Estimation of passive gastrointestinal absorption and membrane retention using PAMPA test, quantitative structure-permeability and quantitative structure-retention relationship analyses of ethylenediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid and 1,3-propanediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid derivatives

Biljana Tubić (Methodology) (Investigation) (Writing - original draft), Vladimir Dobričić (Methodology) (Investigation) (Writing - original draft)<ce:contributor-role>Formal Analysis), Jelena Poljarević (Resources) (Writing - original draft), Aleksandar Savić (Resources) (Writing - original draft), Tibor Sabo (Resources) (Supervision) (Project administration), Bojan Marković (Methodology) (Investigation) (Supervision) (Project administration)

PII: S0731-7085(19)32942-5

DOI: https://doi.org/10.1016/j.jpba.2020.113213

Reference: PBA 113213

To appear in: *Journal of Pharmaceutical and Biomedical Analysis*

Received Date: 3 December 2019

Revised Date: 23 February 2020

Accepted Date: 24 February 2020

Please cite this article as: Tubić B, Dobričić V, Poljarević J, Savić A, Sabo T, Marković B, Estimation of passive gastrointestinal absorption and membrane retention using PAMPA test, quantitative structure-permeability and quantitative structure-retention relationship analyses of ethylenediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid and 1,3-propanediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid derivatives, *Journal of Pharmaceutical and Biomedical Analysis* (2020), doi: https://doi.org/10.1016/j.jpba.2020.113213

Estimation of passive gastrointestinal absorption and membrane retention using PAMPA test, quantitative structure-permeability and quantitative structure-retention relationship analyses of ethylenediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid and 1,3-propanediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid derivatives

Biljana Tubić[a,1,2], Vladimir Dobričić[a,2,*] vladimir@pharmacy.bg.ac.rs, Jelena Poljarević[b], Aleksandar Savić[b], Tibor Sabo[b], Bojan Marković[a]

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

[b] *Department of General and Inorganic Chemistry, University of Belgrade – Faculty of Chemistry, Studentski trg 12-16, 11158 Belgrade, Serbia*

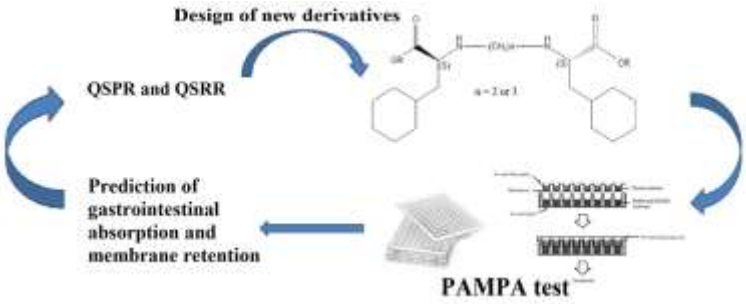
*Corresponding author. Tel.: 00381 11 3951 341

E-mail address: (V.Dobričić).

[1] Present addresses: Agency for Medicines and Medical Devices of Bosnia and Herzegovina, Veljka Mladjenovica bb, 78000 Banja Luka, Bosnia and Herzegovina; Faculty of Medicine - Department of Pharmacy, University of Banja Luka, Save Mrkalja 14, 78000 Banja Luka, Bosnia and Herzegovina

[2] Both authors contributed equally to this work

Graphical abstract



Highlights

- Esters of 1,2-ethane and 1,3-propanediamine-N,N'-di-2-(3-cyclohexyl)propanoic acid;
- Gastrointestinal absorption and membrane retention were predicted using PAMPA;
- Membrane retention is in correlation with previous results of cytotoxic activity;
- QSPR and QSRR analyses were performed;
- New derivatives with favourable absorption and retention properties were designed;

ABSTRACT

Passive gastrointestinal absorption and membrane retention of twelve esters of (*S*,*S*)-ethylenediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid (EDCP) and (*S*,*S*)-1,3-propanediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid (PDCP), as well as of these two non-esterified acids were estimated using PAMPA test. Artificial PAMPA membrane used in this study for the simulation of gastrointestinal barrier was solution of egg lecithin in dodecane (1 % w/v). All tested

compounds belong to class III (high membrane retention and low permeation), whereas EDCP, dipentyl ester of PDCP (DPE-PDCP) and diisopentyl ester of PDCP (DIPE-PDCP) belong to class I (negligible membrane retention and low permeation). Finally, quantitative structure – permeability and structure – retention relationships models were created in order to find quantitative relationships between physico-chemical properties of tested compounds and PAMPA membrane permeability/membrane retention parameters. Statistically the most reliable models were analysed and used for the design of new compounds for which favourable membrane permeability and retention can be expected.

Keywords: (*S*,*S*)-ethylenediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid and (*S*,*S*)-1,3-propanediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid derivatives; PAMPA; membrane permeability; membrane retention

1. Introduction

Chemotherapy is one of commonly used strategies in clinical protocols for treatment of cancer diseases with different localization. However, this therapy is usually associated with adverse side effects and resistance. Therefore, the discovery of new anticancer compounds has become one of the most important goals in medicinal chemistry.

Recent studies have shown that ethylenediamine-type ligands can induce anticancer activity in various types of cancer cell lines [1]. A set of twelve compounds representing ester derivatives of (*S*,*S*)-ethylenediamine-*N*,*N*′-di-2-(3-cyclohexyl)propanoic acid and (*S*,*S*)-1,3-propanediamine-*N*,*N*′-di-2-(3-cyclohexyl)propanoic acid (**Fig. 1**) has been recently synthesized and physico-chemically characterized at University of Belgrade – Faculty of Chemistry, Belgrade, Serbia. For novel ester derivatives of (*S*,*S*)-ethylenediamine-*N,N*′-di-2-(3-cyclohexyl)propanoic acid toxicity towards different cell lines was determined. Methyl, ethyl, and *n*-propyl esters were toxic to HL-60, REH, MOLT-4, KG-1, JVM-2, and K-562 leukemic cell lines, while the non-esterified acid and *n*-butyl ester were devoid of cytotoxicity. The ethyl ester exhibited the highest cytotoxicity on leukemic cell line HL-60 (IC_{50} was in the range 11 μM – 45 μM) [2]. 1,3-propanediamine-*N,N'*-di-2-(3-cyclohexyl)propanoic acid derivatives were toxic to glioma cell lines. *In vitro* antitumor potential was investigated for methyl, ethyl, *n*-propyl, and *n*-butyl esters of (*S*,*S*)-1,3-propanediamine-*N,N*′-di-2-(3-cyclohexyl)propanoic acid on several tumour cell lines: human (U251), and rat (C6) glioma, HL-60, SHSY-5Y, and L929. The *n*-butyl ester showed the highest cytotoxicity to glioma cells, with 24h IC_{50} values lower than those for cisplatin [3].

