Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. *86* (4) 381–391 (2021)
JSCS–5428
Original scientific paper

Prediction of retardation factor of protein amino acids in reversed phase TLC with ethanol–sodium azide solution as the mobile phase using QSRR

SUSAN TORABI[1], FATEMEH HONARASA[2] and SAEED YOUSEFINEJAD[3]*

[1]*Deputy of Food and Drug Control, Shiraz University of Medical Sciences, Shiraz, Iran,* [2]*Department of Chemistry, Shiraz Branch, Islamic Azad University, Shiraz, Iran and* [3]*Research Center for Health Sciences, Institute of Health, Department of Occupational Health Engineering, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran*

(Received 11 June, revised 30 August, accepted 6 October 2020)

Abstract: Due to the importance of amino acids (AAs) as the basic bricks of proteins and their application in the drug and food industries, there is great interest in their separation and identification using simple and inexpensive approaches. Application of predictive models for the determination of the behavior of AAs can reduce trial-and-error experiments. Herein, the retardation factor (*RF*) of 21 protein AAs were studied using the quantitative structure-retardation factor (QSRR) model. The *RF* values of the AAs in ethanol–sodium azide solution as the mobile phase of reversed phase thin layer chromatography (RP-TLC) were correlated with the structural properties of the AAs. The suggested QSRR indicated excellent fitting and prediction ability (R2train = 0.95 and R2test = 0.94). Furthermore, other statistical tests, such as *y*-scrambling, cross validation and the Williams plot confirmed the stability, absence of chance and the suitable applicability domain, respectively. It was shown that the sum of geometrical distances between oxygen and nitrogen atoms in an AA molecule is an important factor for the *RF* values of the AAs in the ethanol––sodium azide.

Keywords: natural amino acids; descriptors; structural property; thin layer chromatography; QSRR.

INTRODUCTION

Finding correlation and relationship between chromatographic retention time or retardation indices and structural parameters of the desired analysts is an interesting subject because it is a way to obtain basic information on the impact of the structural features on the retention/retardation indices and to give an insight into possible mechanisms for separation forces and elution process.[1,2]

* Corresponding author. E-mail: yousefisa@sums.ac.ir
https://doi.org/10.2298/JSC200611065T

The ideal objective of models of quantitative structure-retention relationships (QSRR), as a subfield in quantitative structure-retention relationships (QSPR), is its application to predict the retention behavior of newly identified molecules and similar synthesized derivatives or to estimate the involved mechanisms during various interactions, such as solute–solute, solute–stationary phase and solute–mobile phase.[3,4] The prediction of retention/retardation of compounds in chromatographic systems using QSRR can reduce trial-and-error experiments, which is important in some series of species such as polychlorinated biphenyls (PCBs), polybrominated, diphenyl ethers (PBDEs), polychlorinated dibenzofurans (PCDFs), *etc.* This could be significant in such compounds due to their toxicity or scarce information about some of their derivatives.[5,6]

Different QSRR studies have been conducted on various chromatography methods, such as reverse phase- and normal phase-high performance chromatography (RP-HPLC and NP-HPLC), micellar HPLC, gas chromatography, high performance affinity chromatography, immobilized artificial membranes, and planar chromatography, which are reviewed in different articles.[5,7,8]

RP-TLC is one of the most popular techniques and its retention mechanism is based on partitioning of the compounds between hydrophilic mobile phase and hydrophobic stationary phase. Due to the nature and mechanism of RP-TLC, the lipophilicity of the analytes has an effect on their retention with a strict correlation.[9]

Analysis and identification of amino acids (AAs) is of high importance because AAs are the basic units of biomolecules, enzymes, peptides, and protein and have impacts in the food and drug industries. Separation and identification of AAs (protein or non-protein) have been intensively performed with column chromatography and their separation using simple methods, such as thin layer chromatography (TLC), have been considered in many studies.[10,11] In a previous study, a QSRR model was reported for separation of AAs in Normal Phase TLC (NP-TLC) by considering both the properties of the mobile phase and the structures of AAs.[12] To show the ability of QSRR in modeling the behavior of AAs in RP-TLC, a QSRR modeling was recently designed in two different mobile phases and the potential of structure-chromatography was confirmed in this case.[13] However, because the recent work was performed by considering only the structure of the AAs, more studies are required to show the capability of the QSRR approach in the separation of AAs in RP-TLC in various mobile phases. Thus, in the current study, a QSRR model was developed for the retardation factor (R_F) of 21 protein AAs in RP-TLC using an ethanol–sodium azide solution as a well-known mobile phase to identify the significant features in this elution process.

