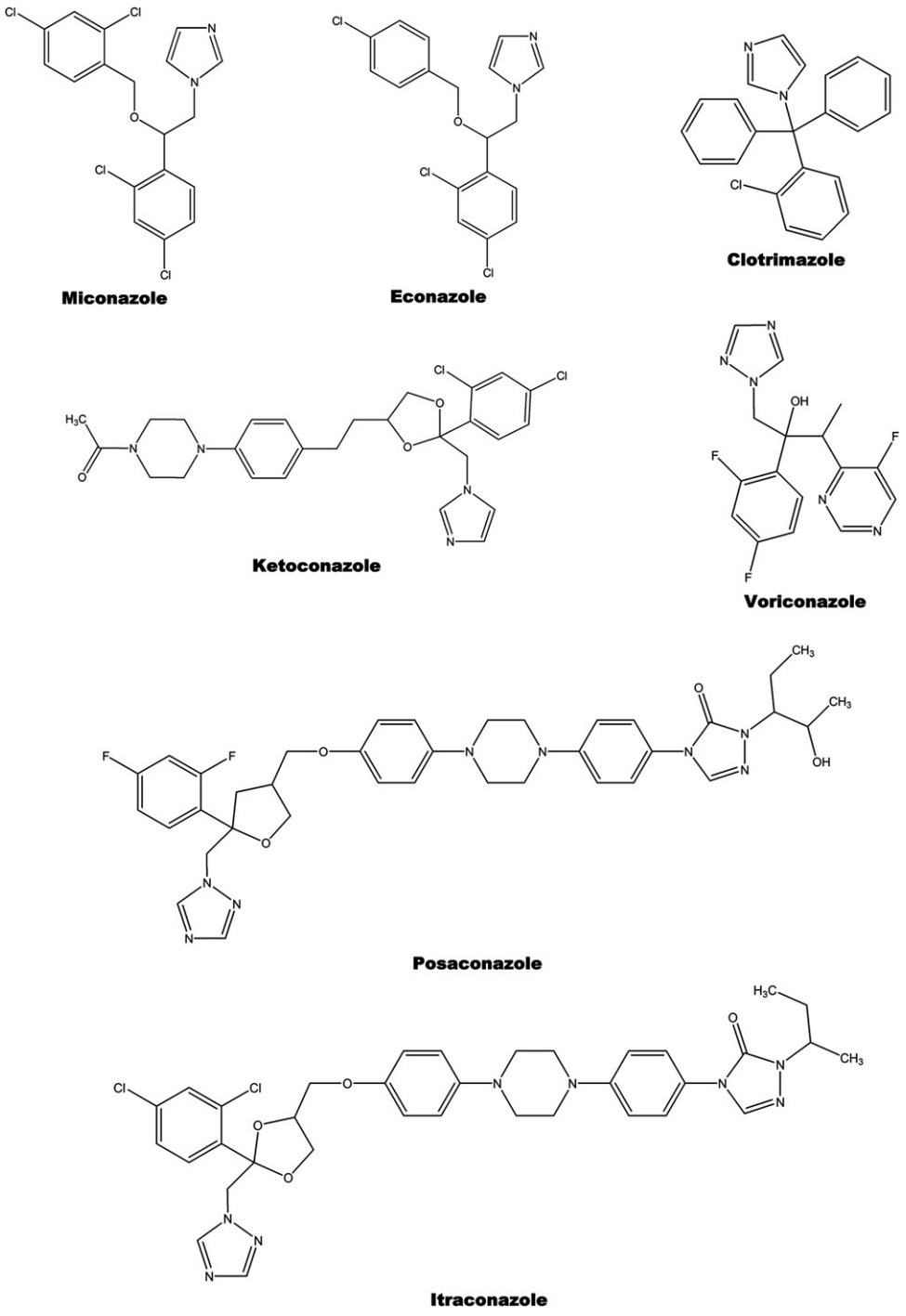

Fig. 1. Chemical structures of the azoles.

with a black box having multiple inputs and multiple outputs which operate using a large number of mostly parallel connected simple arithmetic units. The most important thing to remember about all ANN methods is that they work best if they are dealing with non-linear dependence between the inputs and outputs [18–20]. An important advantage of ANN compared with classical statistical methods is that it does not require preliminary knowledge of the mathematical form of the relationship between the variables [21]. ANN has been chosen as it shows better results in retention prediction than other techniques such as multilinear regression (MLR) [16,17,22].

To the best of our knowledge, no paper about using QSRR-ANN as a predictive tool for UHPLC analysis has been published so far. The study included seven antifungal agents, azoles by structure: miconazole, econazole, ketoconazole, clotrimazole, itraconazole, posaconazole and voriconazole (Fig. 1), which are considered to generate a model. All UHPLC factors which showed a statistically significant influence on the retention behavior of the investigated azoles were included in model building. In previously published papers it can be seen that the percentage of acetonitrile and pH of the mobile phase have been often included in the QSRR model [16,17,22]. This was based on the assumption that usually these factors have the greatest influence on the retention properties of the investigated substances. But, in this work, all potentially relevant chromatographic factors were firstly investigated through fractional factorial design (FFD) and the ones which showed statistically

significant influence on the retention behavior of azoles were further included in the QSRR model.

