Structure-activity relationships of taxoids: a molecular descriptors family approach

Sorana D. Bolboacă[1], Lorentz Jäntschi[2]

[1]"Iuliu Hatieganu" University of Medicine and Pharmacy Cluj-Napoca, 6 Louis Pasteur, 400349 Cluj-Napoca, Romania
[2]Technical University of Cluj-Napoca, 103-105 Muncii Bvd, 400641 Cluj-Napoca, Romania

Submitted: 6 November 2007
Accepted: 20 January 2008

Arch Med Sci 2008; 4, 1: 7–15

Corresponding author:
Sorana D. Bolboacâ
Department of Medical Informatics and Biostatistics
"Iuliu Hatieganu" University of Medicine
and Pharmacy Cluj-Napoca
6 Louise Pasteur,
400349 Cluj-Napoca, Romania
Phone: +4 0264 431697
Fax: +4 0264 593847
E-mail: sbolboaca@umfcluj.ro

Abstract

Introduction: Taxoids, groups of diterpenoid cyclodecanes isolated from the genus Taxus, are known and used as anticancer agents. Starting from the successful results obtained by an original molecular descriptors family on structure-activity relationships (Jäntschi and Bolboacă, 2007), the aim of the research was to investigate and to assess the estimation and prediction abilities of this approach on a sample of taxoids.
Material and methods: The molecular descriptors family on structure-activity relationships approach was used in order to characterize the link between structure of a sample of thirty-four taxoids and associated growth inhibition activity. The chance correlation of obtained models was investigated using random assignment of compounds in leave-one-out analysis and training versus test experiments.
Results: One model with one descriptor and two multivariate models, one with three and the other with five descriptors, proved to have estimation and prediction abilities. Statistical characteristics of the models revealed that the MDF-SAR model with five descriptors has excellent estimation and prediction abilities compared with models with one or three descriptors.
Conclusions: The molecular descriptors family on structure-activity relationships approach proved its usefulness in characterization of growth inhibition activity of studied taxoids. Further studies on new taxoids are necessary in order to assess the robustness and prediction ability of the MDF-SAR model with five descriptors.

Key words: quantitative structure-activity relationship (QSAR), plant extracts characterization, one- and multi-parameter models, regression analysis.

Introduction

Taxoids are groups of diterpenoid cyclodecanes isolated from Taxus baccata [1], Taxus brevifolia [2], Taxus Canadensis [3], Taxus chinensis [4], Taxus cuspidate [5, 6], Taxus floridana [7], Taxus sumatrana [8] or from Taxus wallichiana [9]. Some taxoids act to decrease the critical concentration of tubulin required for assembly [10, 11] through the inhibition of microtubule disassembly [12, 13]. Other taxoids reduce the $CaCl_2$-induced depolymerization of microtubules [14] or increase the cellular accumulation of vincristine in multi-drug resistant tumour cells [14, 15]. Since their discovery, taxoids have been used in treatment of polycystic kidney diseases [16] and neoplasms (ovarian cancer [17], breast cancer [18], lung non-small cell cancer [19], prostate cancer [20], and head and neck cancer [21]). A major disadvantage of these compounds is their poor solubility [22, 23].

Many studies are being conducted using the structure-activity relationships approach and/or synthetic modification in order to increase activity and solubility of new taxoid analogues [24, 25].

Thirty-five cytotoxic taxoids, compounds isolated through chromatographic purification of taxoid fraction from the steam of Taxus cuspidate Sieb. Et Zucc. Var nana Rehder [25, 26], were studied using comparative molecular field analysis (CoMFA) [6]. Statistical characteristics of the model reported by Morita et al. [6] are:

r^2=0.979, $r^2_{cv\text{-}loo}$=0.818

s=0.196, F=267.621, n=35, v=5 Eq (1),

where r^2 – squared correlation coefficient, $r^2_{cv\text{-}loo}$ – squared cross-validation leave-one-out coefficient, s – standard error of the estimation, F – Fisher parameter, n – sample size, and v – number of variables.

Starting from the successful results obtained by an original molecular descriptors family on structure-activity relationships (MDF-SAR) methodology [27-30], the aim of the research was to investigate and to assess the estimation and prediction abilities of the MDF-SAR approach on a sample of taxoids.

Material and methods

Data set: taxoids

A sample of thirty-four taxoids was investigated. The compounds analyzed were as follows: taxol (t01), 10-deaceltyl-taxol (t02), taxol B (t03), 10-deaceltyl-taxol B (t04), taxol C (t05), 10-deaceltyl-taxol C (t06), taxuspinanane A (t07), taxol D (t08), baccatin III (t09), 9-dihydro-14-acetyl baccatin III (t10), taxuspinanane C (t11), 7,9,10-deacetyl baccatin VI (t12); taxuspinanane D (t13), brevifoliol (t14), taxusin (t15); 2α-deacetoxy taxinine J (t16); taxinin (t17); taxa-4(20), 11-diene-2α, 5α, 9α, 13α-pentaol pentaacetate (t18); taxa-4(20), 11-diene-5α, 7β, 9α, 10β, 13α-pentaol pentaacetate (t19); taxa-4(20), 11-diene-5α, 7β, 9α, 10β, 13α-pentaol 7β, 9α, 10β-triacetate (t20); 2α-α-methyl butyryloxy-5α-7β, 10β-triacetyl-(4), 20, 11-taxadine (t21); taxa-4(20), 11-diene-5α, 7β, 10β, 13α-pentaol 7β, 9α, 10β, 13α tetra-acetate (t22); taxinin B (t23); decinnamoyl taxinine J (t24), taxuspinanane K (t25), taxuspine F (t26); taxuspinanane G (t27), taxuspine L (t28); taxchin A (t29); taxinine M (t30), taxgifine (t31); taxa-4(20), 11-taxadiene-2α, 5α, 10β, 14β-(s)2'-methyl butyrate (t32); 1β-hydroxy-baccatin I (t33); and taxuspinanane H (t34). The growth inhibition activity expressed as log $1/IC_{50}$ (where IC_{50} is the concentration of a taxoid that is required for 50% growth inhibition in vitro) was taken from a previously reported study [6]. The generic structures of the taxoids are presented in Figure 1. The abbreviation (Abb.), the substituent (S_i, where i=1, ..., 6) and experimental growth inhibition activity (Y_{obs}) are presented in Table I.

