European Journal of Pharmaceutical Sciences 48 (2013) 542–553

Contents lists available at SciVerse ScienceDirect

European Journal of Pharmaceutical Sciences

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

QSAR studies of macrocyclic diterpenes with P-glycoprotein inhibitory activity

Inês J. Sousa [a], Maria-José U. Ferreira [b], Joseph Molnár [c], Miguel X. Fernandes [a,*]

[a] Centro de Química da Madeira, Centro de Competência de Ciências Exatas e da Engenharia, Universidade da Madeira, Campus da Penteada, 9000-390 Funchal, Portugal
[b] Research Institute for Medicines and Pharmaceutical Sciences (iMed.UL), Faculdade de Farmácia, Universidade de Lisboa, Av. Prof. Gama Pinto, 1649-003 Lisboa, Portugal
[c] Department of Medical Microbiology and Immunobiology, University of Szeged, H-6720 Szeged, Hungary

ARTICLE INFO

Article history:
Received 23 May 2012
Received in revised form 23 October 2012
Accepted 17 November 2012
Available online 7 December 2012

Keywords:
Quantitative structure–activity relationship
P-glycoprotein
Macrocyclic diterpenes
Multidrug resistance
Multiple linear regression

ABSTRACT

Multidrug resistance (MDR) represents a major limitation for cancer chemotherapy. There are several mechanisms of MDR but the most important is associated with P-glycoprotein (P-gp) overexpression. The development of modulators of P-gp that are able to re-establish drug sensitivity of resistant cells has been considered a promising approach for overcoming MDR. Macrocyclic lathyrane and jatrophane-type diterpenes from *Euphorbia* species were found to be strong MDR reversing agents. In this study we applied quantitative structure–activity relationship (QSAR) methodology in order to identify the most relevant molecular features of macrocyclic diterpenes with P-gp inhibitory activity and to determine which structural modifications can be performed to improve their activity. Using experimental biological data at two concentrations (4 and 40 μg/ml), we developed a QSAR model for a set of 51 bioactive diterpenic compounds which includes lathyrane and jatrophane-type diterpenes and another model just for jatrophanes. The cross-validation correlation values for all diterpenes QSAR models developed for biological activities at compound concentrations of 4 and 40 μg/ml were 0.758 and 0.729, respectively. Regarding the prediction ability, we get R^2_{pred} values of 0.765 and 0.534 for biological activities at compound concentrations of 4 and 40 μg/ml, respectively. Applying the cross-validation test to jatrophanes QSAR models, we obtained 0.680 and 0.787 for biological activities at compound concentrations of 4 and 40 μg/ml concentrations, respectively. For the same concentrations, the obtained R^2_{pred} values for jatrophanes models were 0.541 and 0.534, respectively. The obtained models were statistically valid and showed high prediction ability.

1. Introduction

ATP-binding cassette transporters (ABC transporters) are present in all organisms and play a physiological role in a multitude of diverse processes. This super-family is composed of more than 100 membrane transporters/channels that are involved in several functions, namely the extrusion of harmful compounds, uptake of nutrients, transport of ions and peptides, and cell signaling (Stavrovskaya and Stromskaya, 2008). One of the major mechanisms of multidrug resistance (MDR) in cancer is the enhanced ability of tumor cells to actively efflux drugs, leading to a decrease in cellular drug accumulation below toxic levels. Active drug efflux is mediated by several members of the ABC transporter superfamily. Several cancer cell lines with the MDR phenotype overexpress P-glycoprotein (P-gp), which is perhaps the best characterized human ABC transporter (Higgins et al., 1997; Rosenberg et al., 2005). Human P-gp is a 170 kDa polypeptide organized in two homologous halves; each half is constituted of a transmembrane domain (TMD) with 6 TM segments followed by a nucleotide-binding domain (NBD). In the interface between the TMDs and NBDs we have the predicted drug- and ATP-binding sites which couple the energy of ATP hydrolysis to substrate efflux (Higgins et al., 1997; Rosenberg et al., 2005). Thus, P-gp prevents foreign substances from accumulating within cells by transporting a wide variety of structurally and functionally unrelated compounds from the intracellular environment to the extracellular space, hydrolyzing ATP during the process (Rosenberg et al., 2005). Inhibition of P-gp can reverse the MDR phenotype because P-gp modulators, such as verapamil and quinidine (Thomas and Coley, 2003), compete with the cytotoxic agents for transport by the pump. This limits the efflux of the cytotoxic agents, increasing their intracellular concentrations above cytotoxic thresholds in cancer cells. There is no high resolution crystal structure of human P-gp, but the closest homologe, a mouse P-gp obtained with a 3.8 Å resolution (Aller et al., 2009), reveals a very large substrate binding site with approximately 6000 Å^3, which could explain its binding to a variety of unrelated substrates.

Euphorbia species are characterized by an unusual diversity of chemical constituents, which include a wide range of structurally unique jatrophane and lathyrane macrocyclic diterpene polyesters and their polycyclic rearranged derivatives. Macrocyclic

* Corresponding author. Tel.: +351 291705100; fax: +351 291705149.
E-mail address: mxf@uma.pt (M.X. Fernandes).

0928-0987/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejps.2012.11.012

diterpenoids are vastly studied due to their particular molecular architecture and their wide range of biological activities (Kirby et al., 2010; Jassbi, 2006). These diterpenoids have been shown to exhibit microtubule-interacting (Miglietta et al., 2003) and anti-inflammatory activities (Shi et al., 2008). Recently, antiviral activity has been reported for some jatrophane diterpenes. These compounds also act as a down-regulator of HIV receptors (CD4, CCR5, and CXCR4) and as inducers of viral reactivation (Shokoohinia et al., 2011). Some studies showed that jatrophanes as pubescenol and pubescene D, had antiproliferative activity in human tumor cell lines MCF-7 (breast adenocarcinoma), NCI-H460 (non-small cell lung cancer), and SF-268 (CNS cancer) (Valente et al., 2004). It was also shown that lathyranes presented significant cytotoxic activity in Colo250 (colorectal cancer), MT2 (mammary carcinoma) and CEM (leukemia) cell lines (Shi et al., 2008). However, one of the most studied biological activities of these diterpenoids is the MDR modulating activity in cancer cells over-expressing P-gp. Several jatrophane and lathyrane diterpenes, such as Latilagascenes D and E showed significant P-gp inhibitory activity (Duarte et al., 2007). Therefore, the main goal of this work is to optimize macrocyclic diterpenes as leads for overcoming P-gp-mediated MDR.

We have used a rational design approach employing quantitative structure–activity relationship (QSAR) calculations on series of P-gp modulator jatrophane and lathyrane diterpenes to establish a quantitative relationship between their P-gp modulation activity and their molecular descriptors. QSAR studies explored the P-gp modulatory activity of 51 compounds isolated from *Euphorbia* species or obtained by derivatization (Duarte et al., 2007, 2006, 2008; Madureira et al., 2006; Molnár et al., 2006) and whose structures and experimental activity values are shown in Figs. 1a–1c and Table 1. QSAR models here obtained, using the forward feature selection multiple linear regression (MLR) method, are statistically relevant correlating the calculated and the experimental biological activities. They show very good predictive ability, and provide important structural insight regarding, for instance, the introduction of modifications, like the decrease of the number of polar groups in the molecules and the decrease of the number of conjugated systems, to design compounds with improved P-gp modulatory activity.

2. Methodology

2.1. Dataset for analysis

A set of 51 diterpenic compounds (Fig. 1a-c), from *Euphorbia* species was used in the present study. It comprises mainly macrocyclic lathyrane (**1**–**12**) and jatrophane-type diterpenes (**13**–**41**) and some related polycyclic diterpenoids (**42**–**49**), resulting from intramolecular cyclization of the former (named rearranged lathyranes-**42**–**44** and rearranged jatrophanes-**45**–**49**). Two other polycyclic diterpenes (**50–51**) were also included. Most of lathyrane and jatrophane-type diterpenes were found to be P-glycoprotein-mediated MDR reversal agents (Duarte et al., 2007, 2006, 2008; Madureira et al., 2006; Molnár et al., 2006). The QSAR method was applied to all diterpenes (**1**–**51**) and just to the jatrophanes subset which includes the rearranged derivatives (**13**–**41** and **45**–**49**, respectively) Due to the small number of lathyrane derivatives, QSAR was not applied to this subset.

