European Journal of Medicinal Chemistry 44 (2009) 2847–2853

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Juglone derivatives as antitubercular agents: A rationale for the activity profile

Smriti Sharma, Brij K. Sharma [1], Yenamandra S. Prabhakar*

Medicinal and Process Chemistry Division, Central Drug Research Institute, Lucknow 226001, India

ARTICLE INFO

Article history:
Received 18 October 2008
Received in revised form
9 December 2008
Accepted 10 December 2008
Available online 25 December 2008

Keywords:
7-Methyljuglone
Antitubercular activity
QSAR study
Combinatorial protocol in multiple linear regression (CP-MLR)
PLS
MODEL (MOlecular DEscriptor Lab)
descriptors

ABSTRACT

The antitubercular activity and cytotoxicity of juglone derivatives were analyzed with the topological and molecular surface features from a web-based server, MODEL (MOlecular DEscriptor Lab). Analysis of the structural features in conjunction with the biological endpoints in Combinatorial Protocol in Multiple Linear Regression led to the identification of seven descriptors for modeling the activity and six descriptors for that of the toxicity of the compounds. Analysis of these descriptors in PLS highlighted their relative significance in modulating the biological response. They suggested that structures with compact molecular arrangement, partial positive surface areas and increased autocorrelation with small lag values may lead to better antitubercular activity.

1. Introduction

In the current scenario of infectious diseases, tuberculosis (TB) is a leading cause of the morbidity and mortality of a large population worldwide. According to a WHO report, by 2020 AD nearly one billion more people will become new hosts of the TB, 200 million people will get sick and 70 million will die due to the infection if proper steps are not taken to control it [1]. Also, the opportunistic TB infection due to AIDS, and emergence of drug-resistant TB strains have made the chemotherapeutic strategies increasingly intensive [2]. Traditionally, the chemotherapy of TB has relied heavily on a limited number of drugs. The frontline chemotherapeutic agents are isonicotinic acid hydrazide, rifampicin, ethambutol, streptomycin, ethionamide, pyrazinamide, fluoroquinolones etc. [3]. The urgency and complexities involved in combating TB demands a search for new chemical prototypes to control this infection.

The chemical investigation of a South African medicinal plant, *Euclea natalensis*, has led to the identification of diospyrin (Fig. 1) as active constituent against drug-resistant strains of *Mycobacterium tuberculosis* [4]. In the traditional remedies, it has been reported to be practiced against chest complaints [5]. Diospyrin is a dimer of 7-methyljuglone (Fig. 1). Interestingly, 7-methyljuglone has also been reported to be active against *M. tuberculosis* to the extent of streptomycin and ethambutol [6]. In this background, to investigate the structure–activity relationships of this altered natural product, Mahapatra and co-workers prepared several juglone derivatives (Table 1) and evaluated them against *M. tuberculosis* [7]. The quantitative structure–activity relationship (QSAR) studies offer the structural requirements of the probe molecules for an activity. They are successfully used in the development of rationales for divergent biological activities [8–12].

Earlier, we have investigated the quantitative structure–activity relationships (QSARs) in some C-3 alkyl and arylalkyl 2,3-dideoxy hex-2-enopyranosides and highly functionalized heptenol and octenol derivatives for their antitubercular activity [11,12]. These studies have indicated that the compounds reflecting some degree of symmetry, minimum eccentricity and closely placed geometric and electronegativity centers, less branched and saturated structural templates are favorable for activity [11,12]. Recently MODEL (MOlecular DEscriptor Lab), a web-based server, has been introduced to compute a large descriptor base, over 3700 descriptors, of molecules to address various aspects of their properties in QSAR/QSPR modeling studies [13]. These descriptors represent constitutional, charge, physicochemical, topological, geometrical

* Corresponding author. Tel.: +91 522 2612411; fax: +91 522 2623405.
E-mail address: yenpra@yahoo.com (Y.S. Prabhakar).
[1] On study leave from S.K. Government College, Sikar 332 001, India.

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.12.015

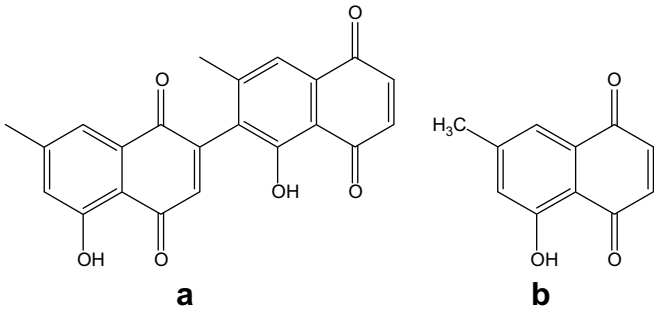


Fig. 1. Structures of (a) Diospyrin and (b) 7-Methyljuglone.

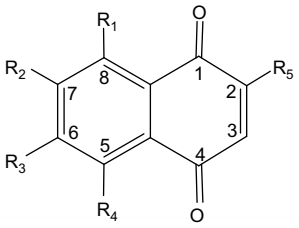
and quantum chemical features of the compounds. In this study we have used it for computing the structural features of the juglone derivatives (Table 1). The features important to the antitubercular activity of the juglone derivatives have been identified through a variable selection approach CP-MLR (Combinatorial Protocol in Multiple Linear Regression) [14]. A partial least squares (PLS) analysis has been carried out on the CP-MLR identified descriptors to derive composite model(s) highlighting their significance to the activity. The results are discussed hereunder.

