Journal of the Serbian Chemical Society
JSCS@tmf.bg.ac.rs • www.shd.org.rs/JSCS

J. Serb. Chem. Soc. *87* (5) 615–628 (2022)
JSCS–5545

Original scientific paper
Published 2 February 2022

Optimization of chromatographic separation of aripiprazole and impurities: Quantitative structure–retention relationship approach

BOJANA SVRKOTA, JOVANA KRMAR, ANA PROTIĆ, MIRA ZEČEVIĆ and BILJANA OTAŠEVIĆ*

Department of Drug Analysis, University of Belgrade - Faculty of Pharmacy, Vojvode Stepe 450, 11221 Belgrade, Serbia

(Received 9 July, revised 21 October, accepted 3 November 2021)

Abstract: A new optimization strategy based on the mixed quantitative structure–retention relationship (QSRR) model is proposed for improving the RP-HPLC separation of aripiprazole and its impurities (IMP A-E). Firstly, experimental parameters (EPs), namely mobile phase composition and flow rate, were varied according to Box–Behnken design and thereafter, an artificial neural network (ANN) as a QSRR model was built correlating EPs and selected molecular descriptors (ovality, torsion energy and non-1,4-van der Waals energy) with the log-transformed retention times of the analytes. Values of the root mean square error (*RMSE*) were used for an estimation of the quality of the ANNs (0.0227, 0.0191 and 0.0230 for the training, verification and test set, respectively). The separations of critical peak pairs on chromatogram (IMP A-B and IMP D-C) were optimized using ANNs for which the EPs served as inputs and the log-transformed separation criteria *s* as the outputs. They were validated by application of leave-one-out cross-validation (*RMSE* values 0.065 and 0.056, respectively). The obtained ANNs were used for plotting response surfaces upon which the analyses chromatographic conditions resulting in optimal analytes retention behaviour and the optimal values of the separation criteria *s* were defined. The optimal conditions were 54 % of methanol at the beginning and 79 % of methanol at the end of gradient elution programme with a mobile phase flow rate of 460 μL min^{-1}.

Keywords: gradient elution; high performance liquid chromatography; artificial neural networks.

INTRODUCTION

In various fields of pharmaceutical research, high performance liquid chromatography (HPLC) is an indispensable tool for analytical testing. Due to their

* Corresponding author. E-mail: biljana.otasevic@pharmacy.bg.ac.rs
https://doi.org/10.2298/JSC210709092S

Available on line at www.shd.org.rs/JSCS/

(CC) 2022 SCS.

widespread use, there is a pronounced need for efficient and improved method development. Whereby the numerous factors affecting HPLC separation, including a large amount of available equipment and the determination of the conditions which ensure optimal method performance, can be extremely challenging. For improving a systematic survey of experimental space for optimal conditions, software-dependent tools can accelerate the development of a given method.[1] The traditionally predominant approach was based on the "trial and error" concept, which was exceedingly time and resources consuming. Nowadays, this approach is considered outdated and is efficiently suppressed by the emergence of more sophisticated chemometric tools. Since chemometric tools are based on reliable mathematical-statistical principles, the experimental work is successfully planned in advance and the resource requirements for its execution are reduced.[2] Chemometric techniques mainly used in HPLC method optimization are design of experiments (DoE) and quantitative structure–retention relationships (QSRR) approach. Within the DoE framework, analyte retention behaviour is mathematically associated with the values of experimental parameters. However, with the help of established DoE models, it is possible to develop an optimal method only for substances for which retention data have been obtained through experiments. In order to achieve a superior predictive dimension of applied modelling technique, it is necessary to involve the chemical–physical properties that are known independently of experimental work. This can be provided with QSRR studies.[3] Explicitly, QSRR models have the possibility of relating chromatographic retention behaviour to numerically expressed chemical information embedded in the molecule structure in the form of molecular descriptors (MDs).[4] This implies that the most common QSRR models are limited to predicting the retention behaviour only at constant values of the experimental parameters (EP). Therefore, when QSRR studies are motivated by practical goals, such as method optimization, EPs have to be included in the model as relevant variables (predictors or inputs).[5]

Models referred to as mixed QSRR models are more complex than the aforementioned classical ones since they explain retention behaviour in context of both, EPs and MDs.[4,6] A significant part of QSRR model building is the technique used for determining the mathematical relationship between MDs and retention behaviour. The most widely applied model-building techniques are multiple linear regression (MLR) and artificial neural networks (ANN).[3] With regards to the introduction of additional inputs, it is reasonable to expect that more sophisticated analytical tools should be used. At the same time, when the specific chromatographic system of interest comprises analytes with different polarities and a gradient elution program is consequently applied, nonlinear behaviour of the system is expected.[7] In such cases, nonlinear tools such as ANN are preferable for QSRR model building. ANN is a biologically inspired machine learning algorithm that mimics the functioning of the human brain.[8] In comparison to

(CC) 2022 SCS.

the classical statistical methods, ANN does not require previous knowledge about mathematical relationships between dependent and independent variables and has better predictive abilities.[4,5,8]

Difficulties in substantial method development for quality assessment purposes were present during the separation of aripiprazole (AR) and its impurities, the structures of which are presented in Fig. 1. Generally, impurities that can originate from raw materials, arise during drug product manufacturing (process-related), or appear as a result of drug degradation can impair the quality of an active substance or product.[9,10] Therefore, a reliable and optimized method for analysis of AR and its impurities had to be introduced. In the literature, IMP A, B, D and E are recognized as degradation products that can arise from various stress conditions,[11–13] while IMP A–D are also identified as process-related.[14] A literature survey revealed that HPLC methods were dominantly developed for the separation of AR and its known impurities.[13–20] It was noticed that octyl (C8) or octadecyl (C18) are the most commonly applied stationary phases. Unfortunately, in all these methods, the overall quality of the chromatographic separations was not quite adequate, because either the peaks of IMP E and AR were significantly close to each other or/and IMP E eluted after the peak of the active pharmaceutical ingredient (API).[13–16]