MATERIALS AND METHODS

The retardation factor (R_F) for the studied AAs using RP-TLC in ethanol–sodium azide were adopted from literature[10] and are given in Table I.

TABLE I. Experimental and predicted R_F of 21 AAs in ethanol–sodium azide and related residual values

Code	Name	R_F (Exp)	R_F (Pred)	Residual
AA 1	Glycine	0.82	0.84	0.02
AA 2	Alanine	0.82	0.78	–0.04
AA 3	Aspartic acid	0.25	0.35	0.10
AA 4	Arginine	0.13	0.14	0.01
AA 5	Proline	0.78	0.65	–0.13
AA 6	Hydroxyproline	0.84	0.88	0.04
AA 7	Lysine	0.02	0.02	0.00
AA 8	Glutamic acid	0.86	0.74	–0.12
AA 9[a]	Serine	0.8	0.74	–0.06
AA 10	Tryptophan	0.83	0.87	0.04
AA 11	Valine	0.83	0.89	0.06
AA 12	Phenyl alanine	0.83	0.90	0.07
AA 13	Isoleucine	0.88	0.84	–0.04
AA 14[a]	Leucine	0.88	0.97	0.09
AA 15	Asparagine	0.62	0.60	–0.02
AA 16	Methionine	0.83	0.84	0.01
AA 17	Cysteine	0.84	0.88	0.04
AA 18[a]	Histidine	0.33	0.35	0.02
AA 19	Threonine	0.83	0.80	–0.03
AA 20[a]	Tyrosine	0.83	0.83	0.00
AA 21[a]	Glutamine	0.77	0.72	–0.05

[a]AA samples used as the test set

The structural properties were produced from 22 categories of descriptors, *i.e.*, topological, constitutional, topological charge indices, geometrical, connectivity, RDF, 3D MoRSE, WHIM, GETAWAY, functional group counts, and some other groups were extracted.[14] The extracted descriptors were from a variety of kinds of features to cover the structural details of natural AAs. The structural descriptors were arranged in a matrix (D) of size of 21×c where c denotes the number of totally utilized descriptors. Then, constant and near constant columns from this matrix was deleted to remove redundant information. Collinear columns in D were also removed after calculating the correlation of the descriptors with the R_F vector and with other descriptors. Finally, among a detected pair of collinear columns, the one with the lowest correlation with the R_F vector was eliminated from the D matrix. After these refining process, 404 descriptors were retained in D (size = 21×404) for further process and variable selection. In the current study, the structures of natural AAs were drawn using Hyperchem software (version 7, Hypercube Inc., http://www.hyper.com, USA) and the AM1 semi-empirical method were applied during optimization. Different categories of the structural features were extracted using DRAGON (http://michem.disat.unimib.it/chm/; Milano Chemometrics and QSAR research group) for all AAs.

In the next step, the constructed data matrix (D) was divided into test and training subsets and the training set was targeted by variable selection and further model development. Stepwise multiple linear regression (SMLR) was applied as the variable selection method and the squared correlation coefficient of training set (R^2_{train}) and cross validation (Q^2_{CV}) were the criteria for choosing the final model.[15-17] In addition to cross validation, different statistical approaches, such as *y*-scrambling and prediction of a small portion of AAs, as the external test set, were utilized to evaluate the prediction ability and stability of the suggested QSRR model in RP-TLC.[18,19]

All statistical and calculations tasks were carried out *via* MATLAB software (version 7.7, R2008b, Math work, Inc., http://mathworks.com, USA). A personal computer under the Windows 7 operating system was used to run all software.