< **Fig. 1** >

In vitro assessment of gastrointestinal absorption could be performed using various methods, such as Parallel Artificial Membrane Permeability Assay (PAMPA) [4-6], biopartitioning micellar chromatography [7,8] and Caco-2 permeability experiments [9,10]. PAMPA has been extensively used for gastrointestinal absorption assessment of early drug candidates. It was firstly introduced by Kansy et al. [4] and this method is based on passive diffusion of tested compounds through artificial membrane. The mostly used artificial membrane is solution of egg lecithin in

dodecane (1% w/v) [4], but also other artificial membranes can be used for this purpose, such as mixture of hexadecane and hexane [6,11]. PAMPA is a high throughput and low cost method, amenable to automation and results obtained enable reliable prediction of *in vivo* passive gastrointestinal absorption. PAMPA is also used to assess membrane retention, which is of particular interest for compounds that target membrane proteins or compounds that are applied locally [12]. PAMPA results could also be used to develop quantitative structure – permeability relationships (QSPR) and quantitative structure – retention relationships (QSRR) models, which could be employed for the design of novel derivatives with improved permeability or membrane retention [12,13]. In general, QSPR refers to quantitative structure – property relationships analyses, but in PAMPA studies permeability is that particular property which is being modelled. QSRR usually refers to quantitative structure – chromatography retention relationships analyses, but in PAMPA studies membrane retention is used for the development of models.

The aims of this study were to estimate potential of (*S*,*S*)-ethylenediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid and (*S*,*S*)-1,3-propanediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid derivatives to be absorbed passively through gastrointestinal tract, to estimate their membrane retention properties and to create QSPR and QSRR models that could be used for the design of novel derivatives for which improved passive gastrointestinal absorption and membrane retention is expected.

2. Experimental

2.1. Chemicals

Investigated compounds (**Fig. 1**) were synthesized at University of Belgrade – Faculty of Chemistry, Belgrade, Serbia. Methanol (HPLC grade) was obtained from JT Baker (Loughborough, UK). Sodium hydroxide (NaOH), phosphoric acid (H_3PO_4), monosodium phosphate dihydrate (NaH_2PO_4 x $2H_2O$) and disodium phosphate dihydrate (Na_2HPO_4 x $2H_2O$) were purchased from Merck (Darmstadt, Germany). Dimethyl sulfoxide (DMSO), trifluoroacetic acid (CF_3COOH), ammonium acetate (CH_3COONH_4) and egg lecithin (99%) were purchased from Sigma-Aldrich (Saint Louis, Missouri, USA). Dodecane was obtained from Acros organics (Geel, Belgium).

2.2. Equipment

Analyses were performed on a UHPLC-MS system consisting of ACCELA UHPLC and TSQ Quantum Access Max triple quadrupole mass spectrometer with a heated electrospray ionization (HESI) interface (Thermo Scientific, Waltham, Massachusetts, USA). The column used in this study was Thermo Scientific Hypersil GOLD aQ (100 mm x 2.1 mm, 1.9 μm particle size). PAMPA was carried out in hydrophobic PVDF 96-well filter plates (MultiScreenTM HTS Millipore, Molsheim, France).

2.3 Parallel artificial membrane permeability assay

Phosphate buffer (pH = 7.4) was prepared by dissolving 13.7 g Na_2HPO_4 x $2H_2O$ and 4.0 g NaH_2PO_4 x $2H_2O$ in water (1000 ml) and pH value of this solution was adjusted with 0.1 M NaOH. Donor solutions were prepared by dissolving investigated substances in DMSO (1 mg/ml). 50 μl of thus prepared solutions were diluted with phosphate buffer (pH = 7.4) in 5 ml volumetric flasks to obtain final solutions (concentrations were approximately 20 μM). Acceptor

solutions were prepared by dissolving DMSO in phosphate buffer pH = 7.4 (concentration of DMSO was 1% v/v).

Artificial membrane was prepared by dissolving egg lecithin in dodecane (1 % w/v) and each well of the donor plate was coated with 5 µl of this solution. Subsequently, in each well of the acceptor plate, 400 µl of acceptor solution was transferred and covered by the donor plate. In each well of the donor plate 300 µl of donor solution was transferred. The system was incubated 16 h at room temperature. After the incubation, concentrations of each investigated substance in starting solution – $C_D(0)$, donor solution after incubation - $C_D(t)$ and acceptor solution after incubation – $C_A(t)$ were measured in triplicate by UHPLC/MS-MS method [14]. Mobile phase A was composed of ammonium acetate (5 mM) - trifluoroacetic acid (99.9:0.1, v/v), whereas mobile phase B was composed of methanol - trifluoroacetic acid (99.9:0.1, v/v). MS analysis was performed as selected reaction monitoring (SRM) for derivatives of (*S*,*S*)-ethylenediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid and single ion monitoring (SIM) for derivatives of (*S*,*S*)-1,3-propanediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid in positive mode. The monitored ions (m/z) are presented in **Table S1** (Supplementary material). The spray voltage was 4000 V, temperature in the capillary was adjusted to 300 °C and vaporizer temperature was set to 300 °C. Sheet gas pressure was 50 units, while the auxiliary valve flow rate was 10 units. MS resolution values were defined to correspond to a mass resolution of 0.7 Da. All data were acquired and processed by Xcalibur software (Thermo Fisher, San Jose, CA, USA).

Gastrointestinal membrane permeability parameters (percent of transport (%*T*) and apparent permeability coefficient (P_{app})) as well as retention factor (*R*) were calculated, according to following equations:

$$\%T = 100 \cdot (A_R \cdot V_A) / (A_{D0} \cdot V_D) \quad (1)$$

$$P_{app} = (V_D \cdot V_A) / ((V_D + V_A) \cdot S \cdot t) \cdot \ln [(100 \cdot V_D) / (100 \cdot V_D - \%T\,(V_D + V_A))] \quad (2)$$

$R = [1 - (C_D(t)/C_D(0))] - [V_A \cdot C_A(t)/V_D \cdot C_D(0)]$ (3)

A_{D0} and A_R - peak areas of the starting solution and the acceptor solution after incubation;

V_A and V_D - volumes of acceptor and donor solutions (ml);

S - surface area of the artificial membrane (0.28 cm^2)

t - incubation time (s);

$C_D(0)$ - concentration of investigated substance in the starting solution (mg/ml);

$C_D(t)$ - concentration of investigated substance in the donor solution after incubation (mg/ml);

$C_A(t)$ - concentration of investigated substance in the acceptor solution after incubation (mg/ml).

2.4. Calculation of molecular descriptors

2D structures of all tested compounds were generated in MarvinSketch program [15], dominant forms of each compound at pH 7.4 were predicted and used for subsequent molecular descriptor calculations. Molecular descriptors were calculated using freely available web-based platform ChemDes [16]. This platform allows users to compute over 3000 molecular descriptors from several open source packages. In this study, 1135 1D, 2D and 3D Chemopy descriptors were calculated (for the calculation of 3D descriptors, ChemDes uses the MOPAC software to optimize each molecule). After the elimination of those without variance, 873 descriptors were retained for further analyses.

2.5. QSPR and QSRR analyses

Descriptor selection as well as multiple linear regression (MLR), partial least squares (PLS) and support vector machine (SVM) modelling were performed in Statistica 13 [17].