Computational methodology

The growth inhibition activity of taxoids was modelled using the MDF-SAR approach [31]. The structure of each taxoid was drawn up using *HyperChem* software [32]. The observed inhibition activity was stored in the *taxoids.txt* file. The molecular descriptors family was generated, the molecular descriptors being calculated strictly based on the information obtained from the compounds' 2D and 3D structures. Each descriptor had an individual seven-letter name expressing its modality of construction [31]. Starting with generated molecular descriptors, an algorithm was applied in order to identify the best MDF-SAR models with one and more than one variable. In identification of the regression models with higher models' goodness-of-fit, a step-wise approach, forward selection was used (it started with one variable in the model, trying out the variables one by one, and including them if the obtained model was statistically significant). The following internal validation approaches were applied for models' assessment: statistical characteristics of the regression model, leave-one-out analysis [33], and correlated

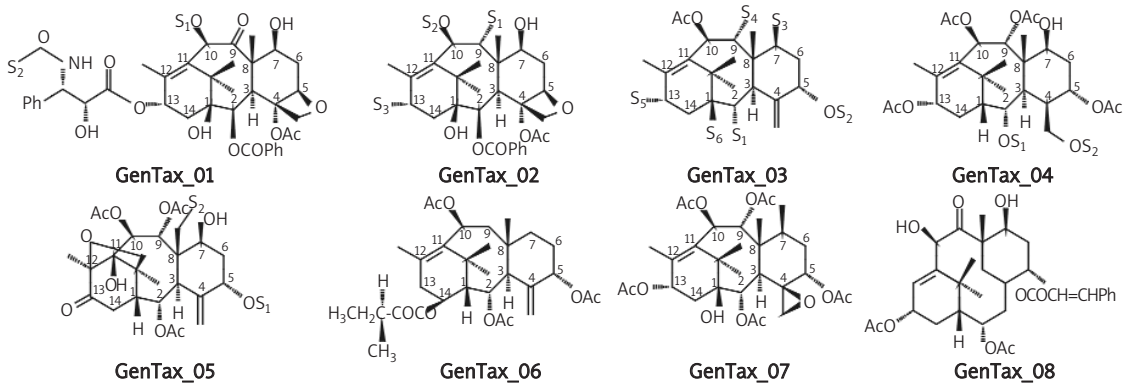

Figure 1. Generic structure of investigated taxoids

Table I. Structure, substituent (S_i, where i = 1, ..., 6), and measured inhibition activity for taxoids (Y_{obs})

GenTax	Abb.	Substituent	Y_{obs}
GenTax_01	t01	S_1 = Ac S_2 = Ph	1.66
	t02	S_1 = H S_2 = Ph	1.37
	t03	S_1 = Ac S_2 = C_4H_7	0.77
	t04	S_1 = H S_2 = C_5H_{11}	1.18
	t05	S_1 = Ac S_2 = n–C_5H_{11}	1.09
	t06	S_1 = H S_2 = n–C_5H_{11}	1.39
	t07	S_1 = Ac S_2 = C_6H_{13}	1.74
	t08	S_1 = Ac S_2 = n–C_3H_7	0.77
GenTax_02	t09	S_1 = =O S_2 = Ac S_3 = OH	–1.20
	t10	S_1 = OH S_2 = Ac S_3 = Oac	–1.28
	t11	S_1 = OH S_2 = H S_3 = =O	–1.00
	t12	S_1 = OH S_2 = H S_3 = OAc	–1.54
	t13	S_1 = OH S_2 = Ac S_3 = =O	–1.32
GenTax_03	t14	S_1, S_2 = H S_3 = OCOPh S_4 = OAc S_5, S_6 = OH S_1, S_3, S_6 = H	–1.60
	t15	S_2 = Ac S_4, S_5 = OAc S_1, S_6 = H	–0.34
	t16	S_2 = COCH=CHPh S_3, S_4, S_5 = OAc S_1, S_4 = OAc S_2 = COCH=CHPh	–0.64
	t17	S_3,S_6 = H S_5 = =O	–2.00
GenTax_03	t18	S_1, S_4, S_5 = OAc S_2 = Ac S_3,S_6 = H	–1.78
	t19	S_1, S_6 = H S_2 = Ac S_3, S_4, S_5 = OAc	–0.62
	t20	S_1, S_2, S_6 = H S_3, S_4 = OAc S_5 = OH S_1 = $OCOC_4H_9$	–1.20
	t21	S_2 = Ac S_3 = OAc S_4, S_5, S_6 = H	–0.48
	t22	S_1, S_2, S_6 = H S_3, S_4, S_5 = OAc S_1 S_3, S_4 = OAc	–1.36
	t23	S_2 = COCH=CHPh S_5 = =O S_6 = H	–2.00
	t24	S_1, S_3, S_4, S_5 = OAc S_2, S_6 = H	–1.90
	t25	S_1, S_2, S_6 = H S_3, S_4 = OAc S_5 = =O	–1.91
	t26	S_1, S_3, S_4 = OAc S_2, S_6 = H S_5 = =O	–1.18
	t27	S_1 = OH S_2 = COCH=CHPh S_3, S_4, S_5 = OAc S_6 = H	–0.59
GenTax_04	t28	S_1 = Ac S_2 = H	–1.85
	t29	S_1 = H S_2 = Ac	–1.91
GenTax_06	t30	S_1 = H S_2 = OCOPh	–1.57
	t31	S_1 = COCH=CHPh S_2 = H	–2.00
GenTax_06	t32		–0.64
GenTax_07	t33		–2.00
GenTax_08	t34		–1.32