2. Material and methods

2.1. Structure database

The juglone derivatives reported in the literature were taken as the dataset of this study (Table 1) [7]. For these analogues the minimum inhibitory concentrations (MIC) against *M. tuberculosis* H37Rv and cytotoxicity against the Vero cell line (inhibitory concentration to produce 50% toxicity on cell line) were reported in the literature [7]. These biological endpoints, in molar doses, have been adopted for the modeling study after their transformation in the form of logarithm of inverse of inhibitory concentration. They are referred as $-\log$ MIC for the antitubercular activity (or simply activity) and as $-\log IC_{50}$ for the toxicity against viro cell line (or simply toxicity). For the QSAR study, the structures of all the juglone derivatives (Table 1) have been generated in the structure builder module of MOE (molecular operating environment) [15]. Compound **6**, being the most active one, has been considered as structural template for rest of the analogues (Table 1). The minimum energy conformation of compound **6** has been identified through its systematic conformational analysis in MOE with default parameter setting followed by energy minimization (forcefield, MMFF94; rms gradient = 0.001). It was further refined in AM1 method (rmsg = 0.001) to obtain the final conformation of

Table 1
Juglone derivatives and their observed and predicted biological activities.



No.	R_1	R_2	R_3	R_4	R_5	$-\log$ MIC						$-\log IC_{50}$				
						Obs[a]	Predicted					Obs[a]	Predicted			
							Eq. (1)	Eq. (2)	Eq. (3)	Eq. (4)	pls1[b]		Eq. (6)	Eq. (7)	Eq. (8)	pls2[b]
1	H	H	H	OH	H	5.24	5.07	5.05	5.09	5.14	5.16	5.16	5.07	5.10	5.09	5.18
2	F	Me	H	OH	H	4.31	4.38	4.26	4.40	4.67	4.49	4.43	4.61	4.70	4.72	4.59
3[c]	Cl	Me	H	OH	H	4.35	4.42	4.41	4.44	4.29	4.36	4.95	4.95	5.03	5.08	4.91
4	Br	Me	H	OH	H	4.73	4.76	4.80	4.76	4.72	4.76	4.87	4.61	4.97	4.89	4.79
5	Cl	H	Me	OH	H	4.65	4.36	4.57	4.35	4.57	4.49	4.68	4.57	4.63	4.82	4.77
6[c]	H	Me	H	OH	H	5.58	4.92	4.89	4.90	4.99	4.97	4.10	4.40	4.29	4.43	4.43
7	H	H	Me	OH	H	4.58	4.50	4.41	4.49	4.42	4.44	4.78	4.83	4.59	4.65	4.63
8	H	Me	H	OAc	H	4.96	5.00	4.81	4.93	4.93	4.94	4.58	4.61	4.52	4.43	4.59
9	Cl	Me	H	OAc	H	4.42	4.54	4.47	4.61	4.28	4.46	4.62	4.63	4.69	4.60	4.70
10[c]	Cl	H	Me	OAc	H	4.42	4.50	4.53	4.48	4.28	4.42	4.10	4.27	4.23	4.24	4.44
11	H	Me	H	OMe	H	4.13	4.23	4.17	4.14	4.13	4.08	4.54	4.76	4.62	4.58	4.51
12	Cl	Me	H	OMe	H	<4.07	3.89	3.77	3.82	3.97	3.80	4.49	4.69	4.56	4.52	4.58
13[c]	Cl	H	Me	OMe	H	4.20	4.13	4.18	4.03	4.27	4.12	4.47	4.36	4.34	4.36	4.51
14	H	Me	H	OEt	H	4.16	4.25	4.22	4.31	4.35	4.29	4.54	4.52	4.42	4.38	4.42
15	Cl	Me	H	OEt	H	<4.10	4.11	4.01	4.15	4.01	4.03	3.84	4.16	4.20	4.21	4.28
16	Cl	H	Me	OEt	H	4.22	4.11	4.17	4.12	4.19	4.14	4.30	4.07	3.90	3.97	4.23
17[d]	H	Me	H	OAc	OAc	<4.27	4.70	4.58	4.68	4.84	4.73	4.34	4.10	4.15	4.11	4.00
18[d]	Cl	Me	H	OAc	OAc	<4.31	4.46	4.52	4.58	4.45	4.52	4.17	3.97	4.17	4.15	3.96
19[d]	Cl	H	Me	OAc	OAc	<4.31	4.32	4.43	4.34	4.27	4.32	3.44	3.73	3.73	3.75	3.72
20	H	Me	H	OH	OH	4.61	4.66	4.66	4.66	4.59	4.61	4.82	4.64	4.63	4.66	4.57
21	H	H	H	H	Me	4.54	4.69	4.79	4.68	4.55	4.68	4.42	4.51	4.38	4.53	4.50
22[c]	H	Me	H	OH	nq1[e]	4.67	4.84	4.75	4.65	4.59	4.62	4.32	4.38	4.50	4.31	4.28
23	H	OH	H	Me	nq2[e]	4.57	4.57	4.74	4.60	4.59	4.59	4.07	4.03	4.13	4.05	4.07

[a] Ref. [7].
[b] Predictions of pls1 and pls2 are from PLS models shown in Tables 3 and 5, respectively.
[c] Test set compound.
[d] In compounds **17**–**19**, [1,4]-diacetoxy napthene is in place of [1,4]-napthoquinone.
[e] nq1 is 5-OH,7-Me-[1,4]naphthoquinon-6-yl; nq2 is 5-OH,7-Me-[1,4] naphthoquinon-8-yl.

compound **6**. The conformations of all other compounds (Table 1) were generated from this template of compound **6**. For this, the template has been appropriately modified using the MOE standard fragment library to result in the conformers of remaining compounds (Table 1). These were energy minimized (AM1; rmsg = 0.001) to result in the conformers used in the study. For the purpose of QSAR study, all active analogues have been randomly divided into training and test sets. Out of the active analogues, five compounds have been placed in the test set for the validation of models (Table 1).

