

Antitumor activity of 3,4-ethylenedioxythiophene derivatives and quantitative structure–activity relationship analysis

Marijana Jukić^a, Vesna Rastija^{b,*}, Teuta Opačak-Bernardi^a, Ivana Stolić^c,
Luka Krstulović^c, Miroslav Bajić^c, Ljubica Glavaš-Obrovac^a

^a Department of Chemistry, Biochemistry and Clinical Chemistry, Faculty of Medicine, J. J. Strossmayer University of Osijek, Croatia

^b Department of Chemistry, Faculty of Agriculture, J. J. Strossmayer University of Osijek, Croatia

^c Department of Chemistry and Biochemistry, Faculty of Veterinary Medicine, University of Zagreb, Croatia

ARTICLE INFO

Article history:

Received 31 August 2016

Received in revised form

24 November 2016

Accepted 24 November 2016

Available online 25 November 2016

Keywords:

Antitumor agents

Heterocycles

Computational chemistry

QSAR

Drug research

Cell culturing

ABSTRACT

The aim of this study was to evaluate nine newly synthesized amidine derivatives of 3,4-ethylenedioxythiophene (3,4-EDOT) for their cytotoxic activity against a panel of human cancer cell lines and to perform a quantitative structure–activity relationship (QSAR) analysis for the antitumor activity of a total of 27 3,4-ethylenedioxythiophene derivatives. Induction of apoptosis was investigated on the selected compounds, along with delivery options for the optimization of activity. The best obtained QSAR models include the following group of descriptors: BCUT, WHIM, 2D autocorrelations, 3D-MORSE, GETAWAY descriptors, 2D frequency fingerprint and information indices. Obtained QSAR models should be relieved in elucidation of important physicochemical and structural requirements for this biological activity. Highly potent molecules have a symmetrical arrangement of substituents along the x axis, high frequency of distance between N and O atoms at topological distance 9, as well as between C and N atoms at topological distance 10, and more C atoms located at topological distances 6 and 3. Based on the conclusion given in the QSAR analysis, a new compound with possible great activity was proposed.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Cancer is a major global health problem and its rates increase every year. In 2012, there were 14.1 million new cancer cases and 8.2 million deaths directly caused by cancer. Cancer treatment research is fundamental to improving outcomes for patients affected by the disease. These efforts include the development of more effective, more selective and less toxic treatments, such as targeted therapies, immunotherapies and cancer vaccines. As part of the research, special attention is paid to the discovery of new molecules that selectively bind to DNA [1,2]. Over the past few decades, a relatively large number of useful anticancer drugs have been discovered or rationally designed based on the principle of nucleic acids recognition [3,4].

Aromatic amidines are structural parts of numerous compounds

of biological interest and form important medical and biochemical agents exhibiting a broad spectrum of significant antimicrobial activity [5–9] and potency, but so far they have been less investigated for their antitumor activity [10,11]. Although the mechanism of action of aromatic amidines has not been fully elucidated, it has been proven that their bioactivity is a direct result of DNA binding and subsequent inhibition of DNA-dependent enzymes or possibly direct inhibition of transcription [12,13]. Despite a broad-spectrum activity of these compounds, only pentamidine has been found to be of significant use in humans, although it displays several adverse side effects. Its toxicity, lack of oral availability and appearance of pentamidine resistance stimulate the development of additional drugs for treatment. A large number of pentamidine analogues have been synthesized and intended to replace the unstable and flexible alkyldiether linker with conformation-restricted five-member heterocycles. Further modifications of two aromatic moieties and terminal amidine group had the objective to improve useful therapeutic properties and reduce undesirable effects. For this reason, over time, our group has synthesized and tested a number of new analogues of pentamidine with stable 3,4-ethylenedioxythiophene as a central linker [7,14–16]. Benzene

* Corresponding author.

E-mail addresses: marijanajukic@mefos.hr (M. Jukić), vrastija@pfos.hr (V. Rastija), tbernardi@mefos.hr (T. Opačak-Bernardi), stolic@vef.hr (I. Stolić), lkrstulovic@vef.hr (L. Krstulović), mbajic@vef.hr (M. Bajić), lgobrovac@mefos.hr (L. Glavaš-Obrovac).

and/or benzimidazole rings were aromatic moieties and unsubstituted, alkyl-substituted or cyclic amidines were used as terminal groups. These structural modifications were meant to increase the stability of the complex formed with DNA minor groove and enhance the biological activity of the compounds. All synthesized compounds were tested on various human cancer cell lines and some of them showed highly significant inhibitory activity.

The aim of this study was to determine the antitumor activity of 27 3,4-EDOT derivatives against six carcinoma cell lines and derive a quantitative structure–activity relationship (QSAR) analysis. The goal of the QSAR analysis was to find out which physicochemical and quantum-chemical molecular properties influence enhanced antitumor activity.

2. Material and methods

2.1. Compounds

The synthesis, physical properties and antitumor activity of symmetrical EDOT's derivatives (**1–18**) have been described previously [14–16]. The synthesis and physical properties of asymmetrical compounds **19–27** were given earlier [7] and their antitumor activity is described in the present study. Structural details of all studied molecules are shown in Fig. 1.

2.2. Cell culturing and 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) test

Antitumor activity for compounds **19** through **27** was tested as follows: tumour cell line derived from MIA PaCa2, CaCo2, HEP2 and HeLa were grown in DMEM medium (Gibco, EU). The NCI H358 and AGS cell lines were grown in RPMI 1640 medium (Gibco, EU). Both media were supplemented with 10% heat-inactivated fetal bovine serum-FBS (Gibco, EU), 2 mM glutamine (Gibco, EU), 1 mM sodium pyruvate (Gibco, EU), 10 mM HEPES (Sigma-Aldrich, USA) and 100 U/0.1 mg antibiotic/antimycotic (Gibco, EU).

Cells were grown on 37 °C, with 5% CO₂ gas in humidified CO₂ incubator (IGO 150 CELLlife™, JOUAN, Thermo Fisher Scientific, Waltham, MA, USA). A trypan blue dye exclusion method was used to assess cell viability. Tested compounds were dissolved in DMSO (dimethyl sulfoxide) as a 1×10^{-2} M (mol dm⁻³) stock solution. Working dilutions were prepared at a concentration range 10^{-3} – 10^{-6} M.

For the MTT test [17] cells were seeded on 96 micro well flat bottom plates (Greiner, Austria) at 2×10^4 cells/ml. After 72 h of incubation with the tested compounds MTT (Merck, Germany) was added. DMSO (Merck, Germany) was used to dissolve the formed MTT-formazane crystals. Absorbance was measured at 595 nm on Elisa micro plate reader (iMark, BIO RAD, Hercules, CA, USA). All experiments were performed three times in triplicates. The IC₅₀ values, defined as the concentration of compound achieving 50% of cell growth inhibition, were calculated, and used to compare cytotoxicity among the compounds.