GenTax – generic structure of taxoid (see Figure 1), Ph – phenyl, Ac – acetyl, S_i – substituent (i = 1, ...6)

correlation analysis (Steiger's and Fisher Z tests) [34]. Pearson, Spearman and semi-quantitative correlation coefficients [35] were calculated for comparison of the MDF-SAR model with higher squared correlation coefficient with previously reported model [36].

Results

One univariate MDF-SAR model and two multivariate models, one with three and the other with five descriptors, proved to have good estimated and prediction abilities. A summary

description of the models is presented in Table II, and the statistical analysis of the models is presented in Table III.

Estimation abilities of the models from Eq(2)-Eq(4) are presented in terms of activity estimated by the model ($\hat{Y}_{Eq(2)}$, $\hat{Y}_{Eq(3)}$, and $\hat{Y}_{Eq(4)}$) and residuals ($R_{Eq(2)}$, $R_{Eq(3)}$, $R_{Eq(4)}$) in Table IV.

In order to test the statistical hypothesis that the correlation coefficient obtained by MDF-SAR models from Eq(2)-Eq(4) are not statistically different, Steiger's Z test at a significance level of 5% was applied. The following results were obtained:

- $r_{Eq(2)}$ vs. $r_{Eq(3)}$: Z=3.4891 (p<0.0001),
- $r_{Eq(2)}$ vs. $r_{Eq(4)}$: Z=5.5845 (p<0.0001),
- $r_{Eq(3)}$ vs. $r_{Eq(4)}$: Z=3.0192 (p<0.003).

The prediction ability of the MDF-SAR model from Eq(4) was assessed by randomly splitting the sample into training and test sets (23 compounds in training and 11 compounds in test). The equation and statistical characteristics are:

$\hat{Y} = -7.41 - 0.30 \times lmPrVQt - 0.03 \times iNMMkQg - 1.10 \times lmPrsCg + 216.40 \times IIMdPQg + 0.75 \times IHDrFHt$

$r^2_{training} = 0.9728$; $F_{training} = 122$ ($p<0.0001$); $r^2_{test} = 0.9752$; $F_{test} = 35$ ($p<0.001$) [5].

The graphical representation of the models in training and test sets when the number of compounds in the training set was of 2/3 of the sample size is presented in Figure 2.

The comparison between the MDF-SAR model with five descriptors and the previously reported CoMFA model [6] was done by applying a correlated correlation analysis using the Pearson, semi-quantitative and Spearman methods [35]. The results expressed as correlation coefficients are presented in Table V.

The graphical representation of the growth inhibition activity measured experimentally and estimated by CoMFA [6] and MDF-SAR model with five descriptors is presented in Figure 3.

Discussion

The aim of the research was reached: the molecular descriptors family on structure-activity relationship proved to be a valid approach in characterization of taxoids' growth inhibition activity based on information obtained strictly from 2D and 3D structures.

Table II. MDF-SARs on taxoids growth inhibition activity: models characteristics

Characteristic	Number of molecular descriptor		
No. of descriptors	**one**	**three**	**five**
MDF SPR equation	$\hat{Y}_{1d} = -8.23 + x \cdot 0.89$ Eq(2)	$\hat{Y}_{3d} = -8.20 + x_1 \cdot (-0.99) + x_2 \cdot 147.51 + x_3 \cdot 0.79$ Eq(3)	$\hat{Y}_{5d} = -7.39 + x_1 \cdot (-0.23) + x_2 \cdot (2.85 \cdot 10^{-2}) + x_3 \cdot 1.11 + x_4 \cdot 193.25 + x_5 \cdot 0.71$ Eq(4)
SAR determination (%)	83	94	98
MDF descriptor (x_i)	*IHDrFHt*	x_1 = *lmPrsCg* x_2 = *IIMdPQg* x_3 = *IHDrFHt*	x_1 = *lmPrVQt* x_2 = *iNMMkQg* x_3 = *lmPrsCg* x_4 = *IIMdPQg* x_5 = *IHDrFHt*
Dominant atomic property	directly bounded hydrogen's (H)	cardinality (C) charge (Q) directly bounded hydrogen's (H)	charge (Q) charge (Q) cardinality (C) charge (Q) directly bounded hydrogen's (H)
Interaction via	bonds (topology)	space (geometry) space (geometry) bonds (topology)	bonds (topology) space (geometry) space (geometry) space (geometry) bonds (topology)
Interaction model	H^2/d^2	C^2/d^3 Q^2 H^2/d^2	Q/d Q^2/d C^2/d^3 Q^2 H^2/d^2
Structure on activity scale	identity	logarithmic identity identity	logarithmic inversed logarithmic identity identity