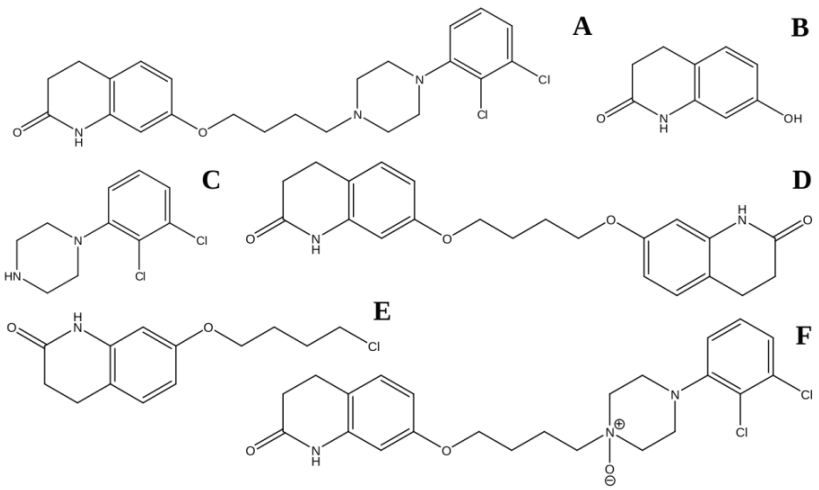

Fig. 1. Chemical structures of: A) aripiprazole, B) impurity A, C) impurity B, D) impurity C, E) impurity D and F) impurity E.

In practice, with such an elution, the tail of the API peak may overlap with a much smaller peak of the following impurity, which eventually occurs with long lifecycle methods and/or during problematic inter-laboratory analytical method transfer and consequently affects the identification and quantification impurity as

(CC) 2022 SCS.

well as the estimation of the overall AR quality. Therefore, separation conditions with different selectivity could be beneficial. In order to meet these demands, an RP-HPLC method was developed for the separation of AR and its impurities using a nonconventional stationary phase, a phenyl-hexyl column, able to produce diverse selectivity. A gradient elution program was employed for overcoming the difficulties in separation of molecules with different log *D* and log *P* features (Table S-I of the Supplementary material to this paper). Such a situation was also recognized by other authors who attempted to address this issue by using different HPLC modes starting from gradient elution in RP-HPLC,[13,16,18–20] then with micellar[6] and ion-pair chromatography.[14,15] The retention mechanisms involved into later separation processes are more complex due to the application of additives to the mobile phase and the consequent presence of secondary equilibriums. In addition, from the practical point of view, these chromatographic systems are considered as less attractive due to prolonged system equilibration and significantly reduced column lifetime. On the contrary, the advantage of the RP mode is reflected in its simpler retention mechanism that enables easier interpretation, better understanding and faster method optimization.

The QSRR strategy was applied in two reported studies. Nikolić *et al.* built a classical QSRR model in order to predict the retention times of impurities while central composite design was used for defining the optimal chromatographic conditions.[21] In such a way, the overall optimization method concept did not rely on the QSRR model; rather its predicting ability was the focus. Snoj Ekmečić *et al*. applied a solvatic retention model.[18] Although this approach gives insight into the physical meaning of the separation process, it has a lower ability in defining the experimental space compared to QSRR with computed MDs. Bearing in mind the optimization of the chromatographic method was the principal goal, the use of QSRR with computed MDs is considered the tool of choice.[3,7] To the best of knowledge, this is the first research where a mixed QSRR-ANN model was used as an optimization tool for the development of a superior method for the separation of aripiprazole and its impurities.

EXPERIMENTAL

Chemicals and reagents

All chemicals used were of analytical grade. Reference standard substances of aripiprazole (7-[4-[4-(2,3-dichlorophenyl)piperazin-4-ium-1-yl]butoxy]-3,4-dihydro-1*H*-quinolin-2-one) and impurities IMP A (7-hydroxy-3,4-dihydro-1*H*-quinolin-2-one), IMP B (1-(2,3-dichlorophenyl)--piperazine), IMP C (1,4-bis[[3,4-dihydro-2(1*H*)-quinolinone]7-oxy]butane), IMP D (7-(4-chlorobutoxy)-3,4-dihydro-1*H*-quinolin-2-one) and IMP E (7-[4-[4-(2,3-dichlorophenyl)-1-oxidopiperazin-4-ium-1-yl]butoxy]-3,4-dihydro-1*H*-quinolin-2-one), purchased from Orchid Chemicals & Pharmaceuticals (Chennai, India), were utilized for the preparation of the working mixture solution for chromatographic analyses. The mobile phase was composed of LC–MS grade methanol (MeOH, Fluka, Switzerland), water HPLC grade purified with simplicity 185 water purification system (Millipore, USA) and acetic acid analytical grade (Sigma–Aldrich).

(CC) 2022 SCS.