2.2. P-gp inhibitory activity data

P-gp inhibitory activity of compounds was previously assessed in L15178 mouse T-lymphoma cell line transfected with the human *MDR1* gene, by flow cytometry (Duarte et al., 2007, 2006, 2008; Madureira et al., 2006; Molnár et al., 2006). A standard functional assay that measures the accumulation in the cells of rhodamine-123, a fluorescent substrate analoge of epirubicine, was used.

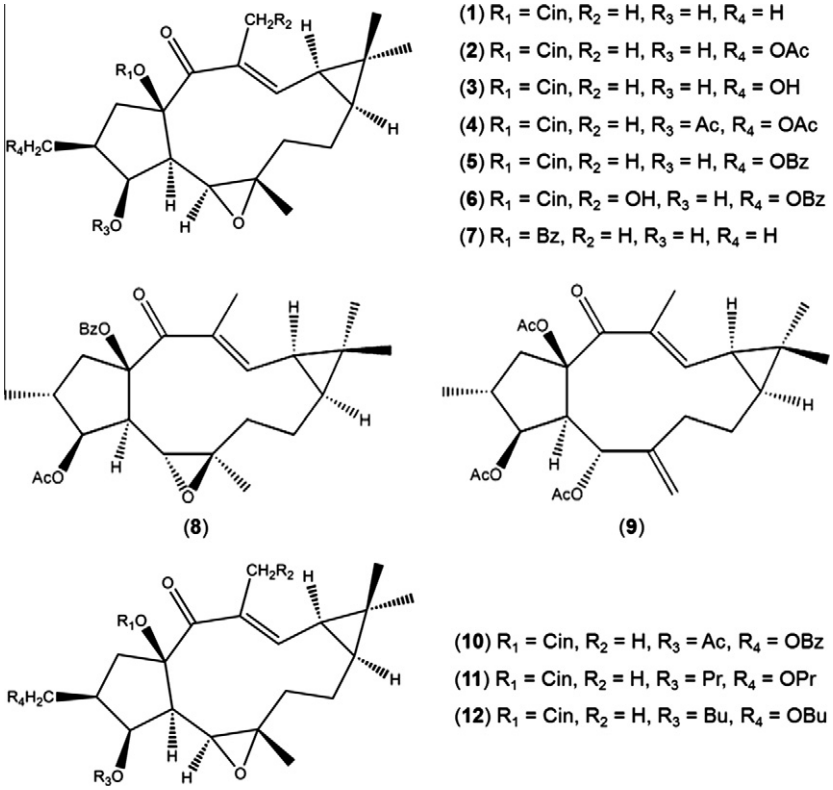

Fig. 1a. Structures of macrocyclic lathyrane-type diterpenes (**1–12**).

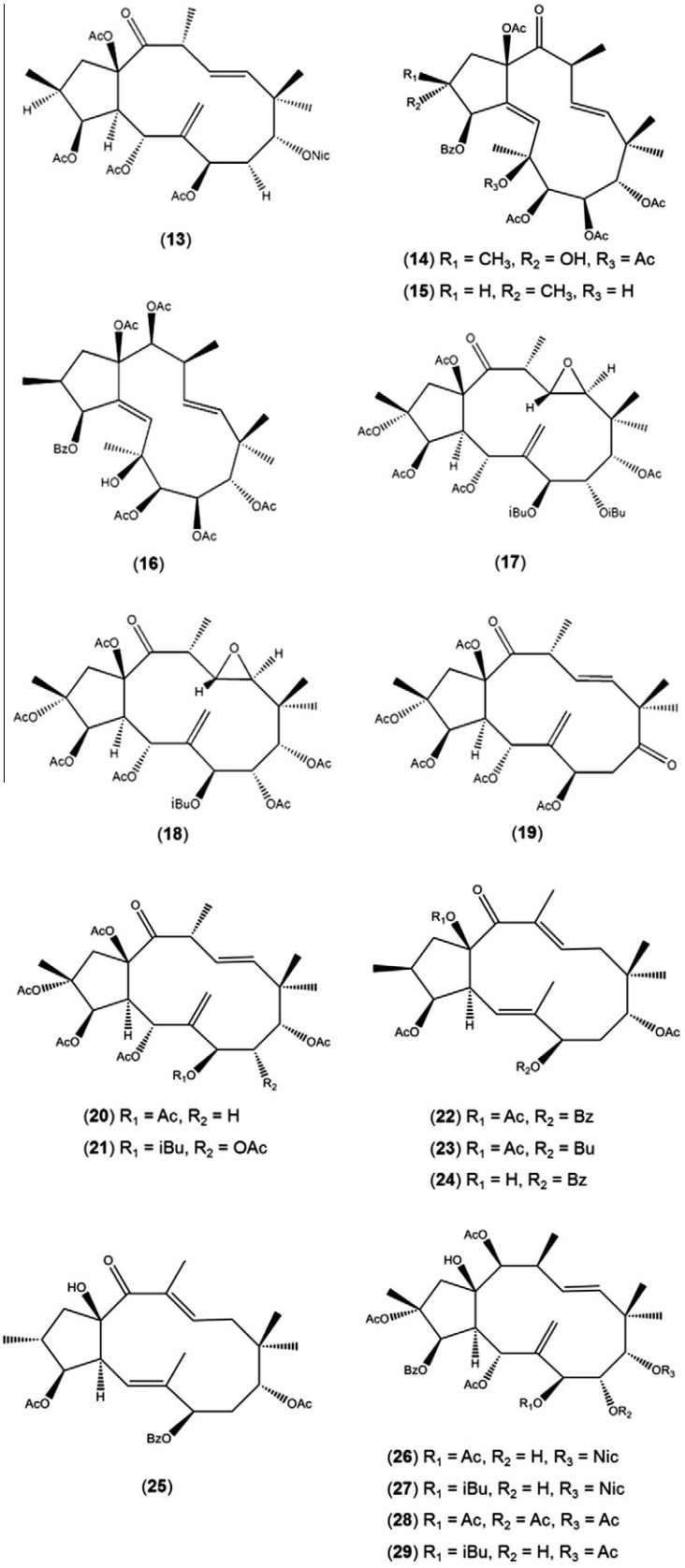

Fig. 1b-1. Structures of macrocyclic jatrophane-type diterpenes (**13**–**29**).

L5178 mouse T-cell lymphoma cells were transfected with pHa *MDR1*/A retrovirus (Cornwell et al., 1987; Pastan et al., 1988). The *MDR1*-expressing cell line was selected by culturing the infected cells with colchicine, thus maintaining the expression of the MDR phenotype in all cells of the population (Choi et al., 1991). L5178 mouse T-cell lymphoma cells (parental, PAR) and