Table III. MDF-SARs on taxoids growth inhibition activity: statistical characteristics of the models

Characteristic (symbol)		Eq(2)	Eq(3)	Eq(4)
Correlation coefficient [95% CI] (r)		0.91 [0.86-0.95]	0.97 [0.94-0.98]	0.99 [0.98-0.99]
Adjusted r squared (r^2_{adj})		0.82	0.94	0.97
Standard error of estimated (s)		0.51	0.31	0.21
Fisher parameter (F)		156*	161*	226*
[95% CI] of intercept		[(–9.47)-(–7.00)]	[(–9.03)-(–7.37)]	[(–8.02)-(–6.77)]
[95% CI] of	x_1	[0.75-1.04]	[(–1.28)-(–0.71)]	[(–0.33)-(–0.14)]
	x_2	n.a.	[98.46-196.59]	[(–0.04)-(–0.02)]
	x_3	n.a.	[0.70-0.89]	[(–1.13)-(–0.92)]
	x_4	n.a.	n.a.	[156.72-229.80]
	x_5	n.a.	n.a.	[0.68-0.82]
Standard error of intercept		0.61	0.41	0.31
Standard error of	x_1	0.07	0.14	0.05
	x_2	n.a.	24.02	0.01
	x_3	n.a.	0.05	0.10
	x_4	n.a.	n.a.	17.84
	x_5	n.a.	n.a.	0.03
t-Stat of intercept		–13.57*	–20.22*	–24.18*
t-Stat of	x_1	12.48*	–7.07*	–4.83*
	x_2	n.a.	6.14*	–5.19*
	x_3	n.a.	16.95*	–11.65*
	x_4	n.a.	n.a.	10.83*
	x_5	n.a.	n.a.	22.23*
Squared cross-validation leave-one-out coefficient (r^2_{loo})		0.81	0.93	0.97
Fisher parameter in leave-one-out analysis (F_{loo})		137*	128*	156*
Standard error of estimated in leave-one-out analysis (s_{loo})		0.53	0.34	0.24

*95% CI – 95% confidence interval; * p<0.0001; n.a. – not applicable*

Due to the possibility of in silico experiments on new taxoids, structure-activity methods are used in order to obtain compounds with increased activity and solubility and decreased toxicity [37-39]. The differences of these methods are at the level of descriptors type, construction and calculation. The MDF-SAR method is unique due to generation and calculation of descriptors based on the topological and geometrical model of compounds.

Analyzing the molecular descriptors used by the MDF-SAR models (Table II) it can be observed that one descriptor appears in all models (*IHDrFHt*), showing that the activity of taxoids is related to compounds' topology, and it is dependent on the number of directly bonded hydrogens. As expected, with increase of the number of descriptors, the estimation abilities increase. The models with three descriptors revealed that the inhibition activity of taxoids is related to compounds' geometry and topology, and it is related to three atomic properties: cardinality, charge and number of directly bonded hydrogens. The MDF-SAR model with five variables showed that the growth inhibition activity of studied taxoids depends on compounds' geometry (*iNMMkQg*, *lmPrsCg*, *IIMdPQg*) as well as on topology (*lmPrVQt, IHDrFHt*). This model also revealed that the partial charges (*lmPrVQt, iNMMkQg, IIMdPQg*), the number of directly bonded hydrogens (*IHDrFHt*) and the cardinality (*lmPrsCg*) are the atomic properties that influence the growth inhibition activity.

All MDF-SAR models were statistically significant (Table III). The estimation abilities of the models is sustained by the values of the correlation coefficient and associated adjusted squared correlation coefficient, which with one exception (for the models with one descriptor) were greater than 0.90. Furthermore, the sum of residuals of the MDF-SAR models were very low (0.0000 for Eq(2), 0.0057 for Eq(3), and 0.0045 for Eq(4), respectively). In statistical terms, it can be concluded that there is a very good level of association between growth inhibition activity and the three descriptors used by the model from Eq(3) and the five descriptors used by the model from Eq(4), respectively. 94% of growth inhibition activity of studied taxoids can be explained by its linear relationship with the variation

Table IV. MDF-SAR descriptors, estimated growth inhibition activity ($\hat{Y}_{Eq(2)}$, $\hat{Y}_{Eq(3)}$, and $\hat{Y}_{Eq(4)}$) and residuals ($R_{Eq(2)}$, $R_{Eq(3)}$, $R_{Eq(4)}$)