RESULTS AND DISCUSSION

After splitting the data into training and test sets of samples using random selection from PCA space (Fig. 1a) and variable selection using stepwise MLR, different models were constructed based on 1 to 8 descriptors of the AAs and the results of fitting and cross validation is represented in Fig. 1b. As can be observed in Fig. 1b, a QSRR containing 5 descriptors was utilized as the optimum one, after auto scaling the X-matrix (descriptors) and *y*-vector (RTLC retardation factor):

$$R_{F(\text{ethanol-sodium azide})} = -2.34\ (\pm 0.584) - 0.19\ (\pm 0.003)\ G_{(N..O)} - \\ -1.594\ (\pm 0.211)\ \text{Mor24u+}\ 6.661\ (\pm 1.080)\ \text{PW2} + \\ 1.018\ (\pm 0.200)\ \text{Mor28u} - 0.619\ (\pm 0.158)\ \text{SEige} \quad (1)$$

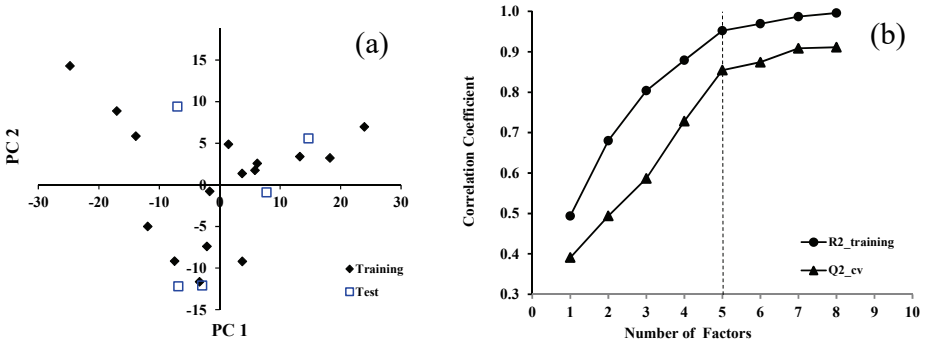

Fig. 1. Two dimensional-PCA plot of the space of the total descriptors in AA chromatographic samples (a) and correlation coefficient of the training set (R^2_{train}) and cross-validation (Q^2) *vs.* the number of descriptors to select the optimum number of factors (b) for the suggested RP-TLC model in ethanol–sodium azide solution.

In the above suggested QSRR in ethanol–sodium azide solution $G_{(N..O)}$ denotes the sum of geometrical distances between the nitrogen and oxygen atoms in the AA structure. Mor24u is the unweighted three dimensional molecular representation of structures based on electronic diffraction (3D-MoRSE) descriptors of signal 24, PW2 denotes the path/walk 2-Ranic shape index, Mor28u corres-

ponds to the unweighted 3D-MoRSE information of signal 28 and SEige is the sum of the Eigenvalue from the electronegativity weighted distance matrix.[14] The definitions and categories of the utilized descriptors are also summarized in Table II.

TABLE II. Class and definition of descriptors used in model 1 and 2

No.	Name	Class	Definition
1	$G_{(N\cdot\cdot O)}$	3D Atom pairs	Sum of geometrical distances between N··O
2	Mor24u	3D-MoRSE descriptors	3D-MoRSE descriptors signal 24 / unweighted
3	PW2	Topological indices	Path/walk 2 - Randic shape index
4	Mor28u	3D-MoRSE descriptors	3D-MoRSE descriptors signal 28 / unweighted
5	SEige	Eigenvalue-based indices	Sum of the eigenvalue from the electronegativity weighted distance matrix

To obtain a better prediction of the R_F values, all five included descriptors and experimental vector of R_F were targeted by mean-centering and scaling (autoscaled).[12,20] After these pre-treatments, the linear model was re-computed based on these prepared data and Eqs. (2) containing standardized MLR-coefficients were obtained:

$$R_{F(\text{ethanol-sodium azide})} = -0.72\ (\pm 0.098)\ G_{(N..O)} - 0.540\ (\pm 0.072)\ \text{Mor24u} + \\ + 0.461\ (\pm 0.075)\ \text{PW2} + 0.457\ (\pm 0.090)\ \text{Mor28u} - \\ - 0.410\ (\pm 0.105)\ \text{SEige} \qquad (2)$$

$$R^2_{\text{train}} = 0.95,\ F = 39.85,\ RMSE_{\text{train}} = 0.22,\ F_{\text{crit}(95\ \%)} = 3.33$$