The quantitative structure-permeability relationships (QSPR) and quantitative structure-retention relationships (QSRR) studies were performed to investigate the correlations between -logP_{app} and R (dependent variables) of the tested compounds and their calculated molecular descriptors (independent variables). In order to perform relevant comparison between different methodologies used to build QSPR models, the same training and test sets were used. For MLR(-logP_{app}), PLS(-logP_{app}) and SVM (-logP_{app}) test set consisted of 4 compounds (EDCP, DE-PDCP, DIB-EDCP and DIB-PDCP), while other compounds were chosen as training set. For MLR(R), PLS(R) and SVM (R) test set also consisted of 4 compounds (DE-PDCP, DB-PDCP, DIB-EDCP and DIPE-PDCP), while other derivatives formed training set. Test sets were formed in the way that -logP_{app} and R of these compounds were homogenously distributed in the whole range of -logP_{app} and R values.

2.5.1 Descriptor selection

Prior to model building, descriptor selection had to be performed. There are several descriptor selection methods, such as genetic algorithm [18], principal component analysis [19] or stepwise MLR [12]. In this study, following methods were utilized: forward stepwise MLR and feature selection and variable screening (FSVS). FSVS was applied for the selection of descriptors prior to MLR(-logP_{app}), SVM(-logP_{app}), SVM(R), PLS(-logP_{app}) and PLS(R) modeling, whereas MLR(R) model could not be created using descriptors selected in this way. Therefore, descriptors for MLR(R) modelling were selected using forward stepwise MLR.

In forward stepwise MLR, independent variables are added one by one into the model, evaluated at each step and finally retained or eliminated based on specified criteria (F to enter and F to remove criteria). In this study, for the selection of descriptors prior to MLR(R) modelling, F

to enter was set to 1, whereas F to remove was set to 0 and following descriptors were selected: Qass and RDFM14.

The methods implemented in the FSVS module are specifically designed to handle extremely large sets of continuous and/or categorical descriptors and evaluate both linear and nonlinear relationships between dependent variable and descriptors. Therefore, descriptors selected in this way are suitable for creation of both linear and nonlinear models. During FSVS analysis, the range of values in each descriptor is separated into k intervals. The type of investigated relationship is defined by the k value. For example, if k = 2 only monotonous relationships between descriptors and dependent variable are investigated. By default, ten intervals turned out to be the most suitable for the majority of analyses (k = 10). In this study, k was set to 10 and following variables were selected: grav and RDFC1 (for MLR(-logP_{app}) and SVM(-logP_{app})), as well as DZ and LabuteASA (for SVM(R)).

2.5.2 QSPR and QSRR model building

MLR was applied to assess linear relationship between calculated molecular descriptors and -logP_{app} and R. Both MLR models were created using forward stepwise multiple regression analysis with following criteria: F to enter = 2 and F to remove = 1 (for MLR (-logP_{app})); F to enter = 1 and F to remove = 0 (for MLR (R)).

Contrary to MLR, PLS modelling is useful when analysing data with collinear, noisy and numerous descriptors. Optimal number of PLS components was selected after the analysis of each component's R^2(Y) value and cumulative R^2(Y) value, which takes into account contribution of all analysed components to the PLS model. The influence of descriptors on created model was evaluated on the basis of their scaled regression coefficient values. Finally,

optimal PLS(-logP_{app}) and PLS(R) models consisted of two components and the most influential descriptors were analysed.

Apart from being used as a classification tool, SVM can also be used to generate real values in regression analysis and as a nonlinear method in QSAR and QSPR modelling [20,21]. In this study, optimal SVM(-logP_{app}) and SVM(R) models were obtained using radial basis function (RBF) Kernel type and regression type 1. Gamma value was optimized on the basis of S.D. ratio for training and test set values and was set to 1 in both SVM models. Subsequently, capacity (C) and epsilon (Ɛ) values were automatically optimized by the software and optimal values were: C = 100 and Ɛ = 0.1 for SVM (-logP_{app}); C = 10 and Ɛ = 0.1 for SVM(R). Finally, optimal SVM(-logP_{app}) consisted of 9 supported vectors (1 bounded) and optimal SVM (R) consisted of 8 supported vectors (4 bounded).

Following statistical parameters were calculated and used for the evaluation of quality of created QSPR and QSRR models: RMSEE (root mean squared error of estimation), RMSEP (root mean squared error of prediction), the F ratio, the p value, r, Q^2 (equation (4)) and R^2_{pred} (equation (5)).

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

$$R^2_{pred} = 1 - \frac{PRESS}{\sum (Y_{obs(test)} - \overline{Y}_{training})^2} \quad (5)$$

$$PRESS = \sum_{i=1}^{n} e_{(i)}^2 \quad (6)$$

RMSEE value was calculated for training, whereas RMSEP was calculated for test set. Q^2 is an internal validation parameter used to assess predictive potential of a model for compounds

similar to training set. This parameter was calculated according to the leave-one-out (LOO) procedure. Each compound of the training set was deleted once, while the remaining ones were used to create a model. The developed model was used to predict the value of the deleted compound. This procedure was repeated until all the training set compounds were deleted once [22]. Q^2 was calculated according to the equation (4). In this equation, $\overline{Y}_{training}$ is average value, whereas $Y_{obs(training)}$ is an observed -logP_{app} or R value of the training set compounds. PRESS was calculated after the completion of the LOO procedure, according to the equation (6). In this equation, $e_{(i)}$ is difference between observed and predicted -logP_{app} or R values. R^2_{pred} is an external validation parameter used to assess predictive potential of a model for compounds that differ in a certain manner from the training set. This parameter was calculated according to the equation (5). $Y_{obs(test)}$ is an observed value of -logP_{app} or R of a test set compound, whereas $\overline{Y}_{training}$ is mean -logP_{app} or R value of the training set compounds. PRESS was calculated for the test set according to the equation (6). High predictive potential of a model could be expected if Q^2 and R^2_{pred} are higher than 0.5 [22,23].

The F-test is based on the ratio MS Regression/MS Residual and evaluates significance of the model. The p-value indicates probability level where a model with this F value may be the result of just chance. Models with p-value lower than 0.05 are considered significant.

3. Results and discussion

3.1. PAMPA test

The values of permeation parameter $C_A(t)/C_D(0)$, apparent permeability coefficient (P_{app}), negative logarithm of P_{app} ($-\log P_{app}$) and retention factor (R) are presented in **Table 1**.

< **Table 1** >

On the basis of relationship between R and $C_A(t)/C_D(0)$ compounds can be classified into four classes – substances with negligible membrane retention and low permeation (class I), substances with low or negligible membrane retention and high permeation (class II), substances with high membrane retention and low permeation (class III) and substances with high membrane retention and high permeation (class IV). All compounds tested in this study belong to class III, apart from EDCP, DPE-PDCP and DIPE-PDCP, which belong to class I (**Fig. 2**).