2.3. Compound delivery

Two delivery systems were tested on compounds **19** and **20**, liposome and albumin based, respectively. Lipofectamine (Sigma, EU) was used to form liposomes containing selected compounds. Selected compounds were diluted in OptiMEM (Gibco, EU) to the concentration of 2×10^{-5} mol dm⁻³ and then, mixed with equal volumes of 10% Lipofectamine in OptiMEM according to the manufacturer's instructions. Final concentration of Lipofectamine was 5% and compound concentration was 10^{-5} mol dm⁻³. Fresh liposomes were made for every treatment.

Albumin particles were made by solvation as previously described [18,19]. Briefly, albumin, fraction V (Roche, Switzerland) was dissolved in sterile water to a concentration of 100 mg/mL, while the selected compound was dissolved to target concentration of 10^{-5} mol dm⁻³ in ethanol. Compound solution was then added dropwise to the albumin solution under constant mixing with final ethanol to albumin volume ratio of 2:1. Glutaraldehyde (8%, v/v, Sigma, EU) was added to enable cross-linking and the solution was left for 24 h with constant mixing. After 24 h particles were centrifuged at 600×g for 10 min and the pellet was resuspended in water in an ultrasound water bath for 10–15 min. Samples were washed 3 times before the resulting solution was stored at 4 °C. The both delivery systems were tested on CML cells (K562) using MTT test as described above.

2.4. Apoptosis induction

To determine apoptosis induction after treatment by selected compounds we measured changes in mitochondrial membrane potential using the Mitochondrial Membrane Potential Kit (Sigma, EU). In brief, K562 cells were plated in 6-well plates (5×10^5 cells/mL) and treated with selected compounds at the concentration 10^{-5} M for 24 h. Cells were then stained according to the kit protocol and analysed by flow cytometry (FacsCanto II, BD Biosciences, USA) using Flowlogic software (Inivai Technologies).

2.5. Regression analysis and validation of models

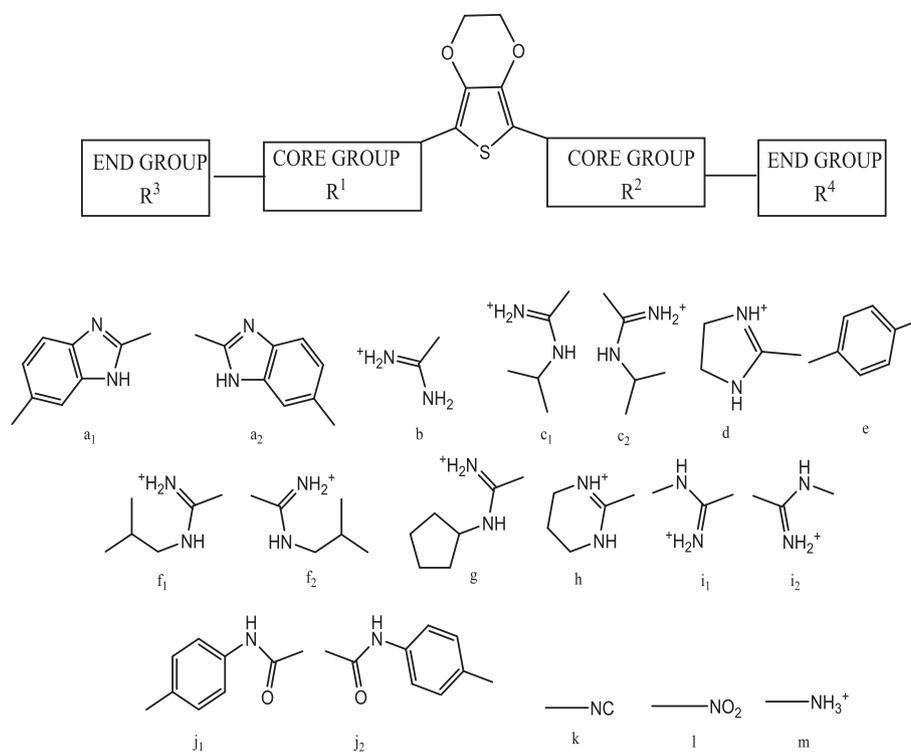
The dataset used for building QSAR models consists of 27 molecules whose antitumor activity was measured and described in present and our previously published papers [14–16]. The synthesis and physical properties of compounds have been described previously [7,14–16]. Antitumor activities expressed as IC₅₀ (μM), were converted in the form of the logarithm (logIC₅₀) and presented in Table 1. For the inactive compounds whose IC₅₀ values are estimated as ">200", log IC₅₀ was set to 2.40.

The 3D structures were optimized using molecular mechanics force fields (MM+) using the HyperChem 8.0 Subsequently, all structures were submitted to geometry optimization using the semi-empirical AM1 method and several physico-chemical and quantum-chemical descriptors were calculated by HyperChem 7.0 (HyperCube, Inc., Gainesville, FL).

The 2D and 3D molecular descriptors used in this study were calculated by applying the online software Parameter Client (Virtual Computational Chemistry Laboratory, <http://146.107.217.178/lab/pclient/>) an electronic remote version of the Dragon program [20]. Seventeen groups of Dragon's descriptors were used to generate QSAR models: constitutional, topological, walk and path counts, connectivity, information, 2D autocorrelations, edge adjacency, BCUT (Burden eigenvalues), topological charge, eigenvalue-based, geometrical, RDF (Radial Distribution Function), 3D-MoRSE (3D-molecular representation of structures based on electron diffraction), WHIM (WeighTted Covariance Matrices), GETAWAY (Geometry, Topology, and Atom-Weights Assembly) descriptors, functional group counts, and molecular properties [21].

The selection of descriptors based on the best-subset method and the multiple regression analysis (MLR) was performed with the use of STATISTICA 12 (StatSoft, Inc. Tulsa, USA). The number of descriptors (*I*) in the multiple regression equation was limited to two, in consideration of the fact that the number of compounds in the training set was 22.