2.2. Descriptors

In MODEL software [13] the conformers of juglone derivatives (Table 1) have resulted in 2882 nonzero 2D- and 3D-descriptors out of the 3778 possible ones to parameterize these compounds (Table 1). A large proportion of these descriptors are due to quantum chemical and spectral characterization of the structures. They include the EVA (EigenVAlues) descriptors calculated from the normal modes of molecular moment in the vibration frequency range of 1–4000 cm^{-1}. In MODEL, EVA descriptors are computed at vibration frequency increments of 2 cm^{-1} thus making them the single largest group among the total indices. The other descriptors are due to constitutional properties, electronic charges, physicochemical, topological and geometrical properties. The QSAR model generation and validation have been done using the combinatorial protocol in multiple linear regression (CP-MLR) [14] and partial least squares (PLS) analysis [16,17]. Prior to the application of CP-MLR procedure, all those descriptors which are intercorrelated beyond 0.9 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This has reduced the total dataset of the compounds to 322 descriptors as relevant ones for the antitubercular activity and 300 descriptors as relevant ones for the toxicity. As the number of descriptors involved in this study is still very large, only those ones identified in the models have been addressed in the discussion. The model development is briefly described in the following section.

2.3. QSAR procedure

The CP-MLR is a 'filter' based variable selection procedure for model development [14]. Its procedural aspects are discussed in some of the recent publications [11,12]. The thrust of this procedure is in its embedded 'filters'. They are briefly as follows: filter-1 seeds the variables by way of limiting inter-parameter correlations to predefined level (upper limit ≤ 0.79); filter-2 controls the variables entry to a regression equation through t-values of coefficients (threshold value ≥ 2.0); filter-3 provides comparability of equations with different number of variables in terms of square-root of adjusted multiple correlation coefficient of regression equation, r-bar; filter-4 estimates the consistency of the equation in terms of cross-validated R^2 or Q^2 with leave-one-out (LOO) cross-validation as default option (threshold value $0.3 \leq Q^2 \leq 1.0$).

As the dataset has limited number of compounds, the CP-MLR search for models has been limited to three descriptors. An initial evaluation of the dataset suggested that there are few two-parameter models to explain the activity of training and test set compounds. Hence, the search was carried out for 3-parameter models. For this the filters – 1, 2, 3 and 4 in CP-MLR were assigned as 0.79, 2.0, 0.85 and $0.3 \leq Q^2 \leq 1.0$, respectively. All identified equations were evaluated for the goodness of fit using the test compounds. Among these, the models showing test set r^2 value beyond 0.60 were reassessed for the chance correlations through one hundred simulation runs with the randomized biological response [18,19]. The descriptors of the CP-MLR models were reused in the PLS procedure to derive composite single-window MLR-like PLS models comprising all identified descriptors.

3. Result and discussion

The antitubercular activity of juglone derivatives (Table 1) has been analyzed with the descriptors from MODEL for 3-parameter equations. In CP-MLR, they have resulted in the following rationales for modeling the activity.

$$\begin{aligned} -\log \mathrm{MIC} &= -0.514 + 4.268(0.967)\mathrm{MATS2r} \\ &\quad + 64.723(9.853)\mathrm{FPSA_3} + 2.055(0.769)\mathrm{RDF_{gq}(3.0)} \\ n &= 13,\ r^2 = 0.835,\ Q^2 = 0.667, \\ Q^2_{\mathrm{L3O}} &= 0.628,\ s = 0.148,\ F = 15.30 \\ r^2_{\mathrm{t}} &= 0.623,\ r^2_{\mathrm{randY}}(\mathrm{s.d.}) = 0.211(0.149) \end{aligned} \quad (1)$$

$$\begin{aligned} -\log \mathrm{MIC} &= -1.383 + 3.933(0.961)\mathrm{MATS2r} \\ &\quad + 0.761(0.308)\mathrm{GATS4m} + 62.380(9.978)\mathrm{FPSA_3} \\ n &= 13, r^2 = 0.824, Q^2 = 0.550, \\ Q^2_{(\mathrm{L3O})} &= 0.624, s = 0.153, F = 14.11 \\ r^2_{\mathrm{t}} &= 0.607, r^2_{\mathrm{randY}}(\mathrm{s.d.}) = 0.221(0.175) \end{aligned} \quad (2)$$

$$\begin{aligned} -\log \mathrm{MIC} &= 0.002 + 3.229(0.946)\mathrm{MATS2r} \\ &\quad + 60.449(10.076)\mathrm{FPSA_3} - 0.104(0.045)\mathrm{RDF_{gu}(2.0)} \\ n &= 13, r^2 = 0.815, Q^2 = 0.576, Q^2_{\mathrm{L3O}} = 0.540, \\ s &= 0.157, F = 13.32\, r^2_{\mathrm{t}} = 0.605, \\ r^2_{\mathrm{randY}}(\mathrm{s.d.}) &= 0.212(0.163) \end{aligned} \quad (3)$$

$$\begin{aligned} -\log \mathrm{MIC} &= 2.919 - 0.071(0.019)\mathrm{MSWHIM\,\widehat{E}3,\text{-}(SAS)} \\ &\quad + 42.565(9.525)\mathrm{FPSA_3} - 0.794(0.226)\mathrm{Mor18u} \\ n &= 13,\ r^2 = 0.808,\ Q^2 = 0.663,\ Q^2_{\mathrm{L3O}} = 0.687, \\ s &= 0.160,\ F = 12.69\, r^2_{\mathrm{t}} = 0.694, \\ r^2_{\mathrm{randY}}(\mathrm{s.d.}) &= 0.220(0.188) \end{aligned} \quad (4)$$

In these and all other regression equations, n is the number of compounds, r^2 is the squared multiple correlation coefficient of regression, Q^2 is cross-validated R^2 from leave-one-out (LOO) procedure, Q^2_{L3O} is cross-validated R^2 from leave 3 compounds out (leave-many-out) procedure (where a group of compounds are randomly kept out side the analysis each time in such a way that all compounds are in the predictive groups for once), s is the standard error of the estimate, F is the F ratio between the variances of calculated and observed activities and r^2_{t} is the test set r^2 value. The r^2_{yrand}(s.d.) is the mean random squared multiple correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values given in the parentheses (in regression equation) are the standard errors of the regression coefficients. It may be mentioned here that compounds **12**, **15**, and **17**–**19** of Table 1 are associated with uncertain activity values (in terms of −log MIC, the active of these compounds are in the order of 4.30–4.00 or less). It makes the total compounds in Table 1 with the antitubercular activity as eighteen, which include five test set compounds. The models 1–4 are validated with an external test set of five compounds (Table 1). The predictions of the test set compounds (Table 1) are found to be satisfactory as reflected in the test set r^2 values (Eqs. (1)–(4)). Also, the activity predictions for the compounds

Table 2
Information content of the descriptors participated in Eqs. (1)–(4).