<**Fig.2**>

Derivatives of (*S*,*S*)-ethylenediamine-*N*,*N*′-di-2-(3-cyclohexyl)propanoic acid possess higher R and P_{app} than derivatives of (*S*,*S*)-1,3-propanediamine-*N*,*N*′-di-2-(3-cyclohexyl)propanoic acid. The highest values of R were determined for diethyl ester of (*S*,*S*)-ethylenediamine-*N*,*N*′-di-2-(3-cyclohexyl)propanoic acid (DE-EDCP), as well as for diethyl and dibutyl esters of (*S*,*S*)-1,3-propanediamine-*N*,*N*′-di-2-(3-cyclohexyl)propanoic acid (DE-PDCP and DB-PDCP), whereas the lowest values were determined for non-esterified acids EDCP and PDCP. These membrane retention data are in good agreement with the results of previous *in vitro* activity studies on various leukemic cell lines, because DE-EDCP and DE-PDCP showed the highest, whereas EDCP and PDCP showed the lowest cytotoxic activity [2,3]. These results indicate that possible mechanism of cytotoxicity might be related to interactions with membrane proteins, cell or

mitochondrial membranes and further design of new derivatives with high membrane retention might result in potent cytotoxicity.

Evaluation of lipophilicity of tested compounds was presented in our previous paper [24]. According to these results, length of alkyl chain on ester groups has significant influence on lipophilicity. Results presented in this paper show that lipophilicity is not well correlated with membrane permeability and length of alkyl chain on ester groups have limited influence on *R*. Therefore, other physico-chemical properties also contribute to membrane permeability and retention, which can be clarified by QSPR and QSRR studies.

3.2. QSPR and QSRR studies

QSPR models (MLR(-logP_{app}), PLS(-logP_{app}) and SVM(-logP_{app})), QSRR models (MLR(*R*), PLS(*R*) and SVM(*R*)) and corresponding statistical parameters are presented in **Table 2**.

<**Table 2**>

According to presented results, PLS(*R*) model should not be taken into further consideration since it didn't pass all validation tests (Q^2 = 0.042). The most reliable QSPR model for -logP_{app} prediction is PLS(-logP_{app}) due to the lowest values of RMSEE (0.127) and RMSEP (0.238), as well as due to the highest value of R^2_{pred} (0.893). Similarly, the most reliable QSRR model for the prediction of *R* is SVM(*R*) due to the lowest values of RMSEE (13.282) and RMSEP (11.161), as well as due to the highest values of Q^2 (0.641), r (0.986) and R^2_{pred} (0.891). Both selected models can be considered to have good predictive capacity due to the Q^2>0.5 and R^2_{pred} >0.5.

3.2.1. Interpretation of PLS(-logP_{app}) model and design of novel derivatives

Descriptors with highest negative influence on -logP_{app} and consequently highest positive influence on permeability are RDFC8, MoRSEC20, MoRSEP10 and MoRSEV10. MoRSEC20, MoRSEP10 and MoRSEV10 are 3D MoRSE descriptors based on atomic charge, polarizability and van der Waals volume, respectively. Interpretation of these descriptors is very difficult and sometimes their meaning is not clear [25-27]. According to Devinyak et al. [28], aromatic rings and unsaturated bonds increase values of these descriptors. Therefore, it can be expected that replacement of cyclohexyl with cyclohexenyl, cyclohexadienyl or phenyl groups could positively affect PAMPA permeability and gastrointestinal absorption.

Descriptors with highest positive influence on -logP_{app} and consequently highest negative influence on permeability are RDFU6, RDFP5, RDFE6 and RDFV5. RDFU6 is an unweighted radial distribution function (RDF) descriptor, whereas RDFP5, RDFE6 and RDFV5 are radial distribution function (RDF) descriptors based on atomic polarizability, electronegativity and van der Waals value, respectively. Numbers 5 and 6 indicate distance of 3-4 Å from the geometrical center, which represents spacer between two –NH groups in tested compounds. To better understand the influence of these descriptors on PAMPA permeability, pairs of tested compounds with the same distance between –NH groups were analysed. Generally, lower values of these descriptors occur within the ethylenediamine in comparison to 1,3-propanediamine derivatives, resulting in higher PAMPA permeability of ethylenediamine derivatives. Therefore, the length of spacer between two –NH groups is important and design of novel derivatives could be based on modifications of length and/or introduction of branched side chain in this part of structure.

Taking into account these conclusions, twelve derivatives were designed, their molecular descriptors were calculated as previously described and -logP_{app} predicted using PLS(-logP_{app})

model. These compounds were designed by modifications of DM-EDCP structure (compound with highest PAMPA permeability) in the following manner: replacement of one or both cyclohexyl groups with one or two unsaturated rings (1-cyclohexenyl, 2-cyclohexenyl or phenyl), reduction of the spacer length between two –NH groups and introduction of branched side chains in this spacer. Structures of all designed derivatives are presented in **Table S2** (Supplementary material), whereas five designed derivatives (**P1, P2, P3, P4** and **P6)** with the highest predicted PAMPA permeability (lowest -logP_{app}) are presented in **Fig. 3**. Predicted -logP_{app} of these derivatives (2.12-2.67) are similar or slightly higher in comparison to DM-EDCP (1.87).

<**Fig.3**>

3.2.2. Interpretation of SVM(R) model and design of novel derivatives

Descriptors which form SVM(*R*) model are DZ and LabuteASA. DZ is Pogliani index and belongs to topological descriptors. It can be defined as the sum over all non-hydrogen atoms of a modified vertex degree calculated as the ratio of the number of valence electrons over the principal quantum number of an atom [29]. Relationship between *R* and Dz values of all tested compounds is presented in **Fig. 4**.

<**Fig.4**>

Presented graph indicates that optimal values of Dz should be within the range 63 (DM-PDCP) - 73 (DB-EDCP and DIB-EDCP). In order to keep Dz values in optimal range, design of new derivatives should be performed by structural modifications of compounds whose Dz values

fall into this range by rearrangement of existing groups and total number of C atoms should not be changed. Introduction of functional groups with heteroatoms could have different effects on Dz values, which cannot be easily predicted, and it should be evaluated for each designed derivative separately.

LabuteASA is Labute approximate surface area and belongs to MOE-type descriptors. Calculation of this descriptor and additional explanation were presented by Labute [30]. Relationship between *R* and LabuteASA values of all tested compounds is presented in **Fig. 5**.

<**Fig.5**>

Similarly to Dz, range of optimal LabuteASA values can be defined: 176 (DM-PDCP) - 210 (DB-EDCP and DIB-EDCP). Therefore, design of new derivatives should be performed in the same manner – structural modifications of compounds whose LabuteASA values fall into this range by rearrangement of existing groups without the change of total number of C atoms. Additionally, isosteric replacements could also be considered.

On the basis of these conclusions, sixteen derivatives were designed, Dz and LabuteASA values were calculated as previously described and *R* predicted using SVM(*R*) model. Structures of all designed derivatives are presented in **Table S3** (Supplementary material), whereas five designed derivatives (**R1, R9, R14, R15** and **R16)** with the highest predicted *R* values are presented in **Fig. 6**. Predicted *R* values of these derivatives (88.40-91.27) are similar to DE-PDCP (91.27).