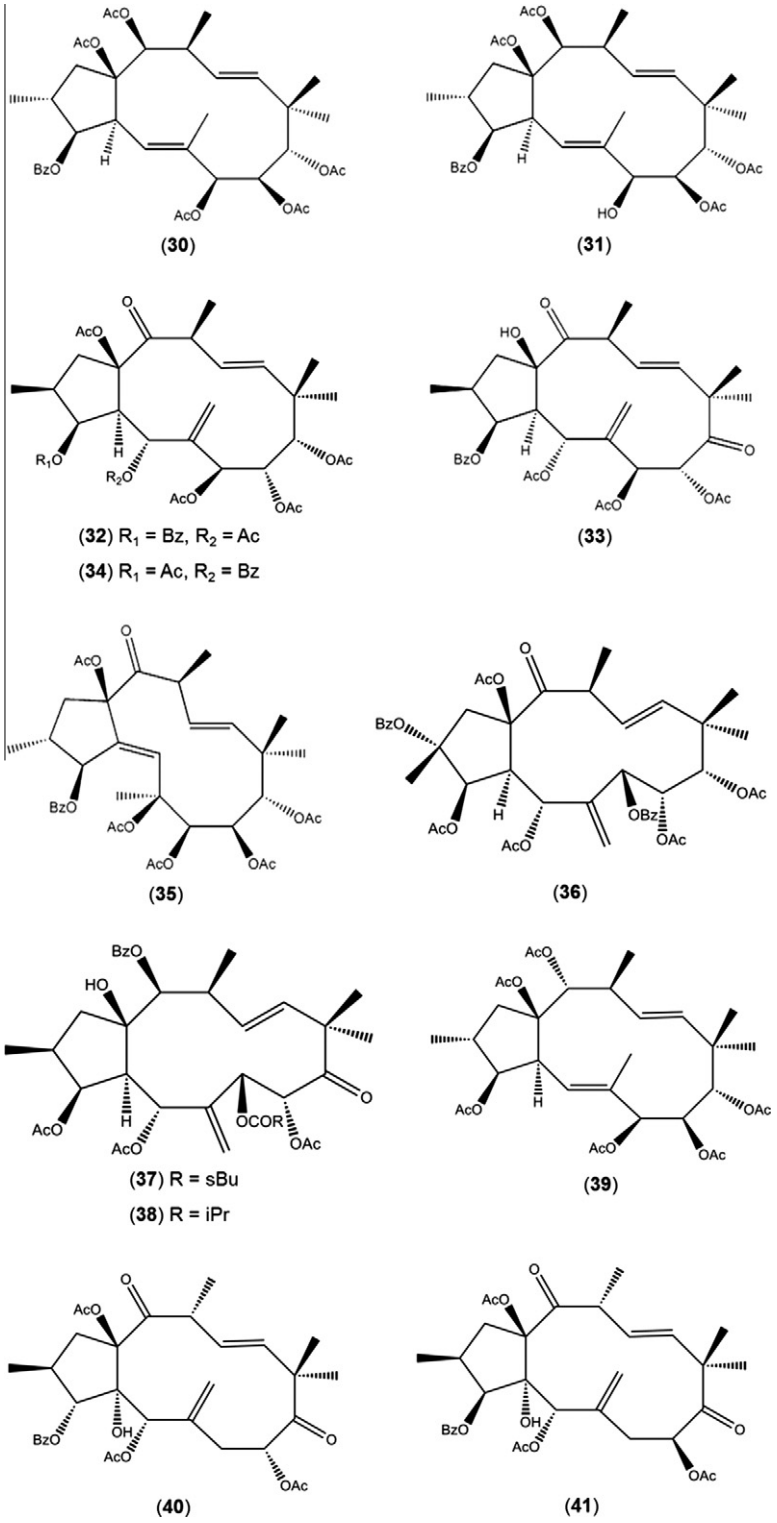

Fig. 1b-2. Structures of macrocyclic jatrophane-type diterpenes (**30**–**41**).

the human *MDR1*-transfected subline (MDR) were incubated with two different concentrations (4 and 40 μg/ml) of a given test compound for 10 min at room temperature. Then, rhodamine-123 substrate was added for additional 20 min at 37 °C. The fluorescence uptake of the cells was measured by flow cytometry. Verapamil was used as a positive control. The mean fluorescence intensity (FL-1) was calculated as a percentage of the control for the parental and MDR cell lines as compared to untreated cells. The fluorescence activity ratio (FAR) was calculated on the basis of the quotient between FL-1 of treated/untreated resistant cell line (MDR mouse lymphoma cells) over treated/untreated sensitive cell line (PAR mouse lymphoma cells), according to:

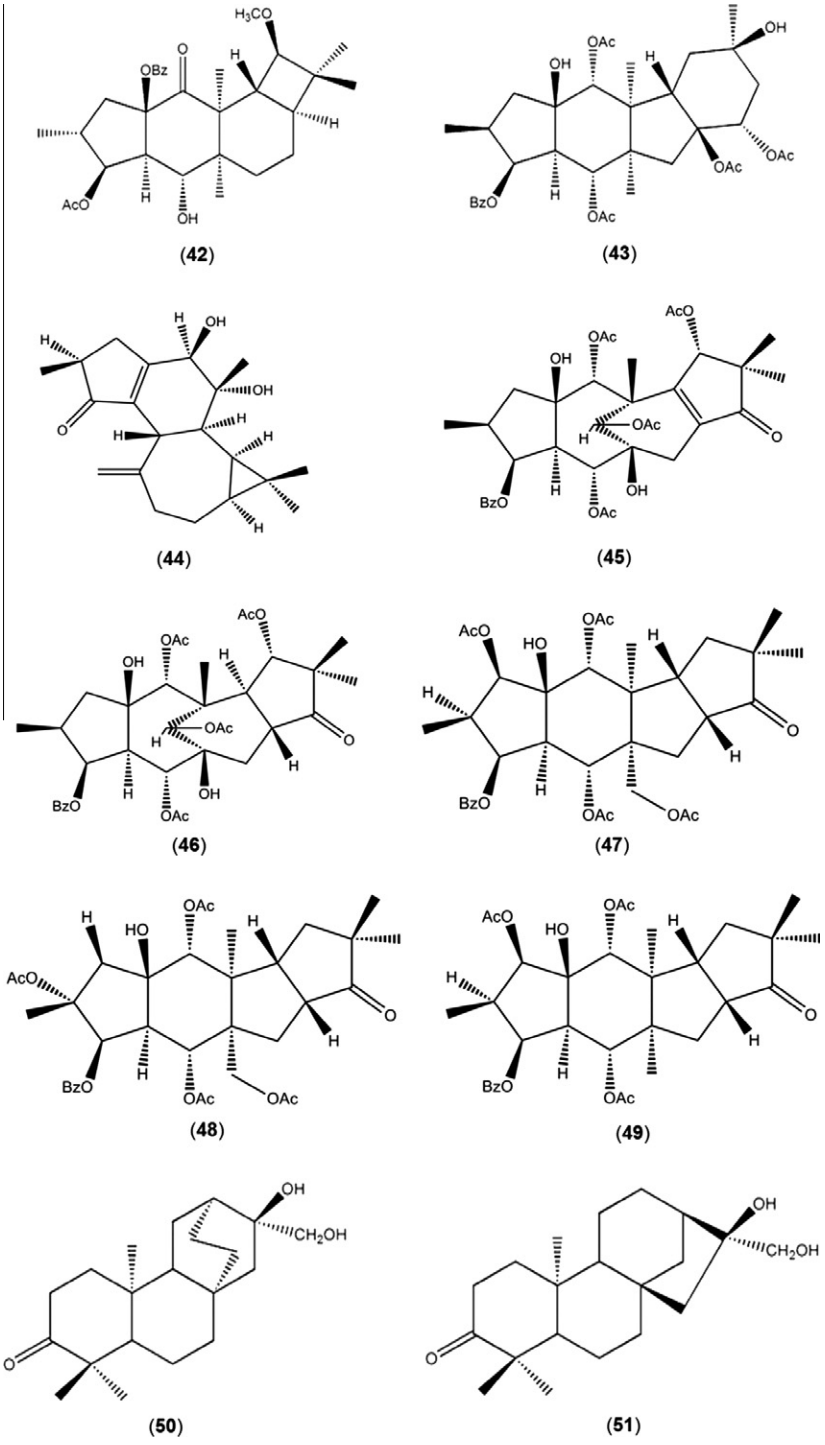

Fig. 1c. Structures of polycyclic diterpenes: rearranged lathyranes (**42**–**44**); rearranged jatrophanes (**45**–**49**); atisane (**50**); kaurane (**51**).

$$FAR = \frac{FL-1\ MDR_{treated}/FL-1\ MDR_{untreated}}{FL-1\ PAR_{treated}/FL-1\ PAR_{untreated}}$$

Compounds with FAR values higher than one were considered to be active as P-gp modulators and those with FAR values higher than 10 were regarded as strong modulators.

2.3. Computational details

The compounds' structures were optimized using the molecular mechanics (MM^+) force field included in HyperChem Release 7.5. (Hypercube, Inc., 2003) Thermodynamic and quantum descriptors were calculated using MOPAC (Stewart, 1993) included in VEGA ZZ 2.2.0 (Pedretti et al., 2004). For the calculation of thermodynamic and quantum descriptors we used the following keywords: "FORCE PRECISE THERMO ROT = X" and "VECTORS BONDS PI POLAR PRECISE ENPART", respectively.

HyperChem structure files, MOPAC output files and additional descriptors (number of H-bond donors and acceptors, and logP) calculated using E-DRAGON (Tetko et al., 2005) were used as input of the CODESSA (Katritzky et al., 1994) program for the calculation of a total of 306 structural descriptors. Five classes of structural

Table 1
Dataset used in QSAR methodology with the corresponding experimental activity values.