S. No.	Descriptor	Descriptor class[a]	Information content[b]
1	MATS2r	2D-AUTO	Moran autocorrelation of lag 2 weighted by VDW radius
2	GATS4m	2D-AUTO	Geary autocorrelation of lag 4 weighted by atomic mass (m)
3	MS-WHIM¦Ê$_3$,-(SAS)	MS-WHIM	Negatively charged molecular electrostatic potential weighted 3rd Cartesian coordinate molecular surface WHIM kurtosis over solvent-accessible surface (SAS)(×1.0e−4).
4	FPSA$_3$	CPSA	Fractional charged partial positive surface area 3 is ratio of the atomic charge weighted positive surface area (PPSA$_3$) and the total solvent-accessible surface (SASA). In this, PPSA$_3$ is the sum of the products of atomic solvent-accessible surface areas and partial charge q_{a+} over all positively charged atoms
5	RDFgu(2.0)	RDF	Un-weighted (u) distance based (2.0 Å) descriptor from radial distribution function.
6	RDFgq(3.0)	RDF	Atomic charge (q) weighted distance based (3.0 Å) descriptor from radial distribution function.
7	Mor18u	3D-MoRSE	Un-weighted (u) 3D-MoRSE signal for scanning distance of 18 Å.

[a] 2D-AUTO: 2D-autocorrelations; MS-WHIM: molecular surface Weighted Holistic Invariant Molecular descriptor; CPSA: charged partial surface area descriptor; RDF: Radial Distribution Function descriptor; 3D-MoRSE: 3D-Molecule Representation of Structures based on Electron diffraction descriptor.
[b] see Ref. [13].

with uncertain activity (compounds **12**, **15**, and **17**–**19**) are found to be with in the reasonable limits of their likely values (Table 1). However, the activity of compound **17** is predicted marginally higher than its likely value.

Eqs. (1)–(4) have shared 7 descriptors among them. The physical meaning of these descriptors is presented in Table 2. In Eqs. (1)–(4), while the regression coefficients of MATS2r, GATS4m, FPSA$_3$ and RDFgq(3.0) have shown positive influence on the activity, the regression coefficients of MS-WHIM¦Ê$_3$,-(SAS), RDF$_{gu}$(2.0) and Mor18u have shown negative influence on the same. A PLS analysis has been carried out on these descriptors to facilitate the development of single-window structure–activity model comprising all of them. It also gives an opportunity to make a comparison of relative significance among the descriptors. The descriptors were autoscaled (zero mean and unit s.d.) to give each one of them equal weight in the PLS analysis. In the cross-validation procedure of the PLS analysis [16,17] of these descriptors, two components were found to be the optimum to explain the activity. The PLS model has explained 89.1 per cent variance ($r^2 = 0.891$, $s = 0.114$, $F = 41.02$) in the activity of the training set compounds and 70.1 per cent variance in the activity of test set compounds ($r^2_t = 0.701$). Fig. 2 shows the plot of observed versus predicted activities of the training and test set compounds from the PLS analysis. The regression coefficients of MLR-like PLS equation from the two-component model are shown in Table 3. Except for RDFgq(3.0), the signs of the regression coefficients of all descriptors in PLS model and Eqs. (1)–(4) are in agreement with each other. In case of RDF$_{gq}$(3.0), the sign of its regression coefficient has become negative in PLS model. Also, the fraction contributions of the MLR-like PLS coefficients of the descriptors suggest that RDF$_{gq}$(3.0) is the least contributing one to the activity. Among the identified descriptors, FPSA$_3$ has significantly influenced (39.1 per cent) to the activity followed by MATS2r and MS-WHIM¦Ê$_3$,-(SAS) (each one about 15 per cent). The plots of the fraction contributions of the individual descriptors to the inhibitory activity are shown in Fig. 3. FPSA$_3$ measures fraction of partial positive surface area of the molecule. Its positive regression coefficient suggests its favorability for the activity. The regression coefficient of MS-WHIM¦Ê$_3$,-(SAS) is associated with negative sign. It hints that negatively charged molecular surface is not favorable for the activity. In these molecules, the quinone moiety and parts of hydroxyl and other electron rich groups contribute to the negatively charged surface area. Minimizing the electron rich substituent groups or modifying their nature towards the electropositive character may favor the activity. It will lead to a favorable condition for FPSA$_3$ as well as MS-WHIM¦Ê$_3$,-(SAS). The regression coefficient of MATS2r indicates the positive influence of Moran autocorrelation of lag 2 weighted by VDW radius on the activity.

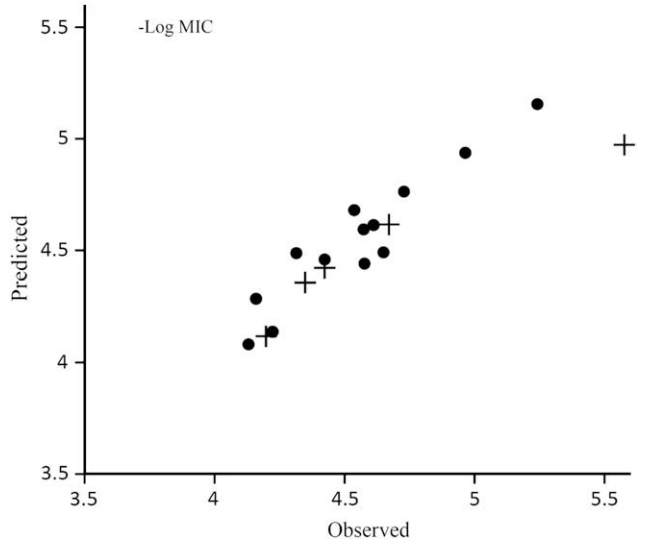


Fig. 2. The plot of observed versus PLS predicted (Table 3) antitubercular activity (−log MIC) of juglone derivatives. The training and test set compounds are shown by (●) and (+), respectively.