Prior to splitting a data set on training and a test set, the initial number of 1280 calculated molecular descriptors and physico-chemical properties were reduced to 35 descriptors. The procedure started with the elimination of variables that are weakly correlated



Cpd.	R ¹	R ²	R ³	R ⁴	Cpd.	R ¹	R ²	R ³	R ⁴
1	a ₁	a ₂	b	b	15	j ₁	j ₂	<i>para</i> -d	<i>para</i> -d
2	a ₁	a ₂	c ₁	c ₂	16	j ₁	j ₂	<i>para</i> -h	<i>para</i> -h
3	a ₁	a ₂	d	d	17	j ₁	j ₂	<i>meta</i> -b	<i>meta</i> -b
4	e	e	b	b	18	j ₁	j ₂	<i>meta</i> -d	<i>meta</i> -d
5	e	e	c ₁	c ₂	19	a ₁	e	b	k
6	e	e	f ₁	f ₂	20	a ₁	e	d	k
7	e	e	g	g	21	a ₁	e	c ₁	k
8	e	e	d	d	22	a ₁	e	c ₁	l
9	e	e	h	h	23	a ₁	e	d	l
10	i ₁	i ₂	e	e	24	a ₁	e	b	l
11	j ₁	j ₂	<i>para</i> -b	<i>para</i> -b	25	a ₁	e	b	m
12	j ₁	j ₂	<i>para</i> -c ₁	<i>para</i> -c ₂	26	a ₁	e	c ₁	m
13	j ₁	j ₂	<i>para</i> -f ₁	<i>para</i> -f ₂	27	a ₁	e	d	m
14	j ₁	j ₂	<i>para</i> -g	<i>para</i> -g					

Fig. 1. Structures of analysed compounds.

with pIC_{50} ($R = 0.30$). Further selection of predictor variables was preceded by the best-subset method using STATISTICA. In the space of individual group of descriptors the 20 best diparametric, and monoparametric QSAR models were built and involved descriptors were selected for further analysis. The criterion for the best model is based on the R^2 values of the obtained models.

Finally, the best QSAR models were obtained by using a Genetic Algorithm (GA) using QSARINS v 2.2 [22]. The models have been

assessed by: fitting criteria; internal cross-validation using leave-one out (LOO) method and Y-scrambling; and external validation. Fitting criteria included: the coefficient of determination (R^2), adjusted (R^2_{adj}), cross-validate R^2 using leave-one-out method (Q^2_{LOO}), global correlation among descriptors (K_{xx}), difference between global correlation between molecular descriptors and y the response variable, and global correlation among descriptors (ΔK), standard deviation of regression (s), and Fisher ratio (F) [23–25].

Table 1Antitumor activity against tumour cell lines: AGS, MIAPaCa2, CaCo2, HEp2, HeLa and NCI H358, expressed as $\log IC_{50}$ ^a

Cpd.	Tumour cell lines						Ref. ^b
	AGS	MiaPaca2	Caco-2	HEp2	HeLa	NCI-H358	
1	2.16	2.4	2.09	1.00	1.83 [#]	1.70	[14]
2	1.9	1.70	2.00	1.00	0.38	1.48	[14]
3	2.04	1.94	2.14 [#]	1.98	1.95	1.23 [#]	[14]
4	2.06	1.00	1.11	0.95	1.51	1.52	[15]
5	1.82	1.11	1.88	1.00	1.86	2.05 [#]	[15]
6	1.77	0.69	1.63	1.00 [#]	1.28	1.86	[15]
7	1.65 [#]	1.04 [#]	1.61	1.79 [#]	1.43	1.79 [#]	[15]
8	0.63	0.95 [#]	0.48	0.48	0.30	0.61	[15]
9	1.91	1.81 [#]	2.00	1.99	1.99	2.00	[15]
10	2.05	2.11	2.13	2.00 [#]	2.15	2.05	[16]
11	2.07 [#]	1.60	1.60	1.65	1.68	1.53	[16]
12	2.13	2.10 [#]	2.11	2.1	2.40 [#]	2.11	[16]
13	1.74	2.04	2.10	2.11	2.11	2.00	[16]
14	2.11	1.95 [#]	2.00 [#]	1.96	1.99	2.09 [#]	[16]
15	2.40 [#]	2.40	2.40	2.11	2.40	2.11	[16]
16	2.02	2.11	2.08	2.13 [#]	2.08	1.94	[16]
17	1.96	1.30	1.85	1.70	1.85 [#]	1.82	[16]
18	1.93	0.95	0.30 [#]	1.00	1.00	0.95	[16]
19	2.40	2.40	2.40 [#]	2.00	1.00	2.08	c
20	1.78	2.40	2.00	1.62	1.41 [#]	2.04 [#]	c
21	1.98	2.40	2.04	1.98	2.00	1.95	c
22	1.98	2.40	2.40 [#]	2.04	2.40	2.40	c
23	2.00 [#]	2.40	2.00	2.04	2.40	2.40	c
24	2.10	2.40	2.40	2.40	2.40 [#]	2.40	c
25	2.00	2.00	2.05	2.00	1.99	2.00	c
26	1.70 [#]	2.40	2.00	2.06 [#]	2.06	1.98	c
27	1.80	2.00	2.00	2.12	2.00	1.97	c

[#] Member of the test set.

AGS (gastric adenocarcinoma), MIAPaCa2 (pancreatic carcinoma), CaCo2 (colon adenocarcinoma), HEp2 (larynx carcinoma), HeLa (cervix adenocarcinoma) and NCI-H358 (bronchioalveolar carcinoma).

^a IC_{50} - drug concentration ($\mu\text{mol dm}^{-3}$) that inhibited cell growth by 50%.^b Antitumor activity was measured and described in previously published papers.^c Antitumor activity was measured and described in the present study.

Internal and external validations also included the following parameters: root-mean-square error of the training set ($RMSE_{tr}$); root-mean-square error of the training set determined through cross validated LOO method ($RMSE_{cv}$), root-mean-square error of the external validation set ($RMSE_{ex}$), concordance correlation coefficient of the training set (CCC_{tr}), test set using LOO cross validation (CCC_{cv}), and of the external validation set (CCC_{ex}) [26], mean absolute error of the training set (MAE_{tr}), mean absolute error of the internal validation set (MAE_{cv}) and mean absolute error of the external validation set (MAE_{ex}) [23], predictive residual sum of squares determined through cross-validated LOO method ($PRESS_{cv}$) in the training set and in the external prediction set ($PRESS_{ex}$). The analysed external validation parameters also include the coefficient of determination (R^2_{ex}). Robustness of QSAR models was tested by Y-randomisation test. New parallel models were developed based on fit to randomly reordered Y-data (Y scrambling), and the process was repeated several times (2000 iterations) [23]. Tools of regression diagnostic as residual plots and Williams plots were used to check the quality of the best models and define their applicability domain using QSARINS.

3. Results and discussion

Compounds **19–27** (Fig. 1) were tested on a panel of cancer cell lines – AGS (gastric adenocarcinoma), MIAPaCa2 (pancreatic carcinoma), CaCo2 (colon adenocarcinoma), HEp2 (larynx carcinoma), HeLa (cervix adenocarcinoma) and NCI-H358 (bronchioalveolar carcinoma). The results of the anticancer activity analysis (Table 1) show that the investigated compounds have differential influence

on tumour cell growth.