As shown in the above results, the significantly higher value of F-statistic in comparison to F_{crit} confirms the significance of the above QSRR developed in ethanol–sodium azide solution. In addition to R^2_{train} of model 2, which was equal to 0.95, Q^2_{LOO} of cross validation was equal to 0.85 that shows suitable fitness and sufficient stability of model 2.[21] In Eqs. (1) and (2), the standard 3of the five descriptors are included because they have a very low value compared to the calculated coefficients. Magnitude of the coefficient of each descriptor of AAs can be used to conclude about the effect of that descriptor on R_F of natural AAs in the studied mobile phase (ethanol–sodium azide solution).

Details and explanation of all structural descriptors in the suggested models will be presented in the next sections. The numerical values of $G_{(N\cdot\cdot O)}$, Mor24u, PW2, Mor28u and SEige are given in the Supplementary material to this paper (Table S-I).

In this model, the $RMSE_{test}$ and R^2_{test} values were 0.26 and 0.94, respectively. Thus, the results showed good agreement between predicted and experimental R_F values for the 21 AAs in the training and test samples. Y-scrambling was applied and the Q^2_{MP} value of this test was 0.25, which confirms that a chance QSRR model in ethanol–sodium azide was not constructed. The predicted values of retardation factor of the 21 free amino acids using the model are pre-

sented in Table I and the statistical parameters of this QSRR model are given in Table III. As was shown in the Fig. 2a, a suitable correlation between experimental and predicted R_F values of the amino acids was obtained. In addition to the denoted statistics, some other criteria suggested by Golbraikh and Tropsha were calculated for the developed QSRR model.[21,22] Some of the important and well-known metrics were calculated for the suggested model are denoted in the following and their threshold values for a valid model are also mentioned in Eqs. (3)–(5):

$$|R_0^2 - R'_0{}^2| = 0.00597 \text{ (Threshold value} < 0.3\text{, Passed!)} \quad (3)$$

$$k = 1.00073 \text{ and } (|R_0^2 - R'_0{}^2|/R^2) = 0.00094$$

$$\text{(Threshold value: } [0.85<k<1.15 \text{ and } |R_0^2 - R'_0{}^2| / R^2 <0.1]\text{, Passed!)} \quad (4)$$

$$k' = 0.99132 \text{ and } (|R'_0{}^2 - R_0^2|/R^2) = 0.00766$$

$$\text{(Threshold value: } [0.85< k' <1.15 \text{ and } |R'_0{}^2 - R_0^2|/R^2 < 0.1]\text{, Passed!)} \quad (5)$$

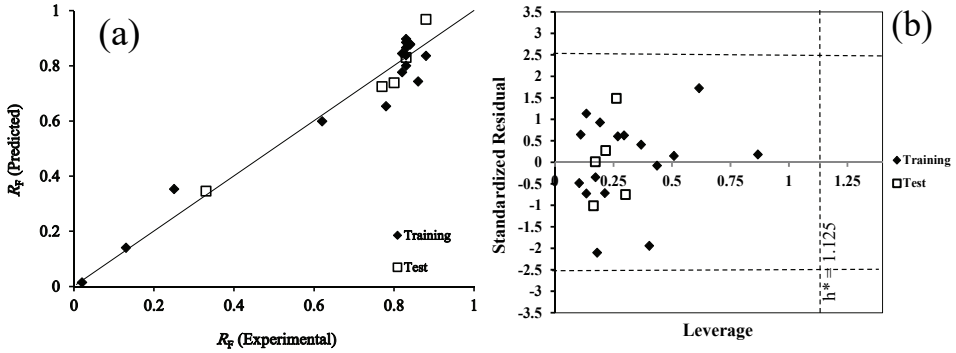

Fig. 2. Plot of predicted R_F *vs*. experimental values for the 21 investigated AA samples using the five parametric RP-TLC model in ethanol –sodium azide (a). The applicability domain of the set of AAs shown by the Williams plot (b). The cut off values of the standardized residual (±2.5 times the standards deviation) and the leverage (h^*) are illustrated by horizontal and vertical dashed lines, respectively. All samples are located within the applicability domain.