Table 3
MLR-like PLS model for the antitubercular activity of juglone derivatives (Table 1) from the 7 descriptors of Eqs. (1)–(4).

S. No.	MLR-like PLS equation	−log MIC
	Descriptor	MLR-like coeff. (f.c.)[a]
1	MATS2r	2.178(0.157)
2	GATS4m	0.205(0.043)
3	MS-WHIM¦Ê$_3$,-(SAS)	−0.033(−0.155)
4	FPSA$_3$	57.028(0.391)
5	RDFgu(2.0)	−0.068(−0.097)
6	RDFgq(3.0)	−0.216(−0.018)
7	Mor18u	−0.353(−0.138)
	Constant	0.705
Regression statistics		
n		13
r^2		0.891
Q^2		0.835
Q^2_{L3O}		0.816
s		0.114
F		41.02
External test set		
r^2_t[b]		0.701

[a] Coefficients of MLR-like PLS equation in terms of descriptors for their original values; f.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.
[b] For 5 test set compounds (Table 1).

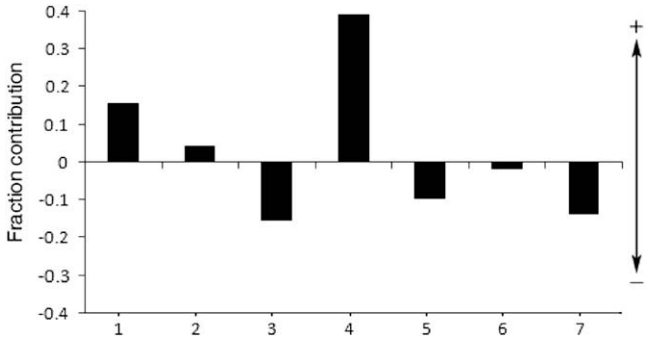


Fig. 3. Plots of fraction contribution of MLR-like PLS coefficients (normalized) (Table 3) of 7 descriptors (Eqs. (1)–(4)) to the antitubercular activity (−log MIC) of juglone derivatives. The horizontal axis refers to the descriptors' numbers as shown in Table 3.

For the juglone derivatives (Table 1) the cytotoxicity against the Vero cell line is also reported in the literature [7]. As the cytotoxicity data is available for all compounds of Table 1, none were omitted from the study. The antitubercular activity and cytotoxicity of the juglone derivatives (Table 1) are poorly correlated ($n = 18$; $r = 0.072$). However, the seven descriptors identified for the activity are investigated for their prospects in explaining the toxicity of these analogues. It resulted in the following equation.

$$\begin{aligned} -\log \mathrm{IC}_{50} &= 3.408 + 23.399(11.169)\mathrm{FPSA}_3 \\ &\quad - 0.814(0.208)\mathrm{Mor18u} \quad n = 18, \\ & r^2 = 0.661,\ Q^2 = 0.479,\ Q^2_{(\mathrm{L3O})} = 0.541, \\ & s = 0.247,\ F = 14.67\ r^2_{\mathrm{t}} = -0.829, \\ & r^2_{\mathrm{randY}}(\mathrm{s.d.}) = 0.108(0.145) \end{aligned} \tag{5}$$

In terms of participating descriptors, Eq. (5) is a subset of Eq. (4) identified for the activity of the compounds. Even though Eq. (5) explained the toxicity of the training set compounds satisfactorily, it has failed to account for the toxicity of test set compounds. The lack of correlation between the activity and cytotoxicity of the compounds ($n = 18$; $r = 0.072$) suggests that the chemical features contributing to these biological responses could be different. This offers a possibility of their modulation for optimum antitubercular activity. In light of this, a search of the descriptors with toxicity as dependent variable resulted in the following models.

$$\begin{aligned} -\log \mathrm{IC}_{50} &= 4.208 - 0.022(0.005)\mathrm{MSWHIM¦\tilde{A}1,\text{-}(VDW)} \\ &\quad + 1.603(0.733)\mathrm{E1s} + 1.292(0.364)\mathrm{Mor22m} \\ & n = 18,\ r^2 = 0.781,\ Q^2 = 0.576, \\ & Q^2_{(\mathrm{L3O})} = 0.580,\ s = 0.180,\ F = 16.75 \\ & r^2_{\mathrm{t}} = 0.735,\ r^2_{\mathrm{randY}}(\mathrm{s.d.}) = 0.133(0.142) \end{aligned} \tag{6}$$

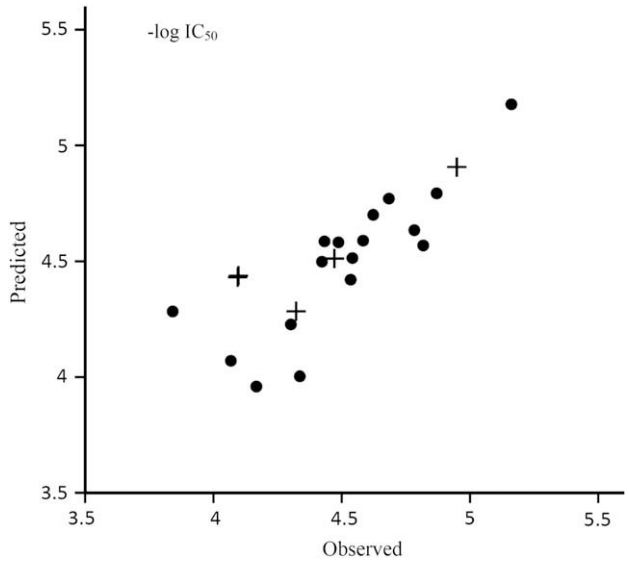


Fig. 4. The plot of observed versus PLS predicted (Table 5) cytotoxicity (−log IC_{50}) of juglone derivatives. The training and test set compounds are shown by (●) and (+), respectively.