Compounds number	FAR[a]	
	Concentration 4 μg/ml	Concentration 40 μg/ml
1	110.40	159.50
2	13.01	46.04
3	28.11	102.07
4	12.18	13.76
5	168.50	175.40
6	216.80	199.30
7	15.30	74.00
8	3.46	50.80
9	8.90	44.16
10	68.90	71.60
11	61.60	78.00
12	62.40	54.20
13	34.74	74.27
14	8.75	12.73
15	8.87	6.70
16	4.31	4.86
17	3.71	3.96
18	2.03	1.93
19	2.58	7.83
20	2.47	4.03
21	5.52	14.75
22	45.94	NT
23	19.76	NT
24	16.51	NT
25	32.93	NT
26	12.75	71.98
27	34.25	78.88
28	16.77	29.48
29	6.30	2.26
30	36.21	51.11
31	20.98	34.79
32	2.60	18.02
33	12.29	22.92
34	2.790	29.29
35	10.57	16.08
36	81.00	106.20
37	39.80	98.90
38	25.00	97.30
39	2.25	1.67
40	2.61	2.62
41	43.52	58.51
42	2.60	17.54
43	4.71	38.37
44	0.80	2.73
45	9.70	2.30
46	1.40	30.70
47	3.21	NT
48	2.12	NT
49	15.59	NT
50	0.60	0.55
51	1.00	1.09

[a] FAR (fluorescence activity ratio) values were calculated by using the equation given in the methodology section.

descriptors were calculated using CODESSA: constitutional, topological, geometrical, electrostatic and quantum-chemical. Constitutional descriptors are related to the number of atoms and bonds in each molecule. Topological descriptors include valence and non-valence molecular connectivity indices, calculated from the hydrogen suppressed formula of molecules, encoding information about size, composition, and the degree of branching of a molecule. The quantum chemical descriptors provide information about binding and formation energies, partial atom charge, dipole moment, and molecular orbital energy levels.

The Heuristic Method, implemented by CODESSA, was used to perform the selection of the molecular descriptors correlated with the biological activity of molecules. This method performs the elimination of descriptors discarding those that satisfy at least one of the following conditions: (a) the descriptor value is not available for every structure; (b) the descriptor has a constant value for all structures. After this elimination, the one-parameter correlation equations for each descriptor are calculated.

To reduce even further the number of descriptors in the initial set, the following criteria are applied and descriptors are eliminated if: (a) the *F*-test's value for the one-parameter correlation with the descriptor is below 1.0; (b) the squared correlation coefficient of the one-parameter equation is smaller than R^2_{min} (0.1); (c) the parameter's *t*-value is smaller than t_1(1.5); (d) the descriptor is highly intercorrelated (above r_{full} (0.8)) to another descriptor with a higher squared correlation coefficient in the one-parameter equations based on these descriptors. All the remaining descriptors are then listed in decreasing order of their regression coefficients for the corresponding one-parameter correlation equations.

2.3.1. Training and test set selection

We investigated two QSAR models: the first one includes all the diterpenic compounds under study and we will call it the general model; and the second model includes only jatrophane-type compounds under study and we call it the jatrophane model. The set of compounds was divided into several subsets according to their biological activity range; the division was performed in 10 subsets for the general model at concentration of 4 μg/ml and in 8 subsets at concentration of 40 μg/ml. For jatrophane model at concentration of 4 and 40 μg/ml, the set of compounds was divided in 7 and 5 subsets, respectively. The compounds integrating the training and test sets were chosen randomly from each subset; the selection of training set compounds was performed in order to ensure the diversity and cover the chemical space of the structures under study and to guarantee that test set compounds were representative of dataset. The compounds were selected keeping in view the training/test set ratio of 4:1.

2.3.2. Forward selection and multiple linear regression modeling

Multiple linear regression (MLR) with forward feature selection was used to establish QSAR models. The F value was used for the analysis of variance and R^2 and RMSE of the training set as criteria for selection of models. We always performed internal validation of the QSAR models using the leave one out (LOO) method. The value of this LOO the cross-validation correlation coefficient is given by q^2. The size of descriptors subset used kept in view that the number of compounds in the training set should not be smaller than five times the number of descriptors.

3. Results and discussion

Jatrophane and lathyrane-type diterpenoids, obtained from *Euphorbia* species, were identified as promising MDR reversing agents (Duarte et al., 2007, 2006, 2008; Madureira et al., 2006; Molnár et al., 2006). The evaluation of MDR-reversal activity of the referred compounds, in non-toxic concentrations, was carried out in L15178 mouse T-lymphoma cell line transfected with the human *MDR1* gene, by flow cytometry, using the rhodamine-123 accumulation assay. The anti-MDR activity of the compounds was evaluated by the fluorescence activity ratio values. It is considered that when the FAR values are higher than 1.0, MDR reversal has taken place, thus higher FAR values mean better activity. In this study, we used FAR experimental values to obtain QSAR models. To get QSAR models, FAR was converted to pFAR, which means that to improve the activity of the compounds we have to lower pFAR values.

3.1. General model (biological activities determined at concentration 4 μg/ml)

The application of Heuristic method to obtain the general model (experimental biological activities of compounds determined at concentration of 4 μg/ml) produced a four parameter QSAR equation. The four molecular descriptors involved in this model are represented in Table 2. The graphical representation of experimental *versus* predicted activity for training and test sets is shown in Fig. 2.

The following equation describes the relation between pFAR and the molecular descriptors are in Table 2; n represents the total number of compounds in the training set, and F is the ratio between the variance explained by the model and the variance not explained by the model. This last parameter reflects the significance of the model.

$$\begin{aligned}\text{pFAR} = {} & (29.364 \pm 4.943) - (5.291 \times 10^{-2} \pm 5.635 \\ & \times 10^{-3})\text{ALOGP2} + (1.027 \pm 0.145)E_R(\text{tot})/\text{NA} - (2.115 \\ & \times 10^{2} \pm 46.986)\text{FPSA} - 3 + (5.691 \pm 1.756)\text{MaxACMO}\end{aligned}$$

$$n = 40, R^2 = 0.806, R^2_{\text{pred}} = 0.765, q^2 = 0.758, F = 36.342, s = 0.243$$

The obtained model shows a correlation coefficient, R^2, of 0.806 and a standard error, s, of 0.243. The internal and external validation of the model was performed using the cross-validation LOO test and the test set method, respectively. The obtained cross-validation value, q^2, has a value of 0.758, and the obtained squared correlation for the test set, R^2_{pred}, has a value of 0.765. The F-test value, F, shows that this model is statistically significant.

In Table 2, X represents the regression coefficient values for each molecular descriptor involved in the model; t-test reveals the significance of each descriptor in the model. t-test values indicate that the most significant descriptor for this model is the square-Ghose–Crippen octanol–water partition coefficient (ALOGP2). This molecular descriptor belongs to the molecular properties descriptors group and describes the lipophilicity of a molecule and is calculated as the sum of atomic lipophilic contributions that depends on the type and chemical neighborhood of the considered atom (D'Archivio et al., 2010).

The second most significant descriptor is the ratio between total molecular 2-center resonance energy and the number of atoms (E_R(tot)/NA). This is a quantum-chemical molecular descriptor and is considered as a measure of the stability of the aromatic systems present in the molecules (Katritzky et al., 2010).

The third descriptor, fractional atomic charge weighted partial positive surface area (FPSA-3) combines information about the atomic contributions to solvent accessible surface area of the molecule with partial charge information and encodes features responsible for polar interaction between molecules. This molecular descriptor belongs to the electrostatic molecular descriptors group (Golmohammadi, 2009).

The fourth descriptor involved in the model is the Maximum anti-bonding contribution of a molecular orbital (MaxACMO) that is a quantum-chemical molecular descriptor. This descriptor is related to the strength of intramolecular bonding interactions and characterizes the stability of the molecules (Yuan et al., 2009).

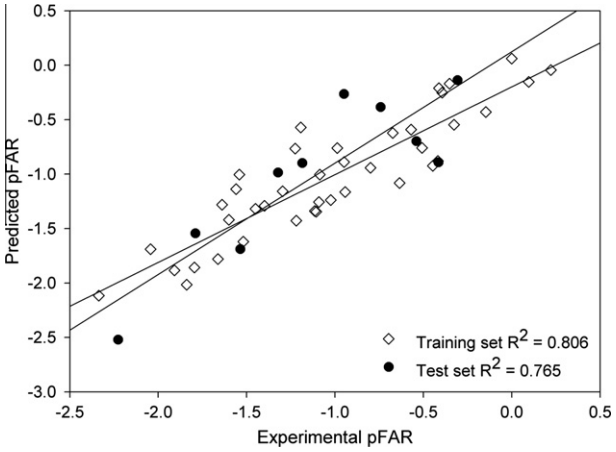

Fig. 2. Plots of experimental *versus* predicted activity values for training and test sets for diterpene compounds (**1**–**51**) at concentration 4 μg/ml.