Details of the calculation of these parameters can be found in the original references[21] but, as can be seen, the constructed QSRR passed these validation criteria. Other criteria were also suggested by Roy *et al.* to show the prediction ability known as the average r^2_m or r^2_m and delta r^2_m or Δr^2_m.[22] r^2_m should be higher than 0.5 and Δr^2_m must be lower than 0.2 in a valid method.[23] For the suggested QSRR shown in Eq. (2), r^2_m and Δr^2_m of test set were 0.845 and 0.0402, respectively, which indicate the validity of the agreement between the predicted and actual solubility in the test set. However, it is noteworthy that the small number of AAs in test set is a limitation but this was because of the limited amount of experimental data.

TABLE III. Statistical performance of the proposed QSRR of the RP-TLC samples of amino acids (in ethanol–sodium azide) using three different random-divided test and training sets

Training-test set	N_{train}[a]	N_{test}[b]	R^2_{train}[c]	$RMSE_{train}$[d]	Q^2_{LOO}[e]	$RMSE_{cv}$[f]	R^2_{test}[g]	$RMSE_P$[h]	Q^2_{MP}[i]
1[j]	16	5	0.95	0.22	0.85	0.35	0.94	0.26	0.25
2[k]	16	5	0.94	0.23	0.84	0.40	0.91	0.35	0.31
3[l]	16	5	0.95	0.23	0.84	0.41	0.91	0.33	0.16

[a]Number of RP-TLC runs in the training set; [b]number of RP-TLC runs in test set; [cc]orrelation coefficient of the training RP-TLC runs; [d]root mean square error of the training RP-TLC runs (Calibration); [e]correlation coefficient of leave-one-out cross-validation; [f]root-mean-square errors of leave-one-out cross-validation; [g]correlation coefficient of the test RP-TLC runs; [h]root-mean-square errors of the test RP-TLC runs; [i]maximum cross--validation correlation coefficient for 30 Y-randomization tests; [j]number of solvents in the test set as indicated in Table I: AA9, AA14, AA18, AA20, AA21; [k]number of solvents in the test set: AA2, AA10, AA12, AA18, AA20; [l]number of solvents in the test set: AA1, AA2, AA12, AA16, AA18

Another important point should be considered in the evaluation of the QSRR is its independency from the AAs that were used as the training set or reserved as the test set. To ensure about this independency, two other subsets of TLC runs were randomly chosen from the RP-TLC samples as the training sets and two rest subsets (containing 5 AAs) as the test sets, which is shown in the results as train--test 2 and train-test 3. As presented in Table III, changing the AAs in the training or test set had no effect on the goodness-of-fit or prediction of the QSRR model for the RP-TLC using ethanol–sodium azide.

Applicability domain and pair/multi correlation

As a well-known recommendation, the suggested QSRR model must exclude any linear dependency between the AAs descriptors. This necessity is because this seriously limits the accuracy in models with collinearity and cannot be useful to justify the chromatographic behavior of analytes in RP-TLC during elution with ethanol–sodium azide. Moreover, the presentation of collinear descriptors led to wrong signs being obtained for the coefficients of the QSRR.[16,24] The correlation between each pair of descriptors in the five utilized AAs descriptors was done and the matrix of pair correlation for the present QSRR model is given in Table IV, which indicates no high correlation in this model.

TABLE IV. Pair correlation matrix for descriptors in the developed QSRR model for ethanol–sodium azide and the related VIF values as an index of multi-collinearity

	$G_{(N\cdot\cdot O)}$	Mor24u	PW2	Mor28u	SEige	*VIF*
$G_{(N\cdot\cdot O)}$	1.00					2.03
Mor24u	0.00	1.00				1.07
PW2	0.00	0.05	1.00			1.17
Mor28u	0.35	0.00	0.00	1.00		1.68
SEige	0.45	0.00	0.03	0.42	1.00	2.30

Not only the existence of pair-correlation, but also multi-collinearity (*i.e.*, collinearity of one with all others) in QSRRs is also a risk for model accuracy and the variance inflation factor (*VIF*) is a good metrics to evaluate such collinearity.[25] The high multi-collinearity can hide some of structural information because of overlapping in independent variables.[25] As could be observed in Table IV, the calculated *VIF* of all utilized descriptors ($G_{(N..O)}$, Mor24u, PW2, Mor28u and SEige) are lower than critical value of 5.0,[25] which shows that the present QSRR does not suffer from risk of multi-collinearity. Thus, the sign of coefficient and their magnitude in suggested QSRR could be trusted to justify the effect of the selected structural properties of the AAs on their R_F value.