The analysed compounds were divided into eight groups according to structure: **i**) bis-benzimidazole- (**1–3**), **ii**) bis-phenyl- (**4–9**), **iii**) biscarboxamidophenyl- (*para* **11–16**), **iv**) bis-carboxamidophenyl- (*meta* **17,18**), **v**) cyano-amidine (**19–21**), **vi**) nitro-amidine (**22–24**), **vii**) amino-amidine (**25–27**) and **viii**) miscellaneous derivative (**10**). Compound **10** was a derivative of compound **4** with the end groups and aromates inversely arranged [7,14–16]. Antitumor potential of groups **i–iv** and **viii** was previously investigated and published [14–16]. In this study, the antiproliferative potential of compounds **19** to **27** was investigated using the same panel of tumour cell lines, with the addition of chronic myeloid leukemia (K562) cells.

Antitumor activities, expressed as IC_{50} (μM), were converted in the form of a logarithm ($\log IC_{50}$) and presented in Table 1. For the inactive compounds whose IC_{50} values are estimated as “>200”, $\log IC_{50}$ was set to 2.40.

Within group **v**, the most efficient compound was **20**, with $\log IC_{50}$ values below 1.8 for three of the six tested cell lines (AGS, HEp2, HeLa). Group **vi** had predominantly low antitumor activity, with $\log IC_{50}$ above 2. Lastly, the compounds in group **vii** compound group showed the widest range of activity. Two members of this group, compounds **25** and **27**, inhibit growth of all cell lines. Overall, only two members of the tested groups (**25**, **27**) inhibit MiaPaCa2 cells and almost all, except one, inhibit AGS and Hep2 lines (**19** and **24**, respectively). The lowest $\log IC_{50}$ value was found after treating AGS cells with compound **26**.

To test possible improvement of antiproliferative capacity of the tested compounds, we used two different delivery systems, lipofectamine and albumin solvation methods. Both methods proved effective in increasing the effectiveness of drugs with poor pharmacokinetics and were successfully used and approved for therapy [27,28]. Additionally, the most effective and the least effective compound (**20** and **19**, respectively) were tested on K562 cells. In this cell line, their cytotoxicity was approximately 20%. Albumin delivery resulted in growth inhibition increase of 5%–10% when compared to the treatment with the compound alone (Fig. 2). Lipofectamine delivery did not increase cytotoxicity of these two compounds. Antitumor candidates should preferably induce apoptosis in target cells to minimize the damage caused to the surrounding tissue [29,30]. Changes in the mitochondrial membrane potential ($\Delta\Psi_m$) were measured in the treated and control K562 cells. The results show that investigated compounds **19** and **20** induced significant disruption of mitochondria (JC-monomers) in the treated K562 cells (Fig. 3). The percentage of

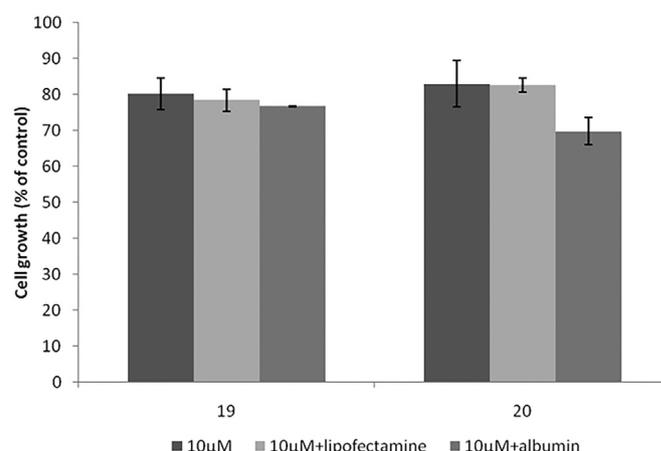


Fig. 2. Growth inhibition with different delivery systems on K562 cells.

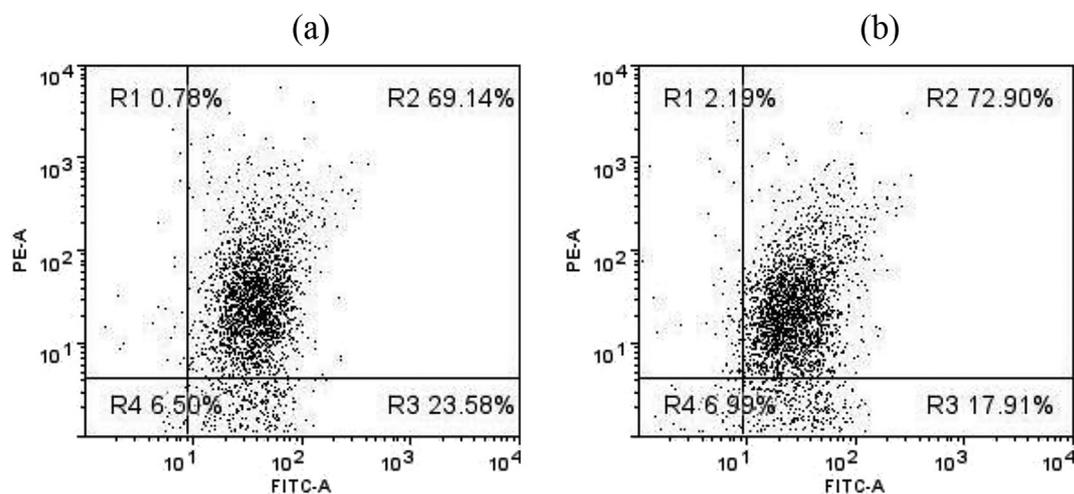


Fig. 3. Flow cytometry analysis of K562 cells after treatment with compounds 19 (a) and 20 (b), FITC-only positive population (R3) are cells in early apoptosis.

apoptotic cells correlates with the percentage of growth inhibition determined by MTT testing shown in Fig. 2.

Dataset used for building of QSAR models consisted of 27 molecules whose antitumor activity was measured and described both in this and our previously published papers [14–16]. Main structures of the analysed compounds illustrated by their major building blocks, 3,4-ethylenedioxythiophene central unit, core groups (R^1 and R^2) and end groups (R^3 and R^4), as well as structural details of 27 molecules are presented in Fig. 1.

Compound **8** showed strong anticancer potential against all six tumour cell lines. Compounds **6** and **7** showed high activity on proliferative capacity of the AGS, MiaPaca2 and Caco-2 cells [15]. These three compounds are isobutyl- (**6**), cyclopentylamidino- (**7**), and imidazolyl- (**8**) bis-phenyl derivatives of 3,4-ethylenedioxythiophene (3,4-EDOT) (Table 1). Isopropylamidino-bis-benzimidazole derivate (**2**) exhibited strong antitumor activity against HeLa cell [14]. Compound (**18**), *meta*-imidazolyl-bis-carboxamidophenyl derivate of 3,4-EDOT, showed excellent activity against Caco-2 cells [16] and also good activity against MiaPaca2 and NCI-H358 cells. These results indicate that the modification of substituents and their position on the 3,4-EDOT central unit resulted in great differences in cytotoxic activity against tumour cells. Evaluation of the molecular descriptors relevance in the obtained QSAR models may help in the elucidation of important physico-chemical and structural requirements for the antitumor activity of the most active compounds.

