

Comparative QSAR Analysis of 3,5-bis (Arylidene)-4-Piperidone Derivatives: the Development of Predictive Cytotoxicity Models

Najmeh Edraki^a, Umashankar Das^b, Bahram Hemateenejad^a,
Jonathan R. Dimmock^b and Ramin Miri^{a, b*}

^bMedicinal and Natural Products Chemistry Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ^bDrug Design and Discovery Research Group, College of Pharmacy and Nutrition, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan S7N 5C9, Canada.

Abstract

1-[4-(2-Alkylaminoethoxy) phenylcarbonyl]-3,5-bis(arylidene)-4-piperidones are a novel class of potent cytotoxic agents. These compounds demonstrate low micromolar to submicromolar IC₅₀ values against human Molt 4/C8 and CEM T-lymphocytes and murine leukemia L1210 cells. In this study, a comparative QSAR investigation was performed on a series of 3,5-bis (arylidene)-4-piperidones using different chemometric tools to develop the best predictive models for further development of analogs with improved cytotoxicity. All the QSAR models were validated by internal validation tests. The QSAR models obtained by GA-PLS method were considered the best as compared to MLR method. The best QSAR model obtained by GA-PLS analysis on L1210, CEM and Molt4/C8 demonstrated good predictively with R²_{pred} values ranging from 0.94-0.80. Molecular density, topological (X₂A) and geometrical indices of the molecules were found to be the most important factors for determining cytotoxic properties.

Key words: Cytotoxicity; 3,5-bis(arylidene)-4-piperidones; QSAR; multilinear regression.

Introduction

The principal aim of this laboratory is the synthesis of conjugated unsaturated ketones as candidate antineoplastic agents. These compounds interact with thiols but in general, they have little or no affinity for amino and hydroxyl groups which are found in nucleic acids (1-3). Hence thiol alkylators may not have the genotoxic properties associated with a number of contemporary anticancer drugs (4).

Initially various series of compounds possessing one conjugated enone group

were developed. However after an initial chemical insult, certain neoplasms are more vulnerable to a subsequent cytotoxic effect than various non-malignant cells (5, 6). Hence by mounting the 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore [AR-C=C-C(O)-C=C-AR] on heterocyclic and cycloaliphatic scaffolds, two sequential interactions with cellular thiols can take place which may be more detrimental to tumours than normal tissues. Such considerations led to the development of 3,5-bis(benzylidene)-4-piperidones 1a-d which demonstrated potent cytotoxic properties with the IC₅₀ values in the low micromolar range against human Molt 4/C8 and CEM T-lymphocytes as well as murine L1210 lymphocytic leukemia cells (7,

* Corresponding author:

E-mail: ramin.miri.15@gmail.com

8). The hypothesis of sequential cytotoxicity was advanced that the 1,5-diaryl-3-oxo-1,4-pentadienyl group interacts at a primary binding site and a side chain on the piperidine nitrogen may align at an auxiliary binding site which could enhance cytotoxic potencies. In order to evaluate this hypothesis, a novel series of N-aryl-3,5-bis(benzylidene)-4-piperidone derivatives 2-7 were synthesized (8). In these compounds, the side chain contains atoms and groups capable of forming hydrogen and van der Waals bonds and also an ionic bond in series 3-6 which could bind at an auxiliary binding site. When the potencies of the compounds in series 3-6 were compared with that of the analogs 1a-d which have the same aryl substituent, the IC_{50} values of the amides 3-6 were lower in 48% of the comparisons made while equipotency was noted in 35% of the cases (8). A number of these amides have submicromolar IC_{50} values.

The objectives of the present investigation are two folds. First, to gain some insight into the physicochemical properties of the 3,5-bis(benzylidene)-4-piperidone derivatives 1-7 that influence cytotoxic potencies, and second, to derive predictive 2D-QSAR models that will be helpful to design new analogs with improved bioactivity.

Experimental

Data set

The *in-vitro* cytotoxic data used in this study were taken from the literature (8). The reported IC_{50} values were converted into logarithmic scale (pIC_{50}) and then were used for QSAR analysis as dependent variables. The data set ($n = 25$) was divided randomly into two groups: calibration set ($n = 19$) and prediction set ($n = 6$). (Table 1) (Asterisk sign in the table demonstrate the prediction set).

Descriptor generation

The chemical structure of molecules was constructed using a Hyperchem package (Version 7; Hypercube Inc., Gainesville, FL, USA) and primarily were energetically minimized (100 steepest descent steps using MM⁺ force field with a gradient convergence value of 0.1 kcal/mol) using HyperChem software. The

Z-matrices of the structures were provided by the software and were then transferred to the GAUSSIAN 98 program. (9). Complete geometry optimization was performed based on most extended conformations as the starting geometries. GAUSSIAN 98 was used for semi-empirical molecular orbital calculations (AM1) of the structures.

A large number of molecular descriptors were calculated using HyperChem, GAUSSIAN 98 and Dragon Packages (10). Different quantum chemical descriptors were calculated using GAUSSIAN 98. Different functional descriptors such as topological, geometrical, and constitutional descriptors were calculated using Dragon software. Moreover, some chemical parameters including molecular volume (MV), molecular surface area (SA), hydrophobicity (logP), heat of formation and hydration energy (HE) were calculated using Hyperchem software.

Data processing and modeling

The calculated descriptors were collected in a data matrix and generated descriptors were scaled using auto scaling procedure. Firstly, the descriptors were checked for constant or near-constant values and those detected were removed from the original data matrix. Then the correlation of descriptors with each other's and with the potency data was determined. Among the detected collinear descriptors ($r^2 > 0.8$), one of them demonstrating the highest correlation with potency was retained, and the rest were omitted. In the model development step, leave-one-out cross-validation was used to optimize the models (e.g., selection of variables or no. of principal components). The final models were validated using a set of 6 molecules in the test set, which did not contribute to the model development.