After the different statistical evaluations of the QSRR, the leverage and standardized residual for the AAs samples were calculated to represent the applicability domain (AD).[26] Considering AD can clarify the limitations and potential of the developed QSRR for these AAs, it might be useful for RP-TLC of similar structures derived from AAs.[19] AA samples with a leverage below the cut-off value and with standardized a residual within the logical range can be considered to be in the normal AD. A standardized residual higher than -3σ or lower than $+3\sigma$ is a suitable value.[27,28] In addition, the cut-off value of the leverage is $h^* = = 3(d+1)/n_{train}$,[27] where d denotes number of descriptors in the QSRR model (here equal to 5) and n_{train} shows the number of AAs in the training subset (here equal to 16). Accordingly, the h^* of the present QSRR was calculated equal to be 1.125.

According to Fig. 2b that is the Williams plot for the representation of both standardized residual and leverage, all the studied chromatographic samples were within the AD of model suggested for RP-TLC using ethanol–sodium azide.

Interpretation of model

In this model, $G_{(N..O)}$ was the first descriptor with a negative effect on the R_F of AAs in ethanol–sodium azide, which confirmed the importance of "sum of geometrical distances between N and O" ($G_{(N..O)}$). It should be emphasized that the effect of $G_{(N..O)}$ was also illustrated in previous research,[12] which was significant in the separation of amino acids in normal phase TLC (NP-TLC) with a negative effect on the R_F of amino acids.

Two unweighted 3D-MoRSE descriptors, named Mor24u and Mor28u (signals 26 and 24),[14] are imported in the model with negative and positive sign of the coefficients. The presence of 3D-MoRSE indices in this model and previous work on this subject show the potential of this category of descriptors in prediction of the retardation factors of amino acids in RP-TLC.[13] On the other hand, different signs of these two 3D-MoRSE descriptors indicate their complex contribution in the retardation of samples using ethanol–sodium azide. The other descriptor in model is *PW*2, which is categorized in topological indices and shows

path/walk 2 – Randic shape index of the amino acids.[14] The positive contribution of *PW*2 in the model illustrates the direct relationship of *PW*2 and R_F of amino acids during separation with RP-TLC using ethanol–sodium azide. SEige is the last parameter in model, which is an eigenvalue-based index with a negative sign. SEige denotes the eigenvalue sum from electronegativity weighted distance matrix[14] and thus, increasing this index in amino acids could enhance their retardation factor in the ethanol–sodium azide system.

CONCLUSIONS

QSRR as a basic field in chromatography is a tool for showing the effect of the molecular structure of analytes on their chromatographic behavior. On the other hand, because of the effect of other parameters such as stationary and mobile phase in separation, this work was focused on modeling the R_F of protein AAs in RP-TLC during elution with ethanol–sodium azide.

One of findings in this study was the impact of the sum of the geometrical distances between N and O on R_F value of AAs in RP-TLC using ethanol––sodium azide. It was found that decreasing the sum of this distance could increase the remaining AAs on the TLC plate and their R_F value in ethanol–sodium azide. This fact was in accordance with previous report on the normal phase TLC of AAs. Eigenvalue sum from electronegativity weighted distance matrix and two 3D-MoRSE properties from AAs also had an important effect on the R_F value in the investigated system.

Moreover, different statistical evaluation on training, cross validation, prediction, *y*-randomization and applicability domain confirmed the stability and accuracy of the suggested QSRR. However, the small number of compounds in training and test sets could be considered a limitation of this work but it is noteworthy that the goal of the current modeling was not only external prediction but also was the chemical/structural description of the chromatographic behavior of AAs. This work could give more information for explaining the separation of AAs, in continuation of previous studies on other mobile phases and could be completed with more studies in future.

SUPPLEMENTARY MATERIAL

Numerical vales of original descriptor used in model Eq. (1) are available electronically at the pages of journal website: https://www.shd-pub.org.rs/index.php/JSCS/index, or from the corresponding author on request.