$$\begin{aligned} -\log \mathrm{IC}_{50} &= 4.249 - 0.032(0.005)\mathrm{MSWHIM¦\tilde{A}1,\text{-}(VDW)} \\ &\quad - 3.191(1.295)\mathrm{Mor21v} + 5.242(1.826)\mathrm{HATS1m} \\ & n = 18,\ r^2 = 0.776,\ Q^2 = 0.586,\ Q^2_{(\mathrm{L3O})} = 0.599, \\ & s = 0.208,\ F = 16.21\ r^2_{\mathrm{t}} = 0.788, \\ & r^2_{\mathrm{randY}}(\mathrm{s.d.}) = 0.142(0.156) \end{aligned} \tag{7}$$

$$\begin{aligned} -\log \mathrm{IC}_{50} &= 3.990 - 0.024(0.006)\mathrm{MSWHIM¦\tilde{A}1,\text{-}(VDW)} \\ &\quad - 2.879(1.284)\mathrm{Mor21v} + 5.204(1.830)\mathrm{HATS0v} \\ & n = 18,\ r^2 = 0.774,\ Q^2 = 0.570,\ Q^2_{(\mathrm{L3O})} = 0.613, \\ & s = 0.209,\ F = 16.06\ r^2_{\mathrm{t}} = 0.690, \\ & r^2_{\mathrm{randY}}(\mathrm{s.d.}) = 0.155(0.133) \end{aligned} \tag{8}$$

The models 6–8 are validated with the same external test set of five compounds (Table 1). They have shown test set r^2 values in the order of 0.70. Together, the models (Eqs. (6)–(8)) have shared six descriptors among them. All these descriptors are different from those identified for the activity. Table 4 gives a brief physical meaning of the participating descriptors. In these equations, the regression coefficients of E1s, Mor22m, HATS1m and HATS0v have positive signs. They suggest that in the compounds decreasing these features would lead to reduced toxicity. The regression coefficients of remaining two descriptors, MS-WHIM¦$\tilde{A}_1$,-(VDW) and Mor21v, are associated with negative signs. They infer that in

Table 4
Information content of the descriptors participated in Eqs. (6)–(8).

S.No	Descriptor	Descriptor class[a]	Information content[b]
8	MS-WHIM¦$\tilde{A}_1$,-(VDW)	MS-WHIM	Negatively charged molecular electrostatic potential weighted 1st Cartesian coordinate molecular surface WHIM skewness over VDW surface.
9	E1s	WHIM	1st directional WHIM density weighted by atomic electrotopological states
10	Mor22m	3D-MoRSE	Atomic mass (m) weighted 3D-MoRSE signal for scanning distance of 22 Å.
11	Mor21v	3D-MoRSE	Atomic VDW volume (v) weighted 3D-MoRSE signal for scanning distance of 21 Å.
12	HATS1m	GETAWAY	Leverage-weighted autocorrelation of lag 1/weighted by atomic masses
13	HATS0v	GETAWAY	Leverage-weighted autocorrelation of lag 0/weighted by atomic van der Waals volumes

[a] WHIM: Weighted Holistic Invariant Molecular descriptor; GETAWAY: GEometry, Topology, and Atom-Weights AssemblY descriptors; see footnote a of Table 2.
[b] see Ref. [13].

Table 5
MLR-like PLS model for the cytotoxicity of juglone derivatives (Table 1) from the 7 descriptors of Eqs. (6)–(8).

S. No.	MLR-like PLS equation	−log IC_{50}
	Descriptor	MLR-like coeff. (f.c.)[a]
8	MS-WHIM¦$\tilde{A}_1$,-(VDW)	−0.016(−0.333)
9	E1s	0.568(0.085)
10	Mor22m	0.884(0.246)
11	Mor21v	−0.561(−0.048)
12	HATS1m	1.345(0.066)
13	HATS0v	3.699(0.222)
	Constant	3.938
Regression statistics		
n		18
r^2		0.781
Q^2		0.647
Q^2_{L3O}		0.631
s		0.199
F		26.71
External test set		
r^2_t[b]		0.548

[a] See footnote of Table 3.
[b] See footnote of Table 3.

the compounds increasing these features would reduce the toxicity. Unlike the seven descriptors identified for modeling the activity (Table 2), these six descriptors involved in the toxicity models did not lead to any kind of equation for the antitubercular activity.

Similar to the case of activity, a two-component PLS model has been developed from the six descriptors identified for the toxicity. It has explained 78.1 per cent variance ($r^2 = 0.781$, $r^2_t = 0.548$, $s = 0.199$, $F = 26.71$) in the training set toxicity. The plot of observed versus predicted toxicities of the training and test set compounds from the PLS analysis is shown in Fig. 4. Table 5 shows the MLR-like PLS equation for the toxicity of the compounds. The signs of the regression coefficients of all descriptors in PLS model and Eq. (6)–(8) are in agreement with each other. The fraction contributions of the individual descriptors to the toxicity of the compounds are shown in Fig. 5. Among these six descriptors, MS-WHIM¦$\tilde{A}_1$,-(VDW) has significantly influenced (33.3 per cent) the activity followed by Mor22m (24.6 per cent) and HATS0v (22.2 per cent). Among the six descriptors, Mor21v is the least contributing one to the toxicity of the compounds. MS-WHIM¦$\tilde{A}_1$,-(VDW) is negatively charged molecular electrostatic potential weighted 1st Cartesian coordinate molecular surface WHIM skewness over VDW surface. Its negative regression coefficient suggests that increasing this characteristic will decrease the toxicity. In other words, it suggests that derivatisation of juglone reduces its toxicity. The Mor22m is associated with positive regression coefficient. It indicates that a decrease in