Two different methods were used for the development of QSAR equations (i) stepwise multiple linear regression (MLR), (iii) genetic algorithm-partial least squares (GA-PLS). These methods are well documented in the literature (11).

In stepwise regression, a multiple-term linear equation was built step by step. The basic procedures involve (i) identifying an initial model, (ii) iteratively <stepping>, that is,

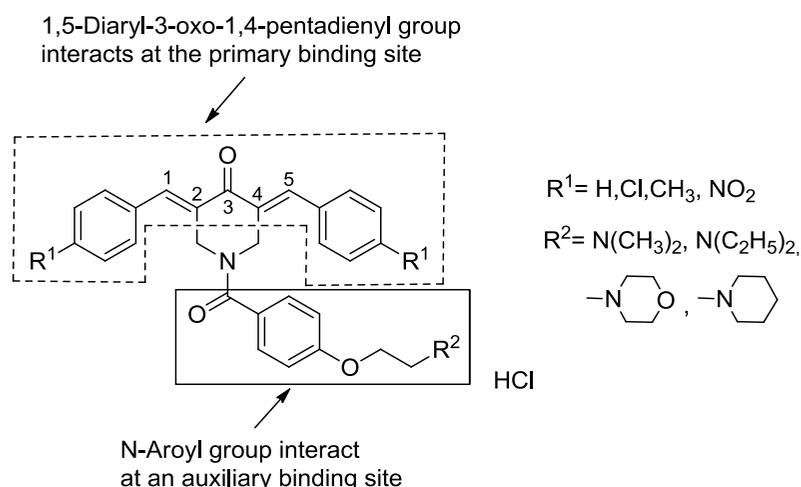


Figure 1. The proposed primary and auxiliary binding sites of cytotoxic N-aryl-3,5-bis(benzylidene)-4-piperidone analogs.

repeatedly altering the model at the previous step by adding or removing a predictor variable in accordance with the «stepping criteria» (in our case, probability of $F = 0.05$ for inclusion; probability of $F = 0.1$ for exclusion for the forward selection method), and (iii) terminating the search when stepping is no longer possible given the stepping criteria, or when a specified maximum number of steps have been reached. Specifically, at each step, all variables are reviewed and evaluated to determine which one will contribute most to the equation. This variable will then be included in the model, and the process starts again. A limitation of the stepwise regression search approach is that it presumes there is a single «best» subset of X variables and seeks to identify it. There is often no unique «best» subset, and all possible regression models with a similar number of X variables as in the stepwise regression solution should be fitted subsequently to study whether some other subsets of X variables might be better. In the present study, MLR with stepwise selection and elimination of variables was applied for developing QSAR models using SPSS 11.5 software. The resultant models were validated by a leave-one-out cross-validation procedure (using MATLAB software) to check their predictive power and robustness.

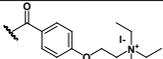
A genetic algorithm uses genetic rules such as reproduction, cross-over, and mutation to build pseudo organisms that are then selected, on the basis of a fitness criterion to survive and pass

information on to the next generation. To select the most relevant set of descriptors, the evolution of population was simulated (12-15). PLS is a generalization of regression, which can handle data with strongly correlated and/or numerous X variables (15). It gives a reduced solution which is statistically more robust than MLR. The linear PLS model finds «new variables» (latent variables or X scores) that are a linear combination of the original variables. To avoid over fitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are non-significant. Cross-validation is a practical and reliable method of testing this significance (14). Application of PLS thus allows the construction of larger QSAR equations while avoiding over fitting and eliminating most variables. PLS is normally used in combination with cross-validation to obtain the optimum number of components (15)19]. In the GA-PLS procedure, the optimum number of latent variable must be determined. Here, for each subset of descriptors (i.e., for each chromosome of the GA), a PLS model was developed separately and consequently the number of latent variables was optimized. The PLS regression method used was the NIPALS-based algorithm found in the chemometrics toolbox of MATLAB software (version 6.5; Math work Inc., Natick, MA, USA). Based on the Haaland and Thomas F-ratio criterion we performed the leave-one-out cross-validation procedure as it was applied in

Table 1. Chemical structures of the N-aryl-3,5-bis(arylidene)-4-piperidones **1-7** used in this study and their experimental and cross-validated predicted cytotoxic potencies (by GA-PLS) toward murine L1210 cells and human Molt 4/C8 and CEM T-lymphocytes.