Acknowledgment. Support of Shiraz University of Medical Sciences (Grant No. 99-01--04-23865) is gratefully acknowledged.

И З В О Д

ПРЕТСКАЗИВАЊЕ ФАКТОРА ЗАДРЖАВАЊА ПРОТЕИНСКИХ АМИНО КИСЕЛИНА У РЕВЕРСНО-ФАЗНОЈ ТАНКОСЛОЈНОЈ ХРОМАТОГРАФИЈИ СА ЕТАНОЛ–НАТРИЈУМ--АЗИДОМ КАО МОБИЛНОМ ФАЗОМ КОРИШЋЕЊЕМ КВАНТИТАТИВНЕ РЕЛАЦИЈЕ СТРУКТУРЕ И ЗАОСТАЈАЊА (QSRR)

SUSAN TORABI[1], FATEMEH HONARASA[2] и SAEED YOUSEFINEJAD3

[1]Deputy of Food and Drug Control, Shiraz University of Medical Sciences, Shiraz, Iran, [2]Department of Chemistry, Shiraz Branch, Islamic Azad University, Shiraz, Iran и [3]Research Center for Health Sciences, Institute of Health, Department of Occupational Health Engineering, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran

Због значаја аминокиселина као основних јединица протеина и њихове примене у индустрији лекова и хране, постоји велико занимање за њихово раздвајање и идентификацију коришћењем простих и јефтиних приступа. Примена предиктивних модела за одређивање понашања АК може скратити експерименте покушаја-и-грешке. Овде су фактори заостајања (R_F) 21 протеинске амино киселине проучавани користећи квантитативни структура-фактор заостајања (QSRR) модел. R_F аминокиселина у раствору етанола–натријум-азида као мобилне фазе танкослојне реверсно-фазне хроматографије (RP-TLC) су корелисани са структурним особинама аминокиселина. Сугерисани QSRR указује на изврсно фитовање и способност предвиђања ($R_{2train} = 0,95$ и $R_{2test} = 0,94$). Надаље, остали статистички тестови као што су „*y*-scrambling“, унакрсна валидација, Williams график, потврђују стабилност, одсуство случајности, односно погодан домен применљивости. Показано је да је збир геометријских удаљености атома кисеоника и азота у аминокиселинама значајан фактор за RF вредности аминокиселина у етанол––натријум-азиду.

(Примљено 11. јуна, ревидирано 30. августа, прихваћено 6. октобра 2020)

REFERENCES

1. S. H. Park, M. De Pra, P. R. Haddad, S. Grosse, C. A. Pohl, F. Steiner, *J. Chromatogr., A* **1609** (2020) 460508 (https://dx.doi.org/10.1016/j.chroma.2019.460508)
2. A. M. Ramezani, S. Yousefinejad, A. Shahsavar, A. Mohajeri, G. Absalan, *J. Chromatogr., A* **1599** (2019) 46 (https://dx.doi.org/10.1016/j.chroma.2019.03.063)
3. J. M. Sutter, T. A. Peterson, P. C. Jurs, *Anal. Chim. Acta* **342** (1997) 113 (https://dx.doi.org/10.1016/S0003-2670(96)00578-8)
4. Y. Marrero-Ponce, S. J. Barigye, M. E. Jorge-Rodríguez, T. Tran-Thi-Thu, *Chem. Pap.* **72** (2018) 57 (https://dx.doi.org/10.1007/s11696-017-0257-x)
5. C. Giaginis, A. Tsantili-Kakoulidou, *Chromatographia* **76** (2013) 211 (https://dx.doi.org/10.1007/s10337-012-2374-6)
6. J. Dai, L. Jin, S. Yao, L. Wang, *Chemosphere* **42** (2001) 899 (https://dx.doi.org/10.1016/S0045-6535(00)00181-8)
7. K. Héberger, *J. Chromatogr., A* **1158** (2007) 273 (https://dx.doi.org/10.1016/J.CHROMA.2007.03.108)
8. R. Kaliszan, *Chem. Rev.* **107** (2007) 3212 (https://dx.doi.org/10.1021/cr068412z)
9. R. Kaliszan, *J. Chromatogr., A* **220** (1981) 71 (https://dx.doi.org/10.1016/S0021-9673(00)98504-2)
10. D. Kaźmierczak, W. Ciesielski, R. Zakrzewski, *JPC - J. Planar Chromatogr. - Mod. TLC* **18** (2005) 427 (https://dx.doi.org/10.1556/JPC.18.2005.6.5)
11. T. Hudaib, S. Brown, D. Wilson, P. E. Eady, *JPC - J. Planar Chromatogr. - Mod. TLC* **29** (2016) 145 (https://dx.doi.org/10.1556/1006.2016.29.2.9)