<**Fig.6**>

4. Conclusions

Passive gastrointestinal absorption and membrane retention of twelve esters of (*S*,*S*)-ethylenediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid and (*S*,*S*)-1,3-propanediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid, as well as of these two non-esterified acids were estimated using PAMPA test, on the basis of permeation ($C_A(t)/C_D(0)$), permeability (P_{app}) and retention (R) parameters. All tested compounds belong to class III (high membrane retention and low permeation), whereas EDCP, DPE-PDCP and DIPE-PDCP belong to class I (negligible membrane retention and low permeation). (*S*,*S*)-ethylenediamine-*N*,*N'*-di-2-(3-cyclohexyl) propanoic acid derivatives possess higher R and P_{app} than derivatives of (*S*,*S*)-1,3-propanediamine-*N*,*N'*-di-2-(3-cyclohexyl)propanoic acid. The highest values of R were determined for DE-EDCP, DE-PDCP and DB-PDCP, whereas the lowest values were determined for non-esterified acids EDCP and PDCP. Finally, QSPR (MLR(-logP_{app}), PLS(-logP_{app}) and SVM(-logP_{app})) and QSRR (MLR(R), PLS(R) and SVM(R)) models were created in order to find quantitative relationships between physico-chemical properties of tested compounds and their permeability/retention parameters. Statistically the most reliable models were selected and on the basis of their interpretation, new compounds with expected favourable permeability and retention were designed.

Dobricic_Author statement

Biljana Tubić: Methodology, Investigation, Writing-Original Draft. **Vladimir Dobričić**: Methodology, Investigation, Writing-Original draft, Formal Analysis. **Jelena Poljarević**: Resources, Writing-Original draft. **Aleksandar Savić**: Resources, Writing-Original draft. **Tibor**

Sabo: Resources, Supervision, Project administration. **Bojan Marković**: Methodology, Investigation, Supervision, Project administration.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia [grant numbers OI172041 and OI172035].

References

[1] M. A. Musa, V. L. Badisa, L. M. Latinwo, Cytotoxic activity of N, N'-Bis (2-hydroxybenzyl) ethylenediamine derivatives in human cancer cell lines, Anticancer. Res. 34 (2014) 1601-1607.

[2] S. Misirlic Dencic, J. Poljarevic, U. Vilimanovich, A. Bogdanovic, A. J. Isakovic, T. Kravic Stevovic, M. Dulovic, N. Zogovic, A. M. Isakovic, S. Grguric-Sipka, V. Bumbasirevic, T. Sabo, V. Trajkovic, I. Markovic, Cyclohexyl analogues of ethylenediamine dipropanoic acid induce caspase-independent mitochondrial apoptosis in human leukemic cells, Chem. Res. Toxicol. 25 (2012) 931-939. doi: 10.1021/tx3000329.

[3] A. Savić, S. Misirlić-Denčić, M. Dulović, L. E. Mihajlović-Lalić, M. Jovanović, S. Grgurić-Šipka, I. Marković, T. J. Sabo, Synthesis, characterization and ROS-mediated cytotoxic action of novel (S, S)-1, 3-propanediamine-N, N′-di-2-(3-cyclohexyl) propanoic acid and corresponding esters, Bioorg. Chem. 54 (2014) 73-80. doi: 10.1016/j.bioorg.2014.04.006.

[4] M. Kansy, F. Senner, K. Gubernator, Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes, J. Med. Chem. 41 (1998) 1007-1010. doi: 10.1021/jm970530e.

[5] M. Kansy, A. Avdeef, H. Fischer, Advances in screening for membrane permeability: high-resolution PAMPA for medicinal chemists, Drug Discov. Today: Technologies. 1 (2004) 349-355. doi: 10.1016/j.ddtec.2004.11.013.

[6] A. Bujard, H. Voirol, P. A. Carrupt, J. Schappler, Modification of a PAMPA model to predict passive gastrointestinal absorption and plasma protein binding, Eur. J. Pharm. Sci. 77 (2015) 273-278. doi: 10.1016/j.ejps.2015.06.023.

[7] M.Molero-Monfort, L.Escuder-Gilabert, R.M.Villanueva-Camañas, S.Sagrado, M.J.Medina-Hernández, Biopartitioning micellar chromatography: an in vitro technique for predicting human drug absorption, J. Chromatogr. B. 753 (2001) 225-236. 10.1016/S0378-4347(00)00546-6.

[8] V. Dobričić, J. Savić, K. Nikolic, S. Vladimirov, Z. Vujić, J. Brborić, Application of biopartitioning micellar chromatography and QSRR modeling for prediction of gastrointestinal absorption and design of novel β-hydroxy-β-arylalkanoic acids, Eur. J. Pharm. Sci. 100 (2017) 280-284. doi: 10.1016/j.ejps.2017.01.023.

[9] P. Artursson, J. Karlsson, Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells, Biochem. Bioph. Res. Co. 175 (1991) 880-885. doi: 10.1016/0006-291x(91)91647-u.

[10] S. Yee, In vitro permeability across Caco-2 cells (colonic) can predict in vivo (small intestinal) absorption in man—fact or myth, Pharm. Res. 14 (1997) 763-766. doi: 10.1023/a:1012102522787.

[11] C. Petit, A. Bujard, K. Skalicka-Woźniak, S. Cretton, J. Houriet, P. Christen, P. A. Carrupt, J. L. Wolfender, Prediction of the passive intestinal absorption of medicinal plant extract constituents with the Parallel Artificial Membrane Permeability Assay (PAMPA), Planta Med. 82 (2016) 424-431. doi: 10.1055/s-0042-101247.

[12] V. Dobričić, B. Marković, K. Nikolic, V. Savić, S. Vladimirov, O. Čudina, 17β-carboxamide steroids–in vitro prediction of human skin permeability and retention using PAMPA technique, Eur. J. Pharm. Sci. 52 (2014) 95-108. doi: 10.1016/j.ejps.2013.10.017.

[13] J. Savić, V. Dobričić, K. Nikolic, S. Vladimirov, S. Dilber, J. Brborić, In vitro prediction of gastrointestinal absorption of novel β-hydroxy-β-arylalkanoic acids using PAMPA technique, Eur. J. Pharm. Sci. 100 (2017) 36-41. doi: 10.1016/j.ejps.2017.01.005.

[14] B. Tubić, B. Marković, S. Vladimirov, S. Ristić, B. Ivković, M. Savić, J. Poljarević, T. Sabo, Highly sensitive UHPLC–MS/MS method for quantification of ethylenediamine-N,N′-di-2-(3-cyclohexyl) propanoic acid derivatives in mouse serum, Acta Chromatogr. 29 (2017) 235-252. doi: 10.1556/1326.2017.29.2.7.

[15] ChemAxon, 2013. MarvinSketch 6.1.3. Budapest, Hungary.

[16] J. Dong, D. S. Cao, H. Y. Miao, S. Liu, B-C. Deng, Y-H. Yun, N-N. Wang, A-P. Lu, W-B. Zeng, A. F. Chen, ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. J. Cheminformatics, 7 (2015) 60. doi: 10.1186/s13321-015-0109-z.