Comp	R1	R2	L1210			CEM			Molt4/C8		
			Exp Pred. PIC50	Pred. PIC50	REP	Exp Pred. PIC50	Pred. PIC50	REP	Exp Pred. PIC50	Pred. PIC50	REPa
1a	H	H	6.90	7.13	0.23	6.23	6.39	0.16	6.22	6.40	0.18
1b*	H	Cl	7.62	8.04	0.42	6.94	7.10	0.16	7.13	7.51	0.38
1c	H	NO2	7.52	7.44	0.08	6.65	6.51	0.14	6.92	6.80	0.12
1d	H	CH3	6.93	6.65	0.28	6.23	6.21	0.02	6.23	6.28	0.05
2a		H	7.08	7.08	0.00	6.21	6.34	0.13	6.23	6.17	0.06
2b*		H	7.58	7.09	0.49	6.70	5.89	0.81	6.38	5.92	0.46
3a		H	7.01	6.85	0.16	6.09	5.95	0.14	5.76	5.78	0.02
3b		Cl	6.86	7.24	0.38	6.04	6.55	0.09	6.16	6.54	0.38
3c*		NO2	6.72	6.47	0.25	5.50	5.67	0.17	6.38	5.92	0.54
3d		CH3	6.85	6.44	0.41	5.95	5.79	0.16	6.14	5.66	0.48
4a		H	6.92	7.04	0.12	6.22	6.29	0.07	6.11	6.33	0.22
4b		Cl	7.67	7.45	0.22	7.09	6.85	0.24	7.26	7.08	0.18
4c		NO2	6.56	6.46	0.10	5.53	5.65	0.12	5.49	5.40	0.09
4d		CH3	6.36	6.60	0.24	6.12	6.28	0.16	6.12	6.19	0.07
5a		H	6.92	7.36	0.44	6.06	6.28	0.22	6.17	6.29	0.12
5b		Cl	8.01	7.87	0.14	7.00	6.94	0.06	7.53	7.41	0.12
5c*		NO2	6.68	6.06	0.62	5.50	5.44	0.06	5.50	5.48	0.02
5d		CH3	6.71	6.82	0.11	6.06	6.08	0.02	5.96	5.96	0.00
6a		H	8.35	7.99	0.36	6.84	6.27	0.57	5.52	5.05	0.47
6b*		Cl	7.91	7.76	0.15	6.92	6.73	0.19	6.93	6.64	0.29
6c		NO2	6.80	6.87	0.07	5.58	5.64	0.06	5.54	5.69	0.15
6d		CH3	7.34	7.44	0.10	6.13	6.06	0.07	6.24	6.35	0.11
7a*		H	8.20	7.81	0.39	8.04	7.46	0.58	8.09	7.20	0.89
7b		Cl	8.86	8.85	0.01	7.85	7.94	0.14	7.75	7.75	0.00

Table 1. Continue.

7c		CH3	7.27	7.35	0.08	7.37	7.24	0.13	7.09	7.13	0.04
----	---	-----	------	------	------	------	------	------	------	------	------

^{a)} REP: Relative error of prediction. ^{*} The compounds are included in the prediction set

our previous article (16). The MATLAB PLS toolbox developed by eigenvector company was used for PLS and GA modeling. The predicted power of obtained equations was validated using cross-validated squared correlation coefficient (Q²). In order to calculate Q² following equation was used.

$$Q^2 = 1 - \frac{\sum(Y_{\text{pred}} - Y_{\text{act}})^2}{\sum(Y_{\text{act}} - Y_{\text{mean}})^2}$$

Where Y_{pred}, Y_{act}, and Y_{mean} are predicted, actual, and mean values of the pIC₅₀, respectively. $\sum(Y_{\text{pred}} - Y_{\text{act}})^2$ is the predictive residual error sum of squares (PRESS) (17). Moreover, the predictive squared correlation coefficients “R²_{pred}” was used for further evaluation of the predictive potential of obtained equations. To calculate the R²_{pred} following equation was used.

$$R^2_{\text{pred}} = 1 - \frac{\sum(Y_{\text{pred}(\text{Test})} - Y_{\text{Test}})^2}{\sum(Y_{\text{Test}} - Y_{\text{Training}})^2}$$

where Y_{pred(Test)} and Y_{Test} are predicted and observed activity values, respectively, of test set compounds, and Y_{Training} is the mean activity value of training set (17).

Variable importance in the projection

In order to evaluate the relative importance of the variable appeared in the final GA-PLS, variable importance in projection (VIP) was employed (18). VIP values reflect the importance of terms in the PLS model. VIP is the sum over all model dimensions of the contributions variable influence (VIN). For a given PLS dimension, A (VIN)_{ak}² is equal to the squared PLS weight (w_{ak})² of that term, multiplied by the explained SS of that PLS dimension (19).

Y-randomization test

Y-randomization test is the technique ensures the robustness of a QSAR model. In this procedure, for each created models, the dependent variable vector pIC₅₀ is randomly shuffled, and a new QSAR model is developed using the original independent variable matrix.

For acceptable specific QSAR model, the new obtained QSAR equation (after several repetitions) is expected to have low R² and q² values (16).

Results and discussion

The structural features and the experimental cytotoxic potencies (represented as pIC₅₀) of the 3,5-bis(benzylidene)-4-piperidone derivatives **1-7** used in this study are shown in Table 1 (8). A SAR study of these compounds demonstrated that the attachment of the N-aroil group to the 3,5-bis(arylidene)-4-piperidones resulted in enhanced cytotoxic potencies toward murine L1210 lymphocyte leukemia cells and two human T-lymphocytes Molt4/C8 and CEM. It was suggested that in the case of various members of series **2-7**, alignment of the N-aroil groups with auxiliary binding sites takes place reinforcing the interaction of the 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore with a primary binding site.

In order to find important structural features of the N-aroil-3,5-bis(benzylidene)-4-piperidones which contribute to the cytotoxic potencies, a QSAR study was performed using different molecular descriptors and chemometrics tools. In this study, the chemometrics methods such as stepwise multi-linear regression (MLR) and genetic algorithm-partial least squares (GA-PLS) were used for modeling the relationship between the biological activity and molecular descriptors and the results were compared to find the best predictive cytotoxicity models.

MLR modeling

Separate stepwise selection-based MLR analyses were performed using different types of descriptors, and then a MLR equation was obtained using the pool of all calculated descriptors. As there are 19 molecules in the training set and according to the rule of thumb (the ratio of 5:1 for molecule / variable; Toplis ratio) (16), MLR models with a maximum

Table 2. The results of multiple linear regression (MLR) analysis for the compound in series 1-7 with different type of descriptors.