The best QSAR models for antitumor activity against six cell lines obtained are given in Table 2. Statistical parameters of the obtained models are given in Table 3. Experimental determined antitumor activities against carcinoma cell lines: AGS, MIAPaCa2, CaCo2, HEP2, HeLa and NCI H358, together with the activities predicted by the best obtained QSAR models, are given in Supplementary File 1 (SF 1).

Satisfaction of fitting criteria implies the following: the closer R^2 values are to unity, the more similar calculated values are to the experimental ones, that is, $R^2 \geq 0.60$. Also, larger F statistic and lower standard deviation means that the model is more significant. In order to avoid overfitting, inter-correlation between the descriptors included in the equation is detected based on K_{xx} and ΔK . According to the QUIK rule [24], only the models with the K_{xy} correlation among the $[X + Y]$ variables greater than the K_{xx} correlation among the $[X]$ variables can be accepted. Therefore, low K_{xx} and $\Delta K \geq 0.05$ implies no chance correlation between descriptors. The minimum acceptable statistical parameters for

internal and external predictivity include the following conditions: $R^2_{ext} \geq 0.60$; $CCC \geq 0.85$; $RMSE$ and MAE close to zero; and $RMSE_{tr} < RMSE_{cv}$. Robust QSAR models should have low $R^2_{y,scr}$ and low $Q^2_{y,scr}$ values and $R^2_{y,scr} > Q^2_{y,scr}$ [31]. In order to investigate the applicability of a prediction model and detect possible outliers, the applicability domain of the selected models was evaluated by a leverage analysis expressed as Williams plot, in which residuals and the leverage values were plotted. Williams plots are given in Supplementary File 2 (SF 2).

Analysis of Table 3 indicates that all six models obtained for different cell lines are statistically robust, and from the internal point of view, all models satisfy the essential conditions and criteria, but have modest predictive power. Low collinearity between the descriptors is also verified by the low values of K_{xx} and $\Delta K \geq 0.05$.

The results of Y-scrambling demonstrated that all models were not obtained by chance correlation. Model (2), obtained for the prediction of antitumor activity against MIAPaCa2 cells, has the best fitting (R^2 , R^2_{adj} , F , CCC_{tr}) and internal validation criteria (Q^2_{LOO} , CCC_{cv}). Model (2) also has good external predictivity regarding the acceptable R^2_{ext} (0.69) and a small difference between $RMSE_{tr}$ and $RMSE_{ex}$, as well as between MAE_{tr} and MAE_{ex} . A scatter plot of experimentally obtained antitumor activity against the pancreatic carcinoma cell line versus the values calculated by model (2) is

Table 2

The best obtained QSAR models for antitumor activity against six cell lines.

Tumour cell lines	Equation	No.
AGS	$\log IC_{50} = 3.87 - 0.89 \text{ BELp6} - 3.30 \text{ G1u}$	(1)
MIAPaCa	$\log IC_{50} = -2.86 + 5.55 \text{ BIC5} - 3.30 \text{ MATS7e}$	(2)
CaCo2	$\log IC_{50} = -26.73 + 16.19 \text{ BELp3} - 0.07 \text{ F10[C-N]}$	(3)
HEP2	$\log IC_{50} = 2.76 - 0.50 \text{ Mor19e} - 0.16 \text{ F09[N-O]}$	(4)
HeLa	$\log IC_{50} = 4.18 - 1.83 \text{ Mor15v} - 3.41 \text{ G1p}$	(5)
NCI-H358	$\log IC_{50} = 0.93 + 32.55 \text{ R4u}^+ - 0.19 \text{ F09[N-O]}$	(6)

BELp6 = highest eigenvalue n. 6 of Burden matrix/weighted by atomic polarizabilities; G1u = 1st component symmetry directional WHIM index/unweighted; BIC5 = bond information content (neighbourhood symmetry of 5-order); MATS7e = Moran autocorrelation - lag 7/weighted by atomic Sanderson electronegativities; BELp3 = highest eigenvalue n. 3 of Burden matrix/weighted by atomic polarizabilities; F10[C-N] = frequency of C-N at topological distance 10; Mor19e = 3D-MorSE - signal 19/weighted by atomic Sanderson electronegativities; F09[N-O] = frequency of N-O at topological distance 09; Mor15v = 3D-MorSE - signal 19/weighted by atomic polarizabilities; G1p = 1st component symmetry directional WHIM index/weighted by atomic polarizabilities; R4u⁺ = R maximal autocorrelation of lag 4/unweighted [21].

Table 3
The statistical results for the QSAR models for antitumor activity.

Tumour cell lines	AGS	MIAPaCa	CaCo2	HEp2	HeLa	NCI-H358
Equation no. ^a	(1)	(2)	(3)	(4)	(5)	(6)
R^2	0.77	0.81	0.70	0.72	0.76	0.70
R^2_{adj}	0.75	0.79	0.67	0.69	0.73	0.67
s	0.16	0.26	0.27	0.30	0.31	0.25
F	31.86	40.20	22.25	24.21	29.56	22.05
K_{xx}	0.27	0.13	0.26	0.11	0.06	0.35
ΔK	0.18	0.33	0.16	0.37	0.38	0.08
$RMSE_{tr}$	0.15	0.24	0.25	0.27	0.29	0.23
MAE_{tr}	0.14	0.21	0.19	0.23	0.22	0.18
CCC_{tr}	0.87	0.89	0.82	0.84	0.86	0.82
s	0.16	0.26	0.27	0.30	0.31	0.25
F	31.86	40.20	22.25	24.21	29.56	22.05
Q^2_{LOO}	0.28	0.74	0.60	0.61	0.69	0.58
$RMSE_{cv}$	0.27	0.28	0.29	0.32	0.33	0.28
MAE_{cv}	0.19	0.25	0.22	0.27	0.25	0.21
$PRESS_{cv}$	1.60	1.71	1.87	2.32	2.34	1.67
CCC_{cv}	0.39	0.86	0.79	0.77	0.82	0.75
R^2_{ySCR}	0.09	0.09	0.10	0.10	0.10	0.10
Q^2_{ySCR}	-0.78	-0.22	-1.82	0.49	-0.39	-0.23
$RMSE_{ext}$	0.46	0.38	0.57	0.42	0.50	0.43
MAE_{ext}	0.39	0.33	0.29	0.34	0.43	0.38
$PRESS_{ext}$	1.07	0.72	1.64	0.88	1.24	0.92
R^2_{ext}	0.79	0.69	0.83	0.80	0.83	0.64
CCC_{ext}	-0.74	0.67	0.36	0.75	0.22	0.47

^a Equations are given in Table 2.

shown in are shown in Supplementary File 3(SF 3).