Preparation of standard and work solutions

The standard stock solution of Aripiprazole was prepared by dissolving the respective amount of the standard substance in MeOH to attain a concentration of 2 mg mL^{-1}. Dissolving was ultrasonically assisted for 5 min using an ultrasonic bath (Fungilab, Barcelona, Spain). For preparing the individual standard stock solutions of the impurities, the respective amounts were dissolved in MeOH assisted with ultrasonication for 5 min. The solutions of the impurities were attained in concentrations of 20 μg mL^{-1}. The working mixture solution was prepared by transferring 1 mL of each standard solution to a 10-mL volumetric flask and filling to volume with a mixture of aqueous part of the mobile phase and MeOH (50:50). The concentration of working solution was 200 μg mL^{-1} of aripiprazole and 2 μg mL^{-1} of each of impurities.

Chromatographic conditions and equipment

A Thermo Accela UHPLC system (Thermo Fisher Scientific Inc., USA) equipped with an autosampler, pump and PDA detector was used for the experimental runs. The chromatographic separations were performed with a Kinetex RP-phenyl-hexyl column (100 mm×4.6 mm, 2.6 μm, Phenomenex Inc., USA). The column temperature was set at 25 °C and the UV detection wavelength was 254 nm. The mobile phase consisting of a methanol/water mixture was composed according to the Box–Behnken design of experiments (Table I) created using Design-Expert 7.0.0 software (Stat-Ease Inc., USA). The pH of the aqueous phase was adjusted to 4.7 using the acetic acid and a pH-meter with a combined electrode, PHM 220 (Radiometer, Denmark). The mobile phase was freshly prepared and filtered before use through Whatman 47 mm Glass/mesh system with membrane carrier and a 0.22 μm pore size membrane filter (Kinesis Inc., USA). The elution program consisted of a 6 min isocratic elution with the initial MeOH (s_{MeOH}) and water mixture, followed by a 1 min gradient that reached final MeOH (e_{MeOH}) and water content that was maintained until the end of the analysis. The chromatograms were gathered using ChromQuest 5.0 software (Thermo Fisher Scientific Inc., USA).

QSRR model development

For QSRR model building, MDs and EPs were determined as independent variables while the dependent variable was the log-transformed retention time of the studied analytes (log t_r). The structures of the molecules were first generated with ChemDraw Ultra 7.0 software (Perkin Elmer, USA). Then, they were utilized for an estimation of the molecular p*K*a and ionization forms using MarvinView 6.1.6 (ChemAxon Ltd., Hungary). Structure energy minimization and computation of MDs were provided with Chem3D® Ultra 7.0 software (Cambridge Soft Corporation, USA). Computation of MDs was performed after minimizing the energy of solute structures *via* the semi-empirical MOPAC/PM3 method. The computations resulted in total of 37 MDs that included: dipole length – *DPLL*; electronic energy – *ElcE;* formal charge – *Charge;* HOMO energy – *HOMO;* LUMO energy – *LUMO;* repulsion energy – *NRE*; total energy –*TotE;* Balaban index – *BIndx*; cluster count – *ClsC*; Connolly accessible area – *SAS*; Connolly molecular area – *MS*; Conolly solvent–excluded volume – *SEV*; diameter – *Diam*; exact mass – *Mass*; molecular topological index – *TIndx*; molecular weight – *MW*; ovality – *O*; principle moment of inertia *X*, *Y* and *Z* – *PMIX*, *PMIY*, *PMIZ*; radius – *Rad*; shape attribute – *ShpA*; shape coefficient – *ShpC*; sum of degrees – *SDeg*; sum of valence degrees – *SVDe*; total connectivity – *TCon*; total valence connectivity – *TVCon*; Wiener index – *WIndx*; bend energy – *Eb*; molar refractivity – *MR*; non–1,4–VDW E – *Ev*; partition coefficient (octanol/water) – *C* log *P*; stretch–bend energy – *Esb*; stretch energy – *Es*;

(CC) 2022 SCS.

torsion energy – *Et*; VDW 1,4 energy – *E14*; total energy – *E* (Table S-II of the Supplementary material). When calculating the values of MDs, the percentage of existing ionization forms present at the mobile phase pH was taken into account. Descriptors were selected taking into account correlation coefficient. That is, in order to avoid multicollinearity, one of two variables were discarded if the correlation between them was higher than 0.9 ($|r| > 0.9$). The MDs correlation was evaluated using Microsoft Office Excel 2010 (Microsoft, WA). Furthermore, MDs with the highest impact on the system behaviour were identified using MLR, with the stepwise inclusion of independent variables in the model. MLR was performed with SPSS software (SPSS, USA).

ANN was used as a tool for *QSRR* model building as well as for modelling of separation criteria (*s*). Multilayer perception (MLP) feed forward neural networks were used consisting of neurons (processing units) arranged in 3 layers: input, hidden and output and connected by bonds with assigned weights. The networks were trained with a backpropagation algorithm. Logistic transformation was applied in the second and third layer. The model validity was estimated with root mean square error (*RMSE*). An increase in the *RMSE* value indicated overfitting which was a signal to stop the network training. Regression statistics was used to evaluate model fitting to experimentally obtained data. Separate MLPs were trained to model separation criteria ($s_{B\text{-}A}$, $s_{C\text{-}D}$). MLPs were trained on a data set consisting of 13 cases obtained during BBD. Each MLP included 3 inputs (EPs), 2 hidden neurons, and 1 output (log $s_{B\text{-}A}$ or log $s_{C\text{-}D}$). In the MLP used for modelling log $s_{B\text{-}A}$, logistic transformation was applied in the hidden and output layer. For MLP used for log $s_{C\text{-}D}$ estimation, logistic transformation was used in the second and hyperbolic transformation in the third layer. Leave-one-out cross-validation (LOO-CV) was used for an estimation of the generalization ability of the algorithms, which was evaluated using LOO-CV *RMSE*. All ANN models were developed and validated with Statistica Neural Network software (StatSoft Inc., USA).