Equation	Cell line	QSAR Model	n	R ^{2a}	SE ^b	RMSE _{CV} ^c	Q ^{2d}	F ^e
1	L1210	Y = 9.037 (±1.56) Molecular density +3.693 (±1.435) HNar -0.007 (±0.003) Heat of formation - 9.712(±3.910)	19	0.70	0.38	0.44	0.50	11.25
2	CEM	Y=7.352 (±0.913) Molecular density + 0.062 (±0.009) Hydration energy +89.840 (±34.561) X4A - 12.102(±4.537)	19	0.85	0.25	0.29	0.77	29.11
3	Molt/C8	Y=7.507 (±1.155) Molecular density -0.493 (±0.084) Homo + 0.09 (±0.031) Dipole X + 6.039(±1.647)	19	0.81	0.31	0.34	0.72	21.31

^{a)} Squared correlation coefficient; ^{b)} Standard error of regression; ^{c)} Root mean square of cross validation; ^{d)} Leave-one-out cross-validation correlation coefficient (R²CV); ^{e)} F-statistics. Squared correlation coefficient (R²) was calculated using following equation: $1 - \frac{\sum (y_{\text{experimental}} - y_{\text{calculated by model}})^2 / \sum (y_{\text{experimental}} - y_{\text{average experimental}})^2}{n - p}$ Equation for calculation of Standard error of regression: $[SE^2 = \sum (y_{\text{experimental}} - y_{\text{calculated}})^2 / (n - p)]$ (n = number of samples, p = number of model parameters)

number of 3 or 4 variables were selected. The results are summarized in Table 2. A small difference between the conventional and cross-validated correlation coefficients of the different MLR equations (Table 2) reveals that none of the models are over fitted, which can be partially attributed to the absence of collinearity between the variables in one hand and using no extra variables on the other hand. The correlation coefficient (r^2) matrix for the descriptors used in different MLR equations shows that no significant correlation exists between pairs of descriptors (Table 3).

L1210 cell line. Stepwise selection and elimination of variables produced a three-parametric QSAR equation 1, having moderate statistical quality ($R^2 = 0.70$, $SE = 0.38$, $Q^2 = 0.50$, $RMS_{CV} = 0.45$ and $R^2_{\text{pred}} = 0.69$) (Table 2 and 4). The selected variables demonstrate that chemical (heat of formation), topological and spatial parameters (molecular density and Narumi harmonic topological index (HNar)) mainly affect the cytotoxic potential of the 3,5-bis(arylidene)-4-piperidone derivatives against the L1210 cell line.

CEM cell line

The MLR equation 2 obtained from the pool

of the calculated descriptors for the CEM cell line possesses good statistical quality ($R^2 = 0.85$, $SE = 0.25$, $Q^2 = 0.77$, $RMS_{CV} = 0.29$ and $R^2_{\text{pred}} = 0.81$) (Table 2 and 4) and demonstrates that the anticancer property of the compounds is mainly dependent on chemical (heat of formation), topological and spatial (molecular density and average connectivity index chi-4:X4A) properties of the molecule. These results show that the parameters that influence anticancer properties of the compounds against the CEM line are similar to those obtained for the L1210 cell line.

Molt/C8 cell line

The three-parametric QSAR equation 3 (Table 2) demonstrates the quantitative effects of the structural parameters on the cytotoxic potential of 3,5-bis(arylidene)-4-piperidone derivatives on the Molt/C8 cell line. This model demonstrated high statistical quality ($R^2 = 0.81$, $SE = 0.31$, $Q^2 = 0.72$, $RMS_{CV} = 0.34$, and $R^2_{\text{pred}} = 0.64$) (Table 2 and 4) which reveals the significant effects of quantum (HOMO, dipole moment X (DMx)) and spatial parameter (molecular density) on the cytotoxic activity of the compounds. It is revealed that the molecular density plays a determinant effect on the cytotoxic potential of

Table 3. Correlation coefficient (r^2) matrix for descriptors used in multiple linear regression equations 1-3.

	Heat of formation	HNar	Molecular density	Hydration energy	X4A	HOMO	Dipole x
Heat of formation	1	-0.352	0.340	-0.573	.334	0.653	0.045
HNar		1	-0.4920	0.232	.288	-0.292	0.180
Molecular density			1	-0.159	-0.121	0.364	0.244
Hydration energy				1	-0.114	-0.667	0.184
X4A					1	0.110	0.192
Homo						1	0.4730
Dipole x							1

the compounds on this cell line. Therefore, the molecular density and topological parameter are important descriptors that should be taken into consideration in the designing of potent cytotoxic 3,5-bis (arylidene)-4-piperidones analogs.

*GA-PLS modeling
L1210 cell line*

In PLS analyses, the descriptors data matrix is decomposed into orthogonal matrices, the scores of which are constrained to have inner relationships with the dependent variables. Therefore the multi-collinearity problem in the descriptors is omitted by PLS analysis. To find a more convenient set of descriptors in PLS modeling, a genetic algorithm was used. To do so, many different GA-PLS runs were conducted using different initial sets of populations. Given 19 calibration samples, the leave-one-out cross-validation procedure was used to find the optimum number of latent variables for each PLS model. For the L1210 cell line, the GA-PLS

model that resulted in the best fit contained 7 indices (two of these indices such as molecular density and heat of formation were also obtained by MLR in this cell line). Moreover, similar to the results of MLR analysis, GA-PLS analysis also showed that topological parameters affect the cytotoxic potencies of the compounds in the L1210 cell line. As per GA-PLS modeling, a combination of chemical (heat of formation, melting point), quantum chemical (dipole moment X), conformational (torsion energy) and spatial (molecular density) descriptors account for the cytotoxic potential of the compounds toward L1210 cells. The resulted GA-PLS model possess very good statistical quality (i.e., $R^2 = 0.86$, $Q^2 = 0.66$). The predictive ability of the model was measured by application to 6 external test set molecules ($R^2_{pred} = 0.81$ and $SE_{pred} = 0.37$). The calculated values of pIC_{50} obtained by the PLS model in the L1210 assay (refined from cross-validation of external prediction set) are shown in Table 1.