2. Experimental

2.1. Solvents and chemicals

Miconazole, econazole, ketoconazole, clotrimazole, itraconazole, posaconazole and voriconazole and fluconazole standards were obtained from Selectchemie AG (Zürich, Switzerland) and were used without further purification. HPLC-grade acetonitrile and HPLC-grade methanol were purchased from Sigma Aldrich Chemie GmbH, Taufkirchen, Germany. Distilled water, obtained from a Simplicity 185 purification system, Millipore (Billerica, MA, USA) was used for preparation of the mobile phase. Ammonium acetate and acetic acid used for preparing buffers were purchased from Sigma Aldrich. pH of the buffer was adjusted by addition of acetic acid by means of a PHM210 Standard pH-meter (Radiometer Analytical SAS, France) equipped with a glass electrode. Triethylamine, added to the mobile phase was purchased from Fisher Scientific UK Limited, UK. Before use, the mobile phase was vacuum filtered through 0.45 μm nylon membranes (Agilent Technologies, Santa Clara, USA).

2.2. Sample preparation

Stock solutions (1 g/L) were prepared by dissolving accurately weighed 10 mg of each azole in 10 mL of HPLC-grade methanol and stored at 4 °C. These solutions were used to prepare the standard working samples by appropriate dilution. In accordance with the corresponding spectral properties of azoles, stock solutions were diluted with mobile phase to attain the following sample concentrations: clotrimazole —0.1 mg/mL, ketoconazole and fluconazole —0.2 mg/mL, while other azoles were analyzed in concentrations of 0.5 mg/mL.

2.3. Instrumentation

Retention of the azoles was investigated by using Thermo Scientific Accela UHPLC apparatus (Thermo Fisher Scientific Inc.), equipped with an autosampler, degasser and photodiode array detector. The analyses were performed on a Hypersil column (50 mm length, 4.6 mm i.d., 1.9 μm, Thermo Fisher Scientific Inc.). Data were recorded and analyzed with the ChromQuest software version 5.0. The injected sample volume was 5 μL.

2.4. Determination of retention parameters

The UHPLC analyses were carried out under isocratic conditions at column temperatures ranging from 25 to 40 °C with a flow-rate of 400 mL/min. The retention behavior of the analytes was investigated as a function of mobile phase composition ranging from (40: 60, v/v) acetonitrile–ammonium acetate buffer to (60:40, v/v) acetonitrile acetate buffer (v/v) acetonitrile by steps of 5%. The absorbance of the analytes during a chromatographic run was collected in the spectral range 210–380 nm. The detection wavelength was the one providing the maximum peak height: 240 nm.

2.5. Computation of molecular descriptors

For all examined compounds, dominant forms at different analytical pH values have been obtained using Marvin Sketch 4.1.13, Chem Axon Ltd. Each structure was subjected to energy minimization by the semi-empirical MOPAC/AM_1 method of Chem 3D® Pro, Cambridge Soft Corporation. The minimum energy structures of the compounds were then used to calculate all of the molecular descriptors used during the prediction model construction. The octanol/water distribution coefficient ($\log D$), polarizability (POL), H-donor sites (H-don) and H-acceptor sites (H-acc) were calculated by means of Marvin Sketch., whereas the Connolly solvent accessible area (SAS), molar refractivity (MR), dipole–dipole energy (DEN), molecular area (MA), solvent-excluded volume (SEV), van der Waals energy (VDW), non-1,4 van der Waals energy (NON VDW), diameter (D), highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) were generated using Chem 3D® Pro, Cambridge Soft Corporation.

Correlations between descriptors were examined by means of a simple linear regression analysis using Microsoft Office Excel 2007.

2.6. Artificial neural network modeling

ANN topologies or architecture are formed by organizing nodes into layers and linking these layers of neurons with modifiable weighted interconnections [23]. Among the different kinds of ANNs, multi-layer feed-forward networks are most often used in the structure–property relationship analysis. Commonly, they consist of three layers: one input layer formed by a number of neurons equal to the number of descriptors, plus a bias term for intercept, one output neuron (providing the model response) and a number of hidden neurons fully connected to both input and output neurons. Information that propagates from input towards output neurons is modulated by modifiable weights associated to each connection. The post-synaptic potential function is the dot product of the weight vector with the input vector plus a bias value. Weighted signals entering the operative (hidden or output) neurons are transformed by an activation function into the neuron output which, in the case of hidden neurons, is transferred to the next layer or, in the case of the output neuron, is the final network response [16]. Network training consists of an iterative progression of algorithm through a number of epochs. The aim of training is maximizing the overall agreement between computed and target outputs for a set of examples (training set). These outputs are compared in each epoch directing the adjustments of weights.