RESULTS AND DISCUSSION

During the development process of the HPLC method, the type of stationary phase and the composition of eluent were selected before optimization of the method. The domain of chromatographic parameters was selected with respect to the different physicochemical characteristics (values of log *P* and log *D* taking into account the respective pH) of the investigated analytes (Table S-I). Gradient elution was introduced to move strongly retained compounds, while having the least hydrophobic analytes well resolved. Additionally, an inspection of chemical structures of all compounds (Fig. 1) revealed that each of them contained at least one aromatic ring for which it was reasonable to expect that π-electrons could potentially be involved in the separation process. Therefore, a phenyl-hexyl column, the retention mechanism of which is based on π–π and hydrophobic interactions, was considered as the proper choice for the separation of this type of analyte.[22] Following the selection of column, a thorough review of the constituents of the mobile phase was performed. It has already been proven that acetonitrile can suppress π–π interactions.[23] Therefore, MeOH was considered as the organic solvent of choice. The additional benefit of using MeOH was also reflected in its toxicological profile since MeOH is a less environmentally hazardous organic solvent compared to acetonitrile.[24]

(CC) 2022 SCS.

Given that the pH value of the aqueous phase also influences the separation process of ionisable compounds, the working pH value was set. The preliminary experiments revealed that a pH of the aqueous phase set at 4.7 contributed to better separation of the compounds compared to the usually applied pH 3.0, due to a shift in the ionization state of IMP E. Molecules ionization states under these conditions are presented in Table S-I. All these considerations pointed out that the following EPs are the most significant for the intended method development: MeOH content at the start of the gradient elution program (s_{MeOH}) within range 50–60 %; MeOH content at the end of gradient elution program (e_{MeOH}) within the 75–85 % and the mobile phase flow rate (F) varied within 400–500 μL min^{-1}.

In order to provide a proper description of experimental space, DoE methodology was applied. The response surface design by means of three level Box–Behnken design (BBD) was favoured as more economical and efficient due to smaller number of runs comparing to other optimization designs (such as central composite design).[25] BBD is rotatable design meaning that all the experimental points are equally distant from the central point and lie on a sphere when presented in 3D space.[8,26] The number of experiments to be performed according to BBD in the presented research resulted in a total of 13 runs (Table I). Although the value of DoE in method optimization is unquestionable, the obtained mathematical models are not completely in agreement with experimental outcomes, which is reflected in the coefficients of determination, R^2 (Table II). Therefore, the data gathered using DoE was further implemented into QSRR-ANN model.[8]

TABLE I. Experimental plan with EPs and experimentally obtained responses

Content, %		Flow rate, mL min^{-1}	Retention time, min						s / min	
s_{MeOH}	e_{MeOH}		IMP A	IMP B	IMP D	IMP C	IMP E	AR	B-A	C-D
50	75	450	2.50	3.63	11.93	13.19	15.08	18.58	0.977	0.860
60	75	450	2.37	2.96	10.77	12.23	14.20	17.25	0.161	1.019
50	85	450	2.50	4.00	11.00	11.54	12.60	15.94	1.156	0.096
60	85	450	2.37	2.96	9.99	10.54	11.38	13.84	0.152	0.273
50	80	400	2.82	4.46	12.34	13.36	15.09	19.47	1.128	0.507
60	80	400	2.79	3.45	11.40	12.53	14.22	18.22	0.397	0.669
50	80	500	2.26	3.49	11.01	11.81	13.13	16.52	1.010	0.389
60	80	500	2.23	2.79	10.27	11.19	12.51	15.56	0.141	0.481
55	75	400	2.48	3.76	13.36	15.49	18.57	23.32	0.846	1.932
55	85	400	2.77	3.65	10.91	11.45	12.44	15.56	0.654	0.197
55	75	500	2.19	3.09	11.74	13.42	15.78	19.50	0.367	1.238
55	85	500	2.19	3.09	10.24	10.76	11.66	14.50	0.329	0.121
55	80	450	2.44	3.37	11.21	11.71	13.63	17.30	0.463	0.460

The overall goal of this study was to develop an optimization strategy that would enable the appropriate chromatographic separation of all peaks in the shortest possible time. At the same time, it was necessary to avoid non-retention

(CC) 2022 SCS.

behaviour of the first eluting peak corresponding to IMP A. Accordingly, the following responses were selected. The total analysis time was estimated according to the retention time of the last eluting peak (t_{rAR}). The retention time of IMP A (t_{rIMPA}) was monitored to establish separation from the peak of the mobile phase. The quality of the separation between neighbouring peaks of analytes was estimated using the separation criterion *s* that has lately been recognized as a superior descriptor of separation in gradient elution methods known for susceptibility to baseline variations and peak shape irregularities.[27] This criterion represents the difference between the time when the first peak ends (t_{r1e}) and the time when next peak begins (t_{r2b}). This can be defined with following equation: $s = = t_{r2b}–t_{r1e}$. In this respect, the baseline separations between IMP A-B and IMP D-C peak pairs were assessed according to the calculated values of $s_{B\text{-}A}$ and $s_{C\text{-}D}$, respectively (Table I) with an appropriate acceptance criterion $s > 0$. The retention times of interest were modelled using QSRR-ANN, and *s* criteria were modelled using ANNs specially developed for the purpose.