Mol	*IHDrFHt*	*lmPrVQt*	*IIMdPQg*	*iNMMkQg*	*lmPrsCg*	$\hat{Y}_{Eq(2)}$	$R_{Eq(2)}$	$\hat{Y}_{Eq(3)}$	$R_{Eq(3)}$	$\hat{Y}_{Eq(4)}$	$R_{Eq(4)}$
1	10.03	1.41	8.82×10^{-3}	10.54	–0.38	0.74	0.92	1.46	0.20	1.60	0.06
2	10.30	1.02	4.92×10^{-3}	10.70	–0.46	0.99	0.38	1.18	0.19	1.24	0.13
3	10.29	0.08	4.95×10^{-3}	33.69	–0.38	0.98	–0.21	1.09	–0.32	0.70	0.07
4	10.64	1.55	5.83×10^{-3}	17.97	–0.40	1.29	–0.11	1.53	–0.35	1.27	–0.09
5	10.64	–0.15	3.10×10^{-3}	14.43	–0.26	1.29	–0.20	0.98	0.11	1.08	0.01
6	10.96	1.27	4.04×10^{-3}	10.08	–0.46	1.57	–0.18	1.58	–0.19	1.52	–0.13
7	10.76	–0.15	4.16×10^{-3}	14.41	–0.38	1.40	0.34	1.35	0.39	1.51	0.23
8	10.15	–0.15	4.34×10^{-3}	13.77	–0.26	0.85	–0.08	0.77	0.00	0.97	–0.20
9	8.32	1.83	4.27×10^{-3}	6.97	–0.09	–0.79	–0.41	–0.86	–0.34	–0.87	–0.33
10	8.11	1.64	3.32×10^{-3}	9.42	–0.09	–0.97	–0.31	–1.16	–0.12	–1.23	–0.05
11	8.20	0.63	2.12×10^{-3}	9.37	–0.11	–0.89	–0.11	–1.25	0.25	–1.14	0.14
12	8.45	1.79	3.59×10^{-3}	10.77	–0.09	–0.67	–0.87	–0.85	–0.69	–1.00	–0.54
13	7.81	0.63	2.96×10^{-3}	8.51	–0.11	–1.24	–0.08	–1.44	0.12	–1.25	–0.07
14	8.16	1.75	2.90×10^{-4}	8.77	–0.20	–0.93	–0.67	–1.47	–0.13	–1.67	0.07
15	7.44	1.37	7.14×10^{-3}	12.49	–0.60	–1.58	1.24	–0.64	0.30	–0.46	0.12
16	7.55	1.29	4.56×10^{-3}	9.00	–0.60	–1.48	0.84	–0.93	0.29	–0.76	0.12
17	7.06	1.31	4.47×10^{-3}	12.35	0.20	–1.91	–0.09	–2.12	0.12	–2.13	0.13
18	7.34	1.39	8.44×10^{-3}	12.08	0.55	–1.67	–0.11	–1.66	–0.12	–1.55	–0.23
19	7.49	2.00	8.05×10^{-3}	12.26	–0.45	–1.53	0.91	–0.61	–0.01	–0.56	–0.06
20	7.84	3.31	3.65×10^{-3}	8.54	–0.25	–1.22	0.02	–1.17	–0.03	–1.57	0.37
21	8.36	–0.01	5.62×10^{-3}	10.44	0.55	–0.75	0.27	–1.26	0.78	–0.96	0.48
22	7.59	1.02	3.22×10^{-3}	9.44	–0.36	–1.44	0.08	–1.32	–0.04	–1.20	–0.16
23	7.20	1.31	4.75×10^{-3}	12.29	0.20	–1.79	–0.21	–1.96	–0.04	–1.97	–0.03
24	7.50	0.99	5.54×10^{-3}	9.87	0.55	–1.52	–0.38	–1.96	0.06	–1.84	–0.06
25	7.19	1.31	7.10×10^{-3}	38.21	–0.22	–1.80	–0.11	–1.22	–0.69	–1.80	–0.11
26	7.09	1.32	8.60×10^{-3}	8.78	0.20	–1.89	0.71	–1.48	0.30	–1.20	0.02
27	7.64	1.46	5.06×10^{-3}	8.98	–0.60	–1.40	0.81	–0.78	0.19	–0.64	0.05
28	7.71	1.02	3.39×10^{-3}	12.06	0.32	–1.34	–0.51	–1.88	0.03	–1.91	0.06
29	7.66	1.78	5.25×10^{-3}	12.04	0.32	–1.37	–0.54	–1.64	–0.27	–1.76	–0.15
30	7.45	1.95	3.67×10^{-3}	10.35	–0.10	–1.57	0.00	–1.63	0.06	–1.76	0.19
31	7.47	2.25	1.43×10^{-2}	10.04	1.89	–1.54	–0.46	–2.02	0.02	–1.95	–0.05
32	8.47	–1.24	4.11×10^{-4}	9.15	–0.43	–0.65	0.01	–0.97	0.33	–0.46	–0.18
33	7.67	0.31	2.67×10^{-3}	28.65	0.04	–1.37	–0.63	–1.75	–0.25	–2.08	0.08
34	8.02	2.92	8.83×10^{-3}	11.58	0.66	–1.05	–0.27	–1.17	–0.15	–1.44	0.12

IHDrFHt, lmPrVQt, IIMdPQg, iNMMkQg, lmPrsCg – the names of molecular descriptors used by the models
$\hat{Y}_{Eq(2)}$, $\hat{Y}_{Eq(3)}$, and $\hat{Y}_{Eq(4)}$ – estimated activity by MDF-SAR model with 1, 3 and 5 descriptors
$R_{Eq(2)}$, $R_{Eq(3)}$, $R_{Eq(4)}$ – residuals for model with 1, 3 and 5 descriptors – see Table II

of the molecular descriptors used by Eq(3) as predictors. A better result is obtained by Eq(4), where 98% of growth inhibition activity of studied taxoids can be explained by its linear relationship with the variation of molecular descriptors of the model.