On the other hand, a network should be applicable on the cases-ability of generalization. At some moment, generalization ability progressively deteriorates as a consequence of overfitting. To avoid this, the predictive power of the network is evaluated after each weight adjustment on unknown data (validation set). The minimum of the validation error is taken as a suitable criterion to define the optimal duration of learning for a given network, or to select among alternative trained networks the one with the expected best predictive capability. After optimization, the actual predictive performance of the trained network was evaluated using an external validation data set, here called test set [16].

In this paper, data of six molecules are used to build the ANN-based model, while the seventh molecule (econazole) is preserved to finally evaluate the model quality. Econazole was selected as a test substance because none of its descriptors had a limit value compared to the other azoles' descriptors. The remainder data points were randomly partitioned to the training or validation data set, so that the test data set included 125 cases, while the validation data set included 25 cases.

3. Results and discussion

3.1. Descriptor selection

If one was capable of predicting the retention of analytes and/or the separation of a mixture on chromatographic systems

relatively well, then the theoretical approach could, in some part, replace the time consuming experimental approach. The QSRR establishes the relations between retention data and molecular structure. Molecular structure presents an important factor for the QSRR model and is encoded by descriptors [13,14].

By descriptor selection we encompassed all major groups of descriptors as physicochemical, quantumchemical, topological and spatial structural descriptors. In addition, we selected only those which are not highly correlated to each other. Descriptors with correlation coefficients higher than 0.990 were not all considered in the ANN analysis. The correlation analysis was carried out to evaluate whether similar chemical information was encoded by two or more descriptors, and, in such cases, to eliminate redundant descriptors. Each of the correlated descriptors was tested for correlation with retention time using the non-linear neural networks model. A couple of them, including MA,

Table 1
Correlation coefficients between descriptor pairs.

	SAS	MA	SEV	MR	DEN	NON VDW	VDW	*D*	H-don	H-acc	LUMO	HOMO	POL	Log *D*
SAS	1													
MA	0.9993	1												
SEV	0.9939	0.9971	1											
MR	−0.1798	−0.1778	−0.1840	1										
DEN	−0.6373	−0.6347	−0.6338	0.2875	1									
NON VDW	0.2940	0.3215	0.3638	0.1674	−0.2521	1								
VDW	0.9931	0.9961	0.9989	−0.1803	−0.6314	−0.2521	1							
D	0.9938	0.9950	0.9918	−0.1580	−0.5765	−0.6314	0.9886	1						
H-don	0.0358	0.0240	0.0133	−0.0665	−0.2516	−0.5765	0.0289	0.0074	1					
H-acc	0.8660	0.8664	0.8663	−0.2367	−0.4517	−0.2516	0.8466	0.8892	−0.0921	1				
LUMO	0.9912	0.9930	0.9946	−0.2104	−0.6620	−0.4517	0.9904	0.9876	0.0344	0.9004	1			
HOMO	0.9912	0.9930	0.9946	−0.2104	−0.6620	−0.6620	0.9904	0.9876	0.0344	0.9004	1	1		
POL	0.9980	0.9986	0.9982	−0.1471	−0.6962	0.1764	0.9983	0.9907	0.0378	0.8295	0.9879	0.9879	1	
Log *D*	0.7273	0.7331	−0.7312	0.7820	−0.4634	0.6855	0.7425	0.7021	−0.5385	0.3683	0.6727	0.6727	0.7435	1

Table 2
Experimental plan of FFD and related retention factors.

Experiment no.	Variables				Retention factors						
	A^a	B^b	C^c	D^d	Miconazole	Econazole	Ketoconazole	Clotrimazole	Itraconazole	Posaconazole	Voriconazole
1	−1	−1	−1	−1	13.15	7.98	1.87	6.01	42.64	11.1	2.26
2	1	−1	−1	1	2.46	1.63	0.53	1.93	3.28	1.26	0.76
3	1	1	−1	−1	7.58	3.55	0.78	2.6	2.39	0.8	0.41
4	1	−1	1	−1	2.47	1.19	0.27	1.21	2.55	0.68	0.3
5	−1	−1	1	1	10.03	6.01	1.81	5.68	31.35	9.45	1.85
6	1	1	1	1	3.46	2.51	0.69	1.78	2.04	0.77	0.37
7	−1	1	−1	1	83.14	36.98	6.58	20.62	38.55	9.58	1.72
8	−1	1	1	−1	46.21	22.45	5.6	13	31.72	9.08	1.51

[a] A: Percentage of acetonitrile in mobile phase: 40% (−1), and 60% (1).
[b] B: pH of water phase: 3.5 (−1), and 6 (1).
[c] C: Column temperature: 20 (−1), and 50 (1).
[d] D: Percentage of TEA in water phase: 0.01% (−1), and 0.1% (1).