TABLE II. Statistically evaluated BBD mathematical models using coded factor values; Equations obtained using Design Expert 7.0.0 software

Mathematical model	R^2	Adj. R^2	Pred. R^2
$t_{rIMPA} = 2.45 – 0.25F$	0.8300	0.8179	0.7563
$\log s_{B\text{-}A} = -0.34 – 0.37s_{MeOH} – 0.14F$	0.9124	0.8989	0.8508
$\log s_{C\text{-}D} = -0.34 – 0.093s_{MeOH} – 0.44e_{MeOH} – 0.083F$	0.9294	0.9118	0.8604
$\log t_{rAR} = 1.23 – 0.058e_{MeOH} – 0.031F$	0.7428	0.7032	0.5800

In order to ensure a reliable building of a QSRR model, selection of MDs was carefully driven. The types of MDs were selected using general knowledge of the RP-HPLC separation mechanism. From initially calculated 37 MDs, 13 MDs (*DPLL*, *HOMO*, *LUMO*, *SAS*, *O*, *PMIX*, *ShpC*, *TVCon*, *Ev*, *C log P*, *Esb*, *Es* and *Et*) were retained based on the level of mutual correlation of the MDs ($|r| < 0.9$). This MDs selection procedure was applied in order to avoid multicollinearity and overfitting of the obtained model. From the pool of all independent variables (MDs and EPs), the statistically influential inputs were extracted using MLR. The set of the most significant independent variables indicated that the chromatographic system was dominantly influenced by 3 EPs (s_{MeOH}, e_{MeOH}, *F*) and 4 MDs (*O*, *Ev*, *ShpC*, *Et*). Further inspection of the MDs values indicated that *ShpC* had identical value for 3 of the 6 analytes, which would eventually negatively affect the training of the ANN. Since the quality and predicting ability of the model can be preserved in reduced QSRR when selection of attributes is good,[27] it was decided to remove *ShpC* from the data set.

The aforementioned independent variables (s_{MeOH}, e_{MeOH}, *F*, *O*, *Et*, *Ev*) were used as inputs for the MLP, while log-transformed retention times of the analytes ($\log t_r$) represent output variable. Thus, the QSRR-ANN architecture consisted of 6

(CC) 2022 SCS.

neurons in the input, 4 neurons in the hidden and 1 neuron in the output layer. It was decided to use the multilayer perception ANN due to its proven simplicity and efficiency. MLP is a feed forward ANN trained with a backpropagation algorithm meaning that it computes the error and adjusts the weights backwards through layers, from output to inputs, until the error is minimized.[4] The output was transformed *via* log-function in order to provide normal distribution of the outcome variable. Even though normal distribution of the measured data is not necessary for an ANN to function, in some cases it can improve the performance of a network. Engagement of log-transformation, with the ability of narrowing the response value interval, resulted in significantly better learning ability of an ANN, lower errors and satisfying overall statistical performance.[6] The network was trained using 78 cases (13 cases of BBD for each of the 6 compounds), divided into training, verification, and test set. The training set contained 52 cases, and the verification and test sets contained 13 cases each. The training and verification sets were formed randomly, while the test set contained only cases related to IMP C. This was done intentionally, so that the real predictive ability of the developed model can be estimated for an analyte previously unseen by the network. The learning rate of the ANN was set to 0.3 and momentum to 0.1. The ANN weights are presented in Table S-III of the Supplementary material. The predictive power of the obtained network was validated using internal and external validation.[28] IMP C was used for the external validation and cross-verification with verification data set was applied as an internal type of validation. Low and balanced *RMSE* values indicated good predictive ability of the network. Furthermore, regression statistics and the correlation coefficient values of the model indicated good fitting of the predicted model to experimental data (Table III).[29]

TABLE III. Statistical performance of the QSRR-ANN developed model

Parameter	Training log (t_r / min)	Verification log (t_r / min)	Test log (t_r / min)
Data mean	0.8634	0.8902	1.0859
Data *S.D.*	0.3442	0.3759	0.0462
Error mean	–0.0005	0.0072	0.0154
Error *S.D.*	0.0229	0.0184	0.0178
Abs E. mean	0.0176	0.0144	0.0194
S.D. ratio	0.0665	0.0490	0.3858
Correlation	0.9978	0.9989	0.9253
RMSE	0.0227	0.0191	0.0230

Considering that $s_{B\text{-}A}$ and $s_{C\text{-}D}$ could not be estimated by the means of the QSRR-ANN model, they were evaluated using other ANNs. Although it is uncommon for ANNs to model small data sets, they still can be considered in circumstances where classical regression modelling tools are not applicable. Due to the questionable reliability of such ANN, the quality of the model needs

(CC) 2022 SCS.

verification applying robustification methods. One of these methods refers to *V*-fold cross-validation, where *V* is the number of folds. Special case where number of cases is equal to number of folds refers to LOO-CV.[30] In this research, the LOO-CV *RMSE* values were 0.0648 and 0.0560 for log $s_{B\text{-}A}$ and log $s_{C\text{-}D}$, respectively (Table S-IV of the Supplementary material). These results indicate good reliability of the whole dataset of the obtained model. Since none of two peak pairs was favoured, the most similar *s* values, still reflecting baseline separation, were striven for. The established ANNs (Tables S-V and S-VI of the Supplementary material) were further used to obtain response surfaces that were then used for visual inspection of the influence of EPs on the responses (Fig. 2).