Model (1), which predicts antitumor activity against AGS cells, has slightly lower R^2 , R^2_{adj} , F , CCC_{tr} and CCC_{cv} values than model (2), and higher values of external validation R^2_{ext} and CCC_{ext} . Also, model (1) has the lowest standard deviation of regression, Q^2_{LOO} , $RMSE_{tr}$, MAE_{tr} , $RMSE_{cv}$, MAE_{cv} , and $PRESS_{cv}$ of all the six models obtained. As can be seen from the Williams plot (SF 2), antitumor activity of compound **8** predicted by model (1) must be used with reserve, because its leverage values are greater than the warning leverage ($h^* = 0.409$). Also, the same model has generated two outliers, because their standardized residuals were greater than ± 2.5 . There are no compounds outside the domain of applicability of the QSAR model (2), only compound **9**, a member of the test set, is detected as an outlier. Models (3–6) have slightly weaker, but satisfactory results of fitting parameters and internal validation. However, model (4), obtained for estimation of antitumor activity against Hep2 cells, has the best results of external validation due to the highest value of CCC_{ext} (0.75), high value of R^2_{ext} (0.80) and a small difference between $RMSE_{tr}$ and $RMSE_{ext}$, and between MAE_{tr} and MAE_{ext} . Williams plot for model (4) (SF 2) reveals two outliers (compounds **10** and **16**), and no compounds outside of applicability domain. Due to low CCC_{ext} values, models (3), (5) and (6) may not be considered as predictive, regardless of the high value of R^2_{ext} for models (3) and (5) (0.91 and 0.86, respectively).

The best QSAR models obtained (Table 2) include the following group of descriptors: BCUT descriptors (BELp6, BELp3); WHIM descriptors (G1u, G1p); information indices (BIC5); 2D autocorrelations descriptors (MATS7e); 3D-MoRSE descriptors (Mor19e, Mor15v); GETAWAY descriptors (R4u⁺); 2D frequency fingerprint (F09[N-O] and F10[C-N]) [18]. The values of these descriptors are shown in Supplementary File 4 (SF 4). Negative signs of regression coefficients of descriptors BELp6, G1u, G1p, BIC5, MATS7e, Mor19e, Mor15v, F10[C-N] and F09[N-O]) imply that higher values of these descriptors mean lower logIC₅₀ (higher antitumor activity). Considering the high value of regression coefficients in Eq. (1), the most relevant structural features for antitumor activity are depicted by the descriptors G1u and G1p in Eq. (5). Both descriptors belong to the WHIM descriptors that capture three-dimensional (3D)

information relevant for molecular conformation during interaction of a cancer cell with DNA. WHIM descriptors (*Weighted Holistic Invariant Molecular descriptors*) are geometrical descriptors based on statistical indices calculated on the projections of the atoms along principal axes. G1u and G1p describe the symmetry of a molecule evaluated on the basis of the number of symmetric atoms with respect to the molecule centre. Highly potent molecules should have as higher as possible values of these two descriptors, precisely, more symmetrical distribution of atoms along the x axis [21].

Descriptor relevant for the estimation of antitumor activity against Hep2 cancer cells is the 3D-MoRSE descriptor Mor19e. These descriptors were generated from electron diffraction studies and reflect the three-dimensional arrangement of atoms in a molecule. Descriptor Mor19e reflects the contribution of 3D distribution of electronegativity (oxygen atoms) [21]. As can be noticed from Supplementary File 4, this descriptor is extremely sensitive to the changes of substituents and their position on the 3,4-ethylenedioxythiophene central unit. According to Eq. (4), it is expected that increasing values of Mor19e tend to predict higher activity. Therefore, the most active compound (**8**) has the highest value of Mor19e (2.49).

F09[N-O] is a 2D frequency fingerprint descriptor, which corresponds to the frequency of occurrence of oxygen and nitrogen atoms at topological distance 9, while F10[C-N] represents the frequency of occurrence of carbon and nitrogen atoms at topological distance 10. Considering the negative coefficients of F09[N-O] and F10[C-N] descriptors in models (3), (4) and (6), highly active compounds should have a high frequency of distance between N and O atoms at topological distance 9 and between C and N atoms at topological distance 10 [21]. Compounds whose structures include benzimidazolyl group (**1–3**), phenyl group (**4–9**), or phenylcarboxamide group (**17, 18**) at the position R¹ and R², in combination of different end groups R³ and R⁴ that introduce described topological distance between N and O atoms, as C and N atoms, have enhanced antitumor effect. BIC5 bonds information content that quantifies binding symmetry. It is related to the differences in the atomic distribution at neighbourhood symmetry of 5-order [21]. Due to the greatest difference in structure for compounds (**19–27**) with two different substituents in the positions R¹ and R², these compounds have the highest values of BIC5. Considering the positive coefficient of BIC5 in model (2), it may be concluded that the asymmetrical structure of 3,4-EDOT derivatives negatively influences antitumor activity of these compounds or that increased antitumor activity implies a more symmetrical molecular shape. Contrary to that, in a QSAR study for *anti*-HIV activities of 79 HEPT derivatives, a negative coefficient of BIC5 in equations indicated that increased biologic activity implies a more asymmetrical molecular shape [32].

Positive coefficient of three-dimensional GETAWAY descriptor, descriptor R4u⁺, included in model (6), implies that lower values of that descriptor are favourable for the exhibition of antitumor activity against NCI H358 cells. R-GETAWAY descriptors combine the information provided from the influence matrix with geometric interatomic distances in the molecule. Descriptor R4u⁺ is derived from the influence/distance matrix, and since it is unweighted, it treats each atom equally at topological distance 4 [33]. Therefore, small molecules, with lower number of atoms at topological distance 4, have higher values of R4u⁺ (Supplementary File 4) and consequently, higher values of logIC₅₀, according to model (6). BCUT metrics is represented by the hydrogen-suppressed connection table of a molecule as a symmetrical $N \times N$ matrix with atomic numbers along the diagonal and bonding information in the off-diagonal elements [34]. BCUT descriptors, BELp6 in model (1) and BELp3 in model (3), incorporate the bond-type for adjacent and

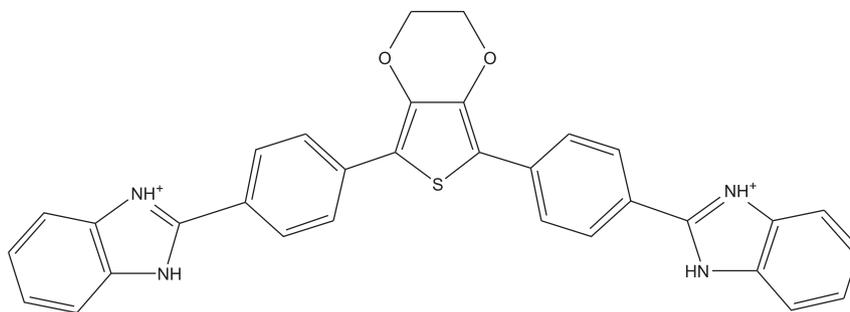


Fig. 4. Structure of the proposed molecule as a promising antitumor agent.

nonadjacent atoms and atomic polarizability [35]. BELp6 and BELp3 represent the lowest eigenvalue number 6 and 3, respectively, weighted by atomic masses.