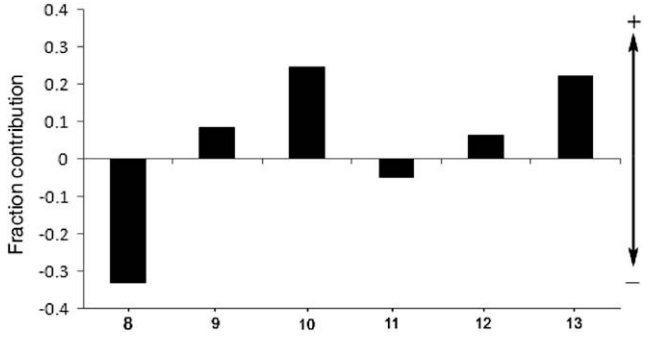


Fig. 5. Plots of fraction contribution of MLR-like PLS coefficients (normalized) (Table 5) of 6 descriptors (Eqs. (6)–(8)) to the cytotoxicity (−log IC_{50}) of juglone derivatives. The horizontal axis refers to the descriptors' numbers as shown in Table 5.

Table 6
MLR-like PLS model for the antitubercular activity and cytotoxicity of juglone derivatives (Table 1) from the 13 descriptors of Eqs. (1)–(8).

S. No.	MLR-like PLS equation	−log MIC	−log IC_{50}
	Descriptor	MLR-like coeff (f.c.)[a]	MLR-like coeff (f.c.)[a]
1	MATS2r	0.100(0.008)	−0.331(−0.042)
2	GATS4m	0.216(0.044)	−0.026(−0.008)
3	MS-WHIM¦$\hat{E}_3$,-(SAS)	−0.033(−0.160)	0.020(0.159)
4	$FPSA_3$	37.871(0.306)	−5.276(−0.070)
5	RDFgu(2.0)	−0.041(−0.060)	−0.019(−0.045)
6	RDFgq(3.0)	0.258(0.0203)	0.325(0.042)
7	Mor18u	−0.037(−0.016)	−0.143(−0.103)
8	MS-WHIM¦$\tilde{A}_1$,-(VDW)	0.001(0.021)	−0.006(−0.171)
9	E1s	−0.212(−0.024)	0.874(0.165)
10	Mor22m	−0.538(−0.093)	0.036(0.010)
11	Mor21v	0.578(0.118)	0.045(0.015)
12	HATS1m	1.440(0.064)	0.575(0.042)
13	HATS0v	1.378(0.065)	1.625(0.125)
	Constant	1.637	4.268
Regression statistics			
n		18	18
r^2		0.645	0.584
Q^2		0.531	0.431
Q^2_{L3O}		0.511	0.495
s		0.240	0.210
F		13.61	10.48

[a] See footnote of Table 3.

the corresponding 3D-MoRSE signal at scanning distance of 22 Å reduces the toxicity. It may be viewed as molecular zone enhancing the toxicity. HATS0v is a GETAWAY descriptor measuring leverage-weighted autocorrelation of lag zero (0) weighted by atomic van der Waals volumes. Its positive regression coefficient suggests that a decreasing 'leverage-weighted zero lag autocorrelation' is preferred molecular arrangement. A compact molecular arrangement may satisfy this condition.

The structural features identified for rationalizing the activity and toxicity give avenues to modulate the structure to a desirable biological end point. For this, the seven descriptors of the activity specific models (Eqs. (1)–(4)) and six descriptors of the toxicity specific models (Eqs. (6)–(8)) have been considered together for the PLS analysis. The activity and toxicity of the compounds are concurrently analyzed with all thirteen descriptors (both X and Y descriptor blocks are in multivariate mode) to identify their relative individual contributions to these biological phenomena. It has been carried out to draw a comparison among these descriptors for their contribution

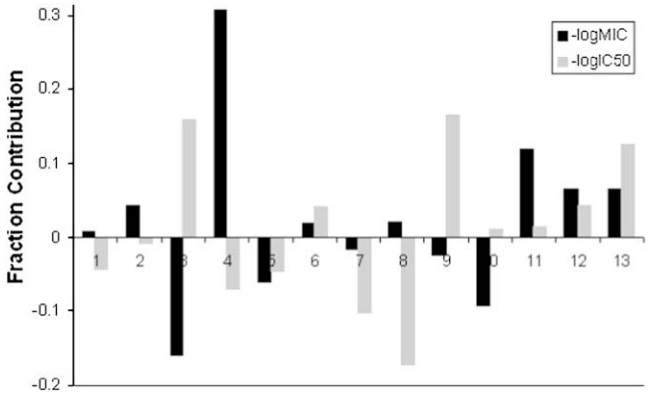


Fig. 6. Plots of fraction contribution of MLR-like PLS coefficients (normalized) (Table 6) of 13 descriptors (Eqs. (1)–(8)) to the antitubercular activity (−log MIC) and cytotoxicity (−log IC_{50}) of juglone derivatives. The horizontal axis refers to the descriptors' numbers as shown in Table 6.

to each biological phenomenon. The MLR-like PLS coefficients of the thirteen descriptors for the activity and toxicity of the compounds from the two-component PLS model are shown in Table 6. Except for a couple of descriptors, the signs of the regression coefficients are in agreement with Eqs. (1)–(8). The 2-component PLS model has explained 64.5 per cent variance ($r^2 = 0.645$, $s = 0.240$, $F = 13.61$) in the activity and 58.4 per cent variance ($r^2 = 0.584$, $s = 0.210$, $F = 10.48$) in the toxicity of the compounds. In these PLS models, the non-relevant features of one biological activity has undermined the explanatory power of the other one. In other words, the features identified for modeling the biological activities have shown high degree of association with respective phenomena. Fig. 6 shows the plots of comparative fraction contributions of the normalized regression coefficients of descriptors to the inhibitory activity and toxicity of the compounds. These fraction contributions are in general agreement with those presented in Figs. 3 and 5. The relative difference in the fraction contributions for the activity and toxicity clearly suggest that GATS4m, MS-WHIM¦Ê$_3$,-(SAS) and FPSA$_3$ are good for modulating the activity of the compounds. Similarly, MS-WHIM¦Ã$_1$,-(VDW), E1s and HATS0v are good for modulating the toxicity of the compounds. These features collectively suggest that the derivatisation of juglone with bulky groups which are not electron rich and other modifications leading to minimizing the electron rich substituent groups may favor the activity of the compounds as well as reduce their toxicity. The regression coefficients of the descriptors clearly differentiate the compounds' requirements for showing the inhibitory activity with reduced toxicity.