The obtained variance inflation factor (VIF) values were calculated using inter-correlation values represented in Table 3. These values are lower than 10, which indicate the absence of collinearity between used descriptors for this QSAR model.

Based on this model, we can observe that to improve the activity we need to increase ALOGP2 and FPSA-3 values, and decrease E_R(tot)/NA and MaxACMO descriptor values.

3.2. General model (biological activities determined at concentration 40 μg/ml)

The application of Heuristic method to obtain the general model (experimental biological activities of compounds determined at concentration of 40 μg/ml) produced a seven parameter QSAR equation. In Tables 4 and 5 we represented the statistical parameters for these descriptors and the obtained equation is the following:

$$\begin{aligned}\text{pFAR} = {} & (8.137 \pm 3.596) - (33.165 \pm 5.569)\delta_C(\text{max}) + (5.523 \\ & \times 10^{2} \pm 1.294 \times 10^{2})\delta_H(\text{min}) - (0.923 \pm 0.220)\text{NN} \\ & + (1.092 \times 10^{14} \pm 3.886 \times 10^{13})\text{MinR}_A\text{O} - (30.664 \\ & \pm 7.263)\text{RNC} + (0.421 \pm 0.118)\text{ALOGPS\_logP} - (1.387 \\ & \pm 0.434)\text{ACIC1}\end{aligned}$$

$$n = 35, R^2 = 0.815, R^2_{\text{pred}} = 0.534, q^2 = 0.729, F = 16.973, s = 0.254$$

This QSAR model shows a correlation coefficient, R^2, for the training set of 0.815 and for test set, R^2_{pred}, of 0.534, as can be observed in Fig. 3. The cross-validation LOO value, q^2, was 0.729 and F-test value, F, 16.973.

Table 2
Statistics for the obtained model for diterpene compounds (**1**–**51**) at concentration 4 μg/ml.

Descriptor	$X \pm \Delta X$	t-Test	Significance (p)	Tolerance	VIF
Intercept	29.364 ± 4.943	5.940	0.000	–	–
ALOGP2	$-5.291 \times 10^{-2} \pm 5.635 \times 10^{-3}$	−9.389	0.000	0.895	1.117
ER(tot)/NA	1.027 ± 0.145	7.101	0.000	0.649	1.540
FPSA-3	$-2.115 \times 102 \pm 46.986$	−4.502	0.000	0.579	1.727
MaxACMO	5.691 ± 1.756	3.242	0.001	0.928	1.078

Table 3
Matrix of the inter-correlation of structural descriptors for diterpene compounds (**1**–**51**) at concentration 4 μg/ml.

	ALOGP2	E_R(tot)/NA	FPSA-3	MaxACMO
ALOGP2	1.000			
E_R(tot)/NA	−0.119	1.000		
FPSA-3	−0.312	0.571	1.000	
MaxACMO	−0.119	−0.032	0.197	1.000

From *t*-test values it is possible to assess the significance of molecular descriptors. The descriptors that contribute more significantly to the model are the Maximum partial charge for a C atom (δ_C(max)) and the Minimum partial charge for a H atom (δ_H(min)). They are electrostatic descriptors that reflect the charge distribution of the molecule and hydrogen atom, respectively (Konoz and Golmohammadi, 2008; Sun et al., 2006).

Relative number of C atoms (RNC) is a constitutional molecular descriptor that provides information about the constitution and the size of molecule (Yao et al., 2004).

Another constitutional descriptor involved in the model is the Number of N atoms (NN) and it is related with the ability to form hydrogen bonds (Papa et al., 2005).

Octanol–water partition coefficient (ALOGPS_logP) belongs to molecular properties descriptors group and is considered as a measure of lipophilicity of molecules (Komsta et al., 2011).

The Average complementary information content (order 1) (ACIC1) is a topological molecular descriptor that gives information about molecular branching and the diversity of the atoms of the branching (Ma et al., 2005).

The least significant descriptor for the model is Minimum 1-electron reactivity index for a O atom (Min R_AO). This quantum-chemical molecular descriptor reflects the reactivity at the site of oxygen atoms in the molecule (Yerramsetty et al., 2010).

This model shows us that compounds' activity can be improved by increasing δ_C(max), NN, RNC, ACIC1 values, and decreasing δ_H(min), Min R_AO and ALOGPS_logP descriptor values.

3.3. Jatrophane model (biological activities determined at concentration 4 μg/ml)

The set of jatrophanes under study was used in this QSAR study and results in a five parameter equation for biological activities determined at concentration of 4 μg/ml, shown below. The molecular descriptors are represented in Tables 6 and 7.

$$\begin{aligned}\mathrm{pFAR} = &-(4.545 \pm 0.542) + (4.714 \times 10^2 \pm 86.584) I_C - (0.434 \\ &\pm 0.103)\Delta H_f/\mathrm{NA} - (1.064 \times 10^5 \pm 3.286 \times 10^4)\mathrm{MinN'_AO} \\ &+ (7.492 \times 10^{13} \pm 2.877 \times 10^{13})\mathrm{MinR_AO} - (24.848 \\ &\pm 11.550)\mathrm{FNSA} - 3\end{aligned}$$

$$n = 28, R^2 = 0.825, R^2_{\mathrm{pred}} = 0.541, q^2 = 0.680, F = 20.793, s = 0.202$$

This model has a correlation coefficient, R^2, of 0.825 and a *F*-test value, *F*, of 20.793. Performing the internal and external validation of the model we obtained: $q^2 = 0.680$ and $R^2_{\mathrm{pred}} = 0.541$, respectively. Fig. 4 represents the plot of experimental *versus* predicted pFAR values for training and test sets.

From *t*-test values we can see that the geometrical descriptor Principal moment of inertia C (I_C) is the most significant molecular descriptor involved in the model, and it is related with the mass

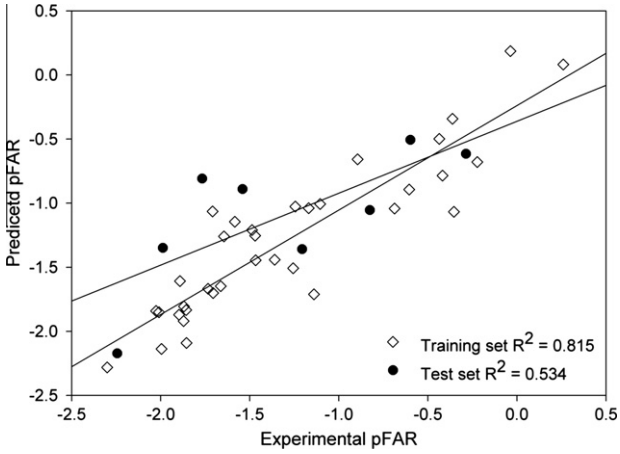

Fig. 3. Plots of experimental *versus* predicted activity values for training and test sets for diterpene compounds (**1**–**51**) at concentration 40 μg/ml.

Table 4
Statistics for the obtained model for diterpene compounds (**1**–**51**) at concentration 40 μg/ml.

Descriptor	$X \pm \Delta X$	*t*-Test	Significance (*p*)	Tolerance	VIF
Intercept	8.137 ± 3.596	2.263	0.015	–	–
δ_C(max)	−33.165 ± 5.569	−5.955	0.000	0.155	6.462
δ_H(min)	$5.523 \times 10^2 \pm 1.294 \times 10^2$	4.269	0.000	0.708	1.413
NN	−0.923 ± 0.220	−4.190	0.000	0.921	1.086
Min R_AO	$1.092 \times 10^{14} \pm 3.886 \times 10^{13}$	2.809	0.004	0.953	1.049
RNC	−30.664 ± 7.263	−4.222	0.000	0.191	5.223
ALOGPS_logP	0.421 ± 0.118	3.564	0.001	0.132	7.548
ACIC1	−1.387 ± 0.434	−3.194	0.002	0.155	6.462

Table 5
Matrix for the inter-correlation of structural descriptors for diterpene compounds (**1**–**51**) at concentration 40 μg/ml.