The greatest values of BELp6 and BELp3 have derivatives of 3,4-EDOT with more C atoms (atom with the highest polarizability) located at topological distances 6 and 3, respectively. According to the negative coefficient of BELp6 in model (1), the enhanced values of this descriptor, which have molecules with larger substituents, positively affect antitumor activity against gastric adenocarcinoma cells. However, according to the positive sign of BELp3 in model (3), higher values of this descriptor are unfavourable for the exhibition of antitumor activity against CaCo2 cells. Earlier QSAR analysis of antitumor activity of heterocyclic amides and quinolones showed that increased volume and the amount of hydrophobic surfaces are highly important for antitumor activity against MIAPaCa2 cells [36].

Based on the conclusion given in the QSAR analysis, a new compound with possible great activity is proposed (Fig. 4). Calculated descriptors are shown in Supplementary File 4. The proposed compound includes two phenyls as core groups (R^1 , R^2) and two benzimidazolyls as end groups (R^3 , R^4). Antitumor activity of the proposed compound has been predicted by means of the predictive model (2) against the MIAPaCa2 ($\log IC_{50} = 0.71$).

4. Conclusion

Structural groups of compounds presented in this paper show cell-line dependent cytotoxicity. The most efficient compounds belong to group vii. The results show that improved delivery can increase effectiveness, even in compounds with low cytotoxicity. In the compounds tested, cytotoxicity is achieved through apoptosis, which makes the tested compounds good candidates for further research and development.

The results of the QSAR analysis suggest that derivatives of 3,4-EDOT with the following structural feature may exhibit great antitumor activity: larger molecules with symmetrical arrangement of substituents that allows as many as possible C and N atoms located at topological distance 10, as well as N and O atoms at topological distance 9. We have constructed two predictive QSAR models of antitumor activity against AGS and MIAPaCa2 and estimated the activity of a future molecule that may exhibit great activity.

Acknowledgment

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix B. Supplementary data

Supplementary data associated with this article can be found in

the online version, at <http://dx.doi.org/10.1016/j.molstruc.2016.11.074>. These data include MOL files and InChIKeys of the most important compounds described in this article.

References

- [1] X. Cai, P.J.J. Gray, D.D. Von Hoff, DNA minor groove binders: back in the groove, *Cancer Treat. Rev.* 35 (2009) 437–450, <http://dx.doi.org/10.1016/j.ctrv.2009.02.004>.
- [2] A. Paul, Y. Chai, D.W. Boykin, W.D. Wilson, Understanding mixed sequence DNA recognition by novel designed compounds: the kinetic and thermodynamic behavior of azabenzimidazole diamidines, *Biochemistry* 54 (2015) 577–587, <http://dx.doi.org/10.1021/bi500989r>.
- [3] S. Neidle, D.E. Thurston, Chemical approaches to the discovery and development of cancer therapies, *Nat. Rev. Cancer* 5 (2005) 285–296.
- [4] O. Seitz, DNA and RNA binders. From small molecules to drugs, 42, in: Martine Demeunynck, Christian Bailly, W. David Wilson (Eds.), *Angew. Chemie Int. Ed.*, vols. 1 & 2, 2003, pp. 4994–4995, <http://dx.doi.org/10.1002/anie.200385976>.
- [5] D.A. Patrick, M.A. Ismail, R.K. Arafa, T. Wenzler, X. Zhu, T. Pandharkar, S.K. Jones, K.A. Werbovetz, R. Brun, D.W. Boykin, R.R. Tidwell, Synthesis and antiprotazoal activity of dicationic m-terphenyl and 1,3-dipyridylbenzene derivatives, *J. Med. Chem.* 56 (2013) 5473–5494, <http://dx.doi.org/10.1021/jm400508e>.
- [6] M.N.C. Soeiro, K. Werbovetz, D.W. Boykin, W.D. Wilson, M.Z. Wang, A. Hemphill, Novel amidines and analogues as promising agents against intracellular parasites: a systematic review, *Parasitology* 140 (2013) 929–951, <http://dx.doi.org/10.1017/S0031182013000292>.
- [7] I. Stolić, H. Cipic Paljetak, M. Peric, M. Matijasic, V. Stepanic, D. Verbanac, M. Bajic, Synthesis and structure-activity relationship of amidine derivatives of 3,4-ethylenedioxythiophene as novel antibacterial agents, *Eur. J. Med. Chem.* 90 (2015) 68–81, <http://dx.doi.org/10.1016/j.ejmech.2014.11.003>.
- [8] W.D. Wilson, B. Nguyen, F.A. Tanious, A. Mathis, J.E. Hall, C.E. Stephens, D.W. Boykin, Dications that target the DNA minor groove: compound design and preparation, DNA interactions, cellular distribution and biological activity, *Curr. Med. Chem. Anticancer Agents* 5 (2005) 389–408.
- [9] W. Zhu, Y. Wang, K. Li, J. Gao, C.-H. Huang, C.-C. Chen, T.-P. Ko, Y. Zhang, R.-T. Guo, E. Oldfield, Antibacterial drug leads: DNA and enzyme multitargeting, *J. Med. Chem.* 58 (2015) 1215–1227, <http://dx.doi.org/10.1021/jm501449u>.
- [10] P.G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M.A. Tabrizi, M.G. Pavani, R. Romagnoli, DNA minor groove binders as potential antitumor and antimicrobial agents, *Med. Res. Rev.* 24 (2004) 475–528, <http://dx.doi.org/10.1002/med.20000>.
- [11] L. Racane, V. Tralic-Kulenovic, S. Kraljevic Pavelic, I. Ratkaj, P. Peixoto, R. Nhili, S. Depauw, M.-P. Hildebrand, M.-H. David-Cordonnier, K. Pavelic, G. Karminski-Zamola, Novel diamidino-substituted derivatives of phenyl benzothiazolyl and dibenzothiazolyl furans and thiophenes: synthesis, anti-proliferative and DNA binding properties, *J. Med. Chem.* 53 (2010) 2418–2432, <http://dx.doi.org/10.1021/jm901441b>.
- [12] P. Peixoto, Y. Liu, S. Depauw, M.-P. Hildebrand, D.W. Boykin, C. Bailly, W.D. Wilson, M.-H. David-Cordonnier, Direct inhibition of the DNA-binding activity of POU transcription factors Pit-1 and Brn-3 by selective binding of a phenyl-furan-benzimidazole dication, *Nucleic Acids Res.* 36 (2008) 3341–3353, <http://dx.doi.org/10.1093/nar/gkn208>.
- [13] W.D. Wilson, F.A. Tanious, A. Mathis, D. Tevis, J.E. Hall, D.W. Boykin, Antiparasitic compounds that target DNA, *Biochimie* 90 (2008) 999–1014, <http://dx.doi.org/10.1016/j.biochi.2008.02.017>.
- [14] I. Stolić, K. Mišković, A. Magdaleno, A.M. Silber, I. Piantanida, M. Bajić, L. Glavaš-Obrovac, Effect of 3,4-ethylenedioxy-extension of thiophene core on the DNA/RNA binding properties and biological activity of bisbenzimidazole amidines, *Bioorg. Med. Chem.* 17 (2009) 2544–2554, <http://dx.doi.org/10.1016/j.bmc.2009.01.071>.
- [15] I. Stolić, K. Mišković, I. Piantanida, M.B. Lončar, L. Glavaš-Obrovac, M. Bajić, Synthesis, DNA/RNA affinity and antitumour activity of new aromatic