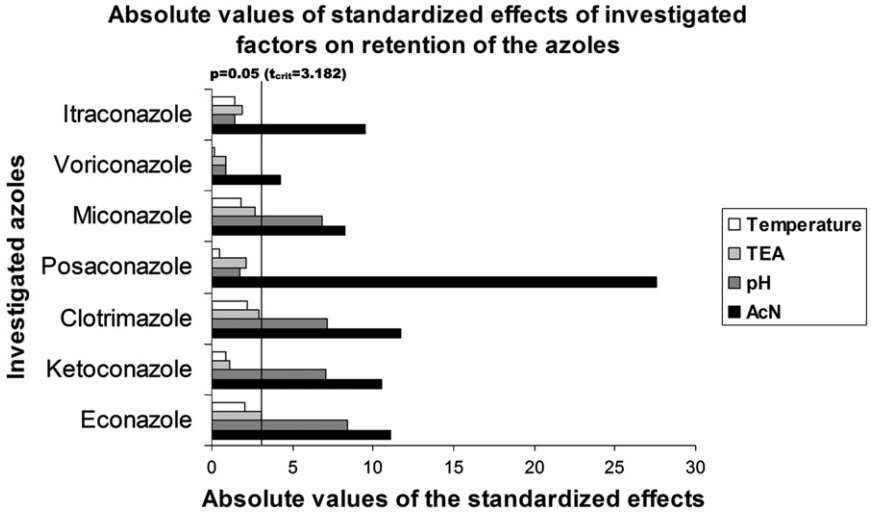

Fig. 2. Absolute values of standardized effects of investigated factors on retention of the azoles.

Table 3
Results of ANN computing and prediction.

Drug	*k*(exp)	Data set	*k*(ANN)	Error
Miconazole	2.59	Training	2.382273	−0.207727
	3.46	Training	2.947798	−0.5122
	4.95	Training	3.717235	−1.232765
	7.66	Training	5.472949	−2.187051
	13.39	Training	12.03887	−1.351132
	5.3	Training	4.711562	−0.5884
	6.77	Training	6.323794	−0.4462
	9.62	Training	8.66004	−0.96
	14.97	Training	13.53277	−1.437227
	25.72	Training	28.03258	2.312584
	7.1	Training	6.832136	−0.2679
	10.45	Training	10.31613	−0.1339
	16.6	Training	16.35945	−0.2406
	27.16	Training	28.53076	1.370756
	49.88	Training	53.08393	3.203934
	7.95	Training	7.379806	−0.5702
	12.71	Validation	12.02064	−0.6894
	21.79	Training	21.72204	−0.06796
	39.63	Training	42.28254	2.652543
	78.79	Training	71.11135	−7.678654
	7.85	Training	6.78699	−1.06301
	12.55	Training	11.2514	−1.298599
	21.64	Training	22.02801	0.3880146
	41.8	Training	47.65346	5.853464
	89.61	Training	77.8926	−11.7174
Ketoconazole	0.59	Training	0.6986098	0.1086098
	0.68	Training	0.679534	−0.000466
	0.87	Training	0.6827213	−0.1873
	1.24	Training	0.7421062	−0.4979
	2	Training	1.115711	−0.8843
	0.9	Training	1.256159	0.356159
	1.09	Validation	1.175481	0.08548
	1.44	Training	1.141235	−0.2988
	2.03	Training	1.259509	−0.7705
	3.22	Training	2.027457	−1.192543
	1.09	Training	1.741918	0.6519177
	1.48	Training	1.75488	0.2748804
	2.14	Training	1.864517	−0.2755
	3.28	Training	2.396478	−0.8835
	5.46	Training	4.971727	−0.4883
	1.22	Training	2.002441	0.7824407
	1.69	Training	2.138994	0.4489937
	2.57	Training	2.4701	−0.0999
	4.17	Training	3.632829	−0.5372
	7.93	Validation	8.944775	1.014775
	1.16	Training	2.08659	0.9265902
	1.7	Training	2.300524	0.6005236
	2.58	Training	2.802796	0.2227958
	4.47	Training	4.546414	0.0764141
	8.74	Training	12.52243	3.782433
Posaconazole	1.37	Training	2.806591	1.436591
	1.97	Training	2.994487	1.024487
	3.15	Validation	3.334811	0.1848111
	5.67	Training	4.366538	−1.303462
	11.72	Training	8.522966	−3.197034
	1.37	Training	2.841467	1.471467
	2.03	Training	3.158004	1.128004
	3.35	Training	3.748208	0.3982078
	6.17	Validation	5.449124	−0.7209
	13.49	Validation	12.25344	−1.236564
	1.36	Training	2.659567	1.299567
	2.07	Training	2.983784	0.9137843
	3.41	Training	3.625022	0.2150224
	6.22	Training	5.558752	−0.6612
	13.29	Training	13.52565	0.2356532
	1.35	Training	2.433038	1.083038
	2.06	Training	2.730347	0.6703472
	3.39	Validation	3.352218	−0.03778
	6.15	Validation	5.320174	−0.8298
	13.49	Training	13.67918	0.1891798
	1.27	Training	2.224438	0.9544384
	1.95	Training	2.486208	0.5362079
	3.17	Training	3.065953	−0.104
	5.84	Validation	4.991913	−0.8481
	12.41	Training	13.40352	0.993517