	δ_C(max)	δ_H(min)	NN	Min R_AO	RNC	ALOGPS_logP	ACIC1
δ_C(max)	1						
δ_H(min)	0.243	1					
NN	0.151	0.193	1				
Min R_AO	0.004	0.150	0.000	1			
RNC	0.315	−0.507	−0.013	−0.057	1		
ALOGPS_logP	0.290	−0.591	0.050	−0.187	0.624	1	
ACIC1	−0.052	0.243	0.151	0.004	0.315	0.290	1

Table 6
Statistics for the obtained model for jatrophane compounds (**13–41**, **45–49**) at concentration 4 μg/ml.

Descriptor	$X \pm \Delta X$	*t*-Test	Significance (*p*)	Tolerance	VIF
Intercept	-4.545 ± 0.542	−8.387	0.000	–	–
I_C	$4.714 \times 10^2 \pm 86.584$	5.444	0.000	0.368	2.716
ΔH_f/NA	-0.434 ± 0.103	−4.218	0.000	0.335	2.984
MinN′$_A$O	$-1.064 \times 10^5 \pm 3.286 \times 10^4$	−3.238	0.002	0.959	1.043
MinR$_A$C	$7.492 \times 10^{13} \pm 2.877 \times 10^{13}$	2.604	0.007	0.699	1.430
FNSA-3	-24.848 ± 11.550	−2.151	0.020	0.708	1.413

Table 7
Matrix for the inter-correlation of structural descriptors for jatrophane compounds (**13–41**, **45–49**) at concentration 4 μg/ml.

	I_C	ΔH_f/NA	MinN′$_A$O	MinR$_A$C	FNSA-3
I_C	1				
ΔH_f/NA	0.704	1			
MinN′$_A$O	−0.036	−0.058	1		
MinR$_A$C	0.214	−0.169	0.149	1	
FNSA-3	0.034	0.381	0.047	−0.367	1

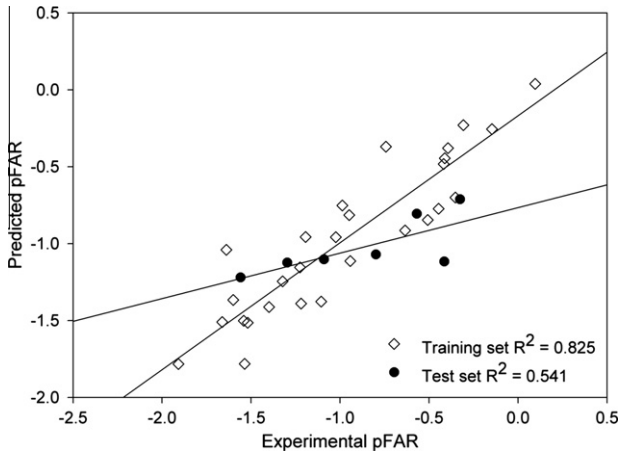

Fig. 4. Plots of experimental *versus* predicted activity values for training and test sets for jatrophane compounds (**14–41**, **45–49**) at concentration 4 μg/ml.

distribution in the molecule (Dashtbozorgi and Golmohammadi, 2010).

The second most significant descriptor for this model is the ratio between final heat of formation and total number of atoms (ΔH_f/NA), a quantum-chemical descriptor that is a measure of the reactive bonds in the molecule (Katritzky and Tatham, 2001).

Another quantum-chemical molecular descriptor involved in this model is the Minimum nucleophilic reactivity index for a O atom (MinN′$_A$O) that reflects the nucleophilicity of the molecule (Du et al., 2008).

The descriptor that presents a lower contribution for the model is the Fractional atomic charge weighted partial negative surface area (FNSA-3) that is an electrostatic descriptor related to the fraction of the total solvent-accessible surface area of the molecule associated with atoms containing a partial negative charge and it is also used to characterize the structural features involved in polar intermolecular interactions (Katritzky et al., 2007; Dyer et al., 2000).

From this model it is possible to conclude that to improve the activity of jatrophanes the values of descriptors ΔH_f/NA, MinN′$_A$O and FNSA-3 have to be increased, and I_C and MinR$_A$C values have to decrease.

3.4. Jatrophane model (biological activities determined at concentration 40 μg/ml)

The Heuristic method was applied to jatrophane-type diterpenes, whose biological activities was experimentally determined at concentration of 40 μg/ml, resulting in the following QSAR model with four quantum-chemical molecular descriptors. The statistics for these descriptors is represented in Tables 8 and 9.

$$\begin{aligned}\text{pFAR} = {} & (19.390 \pm 6.581) - (0.583 \pm 0.142)\Delta H_f/\text{NA} - (1.887 \\ & \times 10^5 \pm 3.411 \times 10^4)\text{MinN}'_A O - (24.890 \pm 6.565)\text{Max}_{\pi-\pi} \\ & - (1.546 \times 10^2 \pm 43.462)\text{MaxN}'_A\text{C}\end{aligned}$$

$$n = 23, R^2 = 0.857, R^2_{\text{pred}} = 0.534, q^2 = 0.787, F = 27.089, s = 0.214$$

The correlation coefficients obtained for the training and test sets were $R^2 = 0.857$ and $R^2_{\text{pred}} = 0.534$, respectively. The plot of the correlation between experimental and predicted activity values is showed in Fig. 5, *F*-test value, *F*, was 27.089 and cross validation test (LOO) value, q^2, was 0.787.

The *t*-test values indicate that Minimum nucleophilic reactivity index for a O atom (MinN′$_A$O) and the ratio between final heat of formation and total number of atoms (ΔH_f/NA) are the most significant descriptors for the obtained model.

The quantum-chemical descriptor Maximum PI–PI bond order ($\text{Max}_{\pi-\pi}$) provides the measure of the extent of sharing π-electrons between two atoms in order to form the most stable bond in the molecule (Kahn et al., 2005).

The least significant descriptor for this model is the Maximum nucleophilic reactivity index for a C atom (MaxN′$_A$C) is a quantum-chemical molecular descriptor that is related with the reactivity of the atoms in the molecule and can be considered as responsible for the reactivity of compounds (Liu et al., 2005).

Table 8
Statistics for the obtained model for jatrophane compounds (**13–41**, **45–49**) at concentration 40 μg/ml.

Descriptor	$X \pm \Delta X$	*t*-Test	Significance (*p*)	Tolerance	VIF
Intercept	19.390 ± 6.581	2.947	0.003	–	–
ΔH_f/NA	-0.583 ± 0.142	−4.110	0.000	0.623	1.606
MinN′$_A$O	$-1.887 \times 10^5 \pm 3.411 \times 10^4$	−5.531	0.000	0.951	1.051
$\text{Max}_{\pi-\pi}$	-24.890 ± 6.565	−3.791	0.001	0.890	1.124
MaxN′$_A$C	$-1.546 \times 10^2 \pm 43.462$	−3.557	0.001	0.700	1.429

Table 9
Matrix for the inter-correlation of structural descriptors for jatrophane compounds (**13**–**41**, **45**–**49**) at concentration 40 μg/ml.

	ΔH_f/NA	MinN′$_A$O	Max$_{\pi-\pi}$	MaxN′$_A$C
ΔH_f/NA	1			
MinN′$_A$O	−0.212	1		
Max$_{\pi-\pi}$	0.299	−0.120	1	
MaxN′$_A$C	0.535	−0.104	0.047	1

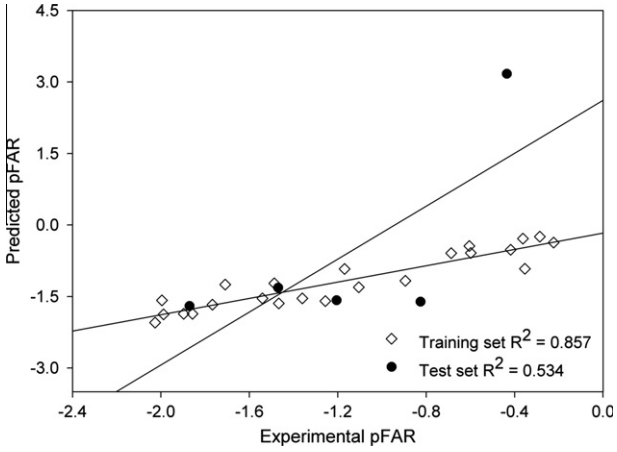

Fig. 5. Plots of experimental *versus* predicted activity values for training and test sets for jatrophane compounds (**14**–**41**, **45**–**49**) at concentration 40 μg/ml.