Table 4. The results of QSAR analysis by MLR^a and GA-PLS^b methods for the compounds in series 1-7

Cell line	Method	R^{2c}	Q^{2d}	SE_{pred}^e	$R^2_{pred}^f$
L1210	MLR	0.70	0.50	0.44	0.69
	PLS	0.86	0.66	0.37	0.81
CEM	MLR	0.85	0.77	0.29	0.81
	PLS	0.87	0.71	0.35	0.83
Molt4/C8	MLR	0.81	0.72	0.34	0.64
	PLS	0.91	0.77	0.32	0.82

^{a)} Multiple linear regressions, ^{b)} Genetic algorithm-partial least square, ^{c)} correlation coefficient of regression, ^{d)} Leave-one-out cross-validation correlation coefficient (R^2CV), ^{e)} Standard error of prediction regression, ^{f)} Correlation coefficient of prediction regression

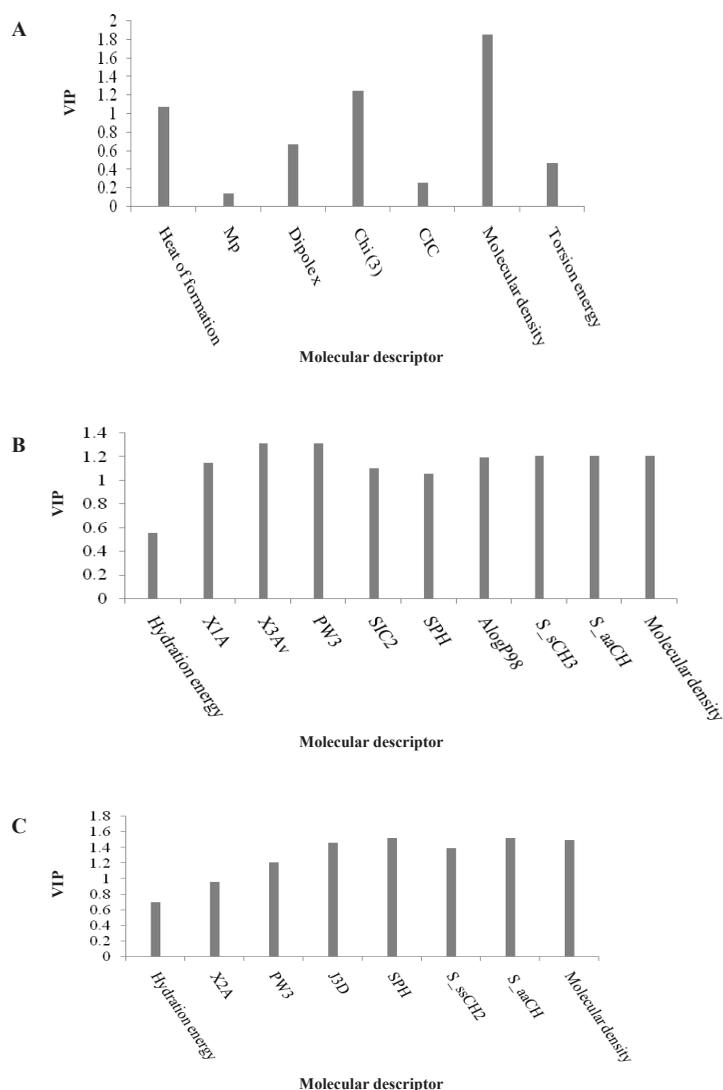


Figure 2. Variable importance for projection (VIP) of the selected variables obtained by GA-PLS model for the QSAR analysis of three different cell lines: (A) L1210, (B) CEM, and (C) Molt/C8.

To measure the significance of the 7 selected PLS descriptors, the variable importance in projection (VIP) was calculated for each descriptor. According to the method described by Eriksson *et al.* (20), X-variable (predictor variable) could be classified according to their relevance in explaining Y-variable (predicted variable) i.e. $VIP > 1.0$ (highly influential), $VIP < 0.8$ (less influential), and $0.8 < VIP < 1.0$ (moderately influential). The VIP analysis of the descriptors of the input variables used in the PLS equation (Figure 2A) shows that the parameters such as topological index (Chi3), molecular density and heat of formation are the

most important indices in the QSAR equation derived by PLS analysis for L1210 cell line. The regression coefficients of the selected variables obtained by GA-PLS model of QSAR analysis for three different cell lines are illustrated in Figure 3.