Table 3 (*continued*)

Drug	*k*(exp)	Data set	*k*(ANN)	Error
Itraconazole	3.38	Training	6.729067	3.349067
	5.62	Training	7.659227	2.039227
	10.26	Training	10.06522	−0.1948
	19.99	Training	18.575	−1.415004
	45.3	Training	45.28886	−0.01114
	3.58	Validation	4.988301	1.408301
	5.9	Training	6.161907	0.2619072
	10.7	Training	9.457813	−1.242187
	21.72	Training	22.00376	0.2837569
	53.4	Validation	54.39898	0.9989835
	3.57	Training	4.600903	1.030903
	6.04	Training	5.775403	−0.2646
	11.04	Training	9.186252	−1.853748
	22.66	Validation	22.40395	−0.256054
	53.34	Training	54.75841	1.418413
	3.56	Training	4.095365	0.5353654
	6	Training	5.185054	−0.8149
	10.96	Training	8.467299	−2.492701
	22.22	Validation	21.50434	−0.7157
	55	Training	53.0129	−1.987097
	3.39	Validation	3.611526	0.2215259
	5.73	Validation	4.598149	−1.131851
	10.32	Training	7.692992	−2.627008
	20.9	Training	20.32327	−0.5767
	49	Training	50.75271	1.752708
Voriconazole	0.74	Training	0.6028858	−0.1371
	1.82	Training	0.6131097	−1.20689
	1.17	Training	0.6387763	−0.5312
	1.61	Training	0.70984	−0.90016
	2.34	Training	0.9812719	−1.358728
	0.74	Training	0.608318	−0.131682
	0.92	Training	0.6137527	−10.3062
	1.19	Training	0.6296436	−0.5604
	1.64	Training	0.6782669	−0.9617
	2.38	Training	0.89013	−1.48987
	0.71	Training	0.6344958	−0.0755
	0.89	Validation	0.6336279	−0.2564
	1.17	Validation	0.6450088	−0.525
	1.6	Training	0.6891859	−0.9108
	2.39	Training	0.8977122	−1.492288
	0.71	Validation	0.730115	0.02011
	0.9	Validation	0.7119707	−0.188
	1.18	Training	0.7169894	−0.463
	1.58	Training	0.7756998	−0.8043
	2.33	Validation	1.081624	−1.248376
	0.7	Training	0.9483476	0.2483476
	0.88	Training	0.9170993	0.0370993
	1.16	Validation	0.925953	−0.234047
	1.61	Training	1.050502	−0.5595
	2.33	Training	1.712228	−0.6178
Clotrimazole	2.06	Training	2.224767	0.1647669
	2.46	Training	2.33927	−0.1207
	3.19	Training	2.709153	−0.4808
	4.27	Training	3.667659	−0.602341
	6.27	Training	6.670405	0.400405
	3.07	Validation	3.734375	0.6643752
	4.11	Training	4.202328	0.09233
	5.93	Validation	5.176426	−0.7536
	8.89	Training	7.558226	−1.331774
	14.44	Training	15.05328	0.6132789
	3.39	Validation	4.431577	1.041577
	4.97	Validation	5.244337	0.2743371
	7.68	Training	6.821554	−0.8584
	12.48	Training	10.70337	−1.776629
	21.98	Training	23.14052	1.160524
	3.48	Training	4.473569	0.9935691
	5.12	Training	5.394808	0.2748083
	8.11	Training	7.166964	−0.943
	13.56	Training	11.67098	−1.889018
	25.9	Training	26.80276	0.9027582
	3.31	Training	4.259353	0.9493532
	4.95	Training	5.14774	0.1977395
	7.82	Training	6.86185	−0.95815
	13.26	Training	11.38475	−1.875253
	25.77	Training	27.46242	1.692417
Econazole	1.78	Test	1.296065	−0.4839

Table 3 (*continued*)

Drug	*k*(exp)	Data set	*k*(ANN)	Error
	2.26	Test	1.552738	−0.7073
	3.16	Test	2.011345	−1.148655
	4.85	Test	3.010152	−1.839848
	8.34	Test	6.474477	−1.865523
	3.1	Test	2.331403	−0.7686
	4.08	Test	2.868004	−1.211996
	5.74	Test	3.893769	−1.846231
	8.57	Test	6.283902	−2.286098
	14.52	Test	14.27614	−0.2439
	4.29	Test	3.512783	−0.7772
	6.29	Test	4.635233	−1.654767
	9.62	Test	6.934774	−2.685226
	14.07	Test	12.66218	−1.407821
	26.99	Test	29.99099	3.000985
	4.8	Test	4.149337	−0.6507
	7.48	Test	5.754592	−1.725408
	12.48	Test	9.374161	−3.105839
	21.51	Test	19.42613	−2.08387
	41.45	Test	47.14899	5.698985
	4.72	Test	4.182316	−0.5377
	7.42	Test	5.8611	−1.5589
	12.46	Test	9.917749	−2.542251
	22.76	Test	22.47101	−0.289
	46.69	Test	56.30841	9.618409