Table 10
Molecular descriptors and predicted biological activity values of initial and modified compound.

		Initial compound	Modified compound
General model at concentration 4 μg/ml			
Descriptors	ALOGP2	10.921	-5.291×10^{-2}
	E_R(tot)/NA	−13.926	1.027
	FPSA-3	0.015	-2.115×102
	MaxACMO	−2.158	5.691
	Calculated activity	−0.879	−1.908
General model at concentration 40 μg/ml			
Descriptors	δ_C(max)	0.084	−33.165
	δ_H(min)	0.016	5.523×10^2
	NN	0.000	−0.923
	Min R_AO	3.073×10^{-16}	1.092×10^{14}
	RNC	0.393	−30.664
	ALOGPS_logP	2.670	0.421
	ACIC1	2.790	−1.387
	Calculated activity	−0.785	−1.901
Jatrophane model at concentration 4 μg/ml			
Descriptors	I_C	1.581×10^{-3}	4.714×10^2
	ΔH_f/NA	−4.342	−0.433
	MinN′$_A$O	-6.754×10^{-6}	-1.064×10^5
	Min R_AC	7.864×10^{-16}	7.492×10^{13}
	FNSA-3	−0.027	−24.848
	Calculated activity	−0.483	−1.735
Jatrophane model at concentration 40 μg/ml			
Descriptors	ΔH_f/NA	−4.342	−0.583
	MinN′$_A$O	-6.754×10^{-6}	-1.886×10^5
	Max$_{\pi-\pi}$	0.963	−24.890
	MaxN′$_A$C	-1.683×10^{-3}	-1.546×10^2
	Calculated activity	−0.518	−2.225

The previous model shows us that the improvement of compounds activity can be achieved by increasing ΔH_f/NA, MinN′$_A$O, Max$_{\pi-\pi}$ and MaxN′$_A$C descriptor values.

Comparing all the obtained QSAR models, we can say that the general model, obtained for compounds' with biological acitivities determined at concentration of 4 μg/ml, and the jatrophane model, obtained for compounds' with biological acitivities determined at concentration of 40 μg/ml, are the best models. First, these models have fewer molecular descriptors involved, which facilitates the conversion of computational model in experimental synthesis and subsequent testing. Additionally, the model obtained for jatrophanes with biological activities determined at concentration of 40 μg/ml, shows the best values of training set correlation coefficient and cross-validation coefficient. Nevertheless, the general model with biological activities determined at concentration of 4 μg/ml has the best prediction ability and the molecular descriptors involved in this model are easier to interpret and, consequently, the findings here reported could be translated into synthesis of novel compounds.

The ultimate goal of QSAR studies is to propose compound modifications which will show improved biological activity. Having in mind that proposed modifications will not always produce a simultaneous change in the right direction for all the descriptors in the QSAR equation, we propose a real structural modification in a compound and determine the change in the predicted biological activity for every model. In Table 10 is shown the molecular descriptors

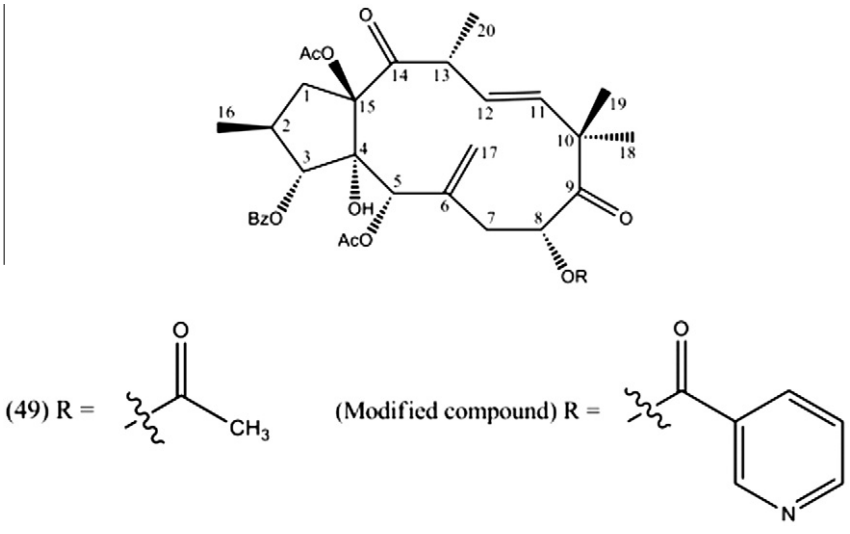

Fig. 6. Structure of compound **40**. C-8 corresponds to position where the substitution of an acetyl group by a nicotinoyl group was made.

and predicted biological activity values of initial and modified compound. As an example, starting from compound **40**, we propose a substitution of an acetyl group by a nicotinoyl at C-8 (Fig. 6). As can be observed in Table 10, this structural feature produces a change in the correct direction in the values of the descriptors used in all 4 QSAR equations. As a result of this change, the calculated activity of the new derivative is substantially improved when compared with that of compound **40** for every QSAR equations established in this study. In the immediate future we intend to obtain this new ester and assess its biological activity experimentally to confirm the predictions here produced.

4. Conclusions

MDR resistance modulators obtained from *Euphorbia* species were used to obtain QSAR models. The analysis of these models allows us to conclude that they are statistically valid and show high prediction ability. The models enabled the determination of the most relevant molecular features for the compounds' P-gp inhibitory activity. The best obtained model, judging from the prediction ability standpoint, is the general model (which includes all diterpenic compounds under study) with compounds concentration at 4 μg/ml. Interpreting the molecular descriptors involved in its QSAR equation it would be possible to improve the biological activity of compounds through structural modifications such as: increase the lipophilicity of molecules; increase the number of atoms with positive charges; and decrease the number of conjugated systems.

Acknowledgements

This research was supported by Fundação para a Ciência e a Tecnologia (FCT) Project PTDC/QUI-QUI/099815/2008, BIOPHARMAC Project (BIOPHARMAC-MAC/1/C104) and Project PEst-OE/QUI/UI0674/2011.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejps.2012.11.012.

References

Aller, S.G., Yu, J., Ward, A., Weng, Y., Chittaboina, S., Zhuo, R., Harrell, P.M., Trinh, Y.T., Zhang, Q., Urbatsch, I.L., Chang, G., 2009. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science 323, 1718–1722.

Choi, K., Frommel, T.O., Stern, R.K., Perez, C.F., Kriegler, M., Tsuruo, T., Roninson, I.B., 1991. Multidrug resistance after retroviral transfer of the human *MDR1* gene correlates with P-glycoprotein density in the plasma-membrane and is not affected by cytotoxic selection. Proc. Natl. Acad. Sci. U.S.A. 88, 7386–7390.

Cornwell, M.M., Pastan, I., Gottesman, M.M., 1987. Certain calcium-channel blockers bind specifically to multidrug-resistant human-kb carcinoma membrane-vesicles and inhibit drug-binding to P-glycoprotein. J. Biol. Chem. 262, 2166–2170.

D'Archivio, A.A., Maggi, M.A., Ruggieri, F., 2010. Multiple-column RP-HPLC retention modelling based on solvatochromic or theoretical solute descriptors. J. Sep. Sci. 33, 155–166.

Dashtbozorgi, Z., Golmohammadi, H., 2010. Quantitative structure-property relationship modeling of water-to-wet butyl acetate partition coefficient of 76 organic solutes using multiple linear regression and artificial neural network. J. Sep. Sci. 33, 3800–3810.

Du, H., Wang, J., Watzl, J., Zhang, X., Hu, Z., 2008. Prediction of inhibition of matrix metalloproteinase inhibitors based on the combination of projection pursuit regression and grid search method. Chemom. Intell. Lab. Syst. 93, 160–166.

Duarte, N., Gyémánt, N., Abreu, P.M., Molnár, J., Ferreira, M.J.U., 2006. New macrocyclic lathyrane diterpenes, from *Euphorbia lagascae*, as inhibitors of multidrug resistance of tumour. Cells Planta Med. 72, 162–168.