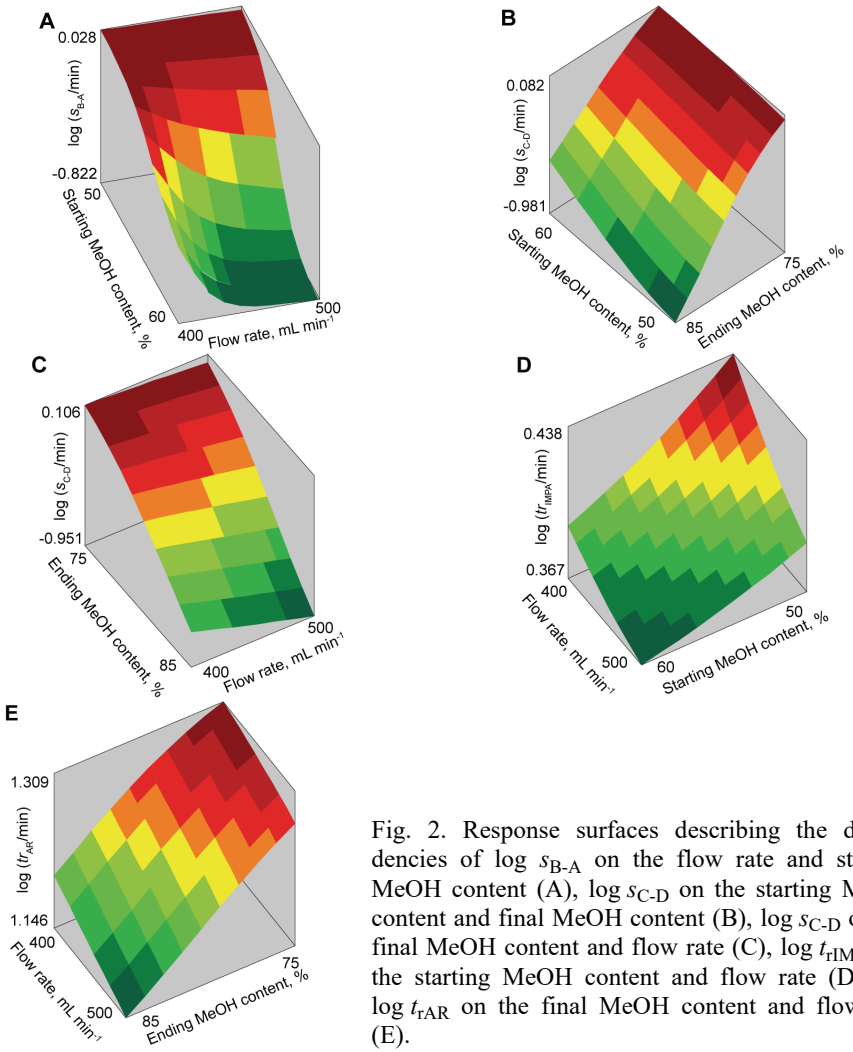

Fig. 2. Response surfaces describing the dependencies of log $s_{B\text{-}A}$ on the flow rate and starting MeOH content (A), log $s_{C\text{-}D}$ on the starting MeOH content and final MeOH content (B), log $s_{C\text{-}D}$ on the final MeOH content and flow rate (C), log t_{rIMPA} on the starting MeOH content and flow rate (D) and log t_{rAR} on the final MeOH content and flow rate (E).

(CC) 2022 SCS.

Choosing the optimal EPs values was a challenging task due to the opposed influences they demonstrated towards all the observed responses. The lower values of all EPs enabled satisfactory baseline separations of neighbouring peaks to be achieved. On the contrary, higher EPs values reduced the analysis time. Among the EPs, the flow rate had a very low influence on both separation criteria (Fig. 2). In the region of the highest EPs values where good peak separation was preserved, the selection of optimal conditions was made mainly to suite the total analysis time. F values higher than 460 μL min^{-1} disrupt an adequate separation between critical peak pairs. Thus, even though this reduces the analysis runtime, a flow rate set to 460 μL min^{-1} was considered as optimal. The factor e_{MeOH} had a greater influence on $s_{C\text{-}D}$ and the maximal value that preserves a good $s_{C\text{-}D}$ was defined to be 79 %. Optimal s_{MeOH} was chosen bearing in mind that values greater than 54 % rapidly decreased $s_{B\text{-}A}$. Furthermore, the selected values of s_{MeOH} and flow rate resulted in satisfactory t_{rIMPA} values. Finally, the values of the EPs providing appropriate meeting of all method optimization goals were: s_{MeOH} = 54 %, e_{MeOH} = 79 % and F = 460 μL min^{-1}. The predicted t_r values were 2.46, 3.31, 12.60, 11.12, 13.89 and 17.83 min for IMP A-E and AR, respectively, while the predicted s values were $s_{B\text{-}A}$ = 0.630 min and $s_{C\text{-}D}$ = 0.591 min. The obtained results were in accordance with experimentally obtained values for t_r (2.49, 3.48, 12.34, 11.37, 13.85 and 17.47 min for IMP A-E and AR, respectively) and s ($s_{B\text{-}A}$ = 0.560 min, $s_{C\text{-}D}$ = = 0.580 min), as presented in Fig. 3.