The analysis of the internal prediction ability of the MDF-SAR models on leave-one-out analysis allows one to calculate the predictive power of the MDF-SAR models. Good predictive power of MDF-SAR models are obtained by Eq(3) and Eq(4), the values of r^2_{loo} being greater than or equal to 0.93 (Table III). The small

difference between squared correlation coefficient and squared cross-validation leave-one-out coefficient of 0.01 (obtained for Eq(3) and Eq(4)) sustained the stability of the multivariate MDF-SAR models.

The correlated correlation analysis revealed that the models with three and five descriptors obtained statistically significantly higher correlation coefficient compared with the models with one descriptor ($p<0.0001$). The models with five descriptors obtained a correlation coefficient statistically significantly higher than the model with three descriptors ($p<0.003$).

The robustness and predictivity assessment [40, 41] of the model with five descriptors (Eq(4)) showed that the model is stable and valid: the intercept and coefficients of the model obtained in training and test set analysis (Eq(5)) fell within the confidence intervals of the intercept and coefficients of the model from Eq(4) (Table III). Moreover, the correlation coefficients in training and test sets are within the 95% confidence interval of the model from Eq(4) (Figure 3, Table III). The MDF-SAR model with five descriptors is satisfactory and stable in training versus test analysis, proving the model's robustness.

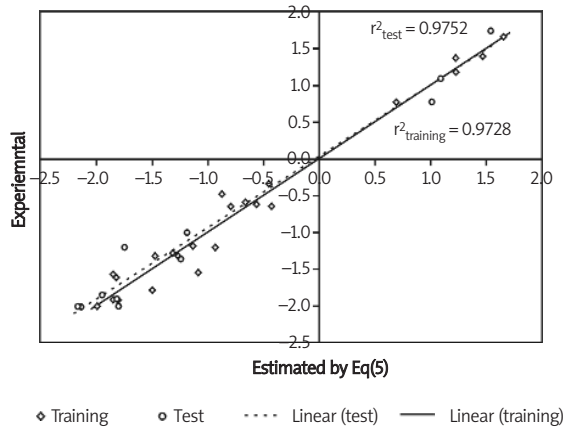

Figure 2. Estimated in training and test sets ($n_{training}$ =2/3n) versus experimental growth inhibition activity

Taking into consideration statistical performances of the MDF-SAR models, it can be

Table V. Comparison between CoMFA model and MDF-SAR model with five descriptors – results

	Pearson			Semi-Q			Spearman		
	CoMFA [6]	$\hat{Y}_{Eq(4)}$	$\hat{Y}_{obs}$	CoMFA [6]	$\hat{Y}_{Eq(4)}$	$\hat{Y}_{obs}$	CoMFA [6]	$\hat{Y}_{Eq(4)}$	$\hat{Y}_{obs}$
CoMFA [6]	1			1			1		
$\hat{Y}_{Eq(4)}$	0.983	1		0.985	1		0.986	1	
$\hat{Y}_{obs}$	0.989	0.988	1	0.989	0.990	1	0.990	0.992	1

CoMFA [6] – growth inhibition activity estimated by CoMFA model [6]
$\hat{Y}_{Eq(4)}$ – growth inhibition activity estimated by MDF-SAR model with five descriptors
$\hat{Y}_{obs}$ – growth inhibition activity measured experimentally [6]
Pearson – Pearson correlation coefficient
Semi-Q – semi-quantitative correlation coefficient [35]
Spearman – Spearman correlation coefficient

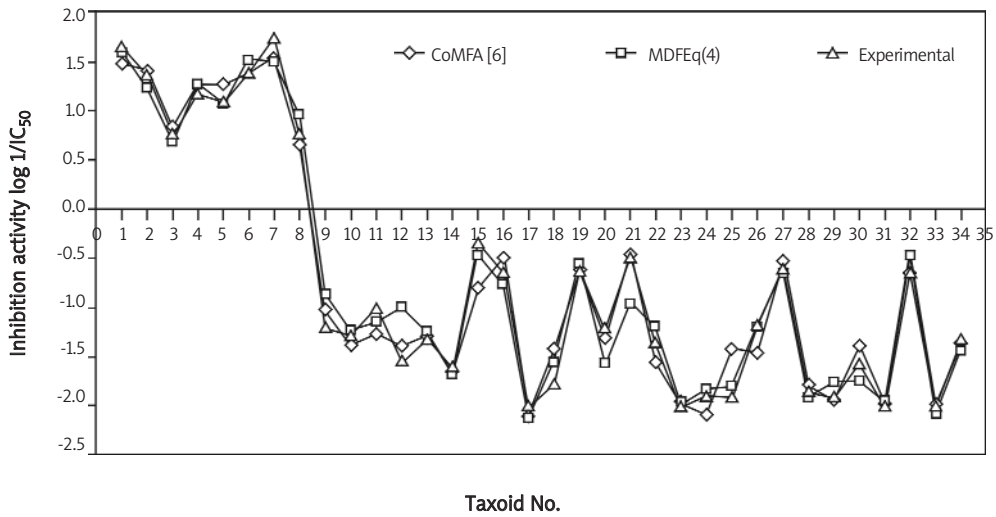

Figure 3. Comparison of CoMFA and MDF-SAR model with five descriptors

concluded that the model with five descriptors is a better model than the models with three and one descriptors, respectively.