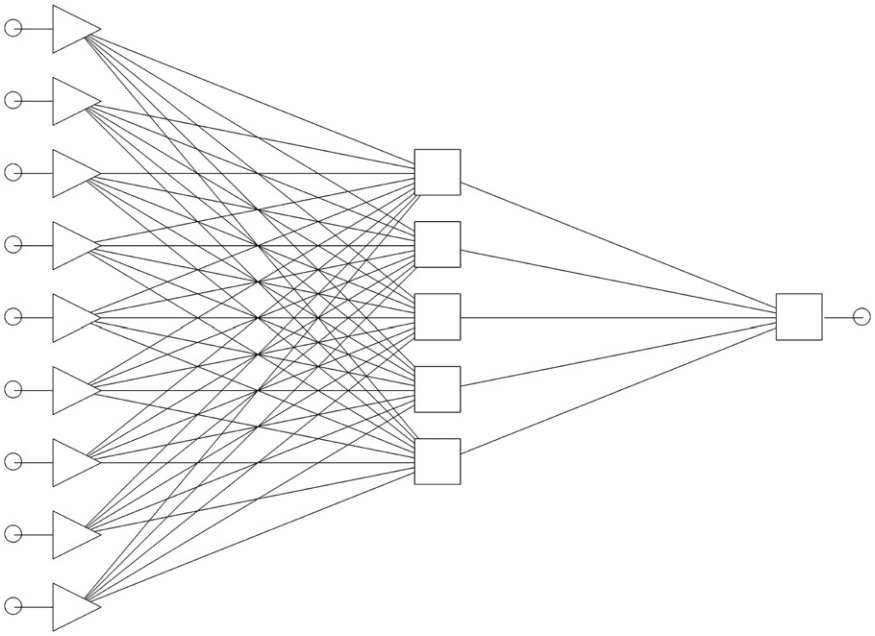

Fig. 3. Graphical presentation of the applied neural network.

showed slightly better results. MA was chosen for further modeling because of it had the highest correlation compared with other correlated descriptors.

Correlation coefficients between descriptor pairs are listed in Table 1. According to this, descriptors which were used in further ANN modeling were Log *D*, MA, MR, DEN, NON VDW, H-don and H-acc.

3.2. Fractional factorial design in selection of statistically significant chromatographic factors

Screening experiments are intended to reveal which factors have the biggest influence on the retention behavior of the analyzed substances in the chromatographic system. In this way, all important factors were included but the number of inputs was also reduced and unnecessary network burdening was prevented. For that purpose, 2^{4-1} fractional factorial design (FFD) was applied. The statistical significance of the investigated factors was estimated towards retention factors (*k*) of azoles as the model output. Percentage of organic eluent in mobile phase, pH of water phase, column temperature and percentage of triethylamine (TEA) as a modifier for basic drug elution in the water phase appeared to potentially affect retention of drugs generally. These assumptions were confirmed in preliminary experiments. On the other hand, the flow rate showed low influence on the retention factors, so it was kept constant at 400 μL/min and was not included in FFD. The factors were varied on two-point levels denoted as +1 and −1, thereby using the

values obtained as suitable in preliminary experiments. The repetition of three experiments at the central point provided a precise estimate of the experimental error and the significance of each variable. The experimental plan of FFD and related retention factors are given in Table 2. The effect of each variable was then tested using Students' t-test with a corresponding *p*-value. According to the obtained retention factors, the estimated effects and then the standardized effects were calculated. The next step was to estimate the importance of the factors. The critical t-value, for $\alpha=0.05$ and 3 degrees of freedom (d.f.), was 3.182 for all substances. All factors whose absolute values of the standardized effects are above the critical t-value are statistically significant and the ones below this value are statistically insignificant. Pareto charts, in which the length of the bars is proportional to the absolute value of the standardized effects, are presented in Fig. 2. The dashed line represents the critical t-value ($p=0.05$) and the importance of the presented factors can be easily noticed. As shown, the percentage of acetonitrile in mobile phase and buffer pH showed statistically significant influence on retention of the azoles, and therefore they were selected as inputs in ANN modeling. Further, it was decided to set column temperature at 20 °C and TEA was added to the mobile phase in a percentage of 0.01 (v/v).

3.3. ANN modeling

Results of ANN computing and prediction are given in Table 3. Percentage of acetonitrile in mobile phase (v/v) was varied from 40% to 65% in steps of 5% (v/v). Buffer pH was varied from 3.5 to 6 in steps of 0.6 or 0.7.