CEM cell line

The best GA-PLS model to predict cytotoxic activities of the compounds in series **1-7** toward the CEM cell line includes 10 indices (three of the indices are similar to that obtained by the MLR analysis). A combination of topological (X_1A), Chi-1(average connectivity index), PW3

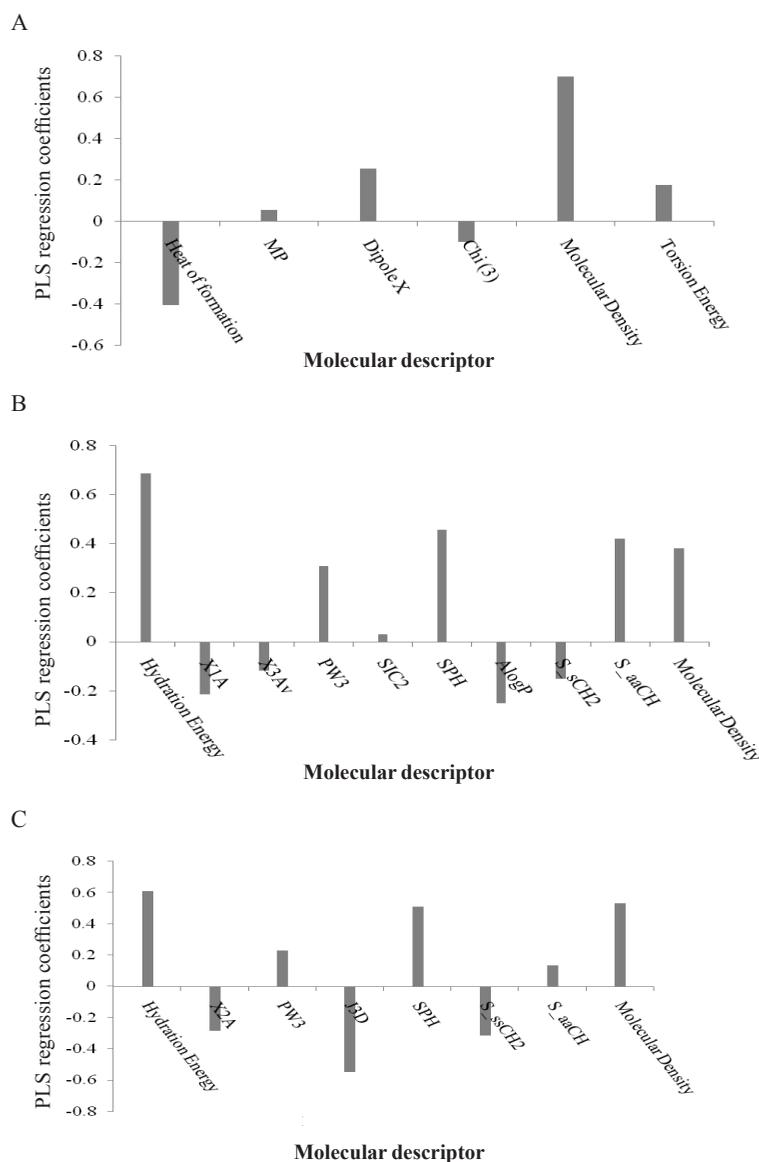


Figure 3. Regression coefficients of the selected variables derived by the GA-PLS model for QSAR analysis of three different cell lines: (A) L1210, (B) CEM, and (C) Molt/C8.

(Randic shape index), X_3Av (average valance connectivity), SIC_2 (structural information content: symmetry 2-order)), geometrical (SPH (sphericity of molecule)), chemical (LogP), spatial (molecular density), electrotopological state (sum of the E-state value for the CH group in the aromatic ring (S_{aaCH}) and sum of the E-state value for the methyl group (S_{sCH_3})) indices account for the cytotoxic properties of the 3,5-bis(arylidene)-4-piperidones toward the CEM cell line. The resulted GA-PLS model

showed high statistical quality (i.e., $R^2 = 0.87$ and $Q^2 = 0.71$). The PLS estimate of the coefficients for these descriptors are given in Figure 3B.

The predictive ability of the model was measured by application to 6 external test set molecules ($R^2_{pred} = 0.83$ and $SE_{pred} = 0.35$).

The pIC_{50} values used in the PLS model (refined from cross-validation of external prediction set) are shown in Table 1. The VIP analysis of PLS equation presented in Figure 2B suggests that among the selected

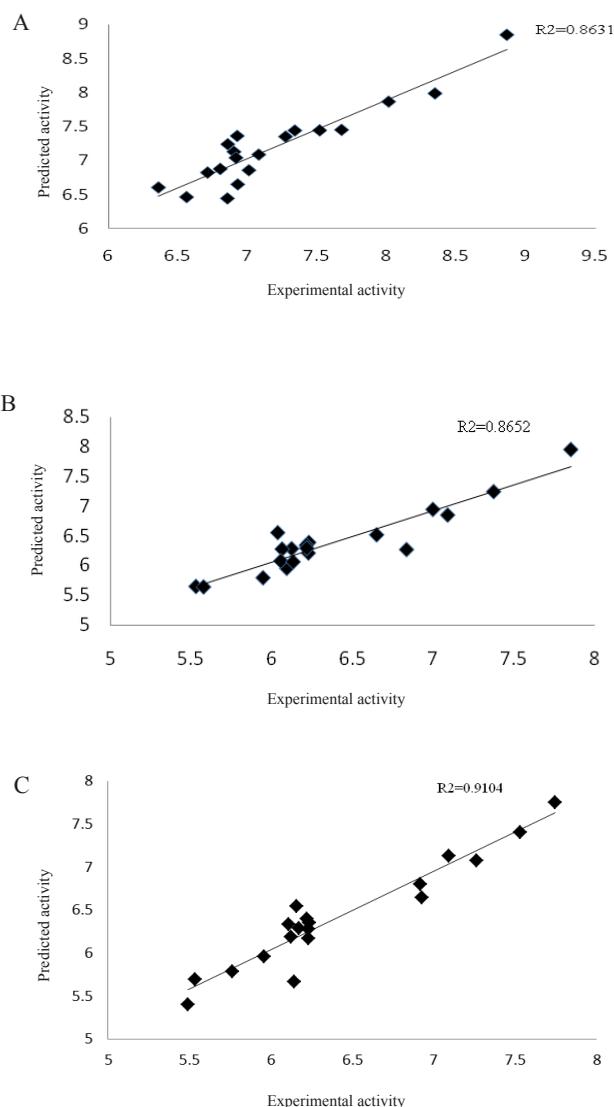


Figure 4. Plots of the cross-validated predicted cytotoxic potency values of activity by GA-PLS against the experimental values for three different cell lines: (A) L1210, (B) CEM, and (C) Molt/C8 cell line.

descriptors, topological and electrotopological parameters such as X_1A , PW_3 , X_3Av and estate keys (S_{sCH_3} and S_{aaCH}) are important influential descriptors. Similar to the L1210 cell line, the molecular density of the compounds is also a determinant parameter that influence cytotoxicity of 3,5-bis(arylidene)-4-piperidones against the CEM cell line.