The comparison of the model with five descriptors with the previously reported model [6] revealed that their abilities are similar (Table V). Analyzing the semi-quantitative and Spearman correlation coefficients, it can be observed that the model from Eq(4) obtained slightly better results in terms of squared correlation coefficient. The absence of statistically significant differences between the CoMFA [6] and MDF-SAR model with five descriptors is seen also in the graphical representation presented in Figure 3. The difference between these models consists of the modality of descriptors generation and calculation and the approach used.

The MDF-SAR model with five descriptors has been included in the MDF-SAR library and could be used to predict the growth inhibition activity of other taxoids [42]. The activity of new taxoids can be obtained by using a virtual environment free of experimental accidents and measurements errors, opening a new pathway in activity characterization of compounds. This environment has a real potential of clinical applications in the first step of knowledge translation (generation of evidence from research [43]) in the design of new drugs with higher curative and lower adverse effects. Any researcher could freely use the predictive environment by drawing the compound as a *.hin file.

Further research will focus on external validation of the model with five descriptors, through assessment of taxoids not included in the process of model development.

In conclusion three molecular descriptors family on structure-activity relationships models, one with one descriptor and the others with three and five descriptors, with good statistical characteristics were obtained. The MDF-SAR model with five descriptors obtained a correlation coefficient significantly greater than the other MDF-SAR models. According to the MDF-SAR model with five descriptors, the growth inhibition activity of studied taxoids is of geometric and topological nature, being related to partial charges of compounds, number of directly bonded hydrogens and cardinality.

Even if the correlation coefficient obtained by the MDF-SAR model with five descriptors is similar to the correlation coefficient obtained by the previously reported model, the applied validation methods demonstrate its stability and reliability.

Acknowledgments

This research was supported by UEFISCSU Romania through grants (ID_458 and ID_1051).

References

1. Khosroushahi AY, Valizadeh M, Ghasempour A, et al. Improved Taxol production by combination of inducing factors in suspension cell culture of Taxus baccata. Cell Biol Int 2006; 30: 262-9.
2. Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 1971; 93: 2325-7.
3. Zamir LO, Zhang J, Wu J, Sauriol F, Mamer O. Five novel taxanes from taxus canadensis. J Nat Prod 1999; 62: 1268-73.
4. Yuangang Z, Yujie F, Shuangming L, Rui S, Qingyong L, Gunter S. Rapid separation of four main taxoids in Taxus species by a combined LLP-SPE-HPLC (PAD) procedure. J Sep Sci 2006; 29: 1237-44.
5. Naill MC, Roberts SC. Culture of isolated single cells from taxus suspensions for the propagation of superior cell populations. Biotechnol Lett 2005; 27: 1725-30.
6. Morita H, Gonda A, Wei L, Takeya K, Itokawa H. 3D QSAR analysis of taxoids from Taxus cuspidate var. nana by comparative molecular field approach. Bioorg Med Chem Lett 1997; 7: 2387-92.
7. Rao KV, Bnakuni RS, Juchum J, Davies RM. A Large scale process for paclitaxel and other taxanes from the needles of taxus x media hicksii and taxus floridana using reverse phase column chromatography. J Liq Chromatogr Relat Technol 1996; 19: 427-47.
8. Shen YC, Lin YS, Hsu SM, et al. Tasumatrols P-T, five new taxoids from Taxus sumatrana. Helv Chim Acta 2007; 90: 1319-29.
9. Joshi BS, Roy R, Chattopadhyay SK, Madhusudanan KP. An NMR and LC-MS based approach for mixture analysis involving taxoid molecules from Taxus wallichiana. J Mol Struct 2003; 645: 235-48.
10. Schiff PB, Fant J, Horwitz SB. Promotion of microtubule assembly in vitro by taxol. Nature 1979; 277: 665-7.
11. Pineda O, Farràs J, Maccari L, Manetti F, Botta M, Vilarrasa J. Computational comparison of microtubule-stabilising agents laulimalide and peloruside with taxol and colchicines. Bioorg Med Chem Lett 2004; 14: 4825-9.
12. Schifft PB, Horwitz SB. Taxol stabilizes microtubules in mouse fibroblast cells. Proc Natl Acad Sci USA 1980; 77: 1561-5.
13. Pazdur R, Kudelka AP, Kavanagh JJ, Cohen PR, Raber MN. The taxoids: paclitaxel (Taxol®) and docetaxel (Taxotere®). Cancer Treat Rev 1993; 19: 351-86.
14. Kobayashi J, Hosoyama H, Wang XX, et al. Effects of taxoids from Taxus cuspidate on microtubule depolymerization and vincristine accumulation in MDR cells. Bioorg Med Chem Lett 1997; 7: 393-8.
15. Kingston DG, Molinero AA, Rimoldi GM. The taxane diterpendoids. In: Progress in the Chemistry of Organic Natural Products. Vol. 61. Herz W, Moore RE, Kirby GW, Steglich W, Tamm C (eds). Springer, Wien 1993; 1-206.
16. Woo DD, Miao SY, Pelayo JC, Woolf AS. Taxol inhibits progression of congenital polycystic kidney disease. Nature 1994; 368: 750-3.
17. Clamp AR, Mäenpää J, Cruickshank D, et al. SCOTROC 2B: feasibility of carboplatin followed by docetaxel or docetaxel-irinotecan as first-line therapy for ovarian cancer. Br J Cancer 2006; 94: 55-61.
18. Steger GG, Galid A, Gnant M, et al.; ABCSG-14. Pathologic complete response with six compared with three cycles of neoadjuvant epirubicin plus docetaxel and granulocyte colony-stimulating factor in operable breast cancer: results of ABCSG-14. J Clin Oncol 2007; 25: 2012-8.