[17] TIBCO Software Inc., 2017. Statistica (data analysis software system), version 13. http://statistica.io.

[18] V. K. Gupta, H. Khani, B. Ahmadi-Roudi, S. Mirakhorli, E. Fereyduni, S. Agarwal, Prediction of capillary gas chromatographic retention times of fatty acid methyl esters in human blood using MLR, PLS and back-propagation artificial neural networks, Talanta 83 (2011) 1014–1022. doi: 10.1016/j.talanta.2010.11.017.

[19] Y. X. Zhang, Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis, Talanta 73 (2007) 68–75. doi: 10.1016/j.talanta.2007.02.030.

[20] K. Heikamp, J. Bajorath, Support vector machines for drug discovery, Expert. Opin. Drug Dis. 9 (2014) 93−104. doi: 10.1517/17460441.2014.866943.

[21] J. Vucicevic, M. Popovic, K. Nikolic, S. Filipic, D. Obradovic, D. Agbaba, Use of biopartitioning micellar chromatography and RP-HPLC for the determination of blood–

brain barrier penetration of α-adrenergic/imidazoline receptor ligands, and QSPR analysis, SAR QSAR Environ. Res. 28 (2017) 235-252. doi: 10.1080/1062936X.2017.1302506.

[22] P. K. Ojha, K. Roy, Comparative QSARs for antimalarial endochins: Importance of descriptor-thinning and noise reduction prior to feature selection, Chemometr. Intell. Lab. 109 (2011) 146–161. doi: 10.1016/j.chemolab.2011.08.007.

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

[24] B. Tubić, B. Marković, S. Vladimirov, A. Savić, J. Poljarević, T. Sabo, A new model to determine lipophilicity of 1,2-ethanediamine-N,N'-di-2-(3-cyclohexyl)propanoic acid and 1,3-propanediamine-N,N'-di-2-(3-cyclohexyl)propanoic acid derivatives with antiproliferative activity by combining shake flask procedure and UHPLC-MS method, Pharmazie 72 (2017) 317-323. doi: 10.1619/ph.2017.6208.

[25] P. R. Duchowicz, M. Fernández, J. Caballero, E. A. Castro, F. M. Fernández, QSAR for non-nucleoside inhibitors of HIV-1 reverse transcriptase, Bioorgan. Med. Chem. 14 (2006) 5876-5889. doi: 10.1016/j.bmc.2006.05.027.

[26] T. Hancock, R. Put, D. Coomans, Y. Vander Heyden, Y. Everingham, A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies, Chemometr. Intell. Lab. Systems 76 (2005) 185-196. doi: 10.1016/j.chemolab.2004.11.001.

[27] R. Todeschini, V. Consonni, Handbook of molecular descriptors, Wiley-VCH, Weinheim, 2000. doi: 10.1002/9783527613106.

[28] O. Devinyak, D. Havrylyuk, R. Lesyk, 3D-MoRSE descriptors explained, J. Mol. Graph. Model. 54 (2014) 194-203. doi: 10.1016/j.jmgm.2014.10.006.

[29] L. Pogliani, Modeling with special descriptors derived from a medium-sized set of connectivity indices, J. Phys. Chem. 100 (1996) 18065-18077. doi: 10.1021/jp961434c.

[30] P. Labute, A widely applicable set of descriptors. J. Mol. Graph. Model. 18 (2000) 464-477. doi: 10.1016/s1093-3263(00)00068-1.

Figure captions

Fig. 1. Chemical structures of tested compounds

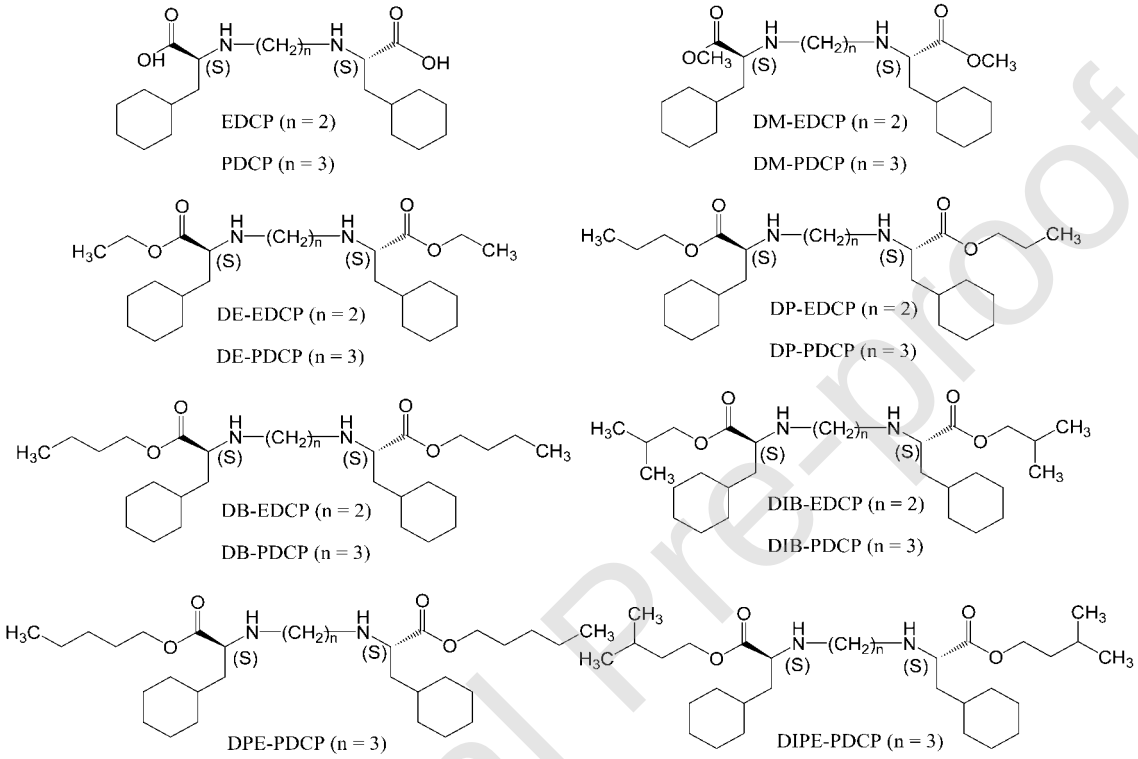


Fig. 2. The relationship between R and $C_A(t)/C_D(0)$ of tested compounds.