Molt 4/C8 cell line

A combination of 8 selected variables such as chemical (hydration energy), topological (X_2A), geometrical (PW_3 , J_3D (3D-balaban index), SPH), spatial (molecular density) and

electrotopological state indices (sum of the E-state value for the CH group in the aromatic ring (S_{aaCH}) and sum of the E-state value for the methyl group (S_{sCH_3})) influence cytotoxic potencies of the compounds in series 1-7 against the Molt 4/C8 cell line as indicated by the best GA-PLS model. Most of these variables contribute significantly to the GA-PLS model as discussed previously for the CEM cell line. The resultant GA-PLS model demonstrated very high statistical quality [$R^2 = 0.91$ and $Q^2 = 0.77$] [Figure 3C]. The predictive ability of the model was measured by applying to 6 external test set

Table 5. R² and Q² values of GA-PLS models after several Y-randomization tests.

Iteration	L1210		CEM		Molt4/C8	
	R ²	Q ²	R ²	Q ²	R ²	Q ²
1	0.11	0.03	0.35	0.12	0.02	0.00
2	0.26	0.06	0.01	0.00	0.11	0.01
3	0.10	0.02	0.18	0.05	0.14	0.00
4	0.34	0.11	0.28	0.03	0.29	0.03
5	0.16	0.00	0.20	0.01	0.33	0.12
6	0.22	0.01	0.08	0.00	0.24	0.10
7	0.03	0.00	0.31	0.14	0.17	0.00
8	0.19	0.03	0.30	0.10	0.36	0.08
9	0.08	0.02	0.17	0.07	0.25	0.02
10	0.38	0.04	0.06	0.00	0.23	0.05

molecules and the squared correlation coefficient for prediction (R^2_{pred}) was found to 0.82.

The VIP analysis of the GA-PLS model presented in Figure 2C showed that the molecular density, topological (X_2A), geometrical (SPH, J_3D) and electrotopological state indices influence the cytotoxicity of the compounds in series 1-7. This result is similar to that obtained for the L1210 cell line.

The plots of the predicted pIC_{50} values (cross-validated) against the experimental values for the L1210, CEM and Molt4/C8 cell lines by GA-PLS analysis showed R^2 values 0.86, 0.87 and 0.91, respectively (Figure 4) which indicates that the QSAR models are reliable to predict cytotoxicity of the 3,5-bis(benzylidene)-4-piperidones 1-7.

All of the generated GA-PLS models were further validated by applying the Y-randomization test. Several random shuffles of the Y vector were performed and the results are shown in Table 5. The low R^2 and Q^2 values indicate that the good results in our original model are not because of a chance correlation or structural dependency of the training set.

The statistical parameters of QSAR analysis by MLR and GA-PLS are demonstrated in Table 4. The results indicate that GA-PLS analysis is more accurate for predicting the cytotoxic potential of 3,5-bis(arylidene)-4-piperidones in all three cell lines. The cross-validation statistics reported in Table 4 suggest the higher prediction ability of the GA-PLS model. This can be a result of the more number of descriptors used

by GA-PLS with respect to MLR analysis. As far as the less parametric model is obtained by MLR analysis, the results of MLR analysis are more descriptive for interpretation of the structure-cytotoxic relationship. However for the prediction of the cytotoxic potential of novel compounds of these series, the GA-PLS model is more useful.

Conclusion

It is clearly understood that some of the spatial, (electro) topological and chemical descriptors are important structural parameters that mainly affects the cytotoxic potencies of these compounds in all three cell lines. The “molecular density” of compounds is one of the most important spatial descriptors that should be considered in the design of novel compounds of these series. As the results of the SAR study of various members of these series suggest, alignment of the N-aroyle groups with auxiliary binding sites takes place reinforcing the interaction of the 1,5-diaryl-3-oxo-1,4-pentadienyl pharmacophore with a primary binding site (8). This result can be interpreted by the important structural descriptor “molecular density”. It is obvious that the orientation of N-aroyle group for better interaction with a corresponding binding site is affected by molecular density. Other topological descriptors also affect this desirable orientation of the ligand and therefore play a determinant role in the binding of the ligands with its target receptors.

The other important electro-topological parameter that should be considered is the electro-topological state indices which encode the electronic state of each atom type in the molecule: "atom-type E-state indices". The results demonstrate that the sum of the electro-topological state of "aromatic CH" (S_{aaCH}) and the sum of the electro-topological state indices of methyl groups (E-state keys: S_{sCH_3}) are among the important influential descriptors in the cytotoxic potential of 3,5-bis(arylidene)-4-piperidones especially toward CEM and Molt/C8 cell lines. The sum of the electro-topological state indices of the aromatic CH and methyl groups especially should be considered in the employment of different N-acyl substituents.