19. LeCaer H, Barlesi F, Robinet G, et al. An open multicenter phase II trial of weekly docetaxel for advanced-stage non-small-cell lung cancer in elderly patients with significant comorbidity and/or poor performance status: The GFPC 02-02b study. Lung Cancer 2007; 57: 72-8.
20. Beer TM, Ryan CW, Venner PM, et al.; ASCENT Investigators. Double-blinded randomized study of high-dose calcitriol plus docetaxel compared with placebo plus docetaxel in androgen-independent prostate cancer: a report from the ASCENT Investigators. J Clin Oncol 2007; 25: 669-74.
21. Samlowski WE, Moon J, Kuebler JP, et al. Evaluation of the combination of docetaxel/carboplatin in patients with metastatic or recurrent squamous cell carcinoma of the head and neck (SCCHN): A Southwest Oncology Group phase II study. Cancer Invest 2007; 25: 182-8.
22. Vaishampayan U, Parchment RE, Jasti BR, Hussain M. Taxanes: an overview of the pharmacokinetics and pharmacodynamics. Urology 1999; 54 (Suppl 6A): 22-9.
23. Hennenfent KL, Govindan R. Novel formulations of taxanes: a review. Old wine in a new bottle? Ann Oncol 2006; 17: 735-49.
24. Guéritte F. General and recent aspects of the chemistry and structure activity relationships of taxoids. Curr Pharm Des 2001; 7: 1229-49.
25. Morita H, Gonda A, Wei L, Yamamura Y, Takeya K, Itokawa H. Taxuspinananes A and B, new taxoids from Taxus cuspidata var. nana. J Nat Prod 1997; 60: 390-2.
26. Morita H, Gonda A, Wei L, et al. Four new taxoids from Taxus cuspidata var. nana. Planta Med 1998; 64: 183-6.
27. Bolboacă SD, Jäntschi L. Molecular descriptors family on structure-activity relationships: modeling herbicidal activity of substituted triazines class. Bulletin of University of Agricultural Sciences and Veterinary Medicine. Agriculture 2006; 62: 35-40.
28. Jäntschi L, Bolboaca SD. Results from the use of molecular descriptors family on structure property/activity relationships. Int J Mol Sci 2007; 8: 189-203.
29. Bolboaca SD, Jäntschi L. Data mining on structure-activity/property relationships models. WASJ 2007; 2: 323-32.
30. Bolboacă SD, Jäntschi L. Modelling the property of compounds from structure: statistical methods for models validation. Environ Chem Lett in press.
31. Jäntschi L. Molecular descriptors family on structure activity relationships 1. Review of the methodology. Leonardo El J Pract Technol 2005; 4: 76-98.
32. ***, HyperChem, Molecular Modelling System [Internet page]; ©2003, Hypercube [about three screens]; [cited 2007 November]. Available from: URL: http://www.hyper.com/.
33. ***, Leave-one-out Analysis. ©2005, Virtual Library of Free Software [cited 2007 May]. Available from: URL: http://l.academicdirect.org/Chemistry/SARs/MDF_SARs/loo/.
34. Steiger JH. Tests for comparing elements of a correlation matrix. Psychol Bull 1980; 87: 245-51.
35. Bolboacă SD, Jäntschi L. Pearson versus Spearman, Kendall's Tau correlation analysis on structure-activity relationships of biologic active compounds. Leonardo J Sci 2006; 9: 179-200.
36. Bolboacă SD, Jäntschi L. Structure versus biological role of substituted thiadiazole- and thiadiazoline- disulfonamides. Studii si Cercetari de Biologie, Seria biologie vegetala 2007; 12: 50-6.
37. Braga SF, Galvão DS. A structure-activity study of taxol, taxotere, and derivatives using the electronic indices methodology (EIM). J Chem Inf Comput Sci 2003; 43: 699-706.
38. Ojima I, Lin S, Chakravarty S, et al. Syntheses and structure-activity relationships of novel nor-seco taxoids. J Org Chem 1998; 63: 1637-45.
39. Fang WS, Liang XT. Recent progress in structure activity relationship and mechanistic studies of taxol analogues. Mini Rev Med Chem 2005; 5: 1-12.
40. Shao J. Linear model selection by cross-validation. J Am Stat Assoc 1993; 88: 486-94.
41. Golbraikh A, Shen M, Xiao Z, Xiao YD, Lee KH, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 2003; 17: 241-253.
42. ***, MDF Predictor, ©2005, Virtual Library of Free Software [cited 2007 October]. Available from: URL: http://l.academicdirect.org/Chemistry/SARs/MDF_SARs/sar/.
43. Lockyer J, Gondocz ST, Thivierge RL. Knowledge translation: the role and place of practice reflection. J Contin Educ Health Prof 2004; 24: 50-6.