4. Conclusions

The QSAR rationales of antitubercular activity and cytotoxicity of juglone derivatives, an altered natural product of South African medicinal plant *E.natalensis*, have suggested the scope of structural modulation for optimum biological response. Analysis of the structural features and biological endpoints in CP-MLR led to the identification of seven descriptors for modeling the activity and six descriptors for that of the toxicity of the compounds. The relative difference in the fraction contributions of PLS regression coefficients for the biological endpoints (activity and toxicity) suggested that MATS2r, MS-WHIM¦Ê$_3$,-(SAS), FPSA$_3$ and Mor18u are good for modulating the activity of the compounds whereas MS-WHIM¦Ã$_1$,-(VDW), E1s and Mor22m are good for modulating the toxicity of the same. These comparative coefficients clearly differentiated the compounds' requirements for showing the inhibitory activity with reduced toxicity. The structures with functional groups leading to partial positive surface areas and increased autocorrelation with small lag values may satisfy the activity requirements. In terms of molecular structure of juglone derivatives, they suggest that while the derivatisation reduces the toxicity, the substituent groups with electropositive character or reduced electronegative character favor the activity.

Acknowledgements

BKS thanks INSA, New Delhi for the financial support in the form of Visiting Fellowship. CDRI Communication No.7648.

Appendix. Supplementary data

Molecular descriptors of Eqs. (1)–(8); PLS factor scores, loadings, weights and sensitivity of independent and dependent descriptors of the PLS models. Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2008.12.015.

References

[1] WHO Tuberculosis Fact Sheet.<http://www.who.int/mediacentre/factsheets/fs104/en> (2007).
[2] W.M. El-Sadr, S.J. Tsiouris, Semin. Respir. Crit. Care Med. 29 (2008) 525–531.
[3] S. Khasnobis, V.E. Escuyer, D. Chatterjee, Expert Opin. Ther. Targets 6 (2002) 21–40.
[4] N. Lall, J.J.M. Meyer, J. Ethnopharmacol. 78 (2001) 213–216.
[5] (a) N. Lall, J.J.M. Meyer, J. Ethnopharmacol. 66 (1999) 347–354;
(b) L.J. McGaw, N. Lall, T.M. Hlokwe, A.L. Michel, J.J. Meyer, J.N. Eloff, Biol. Pharm. Bull. 31 (2008) 1429–1433.
[6] N. Lall, J.J.M. Meyer, Y. Wang, N.B. Bapela, C.E.J. van Rensburg, P.B. Fourie, S.G. Franzblau, Pharm. Biol. 43 (2005) 353–357.
[7] A. Mahapatra, S.P.N. Mativandlela, B. Binneman, P.B. Fourie, C.J. Hamilton, J.J.M. Meyer, F. van der Kooy, P. Houghton, N. Lall, Bioorg. Med. Chem. 15 (2007) 7638–7646.
[8] H.K. Hyun, Y.I. Kwack, Y.D. Lee, Y.H. Park, B.S. Lee, K.K. Chan, Bull. Korean Chem. Soc. 25 (2004) 1801–1806.
[9] Shagufta, A. Kumar, G. Panda, M.I. Siddiqi, J. Mol. Model. 13 (2007) 99–109.
[10] A. Puratchikody, M. Doble, Bioorg. Med. Chem. 15 (2007) 1083–1090.
[11] M.K. Gupta, R. Sagar, A.K. Shaw, Y.S. Prabhakar, Bioorg. Med. Chem. 13 (2005) 343–351.
[12] M. Saquib, M.K. Gupta, R. Sagar, Y.S. Prabhakar, A.K. Shaw, R. Kumar, P.R. Maulik, A.N. Gaikwad, S. Sinha, A.K. Srivastava, V. Chaturvedi, R. Srivastava, B.S. Srivastava, J. Med. Chem. 50 (2007) 2942–2950.
[13] Z.R. Li, L.Y. Han, Y. Xue, C.W. Yap, H. Li, L. Jiang, Y.Z. Chen, Biotechnol. Bioeng. 97 (2007) 389–396.<http://jing.cz3.nus.edu.sg/cgi-bin/model/model.cgi>.
[14] Y.S. Prabhakar, QSAR Comb. Sci. 22 (2003) 583–595.
[15] MOE: The Molecular Operating Environment from Chemical Computing Group Inc., 1255 University St., Suite 1600, Montreal, Quebec, Canada H3B 3X3. <http://www.chemcomp.com>.
[16] (a) S. Wold, Technometrics 20 (1978) 397–405;
(b) N. Kettaneh, A. Berglund, S. Wold, Comput. Stat. Data Anal. 48 (2005) 69–85.
[17] L. Stahle, S. Wold, Multivariate Data Analysis and Experimental Design, in: G.P. Eillis, W.B. West (Eds.), Biomedical Research, Progress in Medicinal Chemistry, vol. 25, Elsevier Science Publishers, B.V., 1988.
[18] Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta, S.B. Katti, QSAR Comb. Sci. 23 (2004) 234–244.
[19] S.S. So, M. Karplus, J. Med. Chem. 40 (1997) 4347–4359.