The results demonstrate that similar structural descriptors affect the cytotoxic potential of N-aryloyl-3,5-bis(arylidene)-4-piperidones in three different cell lines. This result indicates that similar molecular mechanisms are involved in the cytotoxic activity of the compounds in these cell lines. In addition, the previous quantitative-cytotoxic activity study of asymmetrical 2,6-bis(benzylidene)cyclohexanones against MCF-7, SK-N-MC and MDA-MB-231 cell lines (20), reported the spatial and topological parameters as determinant structural descriptors in cytotoxic potential of these compounds against the cell lines. This finding also confirms that the same molecular mechanisms are involved in the cytotoxic activity of different analogues of this scaffold.

References

- (1) Dimmock J, Raghavan S, Logan B, and Bigam G. Antileukemic evaluation of some Mannich bases derived from 2-arylidene-1, 3-diketones. *Eur. J. Med. Chem* (1983) 18: 248-254.
- (2) Mutus B, Wagner JD, Talpas CJ, Dimmock JR, Phillips OA, and Reid RS. 1-p-Chlorophenyl-4, 4-dimethyl-5-diethylamino-1-penten-3-one hydrobromide, a sulfhydryl-specific compound which reacts irreversibly with protein thiols but reversibly with small molecular weight thiols. *Anal. Biochem.* (1989) 177: 237-243.
- (3) Baluja G, Municio A, and Vega S. Reactivity of some α , β -unsaturated ketones towards sulfhydryl compounds and their antifungal activity. *Chem. Ind* (1964) 1964: 2053-2054.
- (4) Okey Ab HP. Principles of Medical Pharmacology, Kalant H, Grant DM, Mitchell J. (eds.) Elsevier; Toronto, Canada (2007) 902.
- (5) Fulda S and Debatin KM. Sensitization for anticancer drug-induced apoptosis by betulinic Acid. *Neoplasia*, New York, NY (2005) 7: 162-170.
- (6) Xu GW, Mymryk JS, and Cairncross JG. Pharmaceutical mediated inactivation of p53 sensitizes U87MG glioma cells to BCNU and temozolomide. *Int. J. cancer.* (2005) 116: 187-192.
- (7) Dimmock JR, Padmanilayam MP, Puthucode RN, Nazarali AJ, Motaganahalli NL, Zello GA, Quail JW, Oloo EO, Kraatz HB, Prisciak JS, Allen TM, Santos CL, Balzarini J, De Clercq E, and Manavathu EK. A conformational and structure-activity relationship study of cytotoxic 3,5-bis(arylidene)-4-piperidones and related N-acryloyl analogues. *J. Med. Chem.* (2001) 44: 586-93.
- (8) Das U, Alcorn J, Shrivastav A, Sharma RK, De Clercq E, Balzarini J, and Dimmock JR. Design, synthesis and cytotoxic properties of novel 1-[4-(2-alkylaminoethoxy)phenylcarbonyl]-3,5-bis(arylidene)-4-piperidones and related compounds. *Eur. J. Med. Chem.* (2007) 42: 71-80.
- (9) Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery Jr JA, Stratmann RE, and Burant JC. Gaussian 98, Revision A.7. Gaussian, Inc., Pittsburgh, PA (1998).
- (10) Todeschini R. Milano Chemometrics and QSAR Group <http://micchem.disat.unimib.it/chm/>.
- (11) Darlington RB, Regression and linear models: McGraw-Hill, New York (1990) 101-280
- (12) Cho SJ and Hermsmeier MA. Genetic Algorithm guided Selection: variable selection and subset selection. *J. Chem. Inf. Comput. Sci.* (2002) 42: 927-36.
- (13) Ahmad S and Gromiha MM. Design and training of a neural network for predicting the solvent accessibility of proteins. *J. Comput. Chem.* (2003) 24: 1313-1320.
- (14) Fan Y, Shi LM, Kohn KW, Pommier Y, and Weinstein JN. Quantitative structure-antitumor activity relationships of camptothecin analogues: cluster analysis and genetic algorithm-based studies. *J. Med. Chem.* (2001) 44: 3254-63.
- (15) Leardi R. Genetic algorithms in chemometrics and chemistry: a review. *J. Chemometr.* (2001) 15: 559-569.
- (16) Khoshneviszadeh M, Edraki N, Miri R, Foroumadi A, and Hemmateenejad B. QSAR study of 4-aryl-4H-chromenes as a new series of apoptosis inducers using different chemometric tools. *Chem. Biol. Drug. Des.* (2012) 79: 442-58.
- (17) Vyas VK, Ghate M, and Katariya H. 2D and 3D-QSAR study on 4-anilinoquinazoline derivatives as potent apoptosis inducer and efficacious anticancer agent. *Org. Med. Chem. Lett.* (2011) 1: 1-11.
- (18) Olah M, Bologna C, and Oprea TI. An automated PLS search for biologically relevant QSAR descriptors. *J. Comput. Aid Mol. Des.* (2004) 18: 437-449.
- (19) Eriksson L, Johansson E, Kettaneh-Wold N, and Wold

S. Multi-and Megavariate Data Analysis: Principles and Applications, Umea, Sweden: Umetrics Academy (2001) chapter 3-4.

- (20) Nakhjiri M, Safavi M, Alipour E, Emami S, Atash AF, Jafari-Zavareh M, Ardestani SK, Khoshneviszadeh M, Foroumadi A, and Shafiee A. Asymmetrical 2, 6-bis

(benzylidene) cyclohexanones: Synthesis, cytotoxic activity and QSAR study. *Eur. J. Med. Chem.* (2012) 50: 113-123.

This article is available online at <http://www.ijpr.ir>

Journal alert and more ...
Visit <http://www.ijpr.ir>
or
[http:// ijpr.sbm.ac.ir](http://ijpr.sbm.ac.ir)