Network architecture and parameters of function were optimized by trial and error to minimize the root mean square (RMSE) for the training and validation data sets. At the start of a training run, the weights and biases were initialized at random values in the range between $+1$ and -1. The best performance showed a multilayer perceptron network consisting of three layers: input, hidden and output layer, with five neurons in the hidden layer. Its graphical presentation is given in Fig. 3. Linear post synaptic potential function operated in both the hidden layer and output neuron. Logistic activation function was set both in hidden and output layers. Learning rate governs the step size as the algorithm alters weights, while the momentum rate helps the algorithm avoid becoming stuck in flat spots and local minima. Optimal results were given by the network trained with back propagation algorithm with learning rate set to 0.6, momentum set to 0.1, while training was carried out for 10,000 epochs. Continued increase in validation error was the criterion for stopping the training. RMSE for training, validation and test were 1.789, 0.7498 and 2.777, respectively. R^2 represents the coefficient of determination between experimentally obtained retention factor values (k(exp)) and values computed by ANN for training, or predicted by ANN for validation and test data sets (k(ANN)). R^2 is 0.9861, 0.9957 and 0.9871 for training, validation and test data sets respectively. High R^2 and low RMSE values for training and validation sets indicate good descriptive ability. On the other hand, high R^2 and low RMSE values for the test set indicate good predictive ability and that no overfitting occurred during the training process.

Agreement between log k(exp) and log k(ANN) is presented in Fig. 4. Logarithmic transformation of responses is likely to be more suitable for plotting because of a broad range of k values. As shown in Table 3 concerning the test data set, as a data set previously unseen in the network, there is a high level of agreement between observed and predicted values in the range of k values, where the test substance, econazole, has a desirable retention ($k=1$–10) and slightly lower (0–1) and higher k values (10–30). On the contrary, ANN is no longer reliable under conditions which make econazole retain longer in the stationary phase. This failure in prediction at higher k values makes test RMSE higher than it would be if dealing with desirable and slightly higher k values. This represents no hindrance, because the main application of this predictive tool is optimization of the HPLC analytical method for a future imidazole- or triazole-derivative antifungal agent, which implies desirable

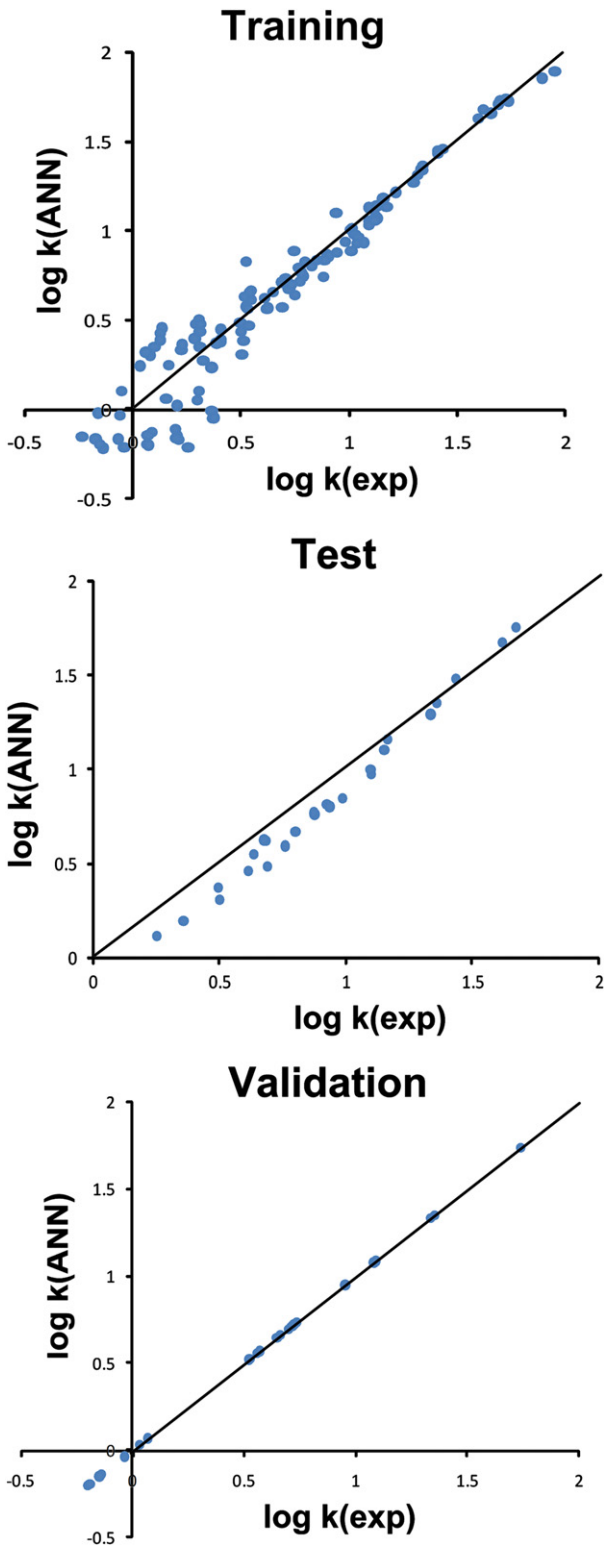

Fig. 4. Agreement between experimentally obtained retention factor values (log k(exp)) and values calculated or predicted by ANN (log k(ANN)).

retention behavior. A broader range of k values with satisfactory prediction (0–30) gives additional robustness to the QSRR-ANN.