Duarte, N., Varga, A., Cherepnev, G., Radics, R., Molnár, J., Ferreira, M.J.U., 2007. Apoptosis induction and modulation of P-glycoprotein mediated multidrug resistance by new macrocyclic lathyrane-type diterpenoids. Bioorg. Med. Chem. 15, 546–554.

Duarte, N., Járdánházy, A., Molnár, J., Hilgeroth, A., Ferreira, M.J.U., 2008. Synergistic interaction between p-glycoprotein modulators and epirubicine on resistant cancer cells. Bioorg. Med. Chem. 16, 9323–9330.

Dyer, S.D., Stanton, D.T., Lauth, J.R., Cherry, D.S., 2000. Structure–activity relationships for acute and chronic toxicity of alcohol ether sulfates. Environ. Toxicol. Chem. 19, 608–616.

Golmohammadi, H., 2009. Prediction of octanol–water partition coefficients of organic compounds by multiple linear regression, partial least squares, and artificial neural network. J. Comput. Chem. 30, 2455–2465.

Higgins, C.F., Callaghan, R., Linton, K.J., Rosenberg, M.F., Ford, R.C., 1997. Structure of the multidrug resistance P-glycoprotein. Semin. Cancer Biol. 8, 135–142.

HyperChem(TM) Professional 7.5, Hypercube, Inc., 20031115 NW 4th Street, Gainesville, Florida 32601, USA.

Jassbi, A.R., 2006. Chemistry and biological activity of secondary metabolites in *Euphorbia* from Iran. Phytochemistry 67, 1977–1984.

Kahn, I., Fara, D., Karelson, M., Maran, U., Andersson, P.L., 2005. QSPR treatment of the soil sorption coefficients of organic pollutants. J. Chem. Inf. Model. 45, 94–105.

Katritzky, A.R., Tatham, D.B., 2001. Theoretical descriptors for the correlation of aquatic toxicity of environmental pollutants by quantitative structure-toxicity relationships. J. Chem. Inf. Comput. Sci. 41, 1162–1176.

Katritzky, A.R., Lobanov, V.S., Karelson, M., 1994. CODESSA: Reference Manual. Version 2, University of Florida.

Katritzky, A.R., Pacureanu, L., Dobchev, D., Karelson, M., 2007. QSPR modeling of hyperpolarizabilities. J. Mol. Model. 13, 951–963.

Katritzky, A.R., Slavov, S.H., Stoyanova-Slavova, I.B., Karelson, M., 2010. Correlation of the photolysis half-lives of polychlorinated dibenzo-p-dioxins and dibenzofurans with molecular structure. J. Phys. Chem. A 114, 2684–2688.

Kirby, J., Nishimoto, M., Park, J.G., Withers, S.T., Nowroozi, F., Behrendt, D., Rutledge, E.J.G., Fortman, J.L., Johnson, H.E., Anderson, J.V., Keasling, J.D., 2010. Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae*. Phytochemistry 71, 1466–1473.

Komsta, L., Radoń, M., Radkiewicz, B., Skibiński, R., 2011. Trilinear multiplicative modelling of thin layer chromatography retention as a function of solute, organic modifier and its concentration. J. Sep. Sci. 34, 59–63.

Konoz, E., Golmohammadi, H., 2008. Prediction of air-to-blood partition coefficients of volatile organic compounds using genetic algorithm and artificial neural network. Anal. Chim. Acta 619, 157–164.

Liu, H.X., Hu, R.J., Zhang, R.S., Yao, X.J., Liu, M.C., Hu, Z.D., Fan, B.T., 2005. The prediction of human oral absorption for diffusion rate-limited drugs based on heuristic method and support vector machine. J. Comput.-Aided Mol. Des. 19, 33–46.

Ma, W., Zhang, X., Luan, F., Zhang, H., Zhang, R., Liu, M., Hu, Z., Fan, B.T., 2005. Support vector machine and the heuristic method to predict the solubility of hydrocarbons in electrolyte. J. Phys. Chem. A 109, 3485–3492.

Madureira, A.M., Gyémánt, N., Ascenso, J.R., Abreu, P.M., Molnár, J., Ferreira, M.J.U., 2006. Euphoportlandols A and B, tetracylic diterpene polyesters from *Euphorbia portlandica* and their anti-MDR Effects in cancer cells. J. Nat. Prod. 69, 950–953.

Miglietta, A., Gabriel, L., Appendino, G., Bocca, C., 2003. Biological properties of jatrophane polyesters, new microtubule-interacting agents. Cancer Chemother. Pharmacol. 51, 67–74.

Molnár, J., Gyémánt, N., Tanaka, M., Hohmann, J., Bergmann-Leitner, E., Molnár, P., Deli, J., Didiziapetris, R., Ferreira, M.J.U., 2006. Inhibition of multidrug resistance of cancer cells by natural diterpenes, triterpenes and carotenoids. Curr. Pharm. Design 12, 287–311.

Papa, E., Villa, F., Gramatica, P., 2005. Statistically validated QSARs, based on theoretical descriptors, for modeling aquatic toxicity of organic chemicals in *Pimephales promelas* (Fathead Minnow). J. Chem. Inf. Model. 45, 1256–1266.

Pastan, I., Gottesman, M.M., Ueda, K., Lovelace, E., Rutherford, A.V., Willingham, M.C., 1988. A retrovirus carrying an MDR1 cdna confers multidrug resistance and polarized expression of P-glycoprotein in mdck cells. Proc. Natl. Acad. Sci. U.S.A. 85, 4486–4490.

Pedretti, A., Villa, L., Vistoli, G., 2004. VEGA – an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. J. Comput.-Aided Mol. Des. 18, 167–173.

Rosenberg, M.F., Callaghan, R., Modok, S., Higgins, C.F., Ford, R.C., 2005. Three-dimensional structure of P-glycoprotein. J. Biol. Chem. 280, 2857–2862.

Shi, Q., Su, X., Kiyota, H., 2008. Chemical and pharmacological research of the plants in genus *Euphorbia*. Chem. Rev. 108, 4295–4327.

Shokoohinia, Y., Chianese, G., Zolfaghari, B., Sajjadi, S., Appendino, G., Taglialatela-Scafati, O., 2011. Macrocyclic diterpenoids from the Iranian plant *Euphorbia bungei* Boiss. Fitoterapia 82, 317–322.

Stavrovskaya, A.A., Stromskaya, T.P., 2008. Transport proteins of the ABC family and multidrug resistance of tumor cells. Biochemistry (Mosc) 73, 592–604.

Stewart, J., 1993. MOPAC Manual, 7ª ed.

Sun, N., He, X., Dong, K., Zhang, X., Lu, X., He, H., Zhang, S., 2006. Prediction of the melting points for two kinds of room temperature ionic liquids. Fluid Phase Equilib. 246, 137–142.

Tetko, I., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V., Radchenko, E., Zefirov, N., Makarenko, A., Tanchuk, V., Prokopenko, V., 2005. Virtual computational chemistry laboratory – design and description. J. Comput.-Aided Mol. Des. 19, 453–463.

Thomas, H., Coley, H.M., 2003. Overcoming multidrug resistance in cancer: an update on the clinical strategy of inhibiting P-glycoprotein. Cancer Control 10, 159–165.

Valente, C., Pedro, M., Duarte, A., Nascimento, M.S., Abreu, P.M., Ferreira, M.J.U., 2004. Bioactive diterpenoids, a new jatrophane and two ent-abietanes, and other constituents from *Euphorbia pubescens*. J. Nat. Prod. 67, 902–904.

Yao, X.J., Panaye, A., Doucet, J.P., Zhang, R.S., Chen, H.F., Liu, M.C., Hu, Z.D., Fan, B.T., 2004. Comparative study of QSAR/QSPR correlations using support vector machines, radial basis function neural networks, and multiple linear regression. J. Chem. Inf. Comput. Sci. 44, 1257–1266.

Yerramsetty, K.M., Neely, B.J., Madihally, S.V., Gasem, K.A.M., 2010. A skin permeability model of insulin in the presence of chemical penetration enhancer. Int. J. Pharm. 388, 13–23.

Yuan, Y., Zhang, R., Hu, R., 2009. Prediction of photolysis of PCDD/Fs adsorbed to spruce [*Picea abies* (L.) Karst.] needle surfaces under sunlight irradiation based on projection pursuit regression. QSAR Comb. Sci. 28, 155–162.