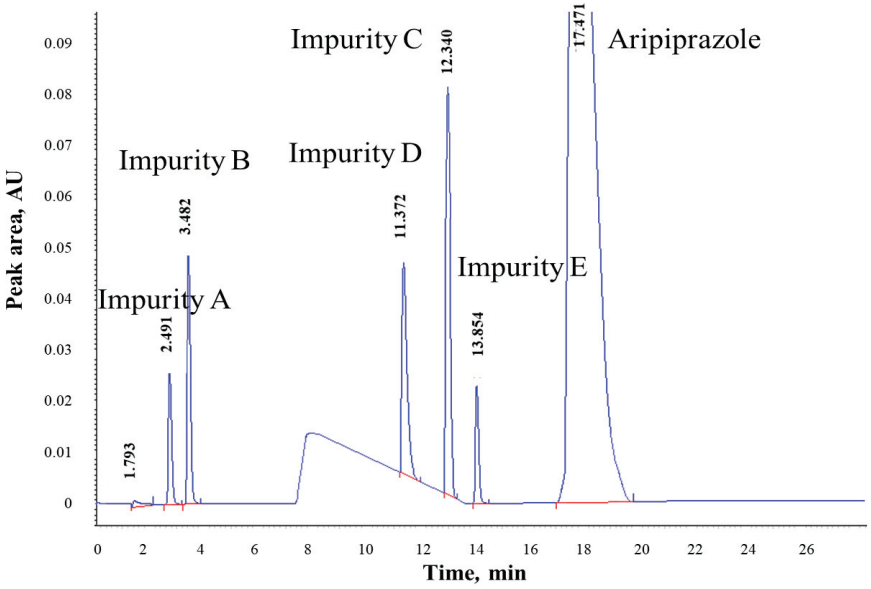

Fig. 3. Chromatogram recorded under optimal conditions.

(CC) 2022 SCS.

CONCLUSIONS

An improved RP-HPLC method intended for the evaluation the purity of AR was developed using a mixed QSRR-ANN model as an optimization tool. The experimental conditions providing adequate separations in a reasonable analysis time were determined (s_{MeOH} = 54 %, e_{MeOH} = 79 % and F = 460 μL min^{-1}). Furthermore, the demonstrated good predictive power of the built models enable future spreading of the current research to other AR related topics, meaning that it is reasonable to suggest that the noted chromatographic behaviour predictions could serve as a starting point in development of analytical procedures aimed at AR stability studies as well as investigations of AR in different biological samples.

SUPPLEMENTARY MATERIAL

Additional data and information are available electronically at the pages of journal website: https://www.shd-pub.org.rs/index.php/JSCS/article/view/10939, or from the corresponding author on request.

Acknowledgment. This research was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia through a Grant Agreement with University of Belgrade – Faculty of Pharmacy No: 451-03-9/2021-14/200161.

ИЗВОД

ОПТИМИЗАЦИЈА ХРОМАТОГРАФСКОГ РАЗДВАЈАЊА АРИПИПРАЗОЛА И НЕЧИСТОЋА: ПРИСТУП КВАНТИФИКОВАЊА ОДНОСА СТРУКТУРЕ И РЕТЕНЦИОНОГ ПОНАШАЊА

БОЈАНА СВРКОТА, ЈОВАНА КРМАР, АНА ПРОТИЋ, МИРА ЗЕЧЕВИЋ и БИЉАНА ОТАШЕВИЋ

Катедра за аналитику лекова, Универзитет у Београду – Фармацеутски факултет, Војводе Степе 450, 11221 Београд

Нова оптимизациона стратегија заснована на грађењу мешовитих модела за квантификовање односа структуре и ретенционог понашања (QSRR) предложена је за унапређење RP-HPLC раздвајања арипипразола и његових нечистоћа (IMP A-E). Експериментални параметри (EP), састав мобилне фазе и брзина протока, варирани су најпре у складу са Box–Behnken дизајном, а затим је награђена вештачка неуронска мрежа као QSRR модел који повезује EP и одабране молекуларне дескрипторе (овалност, торзиона енергија и не-1,4-ван дер Валсова енергија) са логаритамски трансформисаним ретенционим временом аналита. Вредности средње квадратне грешке (*RMSE*) коришћене су за процену квалитета мреже (0,0227, 0,0191 и 0,0230 за тренинг, верификацију и тест сет, редом). Раздвајање критичних парова пикова на хроматограму (IMP A-B и IMP D-C) оптимизовано је коришћењем мрежа за које су EP послужили као улази, а логаритамски трансформисани критеријуми сепарације *s* као излази. Ове мреже су валидиране применом унакрсне валидације изостанка (*RMSE* вредности, редом, 0,065 и 0,056). На основу награђених мрежа, конструисани су дијаграми површина одговора чијом анализом су дефинисани услови при којима се постиже оптимална ретенција аналита, односно вредности критеријума сепарације *s*, а који су подразумевали 54 % метанола на почетку и 79 % на крају програма градијентног елуирања са брзином протока мобилне фазе од 460 mL min^{-1}.

(Примљено 9. јула, ревидирано 21. октобра, прихваћено 3. новембра 2021)

(CC) 2022 SCS.