- diamidines linked by 3,4-ethylenedioxythiophene, *Eur. J. Med. Chem.* 46 (2011) 743–755, <http://dx.doi.org/10.1016/j.ejmech.2010.12.010>.
- [16] I. Stolić, M. Avdičević, N. Bregović, I. Piantanida, Lj. Glavaš-Obrovac, M. Bajić, Synthesis, DNA interactions and anticancer evaluation of novel diamidine derivatives of 3,4-ethylenedioxythiophene, *Croat. Chem. Acta* 85 (2012) 457–467, <http://dx.doi.org/10.5562/cca2141>.
- [17] G. Mickisch, S. Fajta, G. Keilhauer, E. Schlick, R. Tschada, P. Alken, Chemosensitivity testing of primary human renal cell carcinoma by a tetrazolium based microculture assay (MTT), *Urol. Res.* 18 (1990) 131–136.
- [18] A. Jithan, K. Madhavi, M. Madhavi, K. Prabhakar, Preparation and characterization of albumin nanoparticles encapsulating curcumin intended for the treatment of breast cancer, *Int. J. Pharm. Investig.* 1 (2011) 119–125, <http://dx.doi.org/10.4103/2230-973X.82432>.
- [19] K. Langer, S. Balthasar, V. Vogel, N. Dinauer, H. Von Briesen, D. Schubert, Optimization of the preparation process for human serum albumin (HSA) nanoparticles, *Int. J. Pharm.* 257 (2003) 169–180, [http://dx.doi.org/10.1016/S0378-5173\(03\)00134-0](http://dx.doi.org/10.1016/S0378-5173(03)00134-0).
- [20] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko, Virtual computational chemistry laboratory-design and description, *J. Comput. Aided. Mol. Des.* 19 (2005) 453–463, <http://dx.doi.org/10.1007/s10822-005-8694-y>.
- [21] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Handbook of Molecular Descriptors, Wiley-VCH Verlag GmbH, Weinheim, 2000, <http://dx.doi.org/10.1002/9783527613106.ch1a>.
- [22] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: a new software for the development, analysis, and validation of QSAR MLR models, *J. Comput. Chem.* 34 (2013) 2121–2132, <http://dx.doi.org/10.1002/jcc.23361>.
- [23] P. Gramatica, Principles of QSAR models validation: internal and external, *QSAR Comb. Sci.* 26 (2007) 694–701, <http://dx.doi.org/10.1002/qsar.200610151>.
- [24] R. Todeschini, V. Consonni, A. Maiocchi, The K correlation index: theory development and its application in chemometrics, *Chemom. Intell. Lab. Syst.* 46 (1999) 13–29, [http://dx.doi.org/10.1016/S0169-7439\(98\)00124-5](http://dx.doi.org/10.1016/S0169-7439(98)00124-5).
- [25] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, *Mol. Inf.* 29 (2010) 476–488, <http://dx.doi.org/10.1002/minf.201000061>.
- [26] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, *J. Chem. Inf. Model.* 51 (2011) 2320–2335, <http://dx.doi.org/10.1021/ci200211n>.
- [27] M.T. Larsen, M. Kuhlmann, M.L. Hvam, K.A. Howard, Albumin-based drug delivery: harnessing nature to cure disease, *Mol. Cell. Ther.* 4 (2016), <http://dx.doi.org/10.1186/s40591-016-0048-8>.
- [28] G. Zhou, X. Jin, P. Zhu, J.U. Yao, Y. Zhang, L. Teng, R.J. Lee, X. Zhang, W. Hong, Human serum albumin nanoparticles as a novel delivery system for cabazitaxel, *Anticancer Res.* 36 (2016) 1649–1656.
- [29] J.F.R. Kerr, C.M. Winterford, B.V. Harmon, Apoptosis. Its significance in cancer and cancer therapy, *Cancer* 73 (1994) 2013–2026, [http://dx.doi.org/10.1002/1097-0142\(19940415\)73:8<2013::AID-CNCR2820730802>3.0.CO;2-J](http://dx.doi.org/10.1002/1097-0142(19940415)73:8<2013::AID-CNCR2820730802>3.0.CO;2-J).
- [30] L. Qiao, B.C.Y. Wong, Targeting apoptosis as an approach for gastrointestinal cancer therapy, *Drug Resist. Updates* 12 (2009) 55–64, <http://dx.doi.org/10.1016/j.drug.2009.02.002>.
- [31] V.H. Masand, D.T. Mahajan, G.M. Nazeruddin, T. Ben Hadda, V. Rastija, A.M. Alfeefy, Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model, *Med. Chem. Res.* 24 (2014) 1241–1264, <http://dx.doi.org/10.1007/s00044-014-1193-8>.
- [32] C. Duda-Seiman, D. Duda-Seimana, M.V. Putz, D. Ciubotariub, QSAR modelling of anti-HIV activity with HEPT derivatives, *Dig. J. Nanomater. Biostruct.* 2 (2007) 207–219.
- [33] V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors, *J. Chem. Inf. Comput. Sci.* 42 (2002) 682–692.
- [34] B. Pirard, S. Pickett, Classification of kinase inhibitors using BCUT descriptors, *J. Chem. Inf. Comput. Sci.* 40 (2000) 1431–1440.
- [35] M.P. Gonzalez, C. Teran, M. Teijeira, P. Besada, M.J. Gonzalez-Moa, BCUT descriptors to predicting affinity toward A3 adenosine receptors, *Bioorg. Med. Chem. Lett.* 15 (2005) 3491–3495, <http://dx.doi.org/10.1016/j.bmcl.2005.05.122>.
- [36] B. Bertoša, M. Aleksić, G. Karminiski-Zamola, S. Tomić, QSAR analysis of antitumor active amides and quinolones from thiophene series, *Int. J. Pharm.* 394 (2010) 106–114, <http://dx.doi.org/10.1016/j.ijpharm.2010.05.014>.