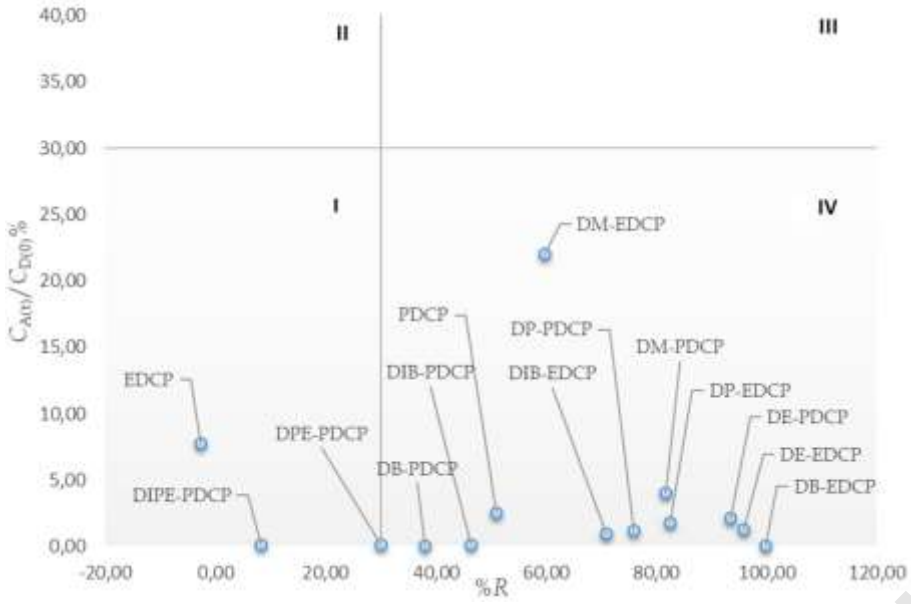


Fig. 3. Chemical structures of five designed derivatives with highest predicted PAMPA permeability (predicted -logP_{app} values are given in brackets).

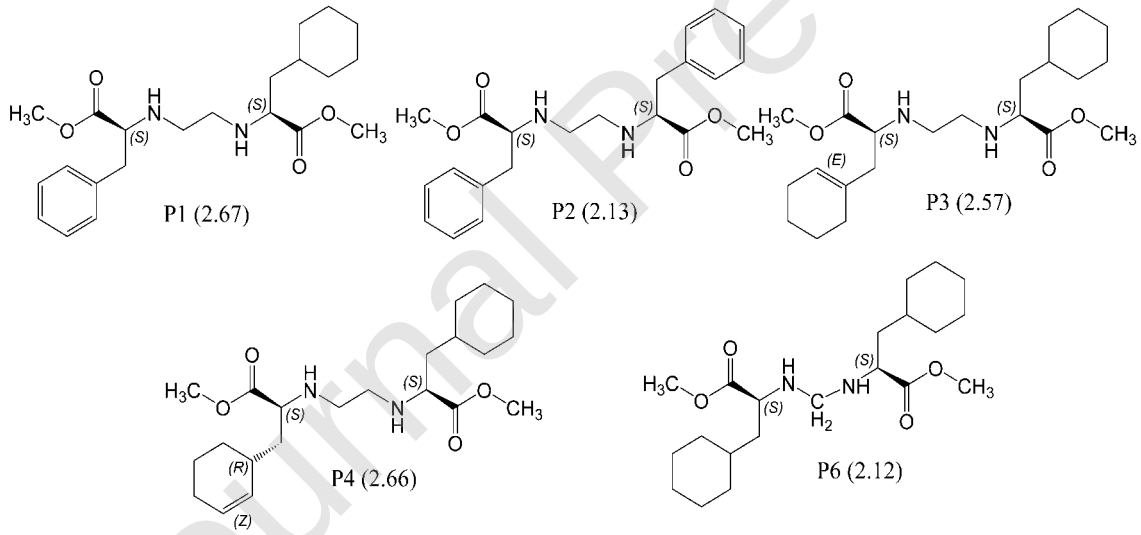


Fig. 4. Relationship between *R* (var3) and Dz (var185).

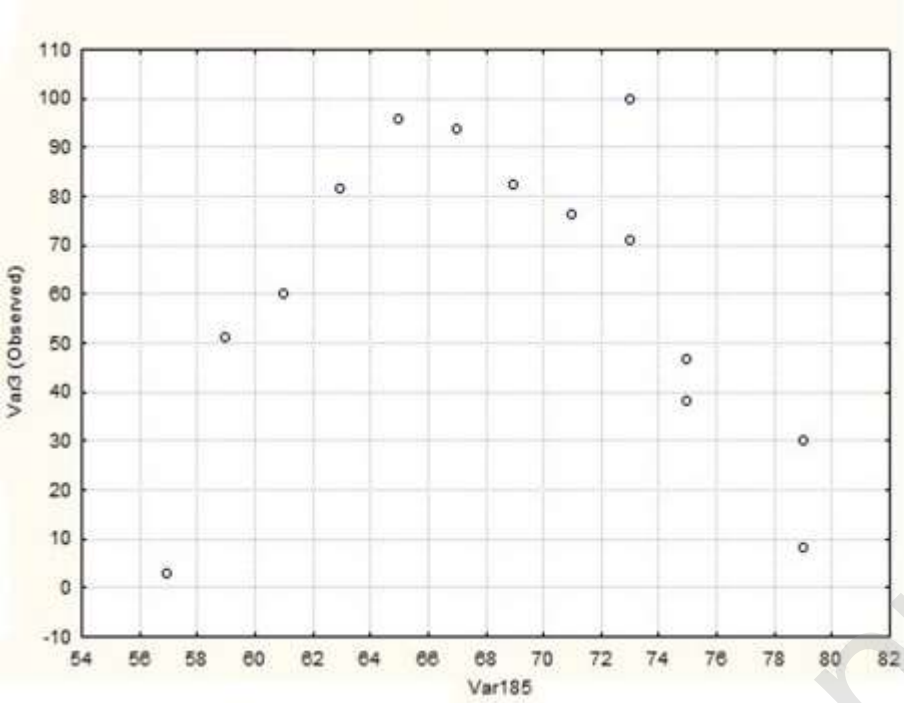


Fig. 5. Relationship between *R* (var3) and LabuteASA (var298).