REFERENCES

1. F. T. Mattrey, A. A. Makarov, E. L. Regalado, F. Bernardoni, M. Figus, M. B. Hicks, J. Zheng, L. Wang, W. Schafer, V. Antonucci, S. E. Hamilton, K. Zawatzky, C. J. Welch, *Trends Anal. Chem.* **95** (2017) 36 (https://doi.org/10.1016/j.trac.2017.07.021)
2. P. K. Sahu, N. R. Ramisetti, T. Cecchi, S. Swain, C. S. Patro, J. Panda, *J. Pharm. Biomed. Anal.* **147** (2018) 590 (https://doi.org/10.1016/j.jpba.2017.05.006)
3. P. R. Haddad, M. Taraji, R. Szücs, *Anal Chem.* **93** (2020) 228 (https://doi.org/10.1021/acs.analchem.0c04190)
4. J. Golubović, A. Protić, B. Otašević, M. Zečević, *Talanta* **150** (2016) 190 (https://doi.org/10.1016/j.talanta.2015.12.035)
5. A. A. D'Archivio, M. A. Maggi, F. Ruggieri, *J. Sep. Sci.* **37** (2014) 1930 (https://doi.org/10.1002/jssc.201400346)
6. J. Krmar, M. Vukićević, A. Kovačević, A. Protić, M. Zečević, B. Otašević, *J. Chromatogr., A* **1623** (2020) 461146 (https://doi.org/10.1016/j.chroma.2020.461146)
7. A. A. D'Archivio, M. A. Maggi, F. Ruggieri, *Anal. Bioanal. Chem.* **407** (2015) 1181 (https://doi.org/10.1007/s00216-014-8317-3)
8. M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, L. A. Escaleira, *Talanta* **76** (2008) 965 (https://doi.org/10.1016/j.talanta.2008.05.019)
9. ICH Topic Q3A (R2), *Impurities in new drug substances*, 2006, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-3-r2-impurities-new-drug-substances-step-5_en.pdf (accessed Feb 16, 2021)
10. ICH Topic Q3B (R2), *Impurities in new drug products*, 2006, https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-3-b-r2-impurities-new-drug-products-step-5_en.pdf (accessed Feb 16, 2021)
11. C. M. El-Maraghy, H. Salem, S. M. Amer, M. Nebsen, *Anal. Chem. Lett.* **9** (2019) 258 (https://doi.org/10.1080/22297928.2019.1585286)
12. K. S. V. Srinivas, R. Buchireddy, G. Madhusudhan, K. Mukkanti, P. Srinivasulu, *Chromatographia* **68** (2008) 635 (https://doi.org/10.1365/s10337-008-0739-7)
13. G. V. R. Reddy, A. P. Kumar, B. V. Reddy, P. Kumar, H. D. Gauttam, *Eur. J. Chem.* **1** (2010) 20 (https://doi.org/10.5155/eurjchem.1.1.20-27.11)
14. N. Djordjević Filijović, A. Pavlović, K. Nikolić, D. Agbaba, *Acta Chromatogr.* **26** (2014) 13 (https://doi.org/10.1556/achrom.26.2014.1.15)
15. M. Rmandić, A. Malenović, *J. Sep. Sci.* **43** (2020) 3242 (https://doi.org/10.1002/jssc.201900985)
16. M. V. V. N. Murali Krishna, S. V. Rao, N. V. S. Venugopal, *J. Liq. Chromatogr. Relat. Technol.* **40** (2017) 741 (https://doi.org/10.1080/10826076.2017.1357572)
17. F. Soponar, M. Sandru, V. David, *J. Liq. Chromatogr. Relat. Technol.* **39** (2016) 70 (https://doi.org/10.1080/10826076.2015.1126729)
18. T. Snoj Ekmečić, I. Kralj Cigic, *Acta Chim. Slov.* **66** (2019) 958 (http://dx.doi.org/10.17344/acsi.2019.5189)
19. *European Pharmacopoeia*, 9th ed., The Council of Europe, Strasburg, 2017
20. *The United States Pharmacopeia*, 42nd revision, United States Pharmacopeia Convention, Rockville, MD, 2019
21. K. Nikolić, N. Djordjević Filijović, B. Maričić, D. Agbaba, *J. Sep. Sci.* **36** (2013) 3165 (https://doi.org/10.1002/jssc.201300477)
22. K. Croes, A. Steffens, D. H. Marchand, L. R. Snyder, *J. Chromatogr., A* **1098** (2005) 123 (https://doi.org/10.1016/j.chroma.2005.08.090)

(CC) 2022 SCS.

23. M. Yang, S. Fazio, D. Munch, P. Drumm, *J. Chromatogr., A* **1097** (2005) 124 (https://doi.org/10.1016/j.chroma.2005.08.028)
24. ICH Topic Q3C (R6), *Residual solvents*, 2006, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ich-guideline-q3c-r8-impurities-guideline-residual-solvents-step-5_en.pdf (accessed Feb 16, 2021)
25. E. Yabalak, O. Gomez, B. G. Sonmez, *J. Serb. Chem. Soc.* **83** (2018) 489 (https://doi.org/10.2298/JSC170909113Y)
26. S. L. Ferreira, R. E. Bruns, E. G. da Silva, W. N. Dos Santos, C. M. Quintella, J. M. David, J. B. de Andrade, M. C. Breitkreitz, I. C. Jardim, B. B. Neto, *J. Chromatogr., A* **1158** (2007) 2 (https://doi.org/10.1016/j.chroma.2007.03.051)
27. B. Otašević, J. Šljivić, A. Protić, N. Maljurić, A. Malenović, M. Zečević*, Microchem. J.* **145** (2019) 655 (https://doi.org/10.1016/j.microc.2018.11.033)
28. M. A. Fouad, E. H. Tolba, M. A. El-Shal, A. M. El Kerdawy, *J. Chromatogr., A* **1549** (2018) 51 (https://doi.org/10.1016/j.chroma.2018.03.042)
29. Y. Cao, S. Yao, X. Wang, T. Yao, H. Song, *J. Serb. Chem. Soc.* **82** (2017) 399 (https://doi.org/10.2298/JSC160725013D)
30. M. De Lozzo, P. Klotz, B. Laurent, *Eng. Appl. Artif. Intell.* **26** (2013) 2270 (https://doi.org/10.1016/j.engappai.2013.07.001).

(CC) 2022 SCS.