For comparison, the possibility of obtaining a linear QSRR model by means of multilinear regression (MLR) was assessed. MLR requires only two data sets: a training set for the construction of the model and a second data set to check the predictive performance. With the purpose of using all the available data, the MLR model was built by fitting the data used for network training–validation (150 data points), thereby leaving the test data set for eventual predictability assessment. Stepwise multiple linear regression method produced poor linear correlation between descriptors and k(exp) with $R^2=0.5647$ and $F=20.18$. Model predictability on the test data set is therefore very poor, with $R^2=0.6918$, which strongly refers to ANN as a method that deals optimally with non-linear dependence between the inputs and outputs.

4. Conclusion

In this paper, molecular descriptors together with organic modifier content and pH of water phase, have been combined in the same QSRR model to predict the UHPLC retention of imidazole- and triazole-derivative antifungal agents. Despite large differences in terms of structures of six antifungal agents: miconazole, clotrimazole, ketoconazole, itraconazole, posaconazole and voriconazole, the optimized ANN network showed good predictability on previously unseen data, corresponding to econazole. This also proved a good selection of representing descriptors in terms of their influence on retention behavior of model substances. In this way, QSRR modeling combined with ANN could be used for prediction of HPLC behavior of a new antifungal agent with imidazole or triazole structure and molecular descriptor values within values used for ANN optimizing.

Acknowledgments

These results are part of the Project no. 172033, financed by the Ministry of Education and Science of the Republic of Serbia.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2012.07.071.

References

[1] J. Ekiert, J. Krzek, P. Talik, Talanta 82 (2010) 1090–1100.
[2] J. Daniel, J. Sheehan, A. Christopher, A. Hitchcock, Carol M. Sibley, Clin. Microbiol. Rev. 12 (1) (1999) 40–79.
[3] P.Y. Khashaba, S.R. El-Shabouri, K.M. Emara, A.M. Mohamed, J. Pharm. Biomed. 22 (2000) 363–376.
[4] T. Peng, Q. Cheng, C.F. Yang., Fresen. J. Anal. Chem. 370 (2001) 1082–1086.
[5] R.J. Ekiert, J. Krzek, J.S. Czekaj, U. Hubicka, Acta Chromatogr. 21 (2) (2009) 273–282.
[6] European Pharmacopoeia, 7th ed., Council of Europe, Strasbourg, France, 2011.
[7] United States Pharmacopeia 31, The United States Pharmacopeial Convention, Rockville, 2008.
[8] British Pharmacopoeia, British Pharmacopoeia Commission, The Stationery Office (TSO), London, United Kingdom, 2011.
[9] A.B. Eldin, A.A. Shalaby, Youssef M. Maamly, Acta Pharm. Sci. 52 (2010) 229–238.
[10] A.S. Low, J. Wangboonskul, Analyst 124 (1999) 1589–1593.
[11] R. Hájková, H. Sklenárová, L. Matysová, P. Svecová P, P. Solich P, Talanta 73 (2007) 483–489.
[12] A.B. Khetre, P.K. Sinha, M.C. Damle, M.R. Mehendre, Indian J. Pharm. Sci. 71 (2009) 509–514.
[13] K. Bodzioch, A. Durand, R. Kaliszan, T. Bączek, Y. Vander Heyden, Talanta 81 (2010) 1711–1718.
[14] Z. Garkani-Nejad, M. Ahmadvand, Chromatographia 73 (2011) 733–742.
[15] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley–VCH Verlag, Weinheim, 2000.
[16] G. Carlucci, A.A. D'Archivio, M.A. Maggi, P. Mazzeo, F. Ruggieri, Anal. Chim. Acta 601 (2007) 68–76.
[17] M. Aschi, A.A. D'Archivio, M.A. Maggi, P. Mazzeo, F. Ruggieri, Anal. Chim. Acta 582 (2007) 235–242.
[18] L. Fausett, Fundamentals of Neural Networks, Prentice Hall, New York, 1994.
[19] J. Zupan, Acta Chim. Slov. 41 (1994) 327–352.
[20] F. Marini F, R. Bucci, A.L. Magrì, A.D. Magrì, Microchem. J. 88 (2008) 178–185.
[21] A.S. Achanta, J.G. Kowalski, C.T. Rhodes, Drug Dev. Ind. Pharm. 21 (1995) 119–155.
[22] F. Ruggieri, A.A. D'Archivio, G. Carlucci, P. Mazzeo, J. Chromatogr. A 1076 (2005) 163–169.
[23] Y.L. Loukas, J. Chromatogr. A 904 (2000) 119–129.