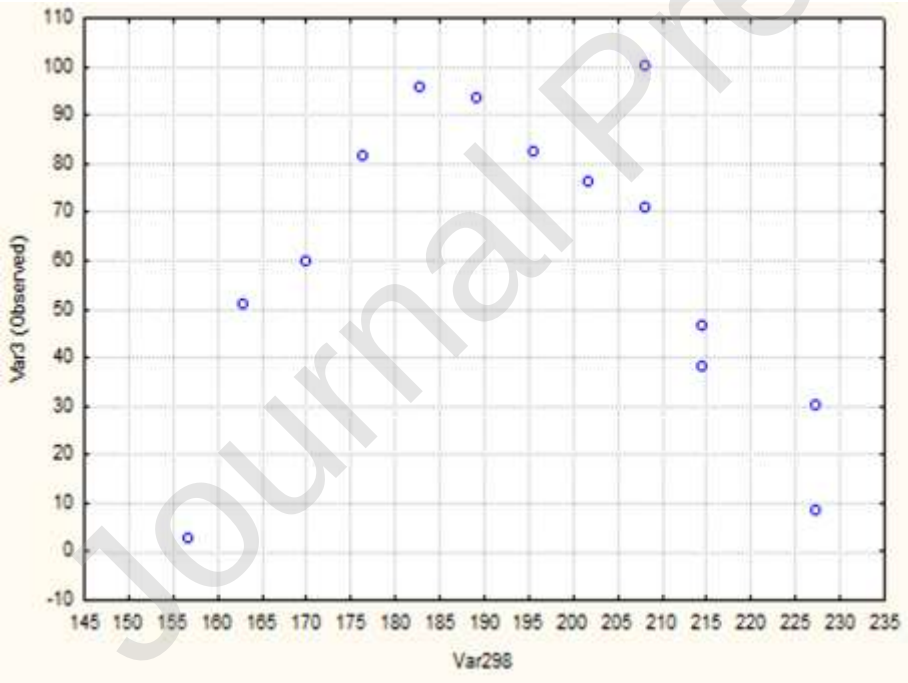
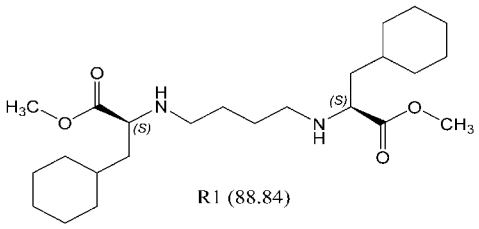
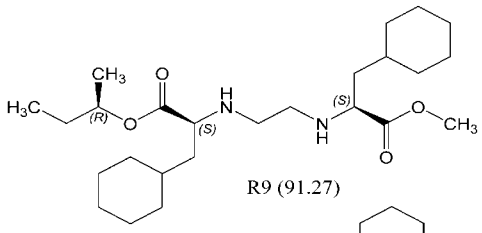


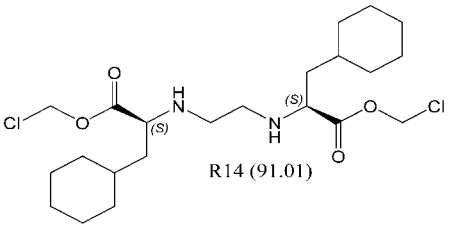
Fig. 6. Chemical structures of five designed derivatives with highest predicted membrane retention (predicted *R* values are given in brackets).



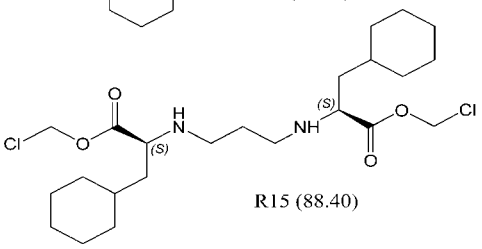
R1 (88.84)



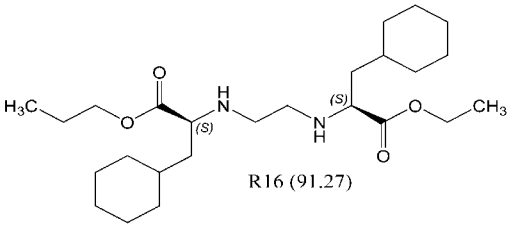
R9 (91.27)



R14 (91.01)



R15 (88.40)



R16 (91.27)

Table 1

Calculated PAMPA parameters of tested compunds.

Compound	$C_A(t)/C_D(0)$	P_{app} (x 10^{-6})	-logP_{app}	R
EDCP	7.73	2749.9	2.56	2.61
DM-EDCP	22.00	11540.0	1.94	59.78
DE-EDCP	1.32	419.3	3.38	95.83
DP-EDCP	1.74	555.8	3.26	82.46
DB-EDCP	0.07	20.6	4.69	99.89
DIB-EDCP	0.93	293.9	3.53	70.94
PDCP	2.50	810.6	3.09	51.08
DM-PDCP	4.02	1335.1	2.87	81.64
DE-PDCP	2.13	686.0	3.16	93.43
DP-PDCP	1.19	376.4	3.42	76.01
DB-PDCP	0.03	9.6	5.02	38.07
DIB-PDCP	0.11	33.6	4.47	46.50
DPE-PDCP	0.12	37.1	4.43	30.09
DIPE-PDCP	0.12	38.6	4.41	8.25

Table 2

QSPR and QSRR models with corresponding statistical parameters.

Model	Regression equations/selected descriptors	RMSEE	RMSEP	Q^2	r	R^2_{pred}	F	p
MLR(-logP_{app})	-logP_{app} = (-5.50784±1.22990) + (0.09753±0.01308) ·grav - (8.33665±3.16406)·RDFC1 [a]	0.288	0.388	0.715	0.970	0.715	31.453	<0.00032
PLS(-logP_{app})	-logP_{app} = f(RDFC8, MoRSEC20, MoRSEP10, MoRSEV10, RDFU6, RDFP5, RDFE6, RDFV5) [b]	0.127	0.238	0.701	0.946	0.893	-	-
SVM(-logP_{app})	-logP_{app} = f(grav, RDFC1) [a]	0.293	0.329	0.389	0.975	0.795	-	-
MLR(R)	R = (712.537±149.637) - (491.037±113.501)·Qass - (6.747±2.454)·RDFM14 [c]	15.128	15.089	0.601	0.900	0.801	9.480	<0.01018
PLS(R)	R = f (RDFC6, MoRSEV24, MoRSEN24, MoRSEM24, MoRSEC23, RDFC4, MoRSEC13, DPSA3) [d]	10.908	20.652	0.042	0.914	0.627	-	-
SVM(R)	R = f (DZ, LabuteASA) [e]	13.282	11.161	0.641	0.986	0.891	-	-

[a] grav – Gravitational 3D index; RDFC1 - Radial distribution function (RDF) descriptor based on atomic charge;

[b] RDFC8 - Radial distribution function (RDF) descriptor based on atomic charge; MoRSEC20 - 3D MoRSE descriptor based on atomic charge; MoRSEP10 - 3D MoRSE descriptor based on atomic polarizability; MoRSEV10 - 3D MoRSE descriptor based on atomic van der Waals volume; RDFU6 - Unweighted radial distribution function (RDF) descriptor; RDFP5 - Radial distribution function (RDF) descriptor based on atomic polarizability; RDFE6 - Radial distribution function (RDF) descriptor based on atomic electronegativity; RDFV5 - Radial distribution function (RDF) descriptor based on atomic van der Waals volume;

[c] Qass - Sum of squares of charges on H,C,N,O and all atoms; RDFM14 - Radial distribution function (RDF) descriptors based on atomic mass;

[d] RDFC6 - Radial distribution function (RDF) descriptor based on atomic charge; MoRSEV24 - 3D MoRSE descriptor based on atomic van der Waals volume; MoRSEN24 - 3D MoRSE descriptor based on atomic number; MoRSEM24 - 3D MoRSE descriptor based on atomic mass; MoRSEC23 - 3D MoRSE descriptor based on atomic charge; RDFC4 - Radial distribution function (RDF) descriptor based on atomic charge; MoRSEC13 - 3D MoRSE descriptor based on atomic charge; DPSA3 - Difference in atomic charge weighted surface area; [e] DZ - Pogliani index; LabuteASA - Labute's Approximate Surface Area.