12. S. Yousefinejad, F. Honarasa, N. Saeed, *J. Sep. Sci.* **38** (2015) 1771 (https://dx.doi.org/10.1002/jssc.201401427)
13. S. Yousefinejad, F. Honarasa, S. Akbari, M. Nekoeinia, *J. Liq. Chromatogr. Relat. Technol.* **43** (2020) 580 (https://dx.doi.org/10.1080/10826076.2020.1774388)
14. R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, 2nd ed., Wiley-VCH, Weinheim, 2009 (ISBN: 9783527318520)
15. P. Gramatica, *Mol. Inform.* **33** (2014) 311 (https://dx.doi.org/10.1002/minf.201400030)
16. S. Yousefinejad, B. Hemmateenejad, *Chemom. Intell. Lab. Syst.* **149** (2015) 177 (https://dx.doi.org/10.1016/j.chemolab.2015.06.016)
17. S. Yousefinejad, F. Honarasa, A. Solhjoo, *J. Chem. Eng. Data* **61** (2016) 614 (https://dx.doi.org/10.1021/acs.jced.5b00768)
18. J. U. N. Shao, *J. Am. Stat. Assoc.* **88** (1993) 486 (https://dx.doi.org/10.2307/2290328)
19. P. Gemperline, *Practical Guide to Chemometrics*, 2nd ed., Taylor & Francis Group, Boca Raton, USA, 2006 (ISBN: 1574447831)
20. F. Honarasa, S. Yousefinejad, S. Nasr, M. Nekoeina, *J. Mol. Liq.* **212** (2015) 52 (https://dx.doi.org/10.1016/j.molliq.2015.08.055)
21. A. Golbraikh, A. Tropsha, *Mol. Divers.* **5** (2000) 231 (https://dx.doi.org/10.1023/A:1021372108686)
22. K. Roy, R. N. Das, P. Ambure, R. B. Aher, *Chemom. Intell. Lab. Syst.* **152** (2016) 18 (https://dx.doi.org/10.1016/j.chemolab.2016.01.008)
23. K. Roy, P. Chakraborty, I. Mitra, P. K. Ojha, S. Kar, R. N. Das, *J. Comput. Chem.* **34** (2013) 1071 (https://dx.doi.org/10.1002/jcc.23231)
24. L. Eriksson, J. Jaworska, A. P. Worth, M. T. D. Cronin, R. M. McDowell, P. Gramatica, *Environ. Health Perspect.* **111** (2003) 1361 (https://dx.doi.org/10.1289/ehp.5758)
25. T. A. Craney, J. G. Surles, *Qual. Eng.* **14** (2002) 391 (https://dx.doi.org/10.1081/QEN-120001878)
26. R. Todeschini, V. Consonni, P. Gramatica, in *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, R. Tauler, B. Walczak, S. D. Brown (Eds.), Elsevier B.V., Amsterdam, 2009, pp. 129–172 (ISBN: 9780444641663)
27. T. I. Netzeva, A. P. Worth, T. Aldenberg, R. Benigni, M. T. D. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C. A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G. Y. Patlewicz, R. Perkins, D. W. Roberts, T. W. Schultz, D. T. Stanton, J. J. M. van de Sandt, W. Tong, G. Veith, C. Yang, *Altern. to Lab. Anim.* **33** (2005) 155 (https://dx.doi.org/10.1177/026119290503300209)
28. S. Yousefinejad, R. Eftekhari, F. Honarasa, Z. Zamanian, F. Sedaghati, *J. Mol. Liq.* **241** (2017) 861 (https://dx.doi.org/10.1016/j.molliq.2017.06.081).